UNIVERSITY OF TURKU

ABSTRACT

	Bachelor's thesis
X	Master's thesis
	Licentiate's thesis
	Doctoral dissertation

Subject	Economics	Date	22.7.2023
Author(s)	Ville Keränen	Number of pages	84
Title	TRAFFIC AND THE ECONOMY How to utilize open traffic data in economic analysis		
Supervisor(s)	Prof. Jouko Vilmunen		

Abstract

This thesis studies the interaction of passenger car traffic and the economy. Finnish law requires a cost benefit -calculation for government-financed infrastructure investments. For this reason, the economic models and literature to model the interactions of infrastructure and the economy has been in high demand in Finland.

The literature review in this thesis looks at traffic economics through the dynamics of congestion, the value of time savings and methods to utilize these valuations in the form of road toll schemes. The literature review also looks at the housing and labour market outcomes from infrastructure investments. To finish off the literature review, there's a comprehensive study of literature around the data used in this study and how it has been utilized.

This literature review generates an understanding of the past and current affairs of how traffic data is used in analysing the economy and what kind of avenues exist to evaluate the impact of transportation infrastructure. These are mainly: population, labour market outcomes, housing market prices and travel demand.

The analysis seeks to validate the use of traffic data as a split dataset between work and leisure times. The literature shows a clear difference in time values and behavior in different types of travel demand. The raw data from Finnish Traffic and Infrastructure Agency is validated and analyzed with simple OLS regressions with different economic variables.

There are distinct differences in work and leisure-time traffic. The split also makes models perform a lot better and when combined they make the models perform very well. However, as the analysis studies all the municipalities along the E18, the analysis ends up showing the differences between municipalities in stead of causal effects of the E18 to local municipalities during the study period.

The analysis finds multiple ways to use and develop the model and how a single piece of infrastructure can be modelled better. These better models can then be utilized in analyzing the efficacy of infrastructure investments with higher precision.

Key words	traffic economics, value of time, congestion pricing, cost benefit analysis, OLS regression, open data

TURUN YLIOPISTO

TIIVISTELMÄ

	Kandidaatintutkielma
X	Pro gradu -tutkielma
	Lisensiaatintutkielma
	Väitöskirja

Oppiaine	Taloustiede	Päivämäärä	22.7.2023
Tekijä(t)	Ville Keränen	Sivumäärä	84
Otsikko	LIIKENNE JA TALOUSTIEDE Kuinka hyödyntaa avointa liikennedataa taloustieteen tutkimuksessa		
Ohjaaja(t)	Prof. Jouko Vilmunen		

Tiivistelmä

Tämä tutkielma tutkii henkilöajoneuvoliikenteen ja talouden välistä yhteyttä. Suomen laki vaatii hyöty-kustannus-analyysin valtiorahoitteisiin infrastruktuuri-investointeihin. Tästä syystä taloudelliset mallit ja kirjallisuus infrastruktuurin ja talouden välisestä vuorovaikutuksesta ovat Suomessa kovan kysynnän kohteena.

Kirjallisuuskatsaus tässä gradussa käy läpi liikennetaloustieteen eri tutkimuskohteita, kuten ruuhkia, ajansäästön arvoa ja tapoja hyödyntää näitä arvoja optimoitujen tietullien avulla. Kirjallisuuskatsaus tarkastelee myös kiinteistömarkkinoiden ja työmarkkinoiden reagointia infrastruktuuri-investointeihin. Lopuksi kirjallisuuskatsaus tekee kattavan koonnin LAM-liikennedatan käyttökohteista ja Suomen liikenneinvestointien vaikutusarvioinneista, jotka hyödyntävät tätä dataa.

Tavoite kirjallisuuskatsaukselle on varmentaa liikennedatan käyttötarkoitukset liikennetaloustieteessä ja antaa kuva eri tavoista arvioida liikenneinfrastruktuurin taloudellisia vaikutuksia. Hyväksi havaittuja tapoja ovat mm. vaikutukset asukaslukuun, työmarkkinoihin, kiinteistömarkkinoihin ja liikenteen määrään.

Analyysiosion tavoite on validoida menetelmä, jossa liikennedata jaetaan työ ja vapaa-ajan välillä, jotta eri aikana ja eri tarkoituksiin tapahtuvan liikenteen kirjallisuudessa havaitut erikoisuudet saataisiin osaksi taloudellista analyysiä. Raakadata Väylävirastolta validoidaan ja analysoidaan OLS-regressiolla erilaisten taloudellisten muuttujien suhteen.

Työ- ja vapaa-ajan liikenteessä havaitaan selviä eroja. Jako myös vaikuttaa regressiomallien suoriutumiseen positiivisesti sekä erillään omina malleinaan että yhdessä samassa mallissa. Koska analyysi ryhmittää datan E18 varrella olevien kuntien mukaan, analyysi tarkastelee enemmän kuntien välisiä eroja kuin ajan kanssa tapahtunutta taloudellista kehitystä E18:n varrella.

Analyysi tunnistaa monia tapoja käyttää ja kehittää mallia varsinkin, kun tarkastellaan yksittäistä infrastruktuurin pätkää. Tätä paranneltua mallia ja sen osia voidaan hyödyntää entisestään, kun analysoidaan infrastruktuuri-investointien vaikutusta paremmalla tarkkuudella.

Avainsanat	Liikennetaloustiede, aika-arvot, ruuhkahinnoittelu, hyöty- kustannusanalyysi, OLS regressio, avoin data.

UNIVERSITY OF TURKU
Turku School of Economics

TRAFFIC AND THE ECONOMY

How to utilize open traffic data in economic analysis

Master's Thesis
in Economics

Author(s):
Ville Keränen

Supervisor(s):
Prof. Jouko Vilmunen

22.7.2023
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Scope

This thesis is limited in its scope to only look at road passenger traffic and analyzing the changes in demand in road passenger traffic over time. This is also reflected in chapters 2 and 3, where the earlier studies analyzed mainly study the costs and benefits regarding passenger traffic and labor-supply and demand. The study also limits the analysis to the E18 only, because of the limitations of the data set.

Chapter two also goes into the economic detail of traffic and how it works because traffic dynamics on a highway are crucial to interpret the research data.

1.2 Research question

Traffic economics is utilized in evaluating the economic impact of infrastructure investments. Feasibility studies that conduct a cost-benefit analysis rely heavily on parameters and values from economic research on traffic demand elasticities, traffic demand predictions and wider economical models.

In this thesis I look at how traffic is viewed in the economics discipline, how it is studied and what kind of data sets are useful. I also investigate the Finnish traffic system and especially how Finnish traffic investments are evaluated. This also includes an in-depth view into the reports from Finnish officials that study and evaluate Finnish traffic system.

The research question is formed around traffic performance data on Finnish highway E18. Over the long-term traffic demand has been coupled with GDP growth. Traffic demand is induced from economic activities and especially passenger vehicle use increases with increased income. Most negative externalities from traffic come from congestion. Congestion is a pricing problem. The value of the activity at the end of the trip tops the costs of sitting in congestion.

Congestion concentrates around the times people travel to and from work. Therefore, these are also the bottlenecks most traffic investments seek to improve. The data from the E18 provides a great opportunity to study long term differences in traffic demand in vehicle travel patterns related to both work and leisure.

The research questions are as follows:

- Is there a difference in leisure and worktime traffic demand? How do they affect the model?
- Is publicly available open data enough to model traffic around an entire road?

2 THEORETICAL BACKGROUND

2.1 Transportation network as a public good

Occasionally, markets fail. This is due to lack of competition, negative externalities, or asymmetric information. In this situation there exists a net social loss from utilizing the free market option. Some products or services are therefore offered through the government or government related parties. Public goods are goods or services, that one cannot be excluded from (non-excludability-condition) and when one uses the good, it is not eating away someone else's share of the public good. This is called the non-rivalry condition. These two conditions define a pure public good. Stiglitz & Walz (2006, 254)

But public goods also distort markets. Government monopolies, for instance, heavily distort the market where the monopoly operates because the prices are not affected by competitive markets. Sometimes this distortion is acceptable and explained by the negative externality the good produces such as alcohol or gambling. Some goods are also by their nature a natural monopoly. This is the case for road infrastructure, sewage, and electricity grid. The private sector can also offer a private good, that is funded by the public sector. A government owned company can operate in the markets, offering a good that is either subsidized or important to the nation.

Public goods can also be offered, because of the free rider problem. Free riding is a situation where some users of a good, who cannot be excluded from using it enjoy the benefits of the good or service, but do not pay for it. A common example is the lighthouse example. One port user may not gain enough benefits from building a lighthouse, but if it can coordinate with other users to pay for it as well, they may see the benefit of building it. However, this type of coordination is difficult and creates additional costs that often end up in private individuals and companies opting to forgo these projects. So even though the overall benefit exceeds the cost of the lighthouse, it will not be built by private operators.

The two conditions that define a public good are extremely strict. Many goods that we see as public goods are not pure public goods. Stigliz & Waltz (2006, 254) construct a two-axis chart that can be used to better visualize the nature of different goods in relation to the two conditions for the public good. This is visualized in fig. 1.

A highway that is not congested can be viewed as a pure public good. It is hard to close it from all users without it reaching everyone, and at the same time for all road users the otherwise massive costs are low. These dynamic changes when the road reaches its maximum capacity. There might be only a small number of additional users for the highway, but the cost to add capacity is high. This is why careful planning goes into road networks and where traffic demand will be in the future. Bearing in mind, that road capacity is mostly for the peak traffic time, learning more about the dynamics of traffic is highly valuable for local and national infrastructure development.

Transportation infrastructure can be viewed as a very common public good. Because roads require a lot of land, acquiring it as a private entity is difficult. This is why almost all highways and main road networks are publicly funded, built and owned. Highway and main road networks are also natural monopolies. It is near impossible to build a competing network privately.

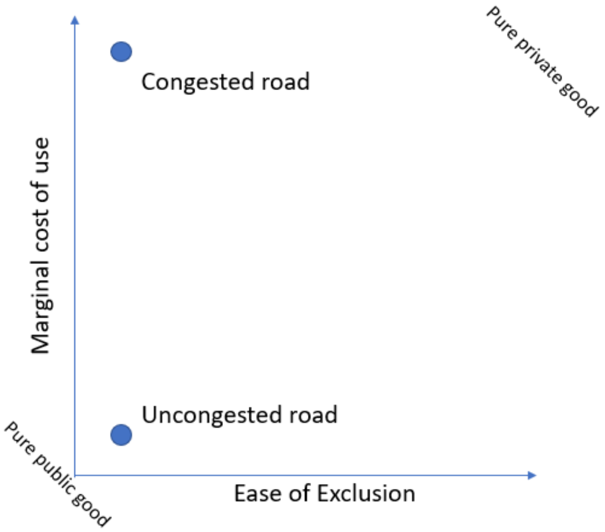


Figure 1 Definition of a public good (Stiglitz & Waltz, 2006, 254)

2.2 Dynamics of congestion

Congestion is of great interest to economists because it is a very classical problem of dividing scarce resources. A congested road is a sub-optimal allocation of public infrastructure. Road supply is limited in the short term and investment in road infrastructure is expensive and it takes up valuable land that could be used in other productive ways. For this reason, other ways to mitigate congestion has been studied widely. Road demand is the sum of road users at a given time. A congested road has a counter intuitive dynamic, where the supply for demanded trips depends on how fast the drivers can drive on the road. Understanding this mechanism is vital to understanding how road pricing schemes are planned.

Anthony Downs goes through several studies regarding vehicle speeds and traffic flows on major roads based on data from highways in California in his book Still Stuck in Traffic – Coping with Peak-Hour Congestion (2004). These studies give a great insight to the dynamics of traffic congestion and how traffic demand works in practice. The technical dynamics are important to understand to make sense of traffic data and how congestion is defined.

Skabardonis et al (2003) uses data collected from California using probe vehicles, where the highest rates of vehicles/lane/hour were achieved between speeds of 40-60 mph (65-96 km/h). In higher speeds the volume decreased. This supports the idea presented in 2.1, that when a road gets congested, the productivity of the road also goes down. Here highest flows were around 2500 vehicles/lane/hour. Chen et. al. (2000) use aggregating 30 second loop detectors along a 7-mile section of an interstate, and concluded, that congestion does not occur identically from day to day, but that there is variation. However, this variation has a roughly normal distribution with a relatively stable mean over time that varies greatly between interstate sections. This means, that recurring congestion has a stochastic element like the surrounding community. The authors also conclude that any delay in estimated travel times caused by an average speed less than 60 mph was caused by congestion. Downs adds that in this data set the max flow rates ranged from 2000 to 2800 vehicles/lane/hour. The authors noticed that the highest flow rates were achieved at around 60 mph. This means that if a 60 mph results in 2800 flow rate of vehicles, according to Downs the average interval of vehicles is only 93 feet (28 meters) or 1.07 seconds. Petty mentions that these results implicate a much shorter spacing for vehicles than the "long-established" two second rule, and far below a safe interval to stop in these speeds. However, these intervals are recorded on California highways like for example in Kockelman (2001).

In 2001 Jia et. al. published another study, that analyzed data from similar loop detectors and concluded, that the most desirable speed to maintain on the tracked freeways was 60mph (96km/h) and a 25% decrease in vehicle speeds from the ideal speeds can result in a 40% loss in highway flow rates. However, it is now known, how many drivers it takes to join the highway to cause this loss is speed.

The intuition Downs derives from these studies is that the higher the speed, the longer the distance between vehicles and this distance controls the flow rate or speed of the vehicles. You can only fit a certain number of cars on a length of road. As the number of vehicles on the road increases, the distance between cars gets smaller. This also makes most drivers drive slower. So, as we have people driving the free flow speed, and the amount of road users increases and the space between cars gets smaller, people drop their speeds and roads get congested. The main variable in this dynamic is the driver's ability to take risks and maintain a high speed while driving close to another vehicle. Now this risk taking is dependent on many factors like weather, road conditions, time of day and seasons. This intuition is described in fig. 2.

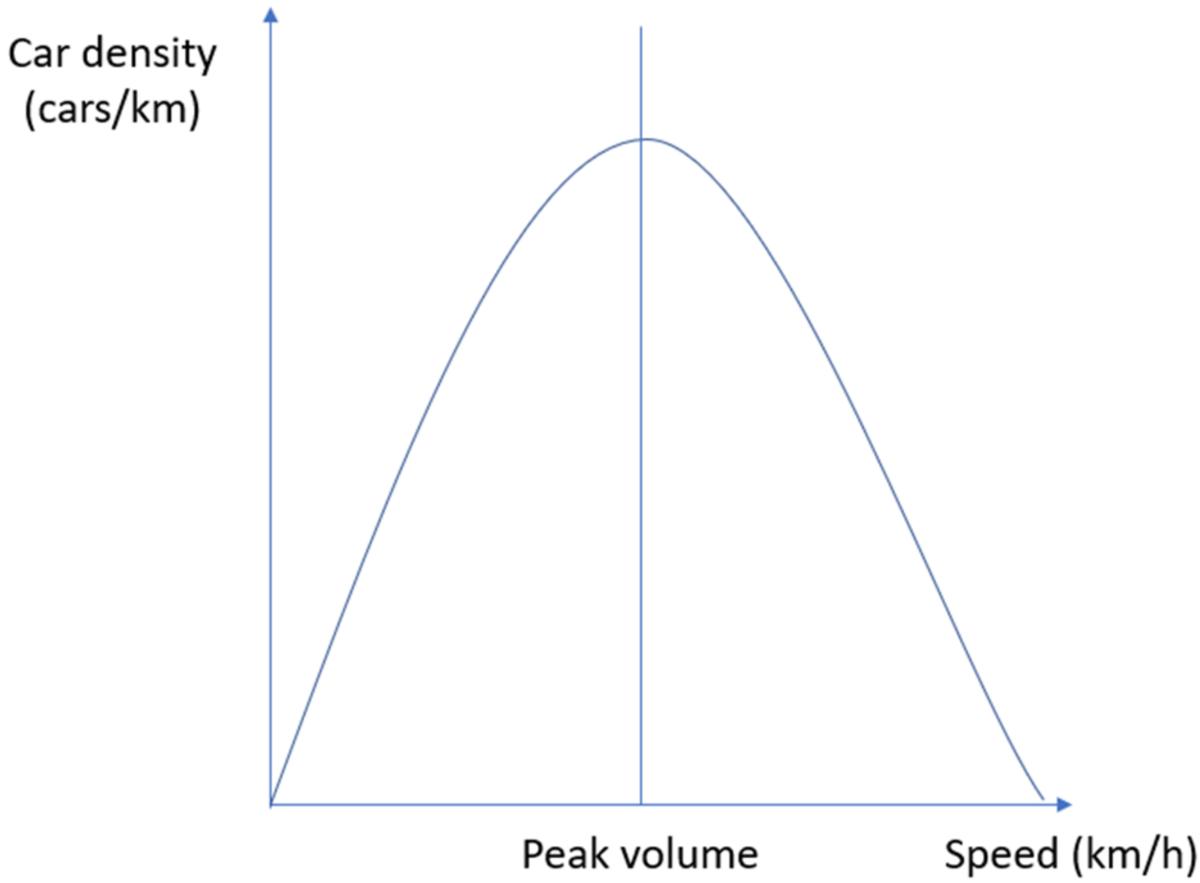


Figure 2, peak volume dependency on vehicle speed and density

Downs also elaborates the dynamics of recovery from congestion. Now that the road is carrying more vehicles than it can optimally hold, the unravelling of this dilemma starts from the beginning. As less vehicles enter the road, and at the front more cars unload to full speeds or to other roads the number of vehicles within a length of road decreases, and the distance between cars gets longer again and speeds return to the free flow speed of the freeway.

2.3 Transport and induced demand

Transport demand is widely referred in the transport economics literature as induced demand. What this means is that introducing more lanes to a highway, increasing the free flow speed through safety improvements, or removing parts of the drivers from the roads will be met with more traffic. This term was first explained by Downs (1962) in his article the law of peak-hour expressway congestion. The law simply states: “On urban commute expressways, peak hour traffic congestion rises to meet maximum capacity.” The article does not formulate this law, nor does it explore the idea further empirically, but it has inspired future research.

This empirical evidence is formulated and empirically tested in depth by Duranton & Turner (2011). They explore the connection between building roads and vehicle kilometers travelled. The model builds on the supply of lane kilometers of road and vehicle kilometers travelled (VKT) and utilizes three statistical methods to estimate VKT elasticity to road supply. The study examines different instrumental variables for road supply to offset the covariance of road kilometers and VKT. This is because roads are built when there is traffic. These methods and transport econometrics are looked at more carefully in chapter 3. Duranton & Turner conclude that the supply of road kilometers increases VKT considerably, and because the instrumental variable approach provides a plausible exogenous variation it is seen as the most confident estimate. They also conclude that public transit has no effect on VKT. This means that adding road capacity and public transit does not decrease congestion, because road expansion just expands traffic, and public transit has no effect on driving behavior when there is available road supply.

Duranton & Turner (2011) also point out, that congestion pricing (increasing the cost of traffic) is the best way to combat congestion.

There is an intuition to this conclusion. The cost of driving is a function of multiple variables. One is of course the actual cost of owning a vehicle and the associated costs of driving like insurance and gas. Driving also puts the driver and passengers in a danger of an accident. The risk of injury and death is a cost to everyone using the road. Driving also takes time. Driving is an activity away from leisure and work. The value of travel time, and the time saved is discussed more in later chapters.

2.4 Congestion pricing

Congestion pricing concerns the most efficient use of the road network. The idea is firs suggested as early as Pigou (1920). Vickrey (1969) formulates traffic congestion in his article Congestion theory and Transport investment. The model has a morning commute with one bottleneck with users that have homogenous time values. In the optimum a road toll finds a similar optimal congestion level that a construction of new lanes would. Only that defining the right level ex ante is difficult, so the construction project would be much more difficult to achieve. This model was developed into a dynamic congestion model that has seen wider use in transport research. But not in economics during the 70's and 80's. Vickrey's model has been used in modelling various tolling and congestion relievement schemes, such as high occupancy tolls and fast lanes. It is a foundation to modelling congestion dynamics. Because this thesis is not focused on modelling congestion itself, so this model will not be studied with any more depth.

Arnott et al (1990) expand on the economics and study coarse tolling, solving for optimal capacities within a single bottleneck. Coarse tolling means a flat fee toll during peak periods and 0 otherwise. What they find is that the road toll causes the congestion to disperse over the peak demand period, leading to higher welfare gains than before. In their next paper Arnott et al (1993) introduce price sensitivities, or elasticities, to the reaction function to the road tolls. They find that a self-financing tolling scheme is possible regardless of the tolling regime. Gasoline tax as a tolling regime would suffice as a road use payment just as well as a toll. They also confirm that the 1990 finding of higher welfare through dispersed traffic instead of congested bottlenecks applies in this case as well. The conditions for this self-financing scheme were first formulated by Mohring & Harwitz in 1962. An optimal toll should finance the capacity increase to maintain optimal capacity for whatever level that may be.

Congestion is best relieved by increasing the cost of driving (Duranton & Turner, 2011). The cost of driving has many aspects that are sought to relieve congestion. These are the price of fuel, parking, and road tolls during congested periods. A simple graph illustrates this below. In figure 3, the inverse demand of traffic, P(VKT), is a downward sloping curve. The area below P(VKT) is also the value of travel to the consumer. Road supply has very steep cost curve in the short time frame. Let the supply of roads be set at S(VKT). As stated in both Downs (2004) and Duranton & Turner (2011) the price for supplying roads and the price for a trip is an upward sloping curve until a point, where congestion occurs. Here the cost curve expands exponentially and causes the supply curve to intersect with the demand curve. This "hockey stick" curve is illustrated on the right-hand side in figure 3. The supply of roads is static in the short term, and therefore congestion occurs. Direct and indirect costs to the road users are caused by both the social costs of lost time staying still in traffic and the externalities that congestion creates. Congestion tolls fix this problem by increasing the price of using the road. This is illustrated on the left-hand side of figure 4.

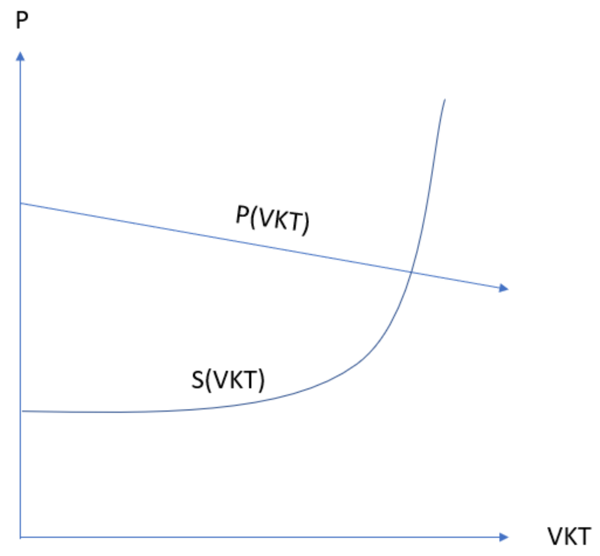


Figure 3 Basic traffic demand

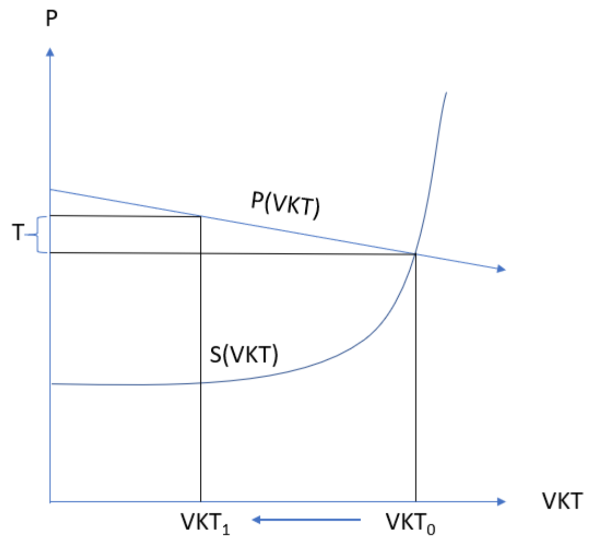


Figure 4 The effect of road tolls on traffic demand

The fundamental law of road congestion dictates, that new road capacity is met with a proportional increase in driving. So, if a social planner is to free up road capacity by removing cars from the roads, increasing public transport supply, or building new roads all they do is shift the supply curve and congestion will still ensue. This was also showed empirically earlier. (Duranton & Turner, 2011) Prices dictate road usage. Tolls increase the cost of travel, here VKT, and therefore lower the demand for driving to ease the problems of congestion in the travel network.

Sweden implemented road tolls in Stockholm and Gothenburg in 2006 and 2013, respectively. Börjessön (2019) discusses the long-term effects of these tolls and cites studies regarding road tolls and traffic volumes in London, Milan, Singapore, and the

United States. All studies have found that road tolls work, but the intensity is more nuanced. The Swedish road-tolls were implemented to two different cities. Stockholm has natural barriers that make tolling easy. Gothenburg is one of the smallest cities in the world that has implemented road tolls and does not have similar natural bottlenecks. This leads to less effective toll scheme, more expensive system with more tolling stations and less popular support. This leads to a conclusion that some cities are more capable of effectively tolling their traffic than others.

Traffic effects in the Swedish road tolls were big, elasticities being -0,87 in Stockholm and -0,69 in Gothenburg in the first year. The tolls are higher in peak hours and lower in off peak hours. Evenings and nights are free. Off peak elasticities are much higher, even though the toll is smaller. When Stockholm increased the toll prices, traffic decreased even during non-tolling hours. This can be attributed to entire shifts to other transportation means. The shift is much more pronounced in Stockholm than in Gothenburg, because Stockholm has a better public transportation network that supports the change to other travel modes. Stockholm also gains from the travel time savings that come lowered congestion levels. Gothenburg does not experience a similar gain, because rat running (diverging traffic around the toll roads to smaller and slower roads) is easier and public transport substitution is lower.

Road tolls are a regressive tax, meaning that because of their flat rate regardless of income, they form a larger percentage share of household income to lower income households. This means, that the decrease in driving seen in Stockholm and Gothenburg are mainly the result of people in lower income households changing to public transportation. In Gothenburg, the effects are smaller because the public transportation network is not as extensive. Börjesson also mentions in her discussion paper, that Gothenburg has higher car dependency amongst lower income households and a lower substitution rate to public transport. This all leads to road tolls being less effective in Gothenburg and they also cause more unwanted negative consequences in negative distribution effects. Negative distribution effect means, that the regressive nature of the tax is more profound, because private car usage is inelastic and therefore lower income households end up paying more as a share of household income when compared to higher income households. In Stockholm this does not occur to the same extent, because lower income households already have a higher rate of public transport ridership, and the substitution effect is larger.

Börjesson's report shows, that road tolls are situational and do not always lead to wanted outcomes. The report and various other studies such as Phang & Toh (1997) in Singapore, Santos & Shaffer (2004) in London and Gibson & Carnovale, (2015) in Milan add to the earlier body of work that shows that road tolls however are very effective in controlling traffic levels. However, the public views schemes such as these less favourably when the cost is borne by lower income households (Börjesson 2019a).

2.5 Economic impact of the traffic system

Traffic system has two ways it affects the surrounding economy. Direct economic impacts through construction projects, maintenance, planning and manufacturing of the equipment used to operate the infrastructure (rails, air control) or use it (train). The second, more difficult to measure impacts are the wider economic impact (WEI) which is more indirect. This is measured though the infrastructures improved ability to nurture economic activities. The avenues that wider economic impacts are introduced into the economy are increased productivity or agglomeration, accessibility and labour markets, market effects and the increased development of land around the road infrastructure. The scope of this thesis looks at the wider economic impacts, especially labour, so the direct economic impacts will not be discussed further.

WEI is being measured, because infrastructure projects are usually funded with public money, and policy makers should be able to explain the need to use public fund for these projects. This also means, that when there are multiple projects to choose from, a policy maker should choose the one with greatest economic impact or welfare gain to price ratio. Public operator is operating under resource scarcity after all.

Assessing WEI to evaluate infrastructure projects is the main reason the impacts are under interest to economists. The main framework to assessing WEI is a cost benefit analysis (CBA). The baseline is generated by calculating a present net value (PNV) of the project, and then assess the non-monetary effects of the project. In Finland authorities are required by law to assess major road and rail projects and their plans. Laki liikennejärjestelmästä ja maanteistä (503/2005) & Ratalaki (110/2007).

The impacts can be measured and viewed through five avenues.

- Impact to the producers through operating costs and ticket/toll revenues
- Impact to the public economy through operational costs and toll revenues
- Impact to traffic safety
- Impact to the environment

- Impact to the users, time savings and money

This thesis is interested on the impact on users, which will be discussed more in depth in its own chapter.

WEI is measured with Pareto optimality. Measuring WEI is not an easy task. The goal is to reach a Pareto optimal result, but how are the benefits measurable? Economists have tackled this issue by studying the unit costs of different traffic related measures, like an hour spent travelling, cost of an accident, ton of pollution or VKT as we saw earlier. These measures are given monetary values through various econometric models that we will explore later.

Traffic models evaluate travel behaviour mathematically and attempt to predict the changes to traffic intensity, direction, type, and location in each network. These computational models are different from the dynamic traffic model from Vickrey that was discussed in 2.4. But dynamic models are often also used to assess WEI. In some cases, general equilibrium models, statistical models from the national accounts and land use models are also used.

Researchers have also utilized qualitative methods such as expert interviews and questionnaires or polls to study the economic impacts. Also studying the projects after their completion is a way of evaluating a future project if a similar project has been evaluated. (Sinha & Labi, 2007)

2.6 Time budgets and travel behaviour

Value of travel time, and the savings in travelled time or value of travel time savings (VTTS) are measurements in welfare that the transportation system and its improvements bring to society. The monetary disutility of congestion is also generally calculated by using the value of time as a metric. Fosgerau & Engelson (2011) Time spent is a disutility and a cost of an activity that an individual must make. (Singleton, 13, 2017) An infrastructure investment is made to make certain groups of people or certain areas more accessible. But not all time is equal. There is a wide field of research in the value of time, and how people perceive time in different modes of transport. In early infrastructure cost benefit analyses this value of time was given a static value of hourly wage of the road users. (Johnson 1966, 135) However, later empirical work has shown, that time values vary for different travel destinations, types of travel, length of travel, and between cohorts. Some of the research is outlined here, and it works as a great motivation to explaining the reason why different types of demand for traffic exist.

In his 1965 paper “A theory of allocation of time” Gary Becker generates a framework for non-work activities, in which travel is also included. (Becker, 1965) The conclusion in the article is, that consumers have a non-monetary budget of time, that contributes to the micro theory of individual value maximizing. This leads to Becker also urge a wider study of time budgets in addition to monetary budgets to better understand changes in market behaviours. The article also discusses traffic and suggests that the assumptions of time value = hourly wage might be more intricate than that. This intuition is continued by Johnson in his 1966 article “Travel time and the price of leisure”, where he argues that wage is a biased measure of travel time value.

Time values are not equal to all people. Values vary between individuals and between modes of travel. In the United Kingdom, a wide meta-analysis of 226 studies by Abrantes & Wardman (2011) establishes a baseline for UK time values by analysing 1749 valuations. From the analysis, the authors display a wide array of time values, that give valuable insight to the traffic operators and public officials about the value of a traveller’s time. In addition to the variation in transport mode, valuations in wait times for public transport and in congestion also have a value or more likely negative value to an individual. Congested roads, as earlier outlined, increase travel times between locations as free flow speeds are decreased. However, as there are not that large scheduling changes and congestion occurs, this means that the timing to get to work is highly valued. Values for being late/early are also measured in multiple studies and they yield high values of travel time. What people value is punctuality. Highest value is placed on the time travelled while working. This is directly connected to the individuals wage, and therefore tightly follows the median hourly wage.

Individually travel time variation is observed through both stated preferences and revealed preferences studies. In stated preferences studies people are asked how much they would value certain trip or decrease on wait time. Stated preferences often have biases, but they are useful, when there are no usable data sets available. Revealed preferences studies the actual observed change in behaviour when there are changes. Revealed preferences are calculated in things like changes in wages, or real-estate values. In the Abrantes & Wardman (2011) analysis, stated preferences studies are very widely used in the UK, covering 96% of the studies in 2005 -2008 period. Stated preferences studies are easier to conduct, but when compared to revealed preferences, they often giver higher estimates.

Income-elasticity of VTTS is the change in the valuation of time when income changes. Income elasticity is crucial, because if the elasticity is more or less than one, then it has wide implications to infrastructure projects. This is because in infrastructure all values are discounted over long time periods like 30 or 50 years, and therefore just small changes around income elasticity of one generate widely different results. Fosgerau (2005) studies Danish income elasticities, and notes that there is a difference when studying before and after-tax incomes. He finds that when after tax-income is used, VTTS income elasticity is much closer to unit elasticity than when before income tax is used. Abrantes & Wardman (2011, 10) also reached a high 0,89 elasticity of travel time and seasonally adjusted GDP.

2.7 Labour market and the traffic system

Gibbons & Machin (2006) outline three effects of transport to labour supply. They are the demand, supply and matching of labour. Demand for labour is generated through better accessibility and saved time in the transport system. More workers can apply for work in a certain area or vice versa, workers can access a wider pool of jobs. Economic development and firm reallocation to more accessible areas that attract more work force create new jobs and demand for labour. Third labour effect is matching, which means that it is easier to demand and supply to meet, or match. This happens through agglomeration, which means increased economic activity in a more accessible or dense economic network.

The first and probably the most important insight from time values and traffic comes from Becker (1965), where the value of time was seen more than just one's hourly salary. This paper was already studied in 2.7 and the intuition is still the same. Additional work such as Evans (1972) and DeSerpa (1971) also expand the idea to more heterogenous time values across various travel types and purposes.

When an employee is looking for work, they will not accept a job under a certain minimum reservation wage. Having excessive transportation costs increases this reservation wage. In labour economics that focus on the decision to participate in the labour force, a major contribution has been made in modelling the effect of travel costs to labour markets. For example, Manning & Petrongolo (2017), Timothy & Wheaton (2001) ja Gutierez & van Ommeren (2010) model the labour market through travel costs.

Timothy & Wheaton (2001) shows in his paper that through linear modelling that people working in similar jobs even out their work -related costs through housing prices. Gutierez & van Ommeren (2010) show, that higher travelling costs increase daily working hours and reduce the days worked. Their results on the supply of labour were not conclusive. Manning & Petrongolo (2017) model working zones as overlapping areas, in which unemployed take in account the travel costs and competition for jobs when they apply. The model shows a strong locality of employment. So, transportation system improvements have only modest effects. This of course depends on how local the labour market is.

The dynamics of productivity, firm location, and the decision to accept work are subjects that are hard to model empirically to find causal relationships. There are empirical studies of road networks effects on the location of firms and the labour supply.

Chandra and Thompson (2000) study United States counties that happened to receive a highway, when one was built to accommodate two larger areas. They find that firms located closer to the highway and brought jobs with them. This was, however, on the expense of other counties that did not benefit from the highway. Holl (2004) also discovered, that between 1990 and 1984 Spanish firms' location was greatly affected by the road infrastructure. This benefit was also borne by the areas further away from the road infrastructure.

Gibbons (2019) discovered, that in London the expansion of highways and creation of bypasses generated new economic activity by introducing new companies into the area. Employment in the existing firms did not increase. The article also cannot state that the overall economic benefit increased, even if it did so locally.

2.8 Agglomeration

There is a notable difference in the wages, housing prices and general economic activity between cities and other areas. Economists explain this effect through agglomeration. Agglomeration is the increased economic activity through the interconnectedness, accessibility and interaction that is involved in businesses working close to each other.

Duranton & Puga (2004) divide the theory of agglomeration into three channels, sharing, matching, and learning. Sharing means, that some high fixed costs are better to be shared between companies. Matching is the increased effectiveness of the markets. Labour markets as well. Employees and companies also learn more effectively in a more interconnected and interacting environment, making both the companies and their employees more valuable.

This is the reason why land values, wages and economic activity often increases with density. Agglomeration requires an effective transportation system to enable an efficient movement of people, though. For this reason, agglomeration and transportation are interconnected. Better transportation system makes an economic zone more accessible to more people and enables agglomeration effects.

Melo et. al. (2017) study agglomeration through the data on 50 largest United States metropolitan statistical areas. They measure agglomeration with the availability of labour in space and time. They focus on wage level to assess local variation in the productivity of the labour force. Economic accessibility is divided into continuous accessibility zones, from which five groups of cities are recognized. They find that agglomeration has a positive effect on productivity, but that the effect is very local.

Börjesson et al. (2019) study agglomeration effects to productivity in Sweden. They have two measures for agglomeration: travel time between the workers living area and all the local jobs. The second measure is the distance between a worker's current job and all the other jobs in the region. The study forms a regression over wages and changes in agglomeration i.e. travel times. They find that the elasticity between a worker's home area is small, whereas the elasticity to their working area to wages is higher (0.035 – 0.028). This means, that working areas benefit from agglomeration. Residential areas might benefit in other ways, but not in increased wages.

3 DATA AND AREA OF STUDY

3.1 Data and statistics

Data and statistics are terms, that are used interchangeably. However, there is a distinct difference and understanding this difference and how it relates to studying any topic, but in this case especially the economics of traffic is important.

Data is the raw dataset, that is saved from operations, observations or other automatic or manual data collection scheme. Often, datasets are also referred to as being machine-readable.

A statistic is the result from refining data to a number, like and average, sum, or other variables. These numbers are the ones that are reported by statistics agencies and researchers. A statistic distils datasets to numbers that answer questions and depict an event or phenomena.

The reason, why this is brought up is because when statistics are studied to understand the topic of research, the statistics are not just reporting of pure data. Data collection often has intrinsic biases or incompleteness, that is corrected by using statistical methods, earlier larger studies related to the data or other modelling. When using anything else than raw data, acknowledging these shortcomings in the research data is crucial.

In addition to the road statistics, the data in this thesis is used widely in annual statistical surveys to Finnish traffic, such as average speeds on Finnish roads Kiiskilä et. al. (2020). Because data on other modes of transport is proprietary, estimating demand between different times of day and times of week between two cities is estimated using the TMS system. For instance, the corridor between Turku and Helsinki can be well estimated with the TMS data. In a report to the Finnish traffic infrastructure agency about train travel passenger demand forecasting and operation models, time variation in traffic demand from TMS data is used as a proxy for travel demand between the two cities. Pitkänen et. al. (2020, 18)

3.2 Finnish traffic system

Almost 85% of passenger kilometers in Finland are travelled by passenger car *Julkisen liikenteen suoritetilasto 2018* (2020). In 2019, domestic road freight was 26 711-ton kilometers Kotimaan tieliikenteen tavarankuljetukset (2020), whereas on rail it was 10 088-ton kilometers. Rautatietilasto (2020)

However, in the face of climate change and the goal to reduce greenhouse emissions, the Finnish government has set out a goal to reduce annual carbon emission levels from traffic to half of 2005 levels by 2030. Traffic consists of 20 % of Finland's carbon emissions. Because road transport is the most common use of transport, main policy goals are aimed at reducing the emissions of road transport. In addition to promoting long term goals, such as new fuel technologies, more efficient fleet and other improvements in technology, Finnish policy aims to reduce traffic and incentivize other modes of transport. This comes with added investment to rail-infrastructure, one of the projects running along the E18 from Turku to Helsinki. The main tools targeted at road use are vignette-legislation, fuel taxation and the promotion of walking and cycling on short trips. (Valtioneuvosto, 2019)

Finland is also forming a 12-year national transport system plan. The plan is a 12-year program of for municipalities and the national government to follow and the projected financing of the transportation system. The program is currently being negotiated between all interested parties. First 12-year period is planned to set between years 2021 and 2032. Laki liikennejärjestelmästä ja maanteistä (2005)

Finnish population and urban density are focused to southern Finland and around the capital. 72 % of Finnish people live in urban areas. Over 2 million people live and over 1,1 million jobs are in inner urban areas. Syke (2020) This density is also illustrated in figure. 7.

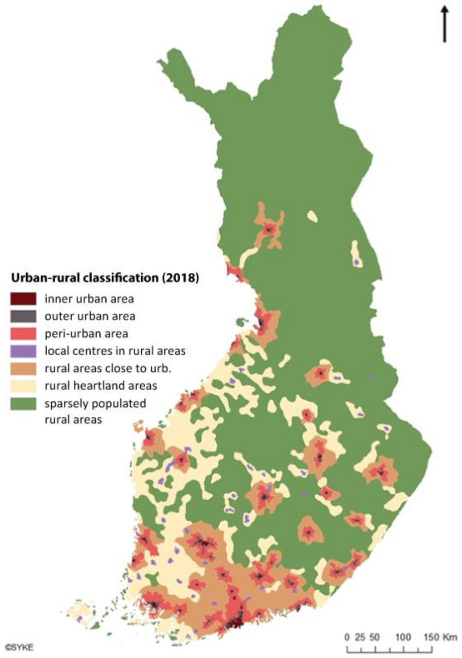
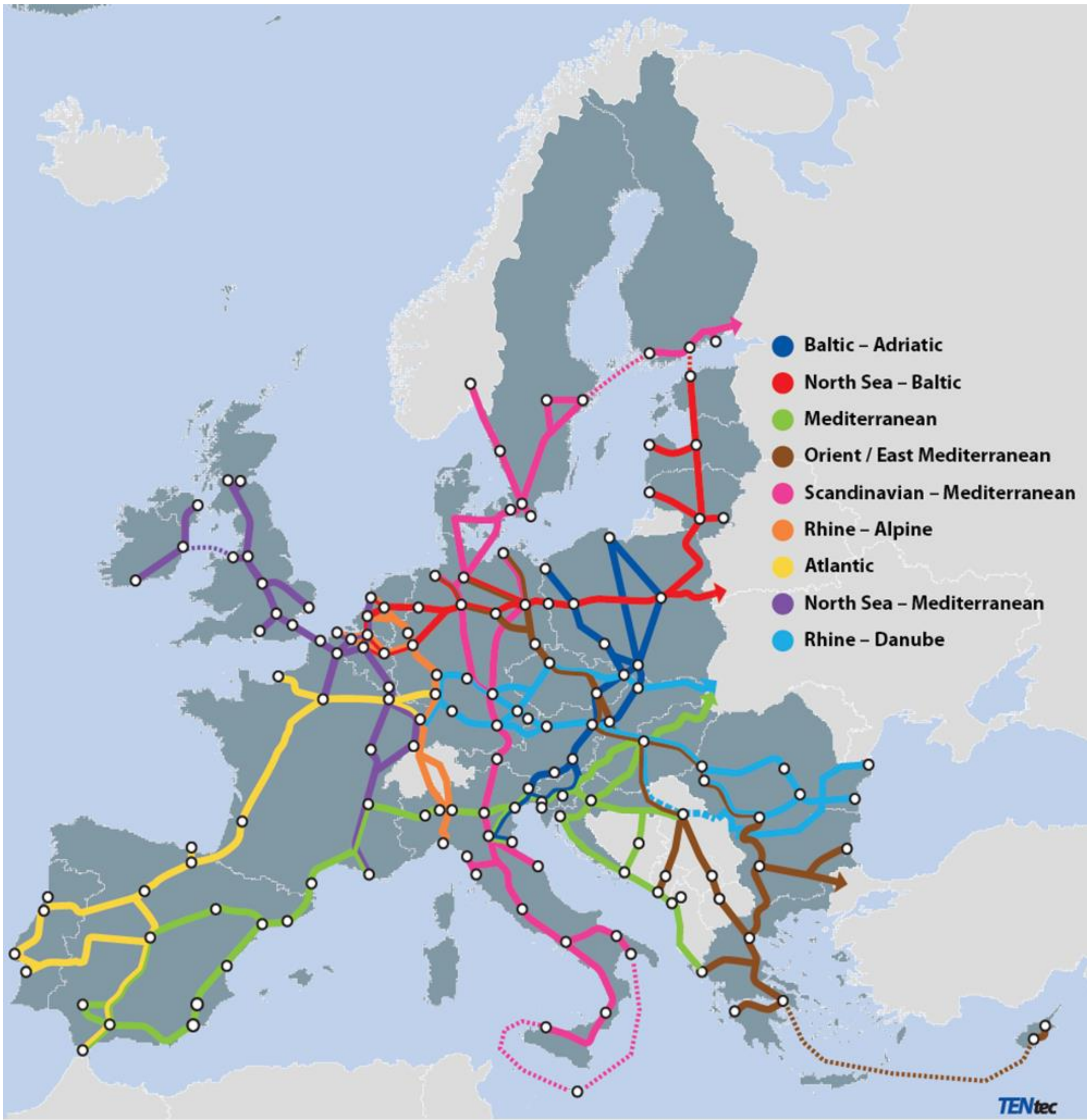


Figure 5 Urban-rural classification SYKE (2020)

3.3 TEN-T network and the E18

The trans-European transport network, or TEN-T in short, is a policy to advance the EU's vision of connected and internationally competitive Europe. The network is comprised of a core network and a comprehensive network. The core network is a collection of the most important collections within the EU, and it is planned to finish in 2030. The comprehensive network is planned to finish in 2050 and includes all European regions. TEN-T policy outlines the goals and projects to reach the goal of a more interconnected Europe. The core network is envisioned in nine corridors, and Finland is part of the Scandinavian – Mediterranean network. FTIA (2020)



Note: the nine TEN-T core network corridors are based on the CEF and TEN-T Regulations (1316/2013 & 1315/2013); they have been created as a coordination instrument to facilitate the completion of major parts of the core network of strategic importance.
Source: European Commission, Directorate-General for Mobility and Transport, TENtec Information System

Figure 6 TEN-T network (FTIA's website, 2020)

These networks are multimodal, meaning that the corridors combine sea, air, rail, and road networks. In the context of this thesis, the road network is the main point of interest. In Finland, the highways part of the E18 is a part of this network.

The E18 is an extensive highway system beginning in Ireland and ending in Eastern Finland. The Finnish part is the longest highway in Finland. It runs through the southern Finland, beginning in the west at Naantali Harbour, running through Turku and Helsinki ring roads and ending at the Russian border in Vaalimaa.

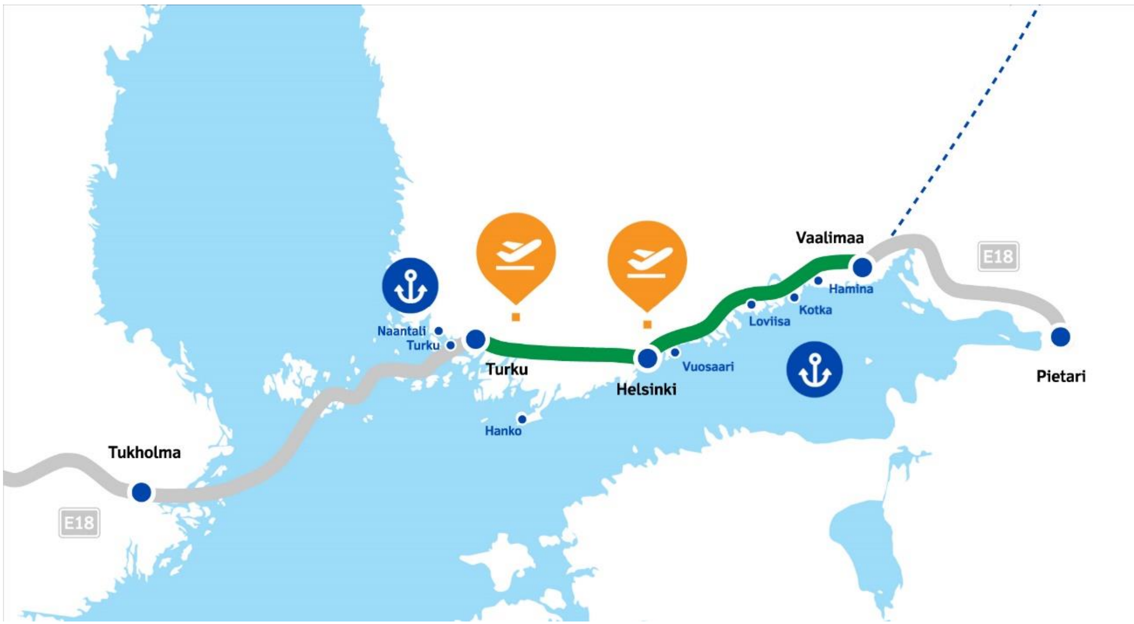


Figure 7 E18 in Finland (FTIA’s website, 2020)

50% of the Finnish GDP is in the vicinity of the highway. It is the longest highway-rated road in Finland, stretching for 344 kilometers. FTIA (2018)

The E18 has seen several improvements from the upgrades to highway one between Turku and Helsinki in the early 2000’s and later in the 2010’s the highway between Loviisa and Vaalimaa was elevated to a four-lane highway. A project evaluation to the Muurla – Lohja section was conducted in 2014. The cost-benefit evaluation concludes that the project had a cost to benefit ratio of 2.3-2.3. Legally all projects should have a ratio of 1.0 or above, so the E18 has been evaluated as a successful public infrastructure investment. There are no economic evaluations on the newer infrastructure projects on the E18 because the evaluation is done after some time has passed to evaluate the outcomes in the long term.

3.4 Traffic measurement system, TMS

The traffic data that both inspired this thesis and that works as the main source of study is the TMS-traffic data, that is hosted as open data by the Finnish Transport Infrastructure Agency.

TMS stands for traffic measurement system, and they are a set of 500 induction loops under the Finnish road system. These loops detect the metallic mass of a vehicle and can detect every vehicle travelling on that road. The location of these measurement stations is mainly in southern Finland. The loops can detect multiple characteristics from the passes, and for every vehicle the system logs the length, lane, direction, vehicle-class, and speed. The vehicle class is divided into 7 different categories. These loops are like the ones described in chapter two.

The TMS service is hosted by Fintraffic Oy. Fintraffic is a government owned company that provides traffic control services for all modes of traffic. These traffic control services include the TMS-system. They are also responsible for developing and providing the data from these services.

Fintraffic also provides insights for real time traffic monitoring through the TMS-system, such as 5min sliding averages of free flow speeds, traffic flow indicators and 1 hr averages of vehicle speed. Finnish Transport Infrastructure agency collects and hosts a database of csv-files and excel files that have all the available historical raw data. This data set is remarkable, because the oldest TMS-locations are from 1995. This means, that studies that utilize the historical models can be refined to models that provide economical estimates on a monthly or even weekly basis.

The potential in this real-time data has been utilized in Fornaro & Luomaranta (2020), where TMS data was used to generate a machine learning forecasting model to predict Finnish economic growth. They provide an extremely fast and reliable estimates to the trend indicator of output and quarterly GPD. This model was also utilized in forecasting the effects of the coronavirus to Finland’s GDP. Lindroos (2020)

For traffic economics, this dataset can be leveraged to metrics that are used in many of the studies in chapter 2. The dataset provides an estimate for traffic demand reaching as far as 1995. This traffic demand can be studied within different subsections. Passenger cars, motorcycles, heavy traffic all can be separated and studied individually. This distinction can be important for studying elasticities of traffic demand, such as Börjessön (2019) uses in studying the effects of road tolls as well as in the forecasting model above.

Because the raw dataset provides every car passing individually timestamped, there is great potential in studying how traffic demand behaves and how the behaviour develops over time. This means, that where congestion studies have approximated the occurrence of traffic or had to use only partial measurement times, this dataset can be leveraged to utilize any timeframe.

In addition to being used in modelling, this dataset is being utilized in official statistics in Finland. The Finnish Transport Infrastructure Agency provides two types of data, automatic TMS data, and a general traffic count, which is provided by a private contractor. TMS-data provides the demand estimate for the road it is measuring. The general traffic count is an annual census of Finnish roads, where the contractor measures an average daily traffic from 3200-3600 locations around Finland. This location sample is planned by the contractor and approved by the Infrastructure Agency. These two measurements offer the estimate of traffic demand on roads in Finland. These two data sets are part of a larger data collection called the road registry, which is available for consultants to browse and utilize for modelling or planning Kiiskilä (2016). The Finnish Traffic and Communications Agency Traficom provides the tangible kilometre demand from inspection data that is reported to Traficom by inspection companies. These two datasets are edited with statistical methods that correct for the discrepancies in the data. The result is the national traffic demand by road type and location Statistics Finland (2020).

3.5 Practical utilization of the TMS system

Parts of the TMS system are widely utilized by the Finnish authorities while measuring, analysing, and predicting the Finnish traffic system. This section aims to outline the usage and methodologies used to analyse TMS data or similar data sets in Finland and how the road system is analysed in economic. The Finnish transport agencies had their operations reorganized and some of the operations between the Finnish Traffic and Infrastructure agency (FTIA) and Fintraffic are different from what they were before.

The Institutional Repository contains all publications from all Finnish ministries from the year 2016 onwards. Some publications from before 2016 are also included. In 2016, Särkkä et. al. writes about the future transport models and forecasts. They reference the road registry, but not TMS network directly. They however mention the future use of big data. The raw sensor data that TMS network collects fits many big data classifications. Salonen et. al. (2018, 11) write about the available data and ongoing projects in the private sector to provide data-driven services and applications. Sensor data like TMS-network is shown to be interesting, and private companies utilize similar approaches through vehicle-implanted sensors to utilize similar data that the TMS-system produces. For this reason, basic research to the dynamics around the TMS-data and more timely and local data sources is called for. The TMS system is part of the digital road infrastructure, which is crucial for autonomous driving. Laakso, Vesanto and Ritari (2018) analyze the entire Digitraffic system and conclude that the digital platform is important to autonomous driving. They also bring out the problem with missing data because the measuring stations occasionally go offline, causing gaps.

Väylävirasto, Finnish Transport Infrastructure Agency or FTIA, is responsible for the Finnish infrastructure system. They also conduct multiple technical research projects, which also require traffic demand estimations. They do this, because by Finnish law, all government-funded infrastructure projects must have their economic impacts studied and FTIA is the responsible agency to both produce the methodology and parameters for this assessment. The documents are more technical and used to support the planning and execution of infrastructure construction. FTIA has produced statistics about the traffic system, but this has moved over to Statistics Finland starting from 2019. FTIA also has predictions and technical base-values about the traffic system.

FTIA has studied novel ways to measure traffic. A report in 2017, titled “Collection and analysis of mobility data” by Niittylä and Varjola studied the feasibility of utilizing the sensor data from smartphones to study individual mobility patterns. The approach was novel within Finnish authorities and the results were promising. Authors and a company developed a phone app, in which authorities could recognize entire mobility chains from different mode to another. Overall, the report states, that to further use this kind of application, more technical, communication and user related issues need to be resolved. In 2019 FTIA studied the development of Finnish road traffic measurements in the report “Preliminary study on the development of general traffic counting” by Laine et al. This specifically targets the TMS system and the general traffic counting service and how to develop them. The report suggests scaling up the measurement network in cities and on the main road network. A new addition in the data collection infrastructure is cellular network data from mobile phone operators to further understand the movement on the road network. These insights can be utilized to further develop the models in calculating the statistics. With these additions mathematical models to extend the knowledge base in traffic demand to daily average traffic information.

The Government's analysis, assessment, and research activities, which are the government and ministries ordering research to study the feasibility and economic impact of planned and enacted policies. This program started in 2015, and here are the road traffic related studies and their results. In 2017, a study looked at the changes in average speeds and speeding after the minimum fines were increased in 2015. Aaltonen & Virtanen (2017). The study utilized TMS data. The researchers found that there is some indication that increasing fines decreases speeding. This is also supported by international studies cited in the study. But because some TMS-locations are also automatic police traffic cameras, further studies are suggested and that they utilize the different stations more effectively. Also, variation in police surveillance and later the decrease in the speeds where police are required to fine a driver changed close to the increasing of the fine, the results are not conclusive. However, this is a very widely used feature of the TMS-system, and studying speeding is quite easy with automated traffic monitoring.

3.6 Research on the Finnish road system

This section outlines studies and reports, which study the Finnish traffic system. Finnish traffic authorities mostly do this research. The studies are often studying multiple modes of transport, but since this thesis focuses on road transport, the reports are cited only on the parts about passenger road transport.

A 2015 report from The Institutional Repository "Development of the Parikkala border crossing point. Traffic forecast" by Kukkonen & Tuominen goes through the economic development around the border crossing. The study draws a very direct line between local economic development in the Republic of Karelia and border crossings. Like the studies from Börjesson, HSL has produced a study on price elasticities of public transport and feasibility studies of road pricing. HSL calculates both long- and short-term elasticities. This is an update to an earlier elasticity study done in 2014. HSL utilizes a price adjustment in 2009 in a 2005-2013 time series. Both logit and time series models are used, and time series models are more representative. Short term price elasticities in value and season tickets are between -0.3 and -0.4. Single ticket elasticity was -0.5. Long term seasonal and long-term tickets saw the elasticity double and single tickets tripled.

In a similar matter FTIA also produces unit values for project evaluations. These include value of travel time savings for vehicles and elasticities for trains. The train demand elasticities are based on the Swedish traffic model and the elasticities derived from it.

FTIA publications also include a series on traffic modelling. The evaluation of the National traffic-prediction model (Moilanen et. al. 2014) is a technical report that outlines the Finnish traffic-prediction models. Authors emphasize, that traffic has grown with GDP, and that in the future multiple factors will show divergence from this trend and more diverse forecasts become more important. Moilanen & Niinikoski continue this work on traffic prediction models in 2017, mapping distinct options for the future of traffic prediction models. This report, however, does not cover the practical execution for the models, and therefore there is also no discussion related to the data sources and their use. In 2017, Moilanen & Niinikoski also wrote the “Developing the calculation of road transport performance.” -report. This is a comprehensive report on various Finnish traffic measuring systems, models, and statistics. It recognizes 6 different statistics, which are the official road statistic, public traffic performance statistic, transport of goods by vehicles, the national travel survey, national traffic emission model and the national vehicle stock. Next in this series of reports Lapp et al. (2018) develop new traffic forecasts of road transport, passenger and freight rail transport and maritime transport volumes. These forecasts are long-term, covering a period until 2050. For the first time in 2017 all modes of transport were forecasted together. Road transport forecasts are based on GDP and population growth projections and forecasts the statistic from the national road statistic. The possibility to differentiate different vehicles from the roads are utilized within the most driven road networks. Traffic intensity between different times of day and rush hours or their forecasts are not explored.

In 2018, the microsimulation model used in public sector called “SISU” was modified to predict changes is vehicle and fuel taxation. Hellman & Lappo (2018). The study did not use TMS data, but Statistics Finland’s statistics on travel demand was used to estimate fuel usage. This auxiliary use is informative, because the developments in the dataset and its accuracy affect simulation models that base traffic demand to them. The traffic component of the SISU-model is used to study the impact of fuel and car taxation. Another study regarding fuel tax subsidies in 2020, but it did not utilize traffic demand at all. Honkatukia et. al. (2020) The report was modelling the changes in the macroeconomy, when diesel-taxes were changed. The model used is a widely utilized dynamic stochastic general equilibrium, or DSGE-model REFINAGE. A DSGE-model is internationally the most widely used policy tool to assess the long-term policy decisions in the economy. The models have national differences. Honkatukia et. al. (2020) This model does not utilize traffic measurements as inputs.

Helsinki region transport (HSL) is the largest regional transport authority. HSL is responsible for planning the Helsinki region transport system and public transportation. Helsinki region is overwhelmingly largest public transit provider in Finland in terms of public transport ridership and investment, therefore it also has resources to study traffic in more depth, than other municipal traffic administrators. HSL utilises TMS data in long term land use planning. HSL also monitors traffic in the Helsinki region and TMS data is utilized on the ring roads and main arterials in the area.

Another traffic modelling system is the HELMET model used by HSL. HELMET models the Helsinki region traffic and generates multiple traffic predictions. In 2020 the demand models were redeveloped entirely, and they now forecast round trips or tours in the Helsinki region. This collection of traffic demand models are logit models. Their inputs are population and land use data as well as travel time, distance, and costs. The last three of these data sets are generated with a traffic assignment software so they are not actual measured results. The model divides the generated tours into seven categories, that are indicative of the national travel survey questionnaire. The model has four modes of travel: walking, cycling, public transport and car. Car is further divided to passenger and driver. HSL (2020)

FTIA also studies the wider economic impacts of the traffic system, and in 2016, Laakso, Kostiainen and Metsäranta published a report “The wider economic impacts of transport projects – Preliminary report.” The report highlights how infrastructure projects are assessed in a wider economic framework, recent literature on the subject and how other countries assess these wider economic impacts. The authors also go through, how Finland should develop the domestic assessment framework. The main finding is that there is no widely adopted framework to assess traffic infrastructure investments, and that both the data and methodology should be further developed in a more centrally planned way. In 2019 Metsäranta et. al. wrote the next installation to this series of reports called “Impacts of road and rail projects on real estate values and municipal contribution: Investigation from the “beneficiary pays” viewpoint”. Authors highlight that the economic benefits of land use stem from the accessibility of an area when the transportation infrastructure is improved. Accessibility also means more demand for the infrastructure. When studying the economic impact of land use, forementioned traffic forecasts to estimate future traffic demand are used. In practice this means, that the traffic is generated either as new trips or transferred trips from elsewhere on the road network. A concrete example of economic value assessment is The Ex-post impact assessment of the Tampere western ring road. FTIA (2019). This report directly utilizes TMS-data to compare actual traffic after the project has finished to projected scenarios in the development phase.

In 2019, Transport and communications ministry in Finland has begun the process to better understand the wider economic impacts of traffic investments. In 2020, multiple reports were developed for this endeavour. The framework for examining economic impacts of transport system development LVM (2020) outlines the scope, terminology and overall methods that model regional economies and how they connect to the traffic system. Methods of regional economics in impact assessment of the transport system development Hokkanen et. al. (2020) extends on what models are available to this evaluation. The report concludes that existing capabilities enable numerical modelling of WEI. The authors also do a test calculation on the E18 highway investment project. The results were calculated from a starting point in 2008 and had three outcomes. One was the expected result without the road project (base scenario). Second a policy scenario, where the effects were evaluated with a model, and a third outcome, which is the researched actual economic situation in the area in 2014.

The model suggests a 1,6-1,9 million euro decrease from 386-1534 million in spending to fuel, but because the model does not account for increases in traffic demand, the actual demand for fuels increased from 14 to 20 million euros. The model also evaluates the increased value add of companies, showing that most of the benefits come from relocation instead of new business. Effects on labour markets are most notable, and the project is shown to have had positive effects. The report concludes that these regional economic models are suitable for the study of WEI.

The Impacts of Transport System Changes on Labour Market and Their Evaluation, Metsäranta et. al. (2019) conclude that it is difficult to evaluate labour market outcomes from infrastructure investments and the few countries that include labour market outcomes set a high bar to the interpretations. In this report, all modes of transport are analysed, and a test study is conducted.

The test study analyses a road expansion project in Lahti. After a brief history on the highway and the connection between Helsinki and Lahti, four possible ways of conducting the study are briefly evaluated. Main issue is the lack of spatial data in a fine enough resolution to give the models more statistical strength. The only gets to use municipal data and cannot evaluate crowding out effects of the project on other areas. Additionally, individual panel data is not available, which makes labour market outcomes harder to assess because selection bias is hard to rule out.

Results from this test study have very wide margins of error, but the test manages to show outcomes to key statistics like population. Because Statistics Finland has more accurate data this approach can be made much more accurate and therefore the report concludes that there are opportunities to better study and evaluate infrastructure projects and their labour outcomes.

3.7 Conclusion from the available literature

The available literature outlines that one cost component of travel is the time taken to travel and this time has a monetary value. These heterogenous valuations of travel time can be utilized to fix congestion issues through road pricing in stead of waiting in traffic. Travel time savings are also a big component in infrastructure cost- benefit analyses conducted around the world. The added productivity of workers and firms through better infrastructure has been widely studied through the labour market.

Labour market and traffic infrastructure have been an interest of study both abroad in Finland. However, the benefits of infrastructure investments to different employment outcomes have been inconclusive. However, differences in travel patterns have been observed based on different variables, such as income, labour market outcomes and living arrangements. Datasets available for these studies varies, but the signal data from individual passes on a network seems to be a very common dataset available to researchers, The data used in this thesis has been used for economic impact analysis in Finland and similar datasets have been used for empirical studies abroad.

4 DATA COLLECTION AND DESCRIPTION

4.1 TMS-data

The TMS-dataset is gathered by the Finnish Traffic and Infrastructure Agency FTIA. The FTIA utilizes measuring stations that measure passing vehicles on highly trafficked Finnish roads. The measuring stations are embedded coils in the road, that can measure passing traffic very accurately. There are currently around 500 measuring stations around Finland, majority of which are in southern Finland. The thesis only uses the 72 measurement points on the E18. The data structure is outlined in table 1. This is not the original data set, but a modified version to make the computational working easier. This data was compressed from around 500gb to 5gb of data and is organized into yearly data sets in csv files.

The dataset is divided into directories by each year by FTIA. These folders have subdirectories for each administrational area. The E18 crosses three areas. These folders contain one file for each measurement point for each day of the year. A R-script was written to download each file to a disk as raw data. Then another script was written to first only take the points on the E18 and select a subset of information from these files and write it into a new file. This made a folder of 30 000 files and 39 Gbit into a file sized at 283 Mbit. This operation was looped over every folder, making the dataset for validation 8 Gbit.

The first analysis for these files is to check for integrity. Measuring stations go offline and get scheduled downtime for maintenance. For this reason, not all stations may have enough days of data to form a feasible data point for analysis. A simple analysis of count of days that contain data was carried out. Table 1 shows the points and their relevant statistics for the selected stations. Out of the 72 stations, 37 have data in each year. Some stations are commissioned after 2010, and some have been decommissioned or have been out of operation for a year or more. Out of these 37, six more are dropped due to having a low number of days in a year. All the 31 selected stations have over 200 days of data. On average all these stations had a minimum on 320 days and at best over 360 days of data.

Table 1 Selected TMS's, days that the stations have data and their location.

	min n of days	max n of days	average	Municipality
lam 1	226	366	334	Porvoo
lam 104	328	366	350	Vihti
lam 112	320	366	357	Porvoo
lam 128	257	365	345	Vantaa
lam 139	303	366	349	Espoo
lam 141	357	366	364	Sipoo
lam 150	350	365	359	Vantaa
lam 160	257	365	346	Vantaa
lam 168	355	365	363	Vantaa
lam 169	286	366	345	Vantaa
lam 191	269	365	346	Kirkkonummi
lam 192	324	365	348	Kirkkonummi
lam 193	301	366	344	Lohja
lam 194	319	366	355	Lohja
lam 227	252	366	336	Kaarina
lam 231	341	366	360	Raisio
lam 233	334	366	361	Raisio
lam 242	234	366	322	Paimio
lam 243	261	366	350	Paimio
lam 244	261	366	350	Paimio
lam 247	221	366	348	Salo
lam 248	251	366	333	Salo
lam 251	309	366	357	Raisio
lam 572	304	366	351	Kotka
lam 1601	349	363	360	Salo
lam 1602	352	367	359	Salo
lam 1603	344	363	357	Salo
lam 1604	352	363	359	Salo
lam 1605	328	367	355	Lohja
lam 1606	351	366	360	Lohja

lam 1607	340	367	359	Lohja

4.2 Filtering and determining leisure/work times.

After finding the points that are good enough for analysis, the dataset is ready to be transformed from annual files into a single dataset. This operation also selects only the columns that are needed and selects personal vehicles. This dataset also only uses hourly intervals.

Next, the day column is used to write a date -variable. A table with all the Finnish national holidays is used to select days, that are full leisure days. From the remaining days, the hours between 7 and 18 are determined as work times. This leisure/work-time division is represented as a classification variable. These hours are similar to Finnish traffic models used by HSL and FTIA. The peak traffic is also between these hours (see figure 8 in 4.4).

Finally, the dataset is run into annual data for each municipality, calculating the average passes for each hour and average passes per hour for leisure and work times. Municipal economic variables from are then added as columns to the dataset.

4.3 Economic indicators

This study only measures the feasibility and applications of traffic data. To widely analyze all the measuring stations with various combinations, the economic indicators used are very basic annual statistics. These statistics were chosen to measure the connection between traffic and the economy on three different axes, population, employment, and land values. These variables are akin to the Finnish reports and studies that utilized the same data.

The use of annual data has drawbacks. Mainly, as the time series is only ten years, regressions have a wide error margin. Problems with annual data are further discussed in 4.6 about the model.

There are five municipal economic variables in total. They are average wages (eur), price of old apartments (eur / square meter), population (number of people), employment rate (ratio of employed persons to the population of the same age) and workplace self-sufficiency. Self-sufficiency means the ratio between the number of labour force living in the municipality and the amount of people working in the municipality. Tilastokeskus (2023) When connected to the traffic data, the final data set looks like this.

Table 2 Variables in the final dataset

Class variables	Analysis variables
Year (2010 – 2019)	municipality population
Municipality (11 municipalities)	municipality employment rate
	municipality workplace self-sufficiency
	municipality price of old apartments
	municipality average wage income
	municipality hourly passes (total)
	municipality hourly passes (leisure)
	municipality hourly passes (work)

The differences in municipality size are massive. Vantaa and Espoo are among the largest cities in Finland with a population of 200 000 and above, but Paimio only has a population of 10 000. During the time of the study, Espoo grew more than 6 other municipalities even have people. At the same time, the traffic averages aren’t that much different. Paimio and Porvoo have a similar size work-time traffic averages even though Porvoo is five times the population of Paimio. Apartment prices also vary widely in different areas. Different municipalities also have very different housing profiles, and the data only shows apartments but not single-family homes.

4.4 Descriptive statistics

Here are some basic descriptive statistics of the dataset that’s used in the analysis. On average, there are 947 passes per hour. During working hours on working days, the average is 1547 passes and leisure time has on average 704 passes.

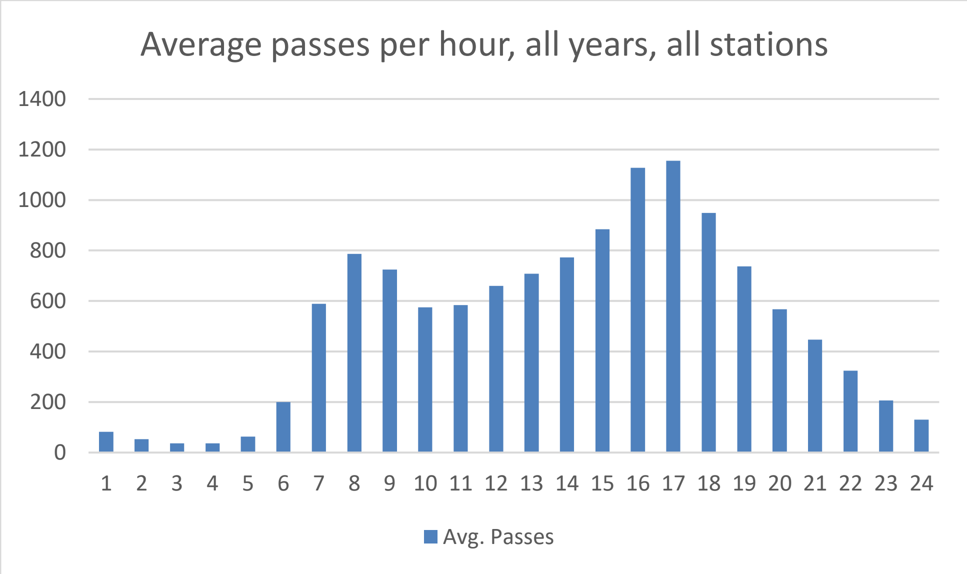


Figure 8 Average hourly passes across all stations and years

Figure 8 shows that the hours between 7 and 18 are the high traffic times. The low hourly averages when compared to the average of working hour averages above comes from the difference in averages run. The graph has weekends and holidays in it. This makes the difference between working time and leisure time traffic less profound.

Economic data and traffic data is presented by municipality. Every municipality has a table of the economic variables, a line graph presenting the traffic data and a map showing the population density, railway, subway, and TEN-T-network roads. The subway is marked in yellow and the railways in teal. Purple lines represent the TEN-T network and the numbers without frames are the road numbers. The purple dots with the framed numbers are the selected TMS-stations. The population density is from Statistics Finland's geodata-API and the population density is a 1x1 kilometre grid divided from light to dark yellow, then from orange dark orange and finally to purple as follows: 1-4, 5-19, 20-49, 50-499, 500-4999, 5000-.

4.4.1 Espoo

Table 3 Economic variables for Espoo

Year	Municipality	Population	Employment	Self-sufficiency	Real-estate prices	Average income
10	espoo	247 970	76	96	3 051	55 137
11	espoo	252 439	76	97	3 150	54 884
12	espoo	256 824	76	96	3 257	54 305
13	espoo	260 753	74	96	3 308	53 568
14	espoo	265 543	74	94	3 280	52 727
15	espoo	269 802	73	93	3 288	53 686
16	espoo	274 583	73	92	3 360	53 605
17	espoo	279 044	74	92	3 448	53 623
18	espoo	283 632	75	92	3 524	54 060
19	espoo	289 731	75	92	3 597	54 330

Espoo has one valid measuring station on the western side of the municipality. It is away from the population centres and away from the commute from Espoo to Helsinki. The traffic in this measuring station has barely changed. In the late 2010's the are saw higher changes in traffic, but nothing that would coincide with the steady development of the economic variables. This point is still the 6th most travelled station in the selected measuring stations.

Between 2010 and 2019, Espoo saw an increase of 17 % in population and 17 % in real-estate prices. Both employment-rate and workplace self-sufficiency decreased. During the 2010's, Espoo saw the opening of new subway stations (yellow line in the bottom right-hand corner in image x) and increased building in these areas. Espoo is also

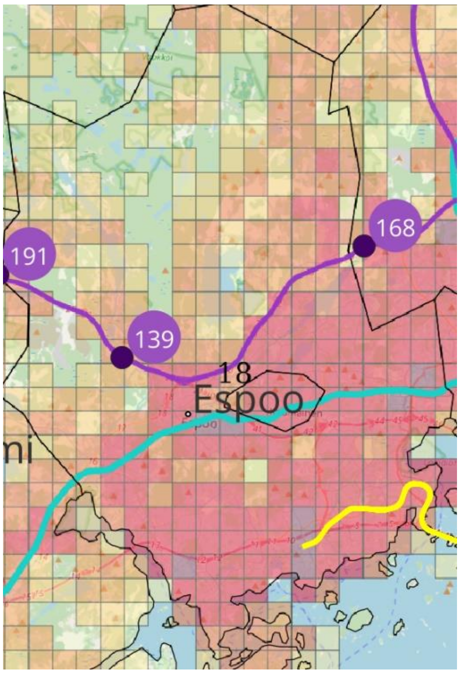


Figure 9 Map of Espoo

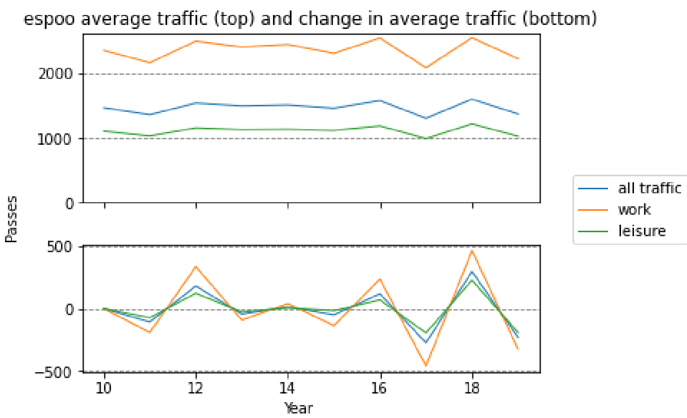


Figure 10 Traffic in Espoo over the study period

served by a train line (light blue). The subway has made Helsinki more accessible and is very likely a contributor to the decrease in workplace self-sufficiency. The roads that lead to Helsinki can be seen on the image in the bottom right corner in the background.

4.4.2 Kaarina

Table 4 Economic variables for Kaarina

Year	Municipality	Population	Employment	Self-sufficiency	Real-estate prices	Average income
10	kaarina	30 911	75	66	1 865	42 405
11	kaarina	31 081	76	66	1 856	42 022
12	kaarina	31 363	76	65	1 820	42 401
13	kaarina	31 798	74	67	1 936	41 009
14	kaarina	32 148	74	67	1 894	40 437
15	kaarina	32 590	74	64	1 946	41 356
16	kaarina	32 738	75	64	1 951	41 563
17	kaarina	33 099	77	64	1 949	41 329
18	kaarina	33 458	79	64	2 049	42 236
19	kaarina	33 937	79	65	2 071	42 843

Kaarina has one measuring station on the eastern side of the municipality. Kaarina does not have railway stations. Again, the E18 is not the commuting route to Turku. Population in Kaarina has developed towards Turku and along the commute corridors. Traffic on this one measuring station is below average but above median.

Kaarina has seen steady population growth of 10 % and prices in real-estate have gone up 11 % during the study period. Employment is higher and has risen 4 %-units closer to 80 %. Kaarina has a much lower workplace self-sufficiency than Espoo. Workplace self-sufficiency has not changed that much during the study period. Average income rose 1 % during the study period.

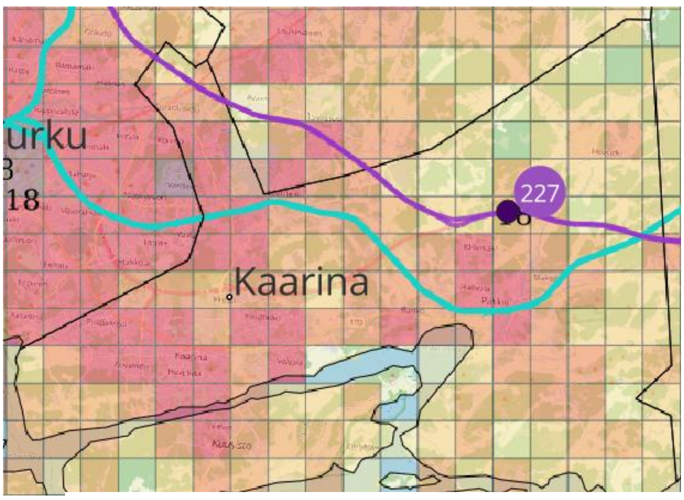


Figure 11 Map of Kaarina

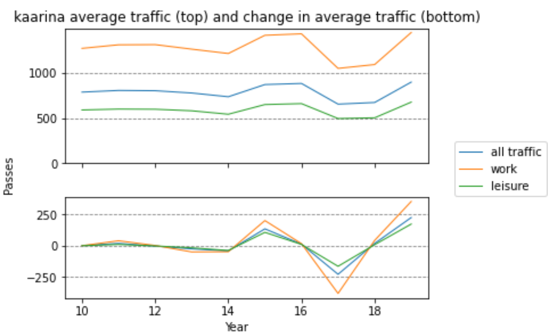


Figure 12 Traffic in Espoo over the study period

4.4.3 Kirkkonummi

Table 5 Economic variables for Kirkkonummi

Year	Municipality	Population	Employment	Self-sufficiency	Real-estate prices	Average income
10	kirkkonummi	36 942	77	61	2 258	55 037
11	kirkkonummi	37 192	78	60	2 327	54 755
12	kirkkonummi	37 567	78	62	2 364	54 254
13	kirkkonummi	37 899	77	61	2 380	53 322
14	kirkkonummi	38 220	76	60	2 421	52 449
15	kirkkonummi	38 649	75	60	2 408	52 630
16	kirkkonummi	39 033	76	60	2 362	52 764
17	kirkkonummi	39 170	77	60	2 439	52 347
18	kirkkonummi	39 262	78	59	2 489	52 293
19	kirkkonummi	39 586	79	59	2 440	52 893

Traffic in Kirkkonummi has grown steadily, but in 2014 there was a lot of variation. The municipality has two measuring stations, that border the northern borders. These points hardly capture the traffic to and from the municipal centre that has a more direct route to Helsinki through a southern highway as shown in the background of image x. Similar situation to Espoo. There's also local commuter train service to Kirkkonummi. Population has centred around these parts, but there's also a small more densely populated area around station 192.

Kirkkonummi has very similar economic and population characteristics to Kaarina, but it has higher real-estate prices. Population growth was 7 % and real-estate prices grew by 8 %. However, the average income decreased by 4 %. Unlike Espoo, Kirkkonummi has a low workplace self-sufficiency. One of the lowest of the study group. This means, that commuting is very common in the municipality but because the E18 is awkwardly located in relation to the population centres, these commute flows are not captured properly.

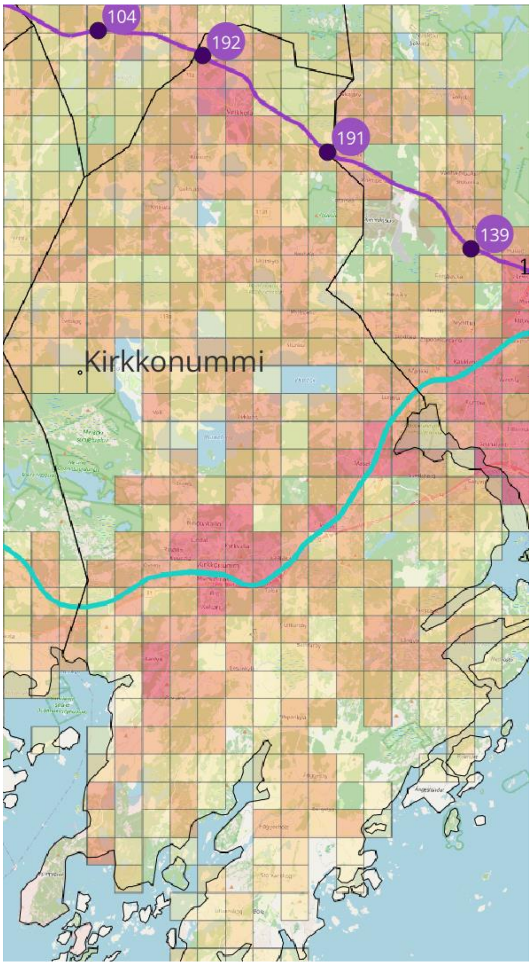


Figure 16 Map of Kirkkonummi

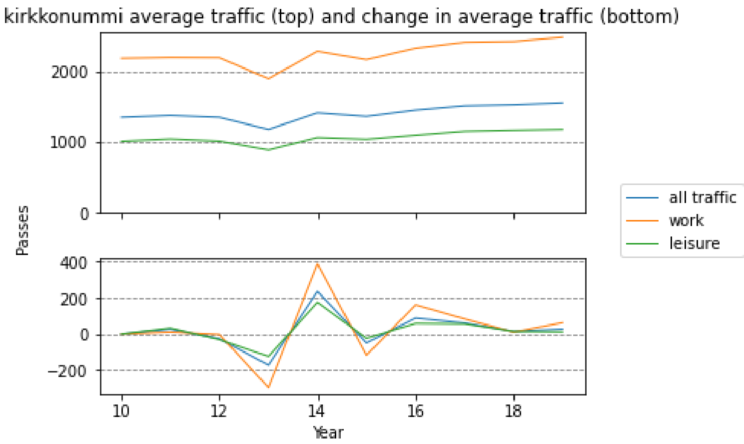


Figure 15 Traffic of Kirkkonummi over the study period

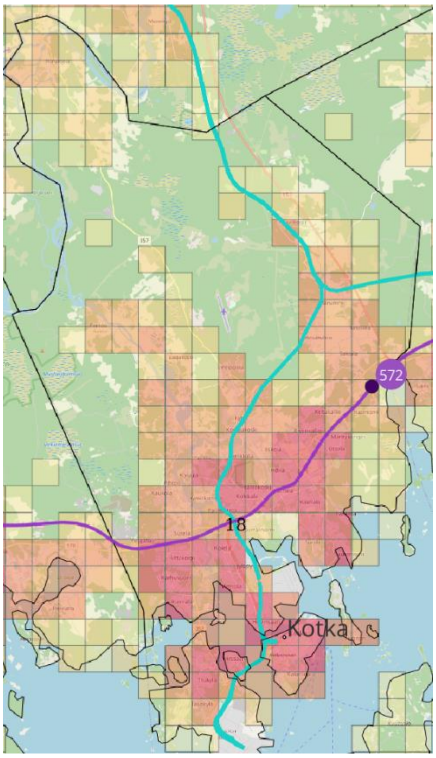


Figure 14 Map of Kotka

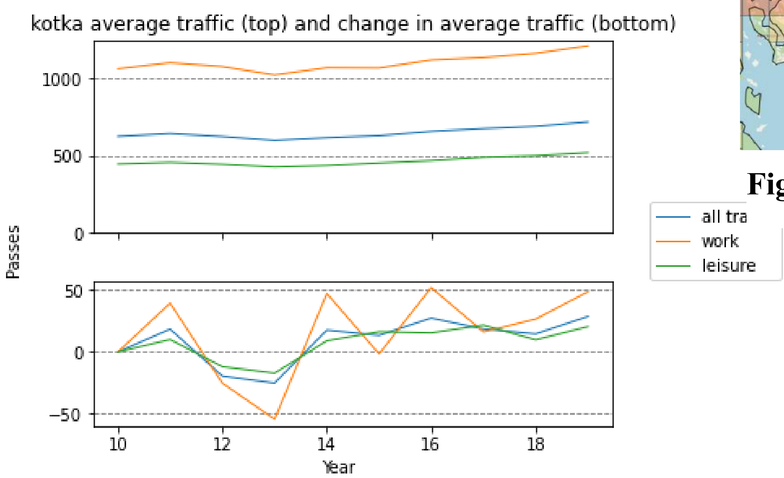


Figure 13 Traffic in Kotka over the study period

4.4.4 Kotka

Table 6 Economic variables for Kotka

Year	Municipality	Population	Employment	Self-sufficiency	Real-estate prices	Average income
10	kotka	54 824	63	110	1 381	28 149
11	kotka	54 831	63	110	1 399	27 918
12	kotka	54 873	62	110	1 456	27 318
13	kotka	54 771	61	110	1 388	26 971
14	kotka	54 518	61	111	1 358	26 549
15	kotka	54 319	59	111	1 341	26 559
16	kotka	54 187	60	111	1 321	26 291
17	kotka	53 539	63	111	1 277	26 297
18	kotka	52 883	65	111	1 261	27 456
19	kotka	52 126	65	112	1 240	27 846

Kotka has seen steady growth in average traffic. This growth is however not uniform. Unlike the traffic in the other municipalities, Kotka's average work time traffic has had very different changes to the total traffic. The eastern measuring station is between Kotka and Hamina, two cities that share a large port. This means, that traffic between the cities is captured very well in the data. Kotka has a railway but no connection to the capital in the west.

Unlike other cities so far, Kotka has experienced population loss of 5 % during the study period. It also has a relatively low employment rate but a workplace surplus that has strengthened during the study period. Low real-estate prices have gone down an additional 10 % during the study period. Kotka has outstanding workplace self-sufficiency. Because the station is located between the two municipalities, it is likely that it captures plenty of commuting traffic.

4.4.5 Lohja

Table 7 Economic variables for Lohja

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	lohja	47 341	73	79	1 844	36 932
11	lohja	47 374	74	79	1 880	36 804
12	lohja	47 516	74	80	1 977	36 758
13	lohja	47 703	72	79	1 939	36 159
14	lohja	47 624	71	80	1 887	35 431
15	lohja	47 353	70	80	1 770	35 530
16	lohja	47 149	72	80	1 862	35 469
17	lohja	46 785	73	81	1 902	35 208
18	lohja	46 296	75	80	1 823	35 501
19	lohja	45 965	75	81	1 861	35 822

Lohja has the second highest number of measuring stations of five on the E18. They are spread evenly across the municipality. The railway that crosses the municipality is not a commute railway and it doesn't have a direct connection to Helsinki. However as discussed earlier, the municipality border and population centre are not the same thing and only one point is on a possible route from the city centre to Helsinki. The city centre can be seen as the densely populated area east of the lake on the south-eastern side of the municipality. Lohja has a good workplace self-sufficiency and the E18 is very well situated to serve the municipality. During the study period, Lohja lost 3 % of its population and real-estate prices remained at a steady level rising only 1 %.

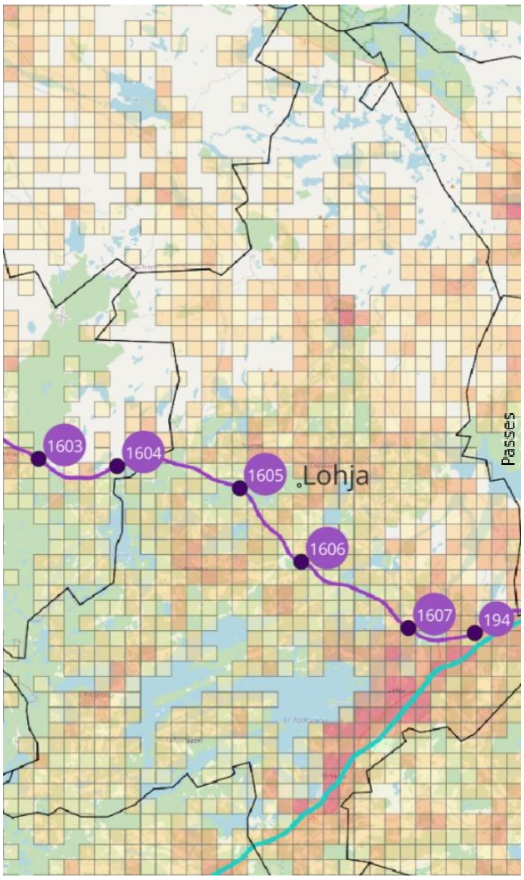


Figure 17 Map of Lohja

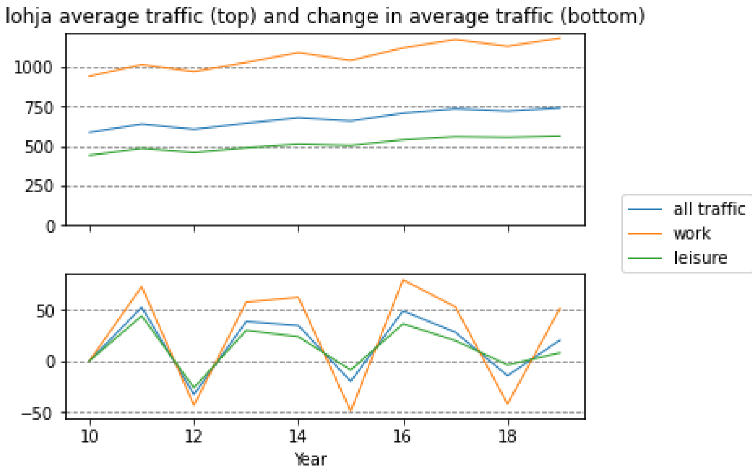


Figure 18 Traffic in Lohja over the study period

4.4.6 Paimio

Table 8 Economic variables for Paimio

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	paimio	10 402	75	75	1 480	36 790
11	paimio	10 471	77	74	1 440	36 906
12	paimio	10 591	75	74	1 465	37 350
13	paimio	10 590	76	75	1 491	36 665
14	paimio	10 628	75	77	1 363	36 030
15	paimio	10 620	75	73	1 388	36 324
16	paimio	10 713	75	68	1 537	36 087
17	paimio	10 730	77	67	1 410	36 047
18	paimio	10 832	80	65	1 413	37 171
19	paimio	10 850	79	65	1 381	37 445

Paimio has two measuring stations spread evenly around the municipality and the E18 goes through the municipality. The municipality is among the best candidates for the E18 to serve as a commuting route. Paimio's population is within a good distance from the E18. The railway that pierces the municipality does not stop in Paimio.

Paimio has seen a population decrease of 4 % and property value decrease of 7 %. Employment is good but Paimio has had a drop of 10 percentage units in its workplace self-sufficiency, most likely driving up commuting from the city. Wages have increased by 2 %.

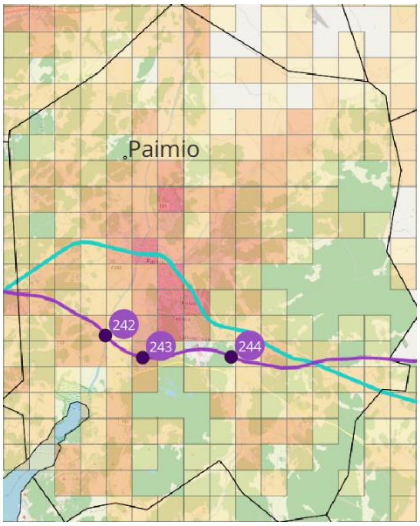


Figure 20 Map of Paimio

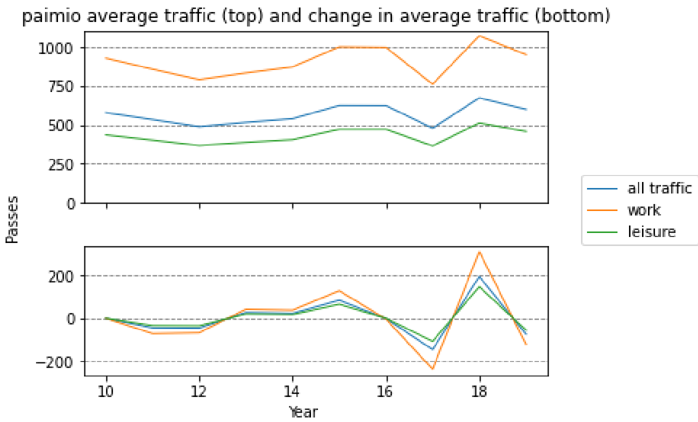


Figure 19 Traffic in Paimio over the study period

4.4.7 Porvoo

Table 9 Economic variables for Porvoo

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	porvoo	48 768	74	90	2 219	41 540
11	porvoo	48 833	74	90	2 387	40 741
12	porvoo	49 028	75	91	2 432	40 793
13	porvoo	49 426	74	92	2 447	40 381
14	porvoo	49 728	73	92	2 412	40 193
15	porvoo	49 928	73	93	2 407	40 591
16	porvoo	50 144	74	93	2 565	40 342
17	porvoo	50 159	75	95	2 506	40 322
18	porvoo	50 262	77	94	2 594	40 578
19	porvoo	50 380	77	95	2 529	41 044

Porvoo has two measuring stations and the E18 goes through the municipality. The railway is a museum railway. Traffic levels have been growing for most of the measuring time. The points have a hude difference in the traffic averages. 112 has almost a double overall average over the station number 1. Helsinki has its pull.

Porvoo has seen a population increase of 3 % and real-estate price increase of 14 %. Workplace self-sufficiency is very good, and the employment rate is ok. Porvoo has an income level in between of the municipalities around Turku and Helsinki. Average incomes have stayed steady in the 2010's.

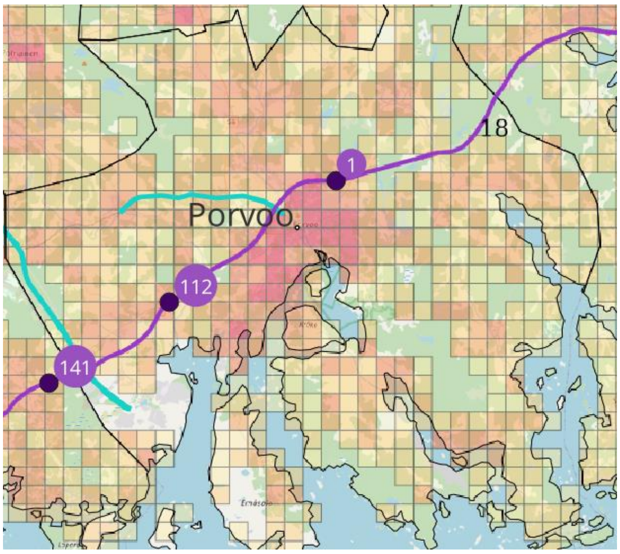


Figure 22 Map of Porvoo

Figure 21 Traffic in Porvoo over the study period

4.4.8 Raisio

Table 10 Economic variables for Raisio

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	raisio	24 427	71	92	1 699	35 461
11	raisio	24 559	73	92	1 680	35 019
12	raisio	24 562	72	94	1 683	35 307
13	raisio	24 565	71	96	1 678	34 421
14	raisio	24 371	69	94	1 675	33 511
15	raisio	24 290	69	94	1 657	33 647
16	raisio	24 283	71	97	1 688	34 095
17	raisio	24 234	73	97	1 659	34 190
18	raisio	24 178	75	99	1 728	35 115
19	raisio	24 056	76	100	1 710	36 029

Raisio has three measuring stations on the E18 that crosses the municipality from west to east. There's also a crossing highway to the north. Population density in Raisio is very similar to Kaarina. It's located close to Turku. The railway has no stops in Raisio.

Points 231 and 251 see very low traffic volumes. Overall traffic on the E18 is also low considering the vicinity to a large population centre. 233 has an above average traffic count. Traffic volumes in Raisio have increased steadily over the study period.

Raisio is a small municipality with a small population within the study group. Population shrank by 2 % during the study period. Employment ratio is very similar to other municipalities and the workplace self-sufficiency is high. Apartment prices have not grown, and income only rose by 2 %.

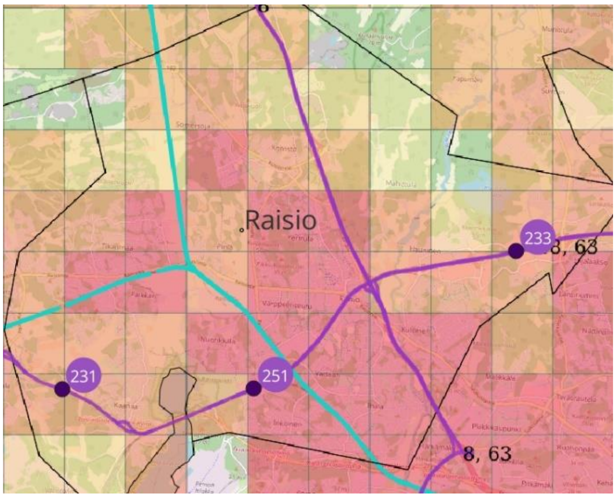


Figure 23 Map of Raisio

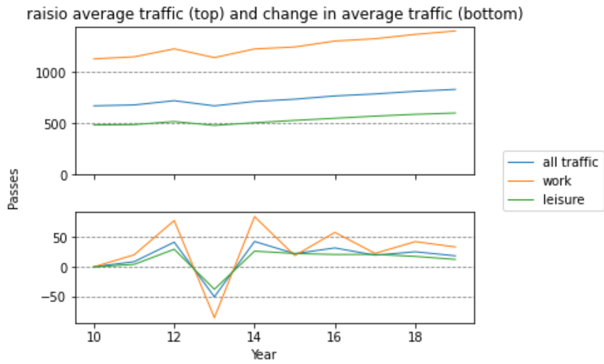


Figure 24 Traffic in Raisio over the study period

4.4.9 Salo

Table 11 Economic variables for Salo

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	salo	55 235	69	105	1 415	29 884
11	salo	55 283	70	102	1 471	29 575
12	salo	54 858	65	99	1 457	30 101
13	salo	54 478	65	96	1 353	27 225
14	salo	54 238	65	94	1 307	26 948
15	salo	53 890	65	93	1 259	27 597
16	salo	53 546	66	91	1 289	27 717
17	salo	52 984	68	90	1 298	26 672
18	salo	52 321	70	89	1 298	27 200
19	salo	51 833	71	87	1 282	27 925

Salo is a large municipality by land area. Salo also has a railway station and rail service to Turku and Helsinki. The six measuring stations are positioned on the eastern side of the municipality. They have the lowest averages of all the selected stations. Salo has the lowest average passes after its neighbour to Paimio to the west. Traffic has grown steadily over the study period.

Salo has had a population decline of 6 % and a drop in the workplace self-sufficiency. Apartment prices have dropped 9 % and income 7 %. All this despite growing traffic numbers. The growth in traffic and drop in workplace self-sufficiency are most likely caused by a large employer leaving the municipality. Nokia had a lot of mobile phone operations in Salo and they were shut down in the 2010's.

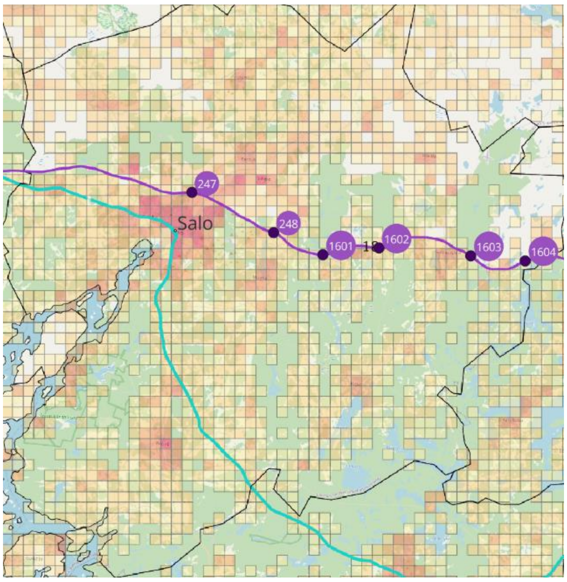


Figure 25 Traffic in Salo over the study period

Figure 26 Map of Salo

4.4.10 Sipoo

Table 12 Economic variables for Sipoo

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	sipoo	18 253	79	58	2 377	50 304
11	sipoo	18 526	80	57	2 422	50 887
12	sipoo	18 739	80	58	2 445	51 471
13	sipoo	18 914	79	58	2 438	51 145
14	sipoo	19 034	79	63	2 450	50 506
15	sipoo	19 399	79	60	2 364	50 844
16	sipoo	19 922	80	60	2 400	51 225
17	sipoo	20 310	81	65	2 380	51 741
18	sipoo	20 666	82	67	2 537	52 140
19	sipoo	21 170	81	68	2 494	51 898

Sipoo has one measuring station on the eastern side of the municipality. Sipoo is populated evenly around the whole municipality. The population centre is located close to the E18 and another measuring station on the western side would serve the municipality's commute traffic. Traffic volumes have grown over the study period.

Sipoo has had a population growth of 16 % over the study period. Employment rate has stayed steady, and the workplace self-sufficiency has risen ten percentage-units. Apartment prices rose by 5 % and income by 3 %.

Figure 27 Traffic in Sipoo over the study period

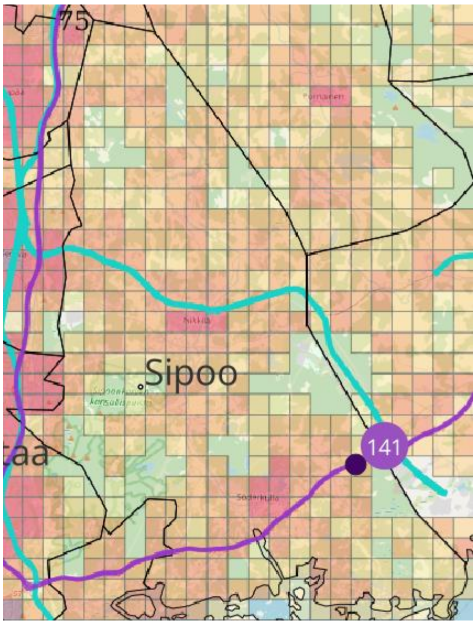


Figure 28 Map of Sipoo

4.4.11 Vantaa

Table 13 Economic variables for Vantaa

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	vantaa	200 055	75	104	2 624	43 070
11	vantaa	203 001	76	105	2 684	42 845
12	vantaa	205 312	76	105	2 756	42 625
13	vantaa	208 098	74	106	2 786	41 888
14	vantaa	210 803	73	108	2 826	41 185
15	vantaa	214 605	73	109	2 630	41 035
16	vantaa	219 341	73	108	2 655	41 011
17	vantaa	223 027	74	109	2 678	41 018
18	vantaa	228 166	76	109	2 735	41 484
19	vantaa	233 775	76	107	2 777	41 892

The E18 serves as the outermost ring-road around Helsinki and it goes through Vantaa. There's also Finland's largest airport next to the E18 and rail lines that serve regional commuters. Vantaa is densely populated around the E18 but also the railway stations. The five measuring are the five most driven roads on the dataset. The averages have grown but not in a steady pace. The variation between leisure and work time traffic is not as profound as in many other municipalities.

Vantaa has had a population growth of 17 %. It has a good employment rate and terrific workplace self-sufficiency. Incomes dropped by 3 % but apartment prices rose by 6 %. Because of the local rail service, a lot of commuting and housing is centred around the stations.

Figure 30 Map of Vantaa

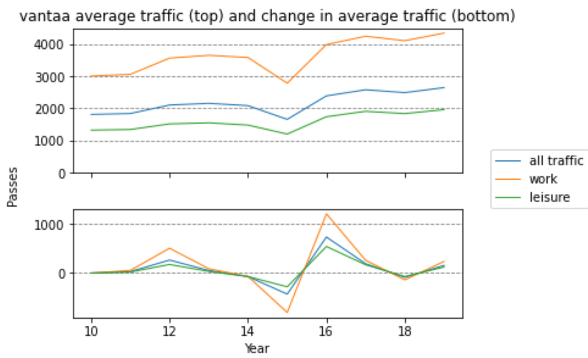


Figure 29 Traffic in Vantaa over the study period

4.4.12 Vihti

Table 14 Economic variables for Vihti

year	kunta	kunta_vaki	kunta_tyol	kunta_omav	kunta_ashinta	tulot
10	vihti	28 311	76	61	2 011	44 829
11	vihti	28 581	77	61	2 038	44 539
12	vihti	28 674	77	60	2 061	44 432
13	vihti	28 929	76	60	2 021	44 060
14	vihti	28 995	75	61	1 894	42 947
15	vihti	28 919	75	62	1 925	43 253
16	vihti	28 967	76	62	2 024	43 173
17	vihti	29 054	77	61	1 958	42 813
18	vihti	29 211	78	62	1 993	43 144
19	vihti	29 158	78	61	1 991	43 352

The E18 barely scratches the southern tip of Vihti, but even though the population density is not high, and it is spread out, the centre is close to the E18, and it does serve as an efficient commuting route to larger population centres. The railway does not serve this municipality. Traffic averages saw a drop in the early 2010’s but have since grown and passed the 2010 average. The placement of the measuring station is very similar to Kirkkonummi, but Vihti has a lot less options when it comes to commuting routes and modes of transport.

Population in Vihti has grown by 3 %. The employment rate is ok and workplace self-sufficiency is low. Vihti is within commuting distance to Helsinki, Espoo and Vantaa. However, because of the E18’s location and spread out population, it is not the primary route for all commuters.

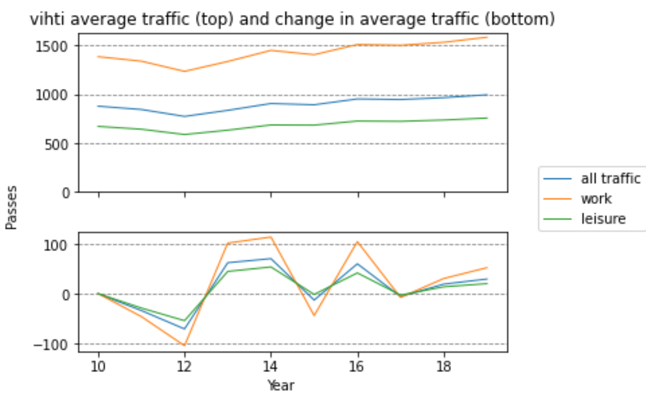


Figure 31 Traffic in Vihti over the study period

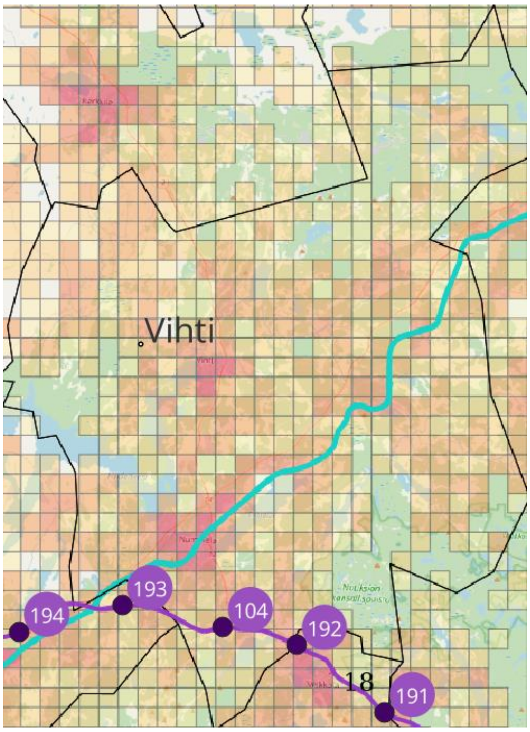


Figure 32 Map of Vihti

Table 15 Average hourly passes for each measuring station

lamid	average hourly passes
150	2775
128	2373
169	1799
160	1746
168	1623
139	1465
191	1453
192	1355
233	1211
112	1045
141	972
194	963
104	899
227	788
193	773
572	649
242	636
231	635
1607	614
243	571
1606	543
1	519
244	490
1604	467
1605	461
1602	457
247	455
1601	454
1603	427
251	373
248	367

Table 16 Average hourly passes for each municipality

kunta	average hourly passes
vantaa	2173
espoo	1465
kirkkonummi	1404
sipoo	972
vihti	899
kaarina	788
porvoo	782
raisio	740
lohja	671
kouvola	649
paimio	566
salo	438

5 MODEL AND RESULTS

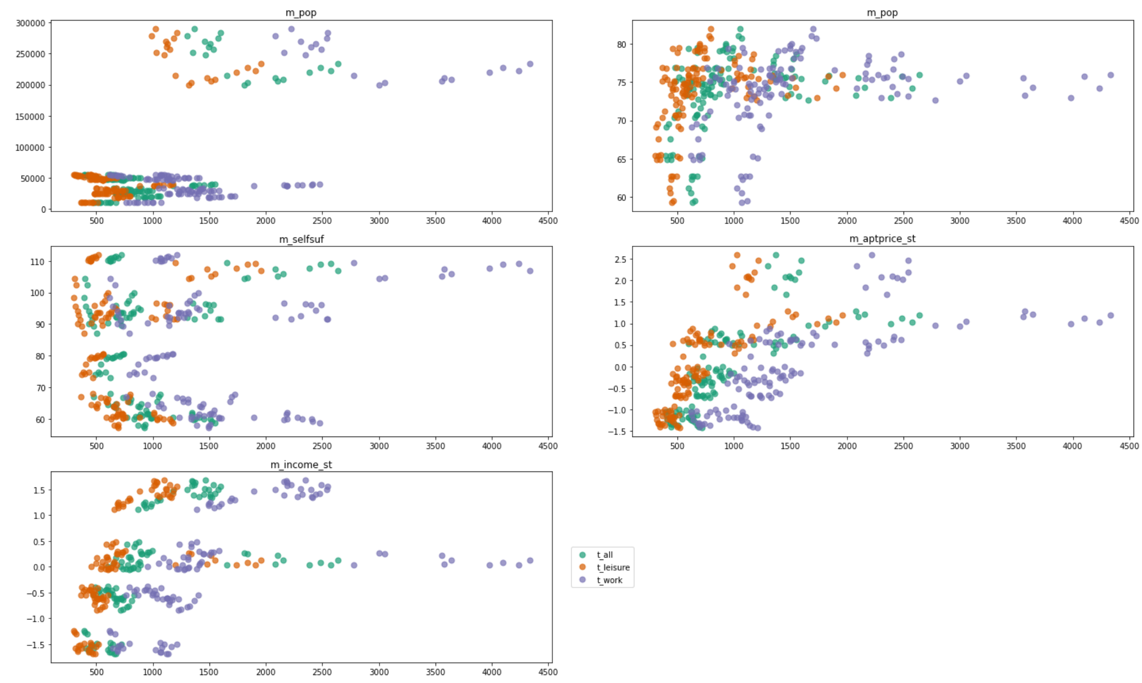


Figure 33 Scatterplot of economic variables and traffic variables

5.1 Model and assumptions

The model aims to compare, how splitting the dependent variable from general traffic average to leisure and work time averages. There are five economic variables, but from the scatterplots in figure 33 it is already clear, that the employment rate and workplace self-sufficiency are variables that have no linear dependence on traffic. There is no reason to run the analysis for these variables.

For population, the location of the measuring stations in many municipalities matters more to the traffic flow than the population number itself. This can be seen with the outliers on the bottom right that show Raisio's measuring station that is located on a commuting route. Overall, it is very hard to say that there is linearity in traffic averages and population.

It's important to acknowledge, that the size and capacity of the E18 is not uniform. This is also not controlled in this analysis.

Both income and apartment prices have a linear trend visible in the scatterplot. Municipalities can be recognized from the bunching, but it's clear that as incomes and apartment prices rise, traffic averages rise as well. The split to leisure and work is clearly visible in the scatterplot.

The first model is an ordinary least squares regression or OLS in short. The first model for the economic variable Y and the combined traffic variable β for every year i and municipality j is as follows:

$$Y_{ij} = \beta_{ij} + \varepsilon_{ij}$$

The second model splits the traffic variable β into two variables, leisure, and work $\beta_{\text{leisure } i\, j}$ and $\beta_{\text{work } i\, j}$,

$$Y_{ij} = \beta_{xij} + \beta_{zij} + \varepsilon_{ij}$$

Then the third and fourth model only look at worktime and leisure averages seperately:

$$Y_{ij} = \beta_{xij} + \varepsilon_{ij}$$

$$Y_{ij} = \beta_{zij} + \varepsilon_{ij}$$

5.2 Limitations in the dataset

This model is for studying the E18 and its passenger car traffic and how it explains economic variables. This means, that for some municipalities that the E18 only barely passes the economic variables covering the entire municipality are not the right match. However, if the model performs at all, it means that there is plenty of potential to run the data for a more targeted area around the E18 and municipal centers around it.

As noted in the municipality analysis, some measuring stations capture more information about commuting than others. This limitation does cause some issues. For example, the stations at the corners of Lohja, Vihti and Kirkkonummi most likely have very similar traffic patterns, but they get assigned the economic variables of their respective municipalities that have huge differences. The municipalities in themselves are very different as well. This is a clear disadvantage in a municipality-based approach.

The split between work and leisure-time traffic on the E18 also does not account the vast differences in public transport between municipalities. Espoo and Vantaa have service levels and population densities to serve entirely different levels of public transport than Paimio, Porvoo or Kotka for example. To correct for this, the dataset should include much more accurate simulations about the accessibility to other modes of transport around the E18. Again, these simulations would have to be done at a similar level the population densities are presented and would require the economic data to be in a similar level. The availability or estimation of economic data on that level is (far) beyond the scope of this thesis.

Overall, the goal in interpreting the results is seeing, how the split affects the analysis. These weaknesses discussed above affect the model, but not the split in traffic data or the traffic itself at all. The traffic data would be the same even if the economic, other transport or accessibility variables and population data would be drilled into a more localized data set.

5.3 Traffic data and normality

The figure 34 already shows that the distribution of the traffic data has a long right tail. This applies to the work and leisure data as well. When looking at the data on an aggregate level and running some transformations, the log transformation does look a good normal fit. The 0 in the log-transformation is an artifact from a +1 added to the values to apply a log transformation. This data is grouped in a less aggregate level incorporating daily aggregate values instead of the annual municipal values that are used in the model.

When looking at the final data in figure 35, the data still shows explicit two humps, but they are improved by the transformations.

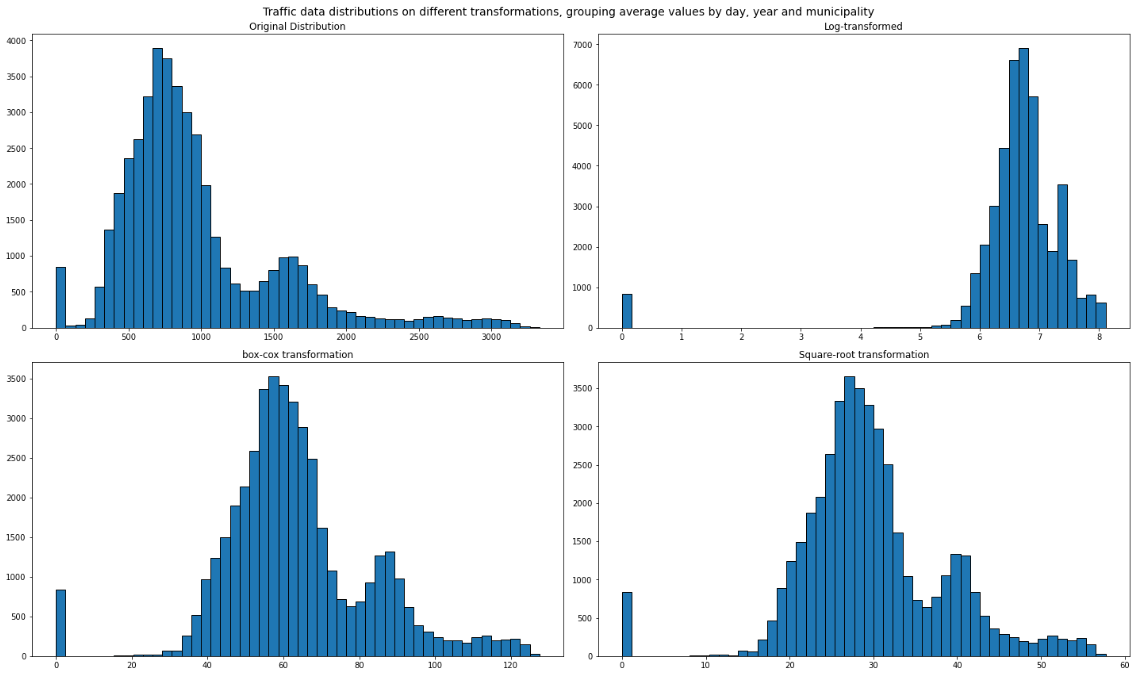


Figure 34 Histograms of traffic data with different transformations

Similar transformations are made to the final data, but +1 is not necessary for box-cox or log-transformation, because the 120 observations in the final data are all non-zero positive values.

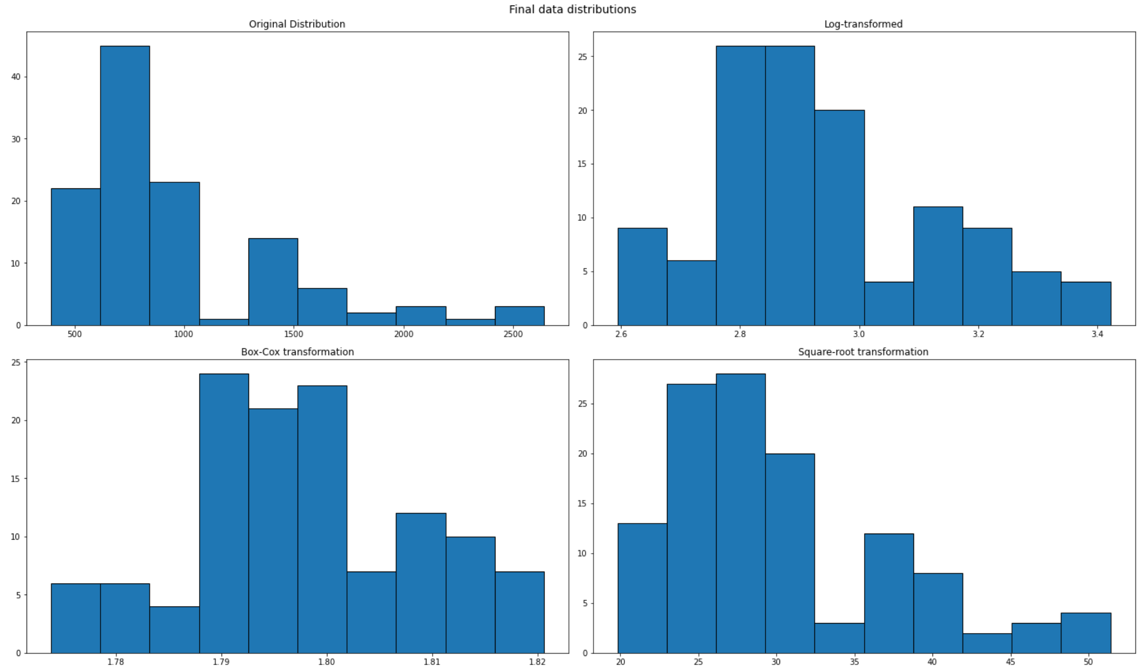


Figure 36 Histograms of grouped traffic data with different transfomations

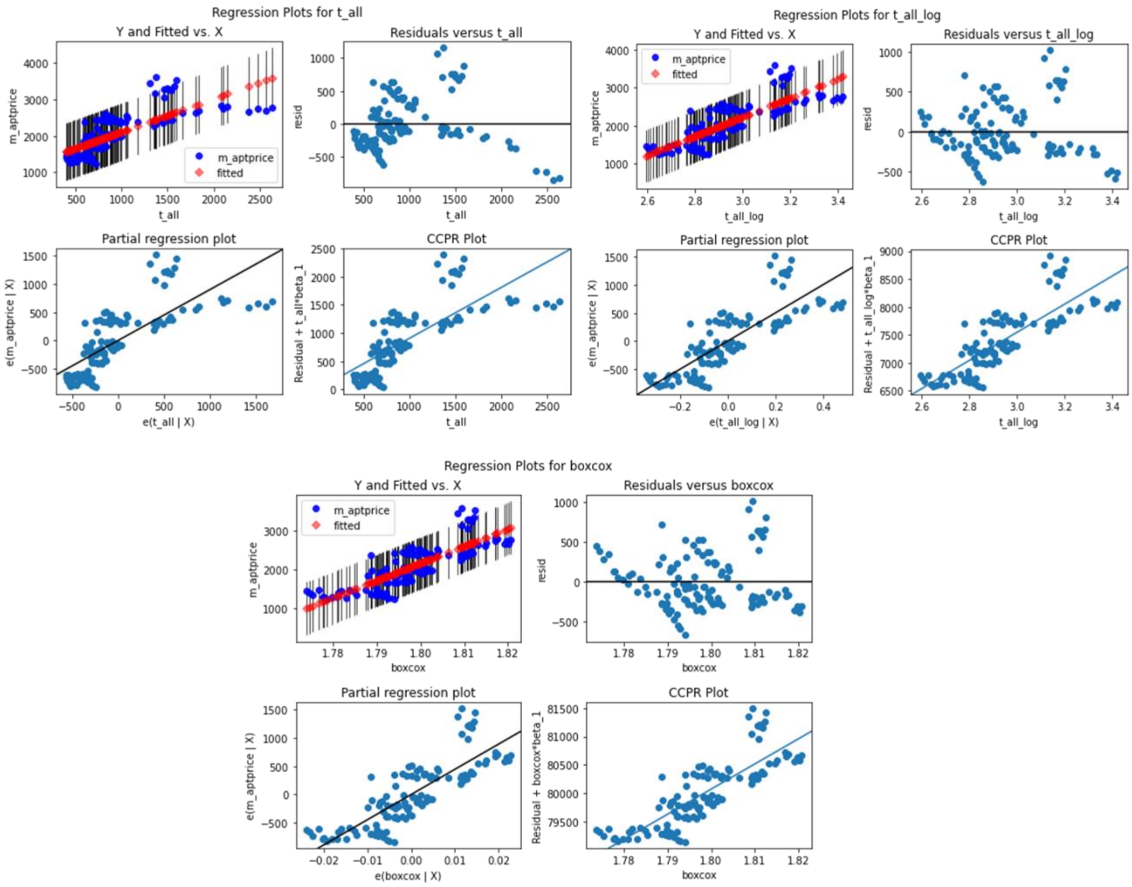


Figure 35 Regression plots for apartment prices and average traffic with and without transformations

Both log-transformation and box-cox are potential in terms of normality. When looking at the regression residuals and overall fit, box-cox is seeming to be the best fit. R-squared values for the regressions also favor a box-cox transformation. For an un-transformed dataset t_all the r-squared is 0,557, log-transformation t_all_log improves this to a 0,667 and box-cox transformation is slightly better at 0,678. Durbin-watson statistic also confirms the interpretation from the residual distribution and box-cox has the least heteroscedasticity, but it’s still not great. values between 1 and 2 are ideal. Box-cox however is has the highest skewness -statistic which means a weaker symmetry and lowest kurtosis -statistic, which implies that box-cox does not handle the outliers from the data in the best way. However, these differences are miniscule.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              m_aptprice   R-squared:                       0.557
Model:                             OLS   Adj. R-squared:                  0.553
Method:                  Least Squares   F-statistic:                     148.3
Date:                 Tue, 27 Jun 2023   Prob (F-statistic):           1.36e-22
Time:                         22:03:24   Log-Likelihood:                -886.07
No. Observations:                  120   AIC:                             1776.
Df Residuals:                      118   BIC:                             1782.
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1207.2041     79.884     15.112      0.000    1049.013    1365.395
t_all          0.9035      0.074     12.178      0.000       0.757       1.050
==============================================================================
Omnibus:                         7.570   Durbin-Watson:                   0.335
Prob(Omnibus):                   0.023   Jarque-Bera (JB):                7.441
Skew:                            0.604   Prob(JB):                       0.0242
Kurtosis:                        3.168   Cond. No.                     2.40e+03
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              m_aptprice   R-squared:                       0.667
Model:                             OLS   Adj. R-squared:                  0.664
Method:                  Least Squares   F-statistic:                     236.0
Date:                 Wed, 28 Jun 2023   Prob (F-statistic):           6.34e-30
Time:                         19:27:53   Log-Likelihood:                -869.00
No. Observations:                  120   AIC:                             1742.
Df Residuals:                      118   BIC:                             1748.
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept  -5315.3023    482.187    -11.023      0.000   -6270.165   -4360.440
t_all_log   2515.5042    163.751     15.362      0.000    2191.232    2839.776
==============================================================================
Omnibus:                        10.127   Durbin-Watson:                   0.377
Prob(Omnibus):                   0.006   Jarque-Bera (JB):               10.664
Skew:                            0.728   Prob(JB):                      0.00484
Kurtosis:                        3.102   Cond. No.                         50.9
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              m_aptprice   R-squared:                       0.678
Model:                             OLS   Adj. R-squared:                  0.676
Method:                  Least Squares   F-statistic:                     248.8
Date:                 Tue, 27 Jun 2023   Prob (F-statistic):           7.71e-31
Time:                         22:05:03   Log-Likelihood:                -866.86
No. Observations:                  120   AIC:                             1738.
Df Residuals:                      118   BIC:                             1743.
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   -7.79e+04   5070.252    -15.364      0.000   -8.79e+04   -6.79e+04
boxcox      4.448e+04   2820.075     15.773      0.000    3.89e+04    5.01e+04
==============================================================================
Omnibus:                        10.531   Durbin-Watson:                   0.406
Prob(Omnibus):                   0.005   Jarque-Bera (JB):               11.241
Skew:                            0.749   Prob(JB):                      0.00362
Kurtosis:                        3.072   Cond. No.                         391.
==============================================================================
```

Figure 37 Regression results for average traffic and apartment prices with and without transformations

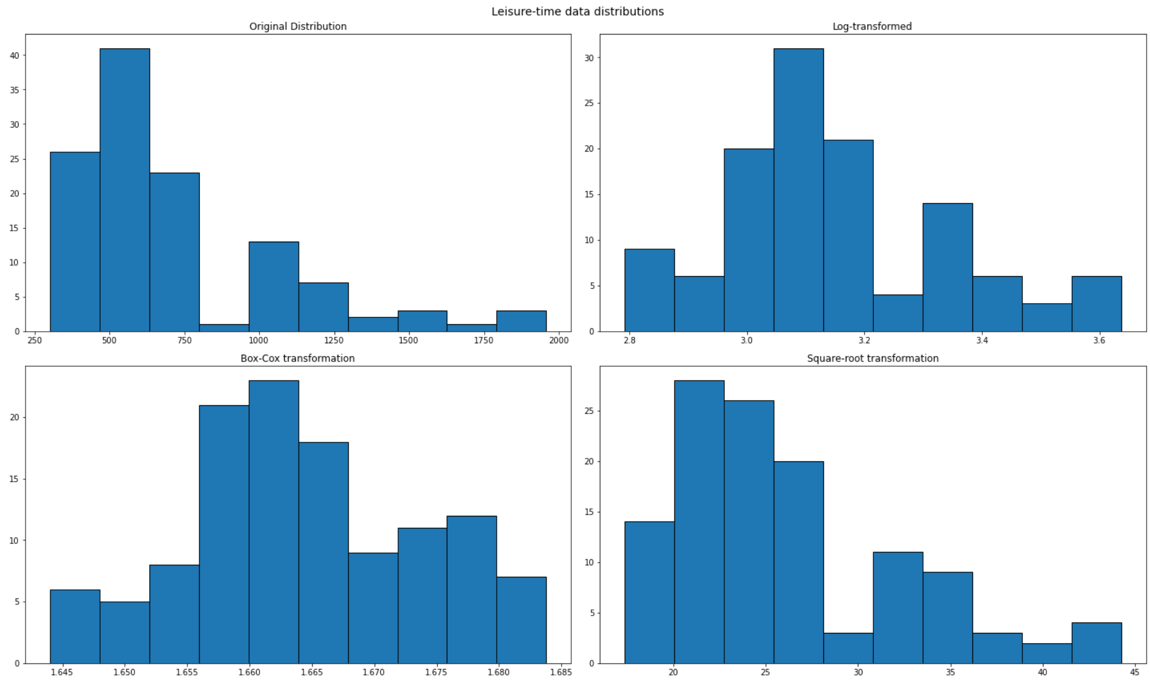


Figure 39 Distribution of leisure-time traffic and its transformations

Does the split into leisure and work times hurt the normality of the traffic data? Leisure time benefits from both log and box-cox transformations. See figure 38. Both data-splits are tested in a similar matter to the apartment prices.

The fit for leisure-traffic and apartment price regression is greatly improved by the transformations. The residuals see an improvement as well. Leisure-time traffic has an r-squared of 0,578 without transformations. Box-cox increases the measure to a 0,702 and a log-transformation has a 0,640 r-squared statistic. Heteroscedasticity is improved in both transformations. Durbin-Watson statistic is a 0,346 untransformed, 0,426 and 0,362 for box-cox and log transforms respectively. Box cox improves the skewness from the untransformed dataset, but log transformation here is slightly better. Kurtosis is not improved, but box-cox makes it less bad.

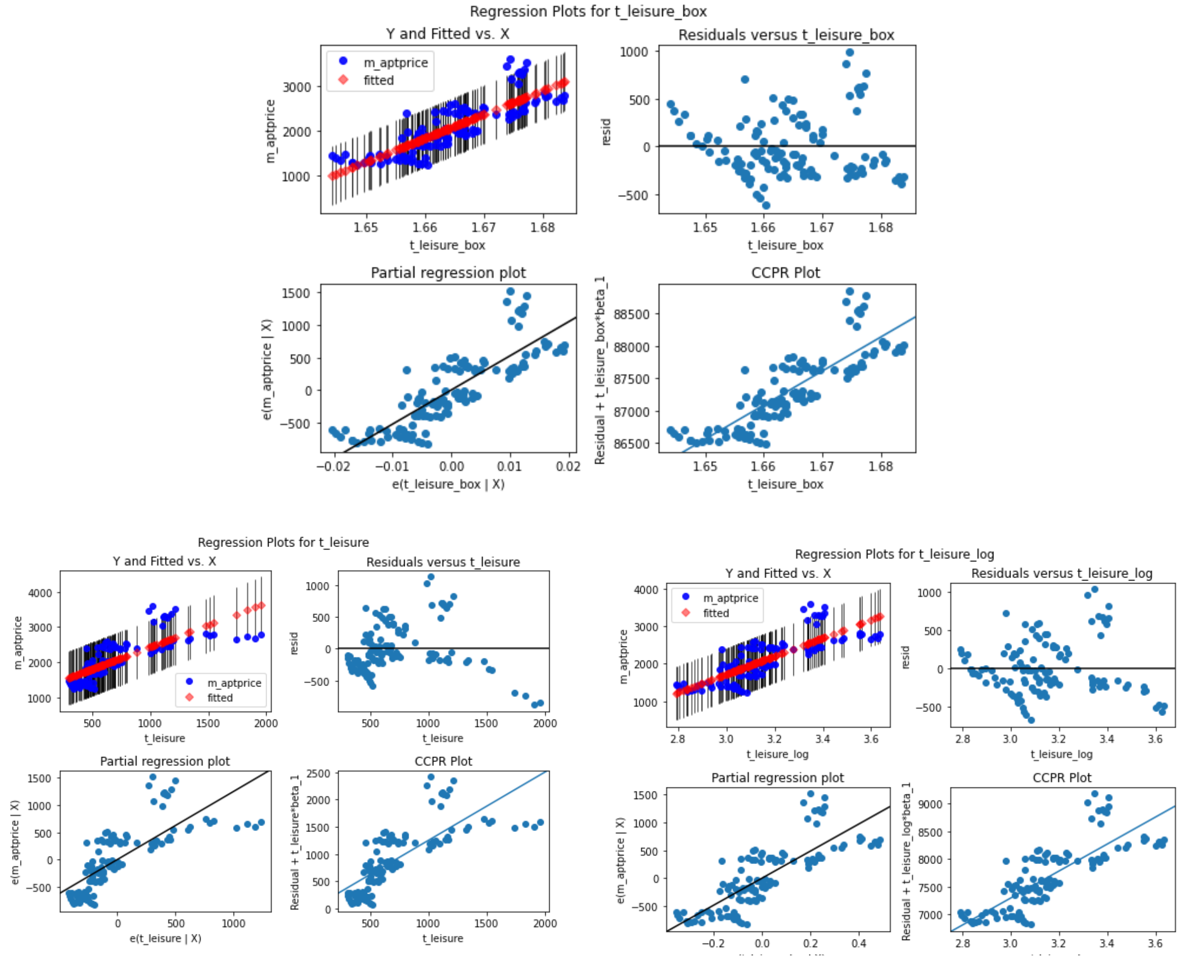


Figure 40 Regression plots for apartment prices and leisure-time traffic with and without transformations

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             m_aptprice   R-squared:                       0.578
Model:                            OLS   Adj. R-squared:                  0.574
Method:                 Least Squares   F-statistic:                     161.5
Date:                Fri, 30 Jun 2023   Prob (F-statistic):           7.65e-24
Time:                        20:10:30   Log-Likelihood:                -883.17
No. Observations:                 120   AIC:                             1770.
Df Residuals:                     118   BIC:                             1776.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1178.0118     78.887     14.933      0.000    1021.795    1334.229
t_leisure      1.2515      0.098     12.709      0.000       1.056       1.446
==============================================================================
Omnibus:                        7.038   Durbin-Watson:                   0.346
Prob(Omnibus):                  0.030   Jarque-Bera (JB):                6.732
Skew:                           0.568   Prob(JB):                       0.0345
Kurtosis:                       3.237   Cond. No.                     1.81e+03
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             m_aptprice   R-squared:                       0.640
Model:                            OLS   Adj. R-squared:                  0.637
Method:                 Least Squares   F-statistic:                     209.6
Date:                Fri, 30 Jun 2023   Prob (F-statistic):           6.30e-28
Time:                        20:10:31   Log-Likelihood:                -873.65
No. Observations:                 120   AIC:                             1751.
Df Residuals:                     118   BIC:                             1757.
Df Model:                           1
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept     -5579.3658    529.857    -10.530      0.000   -6628.627   -4530.105
t_leisure_log  2431.5483    167.972     14.476      0.000    2098.918    2764.178
==============================================================================
Omnibus:                        9.100   Durbin-Watson:                   0.362
Prob(Omnibus):                  0.011   Jarque-Bera (JB):                9.477
Skew:                           0.688   Prob(JB):                      0.00875
Kurtosis:                       3.051   Cond. No.                         56.9
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             m_aptprice   R-squared:                       0.702
Model:                            OLS   Adj. R-squared:                  0.700
Method:                 Least Squares   F-statistic:                     278.5
Date:                Fri, 30 Jun 2023   Prob (F-statistic):           7.62e-33
Time:                        20:10:31   Log-Likelihood:                -862.18
No. Observations:                 120   AIC:                             1728.
Df Residuals:                     118   BIC:                             1734.
Df Model:                           1
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept     -8.524e+04   5231.940    -16.293      0.000   -9.56e+04   -7.49e+04
t_leisure_box  5.246e+04   3143.332     16.690      0.000    4.62e+04    5.87e+04
==============================================================================
Omnibus:                       12.050   Durbin-Watson:                   0.426
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               13.104
Skew:                           0.806   Prob(JB):                      0.00143
Kurtosis:                       3.141   Cond. No.                         403.
==============================================================================
```

Figure 42 Regression results for apartment prices and leisure-time traffic with and without transformations

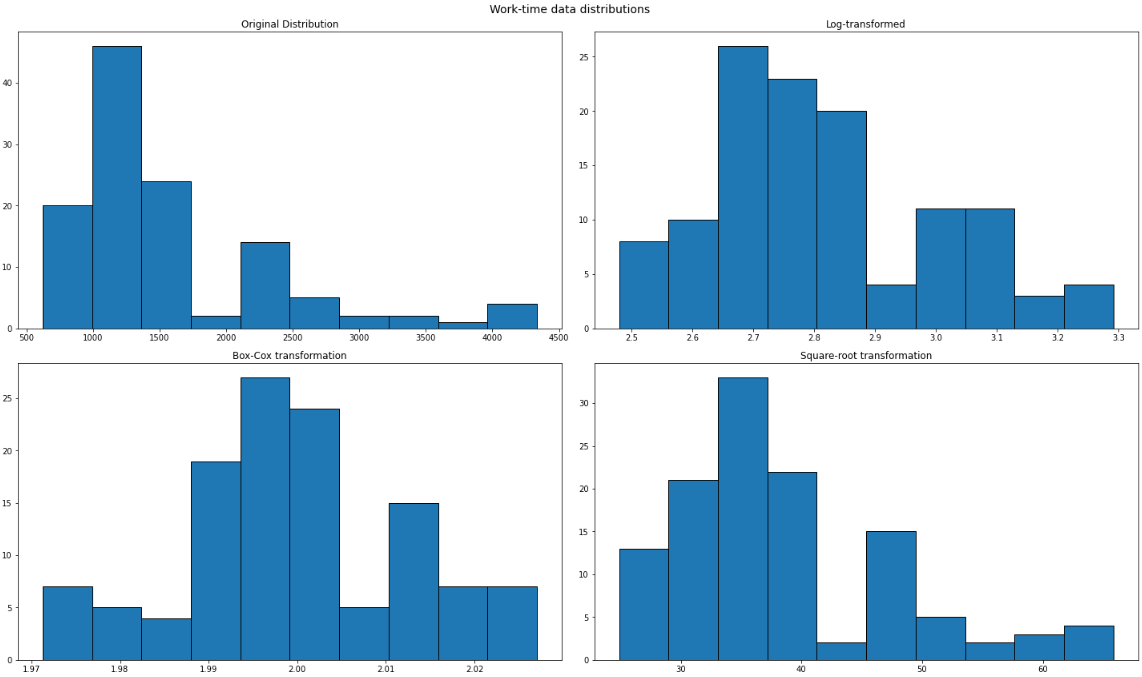


Figure 46 Distribution of worktime traffic and its transformations

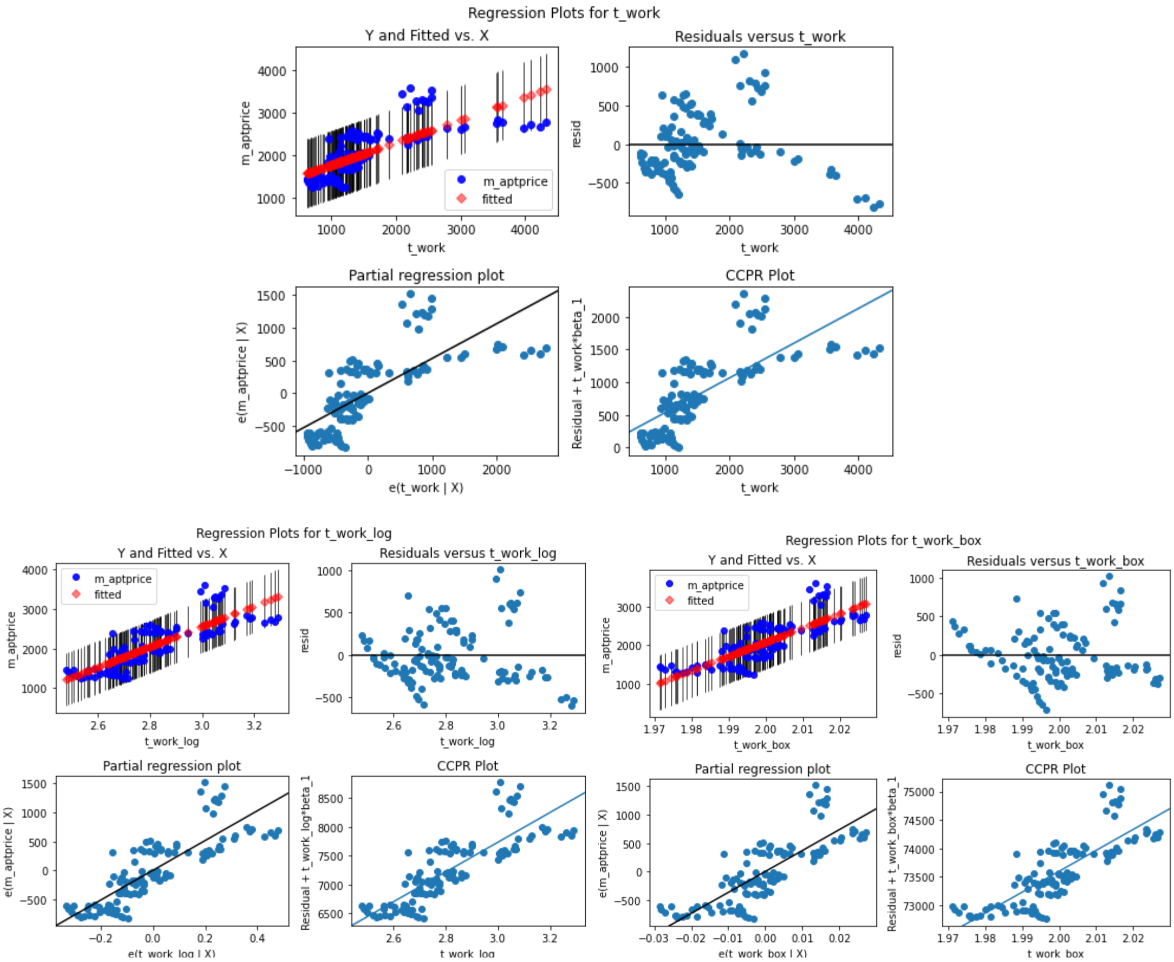


Figure 44 Regression plots for apartment prices and worktime traffic with and without transformations

Worktime data is also improved by both log and box-cox transformations. The fit seems to improve in both transformations and residuals improve visually more on the box-cox transformation. R-squared improves in both transformations, this time to the benefit of log-transformation. Original data has an r^2 of 0,531, log transformed data 0,688 and cox-box transformed data has an r^2 of 0,648. Heteroscedasticity is improved more in log transformations than in the Box-Cox transformation. Skewness is not improved for some odd reason. Kurtosis is improved slightly by the log transformation but not the Box-Cox transformation. Overall, both inspected transformations improve the data in any split. However, Box-Cox does it more frequently than log-transformation. The differences between the transformations are small and mostly in a similar direction.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             m_aptprice   R-squared:                       0.531
Model:                            OLS   Adj. R-squared:                  0.527
Method:                 Least Squares   F-statistic:                     133.8
Date:                Fri, 30 Jun 2023   Prob (F-statistic):           3.78e-21
Time:                        20:40:05   Log-Likelihood:                -889.43
No. Observations:                 120   AIC:                             1783.
Df Residuals:                     118   BIC:                             1788.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1243.3293     80.921     15.365      0.000    1083.084    1403.574
t_work         0.5323      0.046     11.567      0.000       0.441       0.623
==============================================================================
Omnibus:                        8.027   Durbin-Watson:                   0.323
Prob(Omnibus):                  0.018   Jarque-Bera (JB):                8.089
Skew:                           0.634   Prob(JB):                       0.0175
Kurtosis:                       3.101   Cond. No.                     3.86e+03
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             m_aptprice   R-squared:                       0.688
Model:                            OLS   Adj. R-squared:                  0.685
Method:                 Least Squares   F-statistic:                     259.8
Date:                Fri, 30 Jun 2023   Prob (F-statistic):           1.35e-31
Time:                        20:40:05   Log-Likelihood:                -865.10
No. Observations:                 120   AIC:                             1734.
Df Residuals:                     118   BIC:                             1740.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept  -5169.4837    450.601    -11.472      0.000   -6061.796   -4277.172
t_work_log  2576.6242    159.872     16.117      0.000    2260.034    2893.215
==============================================================================
Omnibus:                       10.688   Durbin-Watson:                   0.390
Prob(Omnibus):                  0.005   Jarque-Bera (JB):               11.267
Skew:                           0.746   Prob(JB):                      0.00358
Kurtosis:                       3.161   Cond. No.                         47.5
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             m_aptprice   R-squared:                       0.648
Model:                            OLS   Adj. R-squared:                  0.645
Method:                 Least Squares   F-statistic:                     217.2
Date:                Fri, 30 Jun 2023   Prob (F-statistic):           1.62e-28
Time:                        20:40:05   Log-Likelihood:                -872.28
No. Observations:                 120   AIC:                             1749.
Df Residuals:                     118   BIC:                             1754.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept  -7.152e+04   4994.087    -14.320      0.000    -8.14e+04   -6.16e+04
t_work_box   3.68e+04   2497.150     14.736      0.000     3.19e+04    4.17e+04
==============================================================================
Omnibus:                        8.546   Durbin-Watson:                   0.383
Prob(Omnibus):                  0.014   Jarque-Bera (JB):                8.833
Skew:                           0.664   Prob(JB):                       0.0121
Kurtosis:                       3.032   Cond. No.                         391.
==============================================================================
```

Figure 47 Regression results for apartment prices and worktime traffic with and without transformations

5.4 Regression performance

This chapter depicts, how the regression performance is affected by the traffic split. There are four different ways to run the model on three different economic variables. The chapter goes through each economic variable and how the models perform. P-values for all models are below 0,000, so they are not part of the analysis since the changes are so insignificant. Another measure to test the statistical significance of the model is the p-value, marked as prob f-statistic. The p-values for all models are extremely small. This means that with high confidence the traffic variables and economic variables share a linear connection.

Below, table 17 shows the statistics for the regressions between population and traffic data with its variations.

Table 17 Regression results for population-variable

	t_all	t_combined	t_work	t_leisure
R^2	0.369	0.370	0.370	0.366
Intercept	-7.848e+06	-7.608e+06	-7.396e+06	-8.396e+06
Coef, independent variables	4.404e+06	Leisure: 9.837e+05 Work: 3.02e+06	3.733e+06	5.086e+06
Standard error, independent variables	5.31e+05	Leisure: 4.9e+06 Work: 3.58e+06	4.49e+05	6.16e+05
Omnibus	11.730	12.629	13.204	10.589
Skew	0.802	0.835	0.856	0.758
Kurtosis	3.039	3.091	3.120	2.970
Durbin-Watson	0.276	0.274	0.273	0.278

Population had the weakest visually inspected linear dependence, and the numbers back this up. On the E18, population and traffic have by far the weakest connection out of the three variables. R^2 is not greatly improved by the transformations, the effect is however very statistically significant in every model. Because the numbers on the intercept and coef are so large, the exponential in the scientific notation (+n) is most relevant. Intercept hardly changes for any model and the coefficient, which means the slope of the fitted regression curve is changed by the combined model. In separate models, as leisure or work-time traffic increases, population increases a similar amount. The effect on the joined model with both independent variables is very work-traffic skewed. Sensivity to work time traffic is a bit lower compared to the individual model, but for the average leisure time traffic, the sensitivity is reduced to a fifth from the individual model. Standard errors are smallest in the combined model and largest in the leisure-model. Skew, kurtosis, and heteroscedasticity are all almost unchanged in the change of independent variable.

Table 18 Regression results for apartment prices variable

	t_all	t_combined	t_work	t_leisure
R^2	0.678	0.747	0.648	0.702
Intercept	-7.79e+04	-1.052e+05	-7.152e+04	-8.524e+04
Coef, independent variables	4.448e+04	Leisure: 1.561e+05 Work: -7.63e+04	3.68e+04	5.246e+04
Standard error, independent variables	2820.075	Leisure: 2.31e+04 Work: 1.69e+04	2497.150	3143.332
Omnibus	10.531	10.572	8.546	12.050
Skew	0.749	0.720	0.664	0.806
Kurtosis	3.072	3.310	3.032	3.141
Durbin-Watson	0.406	0.487	0.383	0.426

Table 18 shows the analysis for apartment prices. Overall, traffic and apartment prices work very well together with very good R^2 values in every model. Work time traffic does not improve the models overall fit at all, but leisure does, and the combination of work and leisure time traffic as variables has a good 0.747 R^2 compared to the 0.678 of a model with the average of all traffic. A combined model has a much higher standard error, but the error rate seems to correspond to the increase in the slope of the coefficient of independent variables. In the combined model leisure has a three times larger impact on apartment prices, and the work-time traffic has a larger and inverted relationship to the average apartment price of a municipality. Statistics for the fitness of the model are overall almost the same, but the work-time traffic does not perform alone nearly as well as other models do.

Table 19 Regression results for income variable

	t_all	t_combined	t_work	t_leisure
R^2	0.551	0.708	0.511	0.585
Intercept	-1.015e+06	-1.612e+06	-9.161e+05	-1.126e+06
Coef, independent variables	5.871e+05	Leisure: 3.228e+06 Work: -1.86e+06	4.785e+05	7.011e+05
Standard error, independent variables	4.88e+04	Leisure: 3.63e+05 Work: 2.65e+05	4.31e+04	5.43e+04
Omnibus	8.455	3.892	9.767	7.083
Skew	-0.523	-0.269	-0.521	-0.521
Kurtosis	2.387	2.430	2.314	2.491
Durbin-Watson	0.339	0.448	0.323	0.355

Table 19 shows the results from income as an independent variable. The results are similar to apartment prices. The R^2 statistic is lower than t_all for t_work but is improved at t_leisure and even more in t_combined. As with apartment prices, the combined model makes the regression function steeper. The combined model also makes leisure a more sensitive variable than work, like the separate models. Income behaves in a similar way to apartment prices and worktime traffics effect is inverted in the combined model. A steeper curve also makes the standard error larger and lowers the intercept. The combined model has a better fitness in almost all measurements compared to other dependent variables.

6 CONCLUSIONS

The main goal in this thesis was to approach traffic analysis as a key part of the economy, and how the relationship between economic indicators and traffic in different times work.

6.1 Literature and available data

There is a wide range of literature concerning traffic and the economy. Much of the research is around traffic and the labour market. The accessibility that transportation infrastructure enables and the time savings that benefit the users of the infrastructure are a main component in most benefit cost analyses conducted on infrastructure projects. Traffic is also a consumption good. In some areas leisure time activities around holidays and non-work times contribute a lot to the local economy. These types of differences cannot be properly controlled for when using aggregated traffic averages or counts from long time periods. It also makes forecasting from the traffic indicators much harder.

For this reason, the hypothesis of this thesis is that there is a difference in the data when over all traffic averages are split in to work time and leisure averages and that these differences and their use is a viable method to make linear models better.

Publicly open traffic data from Finnish highways is excellent for studying traffic patterns on different vehicles on different times of day. This data is unfortunately very large in size and requires either a lot of computing power or predetermined summation or filtering to cut back on both computational infrastructure and time requirements for the analysis. However, local economic data is not publicly available. Population densities used in the municipality analysis have a dense enough grid to enable analysis on a level that would enable more robust analyses. This kind of data or even postal-code level data is not publicly available. Income and other population statistics are available for researchers. So are real-estate prices. Finnish registries also offer a chance to use a more accurate time series than annual data. Variables that measure economic activity are harder to come by. GDP is only measured on a county level, and even municipal level is not accurate enough to use data from the measuring stations to their full potential.

The very rough municipal statistics offer a good starting point to test the validity of the approach outlined in the thesis.

6.2 Analysis and findings

In the analysis, work and leisure time traffic averages vary in every municipality, develop in different ways and have different distributions. Municipalities are geographically in very different situations to benefit from the E18 infrastructure. There are also multiple variables that are not controlled for. These include access to public traffic, the number of lanes and maintenance level of E18 in the municipality and the amount of people living in apartments vs. detached houses. These three variables have a chance to explain the differences between municipalities.

The analysis does give enough evidence to conclude, that even on a very aggregate level, there is a linear link between traffic on the E18 and municipal average income and apartment prices. There is a connection in population as well, but that connection is much weaker. These linear connections are affected by the split to leisure and work times. The split makes both income and apartment price models a lot more sensitive to changes in average traffic volumes on the E18 in the 2010's. In both apartment prices and income, the combined model is more sensitive to leisure time traffic than to worktime traffic. The effect of leisure time traffic is in both instances twice the worktime traffic. The effect of worktime traffic on both apartment prices and income in the combined model is negative. If the share of worktime to leisure time tips toward leisure (leisure time traffic average increases or worktime decreases), the municipality more likely has higher apartment prices and incomes. If the municipality sees worktime averages rising or leisure time averages decreasing, this indicates lower numbers for the economic variables.

6.3 Limitations and future studies

This model most likely is not showing the dynamics of economic variables and traffic, but the differences between the municipalities. The differences in municipalities are so large, that even though the connections between traffic and economic variables is consistent, interpretations from the models are more about the differences in municipalities. This is because the lack of control for geographic differences, for instance the location of population centres from the E18 and the accessibility of public transport.

To better study passenger car traffic and economic outcomes, the model should have control-variables to take account differences in the level of infrastructure where the measurements are taken that affects the capacity of the road. These include speed limit, maintenance level, number of lanes and the highway classification for the road. These control variables can be included in the model, or the traffic averages can be weighted according to these variables. The best way to approach this is a topic for a study of its own.

Instead of relying on municipal data, the data should reflect the areas directly served by the infrastructure and measuring station where the traffic is observed. In this thesis, the distance from the measuring station to the furthest part of the municipality is tens of kilometres. At the same time the station might be next to population centres of a neighbouring municipality, which is not accounted for in the data connected to the station. In this analysis the length and design of the E18 helps to alleviate this neighbouring issue through the fact that the E18 splits many of the municipalities in a way that it does not cause the neighbouring issue as much as a more winding road would. For a model that aims to utilize a more scattered collection of traffic measuring stations these issues will come up in a much more profound way.

A dataset that includes grid-level data of the incomes, employment, real-estate prices and economic activities of each grid would enable the model to perform towards the goal originally set in this thesis. There are two ways to assign these grids to the measuring stations. Either by assigning areas within a circle around the measuring station or assigning areas within a set travel time from the measuring station. Studying the effects of the differences between these two approaches and the appropriate distance to set is a topic for another thesis.

The datasets to study local economies are well known but the access to this data is challenging even for researchers. Metsäranta et. al. (2019) There are programming interfaces to map out the accessibility of each measuring station and accessibility is already being measured by Finnish officials. Tilastokeskus, (2023)

The stations do not produce data consistently. They are shut down for maintenance or other reasons. The infrastructure is also repaired and worked on. These changes and gaps in data require more addressing when a dataset with more accuracy is used for analysis. However, this thesis already shows that the amount of data available is enormous and the problem is not the lack of data, but the tools and methods to clean it. The difference between published values and key indicators from the data and the raw signal data from the passes is big. The data needs to be thoroughly vetted when studied more closely. When these studies are made the cleaned data and the methods to edit and vet it should be made available for other researchers to make use of it in their analyses as well.

6.4 Conclusion

This thesis set out to do the following things: explore the usability of the TSM data in research setting and the possible challenges and opportunities it serves. In practice this was done through a small regression analysis studying how the split in daily traffic averages on the E18 affects the performance in explaining the change in certain economic variables.

To summarize, this thesis made the following observations:

- There is a strong connection between the apartment prices and income and traffic on the E18 in the municipalities that the road crosses.
- When the data is split from daily average to daily average during work-time and leisure time separately, the models perform a lot better.
- Municipal boundaries are not the best way to classify economic activity and other variables to traffic measuring stations. This is because traffic on different stations varies greatly within municipalities. Using more local data enables the use of individual stations.
- The analysis without proper control variables ends up analysing differences between municipalities regarding economic variables and traffic averages.

Future studies and further work to make the models work recognized in this thesis were:

- Implement control variables to account for the differences in road infrastructure and other modes of transport.
- Use an economic dataset based on a geographical grid surrounding the measuring station.
    - Study the differences in the performance of the model using a set radius or accessibility within a certain time-frame routing approach for the economic variables.
- A holistic analysis of all the TMS-stations and their raw data and editing the data to a more usable dataset for research.
- Apply this model to an area to measure the economic development around natural experiments such as infrastructure investments, housing development or industrial investments.

REFERENCES

Aaltonen, Mikko, - Virtanen, Miialiila (2017) *Yli- ja keskinopeudet rikesakkojen korotuksen jälkeen,* Valtioneuvoston selvitys- ja tutkimustoiminan julkaisusarja 20-2020, Valtioneuvoston kanslia, Helsinki.

Abrantes, Pedro, A. L. – Wardman, Mark, R. (2011) Meta-analysis of UK values of travel time: An update, *Transportation Research Part A,* Vol. 45, p. 1-17.

Arnott, Richard – De Palma, Andre – Lindsey, Robin (1990) Economics of a Bottleneck, *Journal of Urban Economics,* Vol. 27, p. 111-130.

Arnott, Richard – De Palma, Andre – Lindsey, Robin (1993) A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand, *American Economic Review,* Vol. 83, No 1, p. 161-179.

Becker, G.S. (1965) A theory of the allocation of time, *Economic Journal*, Vol. 75, 493-517.

Börjesson, Maria (2018) Assessing the Net Overall Distributive Effect of a Congestion Charge, *International Transport Forum Discussion Papers,* OECD Publishing, Paris.

Börjesson, Maria – Isacsson, Gunnar – Anderson, Matts – Anderstig, Christer (2019) Agglomeration, productivity and the role of transport system improvements, *Economics of Transportation,* Vol 18 (1), p. 27-39.

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

Chao, Chen – Petty, Karl, Skabardonis, Alexander – Pravin, Varaiya – Lia, Zhangfeng (2001) Freeway Performance Measurement System: Mining Loop Detector Data, *Transportation Research Record: Journal of the Transportation Research Board*, Vol 1748 (1), Transportation Research Board, NRC, Washington, D.C.

DeSerpa, A. C. (1971) A Theory of the Economics of Time, *The Economic Journal*, Vol. 81 (324), p. 828-846.

Down, Anthony (1962) The Law of Peak-Hour Expressway Congestion, *Traffic Quarterly,* Vol 16 (3) p. 393-409.

Downs, Anthony (2004) *Still stuck in traffic* Brookings Institution Press. Washington, D.C.

Duranton, Gilles – Puga, Diego (2004) Chapter 48 - Micro-Foundations of Urban Agglomeration Economies, *Handbook of Regional and Urban Economics*, Vol. 4, p. 2063-2117.

Duranton, Gilles – Turner, Matthew (2011) The Fundamental Law of Road Congestion: Evidence from US cities, *American Economic Review* 101, p. 2616-2652.

E18 moottoritie vihitään käyttöön tänään. < https://vayla.fi/-/e18-moottoritie-vihitaan-kayttoon-tanaan>, haettu 19.7.2023.

Euroopan laajuinen liikenneverkko TENT https://vayla.fi/vaylista/liikennejarjestelma/tent, haettu 19.7.223.

Evans, Alan W. (1972) On the theory of the valuation and allocation of time, *Scottish Journal of Political economy,* Vol 19 (1), p. 1-17.

Ex-post impact assessment of the Tampere western ring road (2019) Publications of the Finnish Transport Infrastructure Agency 40/2019, Finnish Transport Infrastructure Agency, Helsinki.

Fornaro, Paolo – Luomaranta, Henri (2020) Nowcasting Finnish real economic activity: a machine learning approach, *Empirical Economics,* Vol 56 (1), p. 55-71.

Fosgerau, Mogens (2005) Unit Income Elasticity of the Value of Travel Time Savings, European Transport Conference, Strasbourg 3.10.2005.

Fosgerau, Mogens – Engelson, Leonid (2011) The value of travel time variance, *Transportation Research Part B*, 45 p. 1-8.

Framework for examining wider economic impacts of transport system development (2020) Publications of the Ministry of Transport and Communications 2020:5, Ministry of Transport and Communications, Helsinki.

Gibbons, Stephen – Lyytikäinen, Teemu – Overman, Henry G. – Sanchis-Guarner. Rosa (2019) New road infrastructure: The effects on firms, *Journal of Urban Economics,* Vol. 110, p. 35-50.

Gibbons, Stephen – Machin, Stephen (2006) Transport and Labour Market Linkages: Empirical Evidence, Implications for Policy and Scope for Further UK Research, Background paper for the Eddington Report to the Department of Transport.

Gibson, Matthew – Cornovale, Maria (2015) The effects of road pricing on driver behavior and air pollution, *Journal of Urban Economics,* Vol. 89, p. 62-73.

Gutierez, Eva – van Ommeren, Jos (2010) Labour supply and commuting, *Journal of Urban Economics,* Vol. 68 (1), p. 82-89.

Hellman, Sanni-Sandra – Lappo, Sampo (2018) *Liikenteeseen ja autoiluun liittyvien verojen simulointimallin rakentaminen,* Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 71/2018, Valtioneuvoston kanslia, Helsinki.

Hokkanen, Joonas – Savikko, Heikki – Honkatukia, Juha – Metsäranta, Heikki - Sirkiä, Ari – Haapanen, Mika – Tohmo, Timo (2020) Methods of regional economics in impact assessment of the transport system development, Publications of the Ministry of Transport and Communications 2020:7, Ministry of Transport and Communications, Helsinki.

Holl, Adelheid (2004), Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal, *Journal of Regional Science,* Vol 44 (4), p. 693-712.

Honkatukia, Juha – Koskinen, Peetu – Ruuskanen, Olli-Pekka – Villanen, Juuso (2020) *Dieselin verotuen vaikutusten arviointi,* Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2020:4, Valtioneuvoston kanslia, Helsinki.

Johnson, Bruce M. (1966) Travel time and the price of leisure, *Economic Inquiry,* Vol. 4 (2), p. 135-145.

Joukkoliikenteen hintajoustotutkimus 2014 (2015) HSL Helsingin seudun liikenne, Helsinki.

Julkisen liikenteen suoritetilasto 2018 (2020) Traficomin tilastojulkaisuja 2/2020, Liikenne- ja viestintävirasto Traficom, Helsinki.

Kaupunki–maaseutuluokitus päivitetty: Suomen kaupungistumisaste noussut yli 72 prosentin (2020) Suomen ympäristökeskus SYKE, < Suomen ympäristökeskus > Kaupunki–maaseutuluokitus päivitetty: Suomen kaupungistumisaste noussut yli 72 prosentin (syke.fi)> haettu 18.7.2023.rti

Kiiskilä, Kati – Mäki, Ville – Saastamoinen, Kimmo (2020) *Ajonopeudet maanteillä 2019,* Väyläviraston julkaisuja 33/2020, Helsinki.

Kiiskilä, Kati – Tuominen, Janne – Saastamoinen, Kimmo (2016) Liikenneviraston liikennelaskentajärjestelmä: Päivitetty järjestelmäkuvaus, Väyläviraston julkaisuja 36/2016, Helsinki.

Kockelman, Kara (2001) Modeling Traffic's Flow-Density Relation: Accommodation of Multiple Flow Regimes and Traveler Types, *Transportation* 28 (4):363-374.

Kotimaan tieliikenteen tavarankuljetukset (2020a) Tilastokeskus, Viitattu 18.7.2023.

Kukkonen, Reijo – Tuominen, Mika (2015) *Parikkalan rajanylityspaikan kehittäminen, Liikenne-ennuste,* Liikenne- ja viestintäministeriön julkaisuja 5/2015, Liikenne- ja viestintäministeriö, Helsinki.

Lindroos, Pontus (2020) Rekkarallin hiipuminen kertoo tuotannon romahduksesta pika-ennakossa, Tilastokeskus, Tieto ja Trendit blogi, haettu 19.7.2023.

Laakso, Juha – Vesanto, Aleksi – Ritari, Jarno (2018) *Selvitys automaattiajamisen edellyttämistä tiedoista ja kehittämistarpeista,* Liikenne- ja viestintäministeriön julkaisuja 19/2017, Liikenne- ja viestintäministeriö, Helsinki.

Laakso, Seppo – Kostiainen, Eeva, Metsäranta, Heikki (2016) *The wider economic impacts of transport projects – Preliminary report,* Research reports of the Finnish Transport Agency 38/2016, Finnish Transport Agency, Project Planning Department. Helsinki.

Laine, Tomi – Käsnänen, Saku – Niinikoski, Miikka – Hollestelle, Martin – Ilmola, Roni (2019) *Esiselvitys yleisen liikennelaskennan kehittämisestä,* Väyläviraston julkaisuja 44/2019, Väylävirasto, Helsinki.

Laki liikennejärjestelmästä ja maanteistä (503/2005) Helsinki: Liikenne- ja Viestintäministeriö.

Lapp, Tuomo – Iikkanen, Pekka – Ristikartano, Jukka – Niinikoski, Miikka – Rinta-Piirto, Jyrki – Moilanen, Paavo (2018) *Valtakunnalliset liikenne-ennusteet,* Liikenneviraston tutkimuksia ja selvityksiä 57/2018, Liikennevirasto, Helsinki.

Liikenneverkon peittävyys pysyi lähes ennallaan vuonna 2022 (2023) Tilastokeskuksen kokeelliset tilastot, Tilastokeskus, Helsinki.

Manning, Alan – Petrongolo, Barbara (2017), How Local Are Labor Markets? Evidence from a Spatial Job Search Model, *American Economic Review,* Vol. 107, (19), p. 2877-2907.

Melo, Patricia C. – Graham, Daniel J. – Noland, Robert B. (2009) A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics,* Vol. 39 (3), p. 332-342.

Metsäranta, Heikki – Riukula, Krista - Kauhanen Antti – Fornaro, Paolo (2019) *The impacts of transport system changes on labor market and their evaluation,* Publications of the Government´s analysis, assessment and research activities 2019:69, Prime Minister's Office, Helsinki.

Metsäranta, Heikki – Viitanen, Kauko – Falkenbach, Heidi – Ekroos, Ari (2019) *Impacts of road and rail projects on real estate values and municipal contribution:*

Investigation from the "beneficiary pays" viewpoint, Publications of the Finnish Transport Infrastructure Agency 6/2019, Finnish Transport Infrastructure Agency, Helsinki.

Mohring, Herbert – Harwitzs, Mitchell (1962) *Highway Benefits; an analytical framework*, Northwestern University Press, Evanston, IL.

Moilanen, Paavo – Niinikoski, Miikka – Rinta-Piirto, Jyrki – Koponen, Ville – Haapamäki, Taina (2014) *Valtakunnallinen liikenne-ennustemalli*, Liikenneviraston muut julkaisut, Liikennevirasto, Helsinki.

Niinikoski, Miikka – Moilanen, Paavo (2017) Tieliikenteen suoritelaskennan kehittäminen, Liikenneviraston tutkimuksia ja selvityksiä, Liikennevirasto, Helsinki.

Niittylä, Laura – Varjola, Mika (2017) *Liikkumistiedon keruu ja analysointi,* Liikenneviraston tutkimuksia ja selvityksiä, Liikennevirasto, Helsinki.

Osallistava ja osaava Suomi – sosiaalisesti, taloudellisesti ja ekologisesti kestävä yhteiskunta (2019) Valtioneuvoston julkaisuja 2019:31, Valtioneuvosto, Helsinki.

Pastinen, Virpi – Piipponen, Johanna – Supponen, Atte – Vuorio, Lauri – West, Jens – Rantala, Annika – Niinikoski, Miikka – Rinta-Piirto Jyrki – Elolähde, Timo (2020) *Demand models for the Helsinki region commuting area traffic forecasting system 2020,* HSL Helsinki Region Transport, Helsinki.

Phang, Sock-Yong – Toh, Rex S. (2017) From manual to electronic road congestion pricing: The Singapore experience and experiment, *Transportation Research Part E: Logistics and Transportation Review*, Vol 33 (2), p. 97-106.

Pigou, Arthur (1920) Co-operative Societies and Income Tax, *The Economic Journal,* Vol. 30, No. 118 (p.156-162).

Pitkänen, Jukka-Pekka – Pesonen, Hannu – Salminen, Antti – Elmnäinen, Eeva – Musto, Maija – Rinta-Piirto, Jyrki (2020) *Helsinki-Turku-käytävän junaliikenteen matkustusennusteet ja liikennöintimallien vertailu,* Väyläviraston julkaisuja 26/2020, Helsinki.

Ratalaki (110/2007) Helsinki: Liikenne- ja viestintäministeriö.

Rautatietilasto (2020b) Tilastokeskus, viitattu 18.7.2023.

Salonen, Noora – Teittinen, Lari – Niittylä, Laura – Varjola, Mika (2018) *Selvitys yksityisen sektorin käytössä olevista liikkumistiedoista,* Liikenne- ja viestintäministeriön julkaisuja 20/2017, Liikenne- ja viestintäministeriö, Helsinki.

Santos, Georgina – Shaffer, Blake (2004) Preliminary Results of the London Congestion Charging Scheme." *Public Works Management & Policy*, Vol. 9 (2), p. 164–181.

Sarkka, Tapani – Kalenoja, Hanna – Tefke, Joni (2016) *Tulevaisuuden liikennemallit ja ennusteet: Kirjallisuusselvitys*, Liikenne- ja viestintäministeriön julkaisuja 11/2016, Liikenne- ja viestintäministeriö, Helsinki.

Singleton, Patrick, Allen (2017) *Exploring the Positive Utility of Travel and Mode Choice*, Dissertation and Thesis. Paper 3780, Portland State University, Portland.

Sinha, Kumares C. – Labi, Samuel (2007) *Transportation Decision Making Principles of Project Evaluation and Programming,* John Wiley & Sons, Inc, Hoboken, New Jersey.

Skabardonis, Alexander – Varaiya, Pravin – Petty Karl. F (2003) Measuring Recurrent and Nonrecurrent Traffic Congestion, *Transportation Research Record: Journal of the Transportation Research Board*, Vol 1856 (1), Transportation Research Board, NRC, Washington, D.C.

Stiglitz, Joseph – Carl Walsh (2006) *Economics*. Fourth edition W.W.Norton & Company, New York, London.

Tietilasto: tilaston dokumentaatio, Statistics Finland < https://stat.fi/tilasto/dokumentaatio/tiet>, haettu 19.7.2023.

Tie- ja rautatieliikenteen hankearvioinnin yksikköarvot 2018 (2022) Väyläviraston ohjeita 40/2020, Väylävirasto, Helsinki.

Timothy, Darren – Wheaton, William C. (2001) Intra-Urban Wage Variation, Employment Location, and Commuting Times, *Journal of Urban Economics,* Vol. 50 (2), p. 338-366.

Vickrey, William, (1969) Congestion Theory and Transport Investment, *The American Economic Review Papers and Proceedings of the Eighty-first Annual Meeting of the American Economic Association*, Vol. 59, No. 2 p. 251-260.

Example, Emily (2012) Title ipsum dolor sit amet, consectetur adipiscing elit. *Journal of Something*, Vol. 54 (3), 101–125.

Eka, Timo – Toka, Tuomas (2014) *Book and journal names italics*. University of Lugano, Department of Marketing, Working papers A:5.