Magnus Bjørnøy
Francis Emil Westerlund

Are Public Transport Subsidies Progressive?

An Empirical Analysis on the Distributive Effects of Public Transport Subsidies in Oslo and Akershus

Master's thesis in Economics
Supervisor: Colin Green
Trondheim, June 2018

Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

NTNU
Norwegian University of
Science and Technology

Preface

This thesis marks the end of our 2-year master at NTNU. We would like to express our deepest gratitude to our supervisor, Colin Green, for his insightful comments and constructive feedback. A special thanks goes to our two secondary supervisors, Jørgen Aarhaug and Nils Fearnley at the Institute of Transport Economics (TØI) for letting us be a part of their research project "Distributional Effects in Transport" and for providing the data set. They inspired our field of research and helped us along the way. Furthermore, we would like to thank Truls Angell at Ruter for providing swift and valuable answers to all our questions.

In addtion, we want to give tribute to our fellow students, and especially to the ones in our small reading room. It has been a pleasure to work, discuss and laugh by your side. Finally, we want to thank our families for always believing in us and for always being there.

This thesis is a joint work and any mistake is our own.

Magnus Bjørnøy and Francis Emil Westerlund.
Trondheim, June, 2018.

Abstract

The goal of this thesis has been to answer the question; Are public transport subsidies progressive? With data from the Norwegian Travel Survey 13/14 we disaggregate the area of analysis to Oslo and Akershus. By dividing the sample into groups based on income, we examine the distributive effects with a three-step methodology. We study progressiveness using a descriptive, econometric and a calculative method, under the assumption that distribution of subsidies depend on public transport usage.

First, the descriptive propensity approach finds the poorest group to be the most dependent on public transport. Our demand model demonstrates similar results both before and after demographic controls are included. These results pass our robustness checks. Estimated public transport use by income group is then used as an input to calculate the subsidies per group. This approach estimates that the poorest receive the most subsidies, especially when fare discounts are taken into account. Moreover, the poorest group receives on average 2 624 NOK in public transit subsidies per year, which is 38% more than the richest group. The yearly amount of subsidies received for the poorest group corresponds to 1.15% of their average income.

Each method demonstrates that the poorest group is the most dependent on public transport and receives the most subsidies. In addition, the demand model suggests that the dependency is driven by factors related to income and not solely demographic characteristics in each group. Therefore, our analysis indicates that public transport subsidies in Oslo and Akershus are progressive.

Sammendrag

Målet med denne oppgaven har vært å svare på dette spørsmålet; Er subsidier til kollektivtransport progressive? Med data fra den nasjonale reisevaneundersøkelsen 13/14, avgrenser vi analyseområdet til Oslo og Akershus. Ved å inndele utvalget i grupper basert på inntekt er vi i stand til å analysere fordelingseffektene med en tre-stegs metodologi. I oppgaven forutsetter vi at subsidiene avhenger av kollektivtransportbruk. Det gir at de med flest reiser mottar mest subsidier, slik at kan vi måle progressiviteten ved en deskriptiv, økonometrisk og kalkulativ metode.

Vår deskriptive metode viser at den fattigste gruppen bruker mest offentlig transport. Etterspørselsmodellen finner deretter tilsvarende resultater, både før og etter vi kontrollerer for demografiske variabler. Disse resultatene består så våre robusthetstester før vi bruker disse funnene til å estimere subsidiene til hver gruppe. Metodens resultater demonstrerer at de fattigste mottar mest subsidier, spesielt når billettrabatter er inkludert i utregningen. Den finner også at den fattigste gruppen mottar i gjennomsnitt 2 624 kroner i subsidier til kollektivtransport i året, som er 38% mer enn den rikeste gruppen. Den årlige mengden subsidier mottatt av den fattigste gruppen tilsvarer til 1.15% av deres gjennomsnittlige inntekt.

Metodene demonstrerer hver for seg at den fattigste gruppen har høyest bruk av offentlig transport og dermed mottar mest subsidier. I tillegg viser etterspørselsmodellen at bruken av kollektivtransport er drevet av inntektsrelaterte faktorer og ikke bare demografiske karakteristikker i hver gruppe. Vår analyse indikerer derfor at subsidier til offentlig transport i Oslo og Akershus er progressive.

Table of Contents

1 | Introduction

The Norwegian government‘s equity goal is to create a fair resource allocation and good living conditions for all (Ministry of Finance 2001, p. 42). Financial policy such as taxes and subsidies are often used as tools to achieve these goals. By allocating these subsidies to public transport, the government‘s aim is to increase mobility for people and goods in the whole country, reduce transport accidents in line with Vision Zero[1] and reduce other negative environmental consequences (Norwegian Ministry of Transport and Communications 2016, p. 10). Efficiency and environmental effects of these subsidies are widely researched, however the equity effects have not gained as much attention in the economic literature, despite the governments goals of ensuring a fair resource allocation. We contribute to the literature by studying the distributional effect of subsidies to public transport.

Our analysis employs cross-sectional data from the national travel survey in Norway from 2013-2014. The scope of this analysis is Oslo and Akershus, because these areas reflect favorable characteristics for studying our topic. First, the high proportion of public transport travellers allow for comparisons across different population groups. Second, the large variety of different transport modes, for example bus, tram, metro and train, makes it possible to examine whether some transport modes attract specific demographic or income groups. Third, our disaggregated geographic area is more suitable to study than Norway as a whole. One reason is that Norway differs substantially in population density and topography which has a major influence on the general access to public transport, and hence travel habits. Finally, each county is responsible for the provision of public transport, but only a few, such as Oslo and Akershus (through Ruter) provide detailed data regarding transit subsidies.

1.1 Public Transport in Oslo and Akershus

In order to examine the distributive effects of public transit subsidies in Oslo and Akershus, we must first know how the system is organized.

The infrastructure of national railway is managed and maintained by BANE NOR, while The Norwegian Railway Directorate coordinates all transport by trains and operating train companies, such as NSB and Flytoget[2]. NSB manages train transportation of passengers and goods in the whole country, while Flytoget only manages passenger transport between the Oslo Area and Gardermoen. All of these companies and directorates are state-owned and governed

[1] "The vision for the road safety work in Norway is to strive for zero fatalities and zero serious injuries. The policy is referred to as the Vision Zero" (Langeland 2009, p. 7).

[2] Also known as the Airport Express Train.

by the Ministry of Transport and Communications which are responsible for the strategic management, except Flytoget, which is governed by the Ministry of Trade and Industry.

Bus, tram and metro are managed by Ruter AS, which is a publicly owned management company for public transportation that plans, coordinates, orders, and markets public transport. Ruter does not operate public transport themselves, but have permanent contract partners for operating the metro and tram, in addition to agreements with other bus operating companies that are chosen through competitive tendering. The company that can provide the best total solution[3] will enter into a gross costs contract, which means that Ruter pays the operating company a fixed amount for their services. Ruter still receives the full revenue from tickets sales and decide ticket types and prices.

The ticket price depends on the how many geographical zones are passed, how long the ticket is valid and if a person is eligible for discounts based on age or occupational status. A cooperation agreement between Ruter and NSB gives ticket holders access to all public transport modes in Oslo and Akershus, including trains. NSB have a different price structure than adopted in Oslo and Akershus county. As a result, these counties compensate NSB (through Ruter) with a yearly amount to cover the loss in revenue. In 2013, Ruter's total revenue was 6.518 billion NOK in which 52% was generated from traffic revenues and 48% from subsidies. The latter is allocated by Akershus and Oslo county.

1.2 Research Question

This thesis analyses the distributional effect of subsidies to public transport. By using travel survey data from Oslo and Akershus, we examine if subsidies are progressive (pro-poor) or regressive (pro-rich). We assume that subsidies received are dependent on public transport usage, i.e. the most frequent users receive the most subsidies. By then dividing people into income groups, we can explore whether public transport dependency is related to income or driven by other factors.

The research is based on a three-step methodology. First, a descriptive approach is used to measure how public transport dependency differs between income groups. Second, econometric models are used to measure the same, before and after controlling for demographic variables. The goal is to examine whether the results in the first approach still hold or if demographic characteristics drive the public transport dependency, and a potentially progressive subsidy system. As part of this we evaluate the robustness of our results by performing several checks. Third, we calculate an estimated subsidy per group based on these

[3]The best total solution is picked based on price, quality, and service.

results, in order to demonstrate the degree of progressivity and the magnitude of the distributional effects that are generated by public transit subsidies.

The thesis is structured into 8 chapters. Chapter 2 covers a review of previous research in the field and economic theory of rationale for public transport subsidies. Chapter 3 and 4 cover the data material used and the methods used to analyze distributional effects. Results and robustness analysis are presented in Chapter 5 and 6, before we calculate the subsidies and distributive effects in Chapter 7. At last, we present a discussion and our final conclusion in Chapter 8.

2 | Theory and Literature

In this chapter, we first present and discuss the rationale for allocating subsidies to the public transport sector. Additionally, the previous literature in the field is reviewed and discussed.

2.1 Rationale for public transport subsidies

There is no such thing as a universally optimal price of a good or service. The optimal price depends on the goals and objectives of individuals, groups or governments. In traditional economics, the standard assumption is that firms set the price such that profits are maximized. Our focus is on welfare economics[1], where prices are used as a method to achieve a resource allocation that maximizes social welfare (Elgar & Kennedy 2005, p. 72). Taxes and subsidies are typically used to incentivize or control prices such that the policy goals related to social welfare can be achieved. Social welfare can, in general terms, be maximized by maximizing the equation below (Elgar & Kennedy 2005, p. 72),

$$Social\ welfare = Total\ revenue + Consumer\ surplus - Total\ costs$$

$$\begin{aligned} \text{where:}\quad Total\ revenue &= \text{Total revenue for producer} \\ Consumer\ surplus &= \text{Consumers willingness to pay} - \text{Actual purchasing price} \\ Total\ costs &= \text{Variable costs + fixed costs + external costs} \end{aligned}$$

Social welfare is maximized when the price is equal to social marginal costs. This implies that the prices reflect both the producers marginal costs and the external costs. External costs in the public transport sector can be related to queuing, pollution, wear and tear or noise. In the following, we will present the rationale for subsidies and discuss whether it is an efficient tool to reduce car use.

Economies of scale

One of the justifications for subsidizing public transport is it‘s natural monopoly. Railway companies are potentially monopolies due to economies of density that arise from the fixed costs of infrastructure. With scale economics, the average costs and marginal costs are declining

[1] "Welfare economics is the economic study of the definition and the measure of the social welfare; it offers the theoretical framework used in public economics to help collective decision making, to design public policies, and to make social evaluations" (Baujard 2013)

with increased output, and marginal costs are always below average costs. A market solution, where price equals average costs, will produce a dead weight loss. In a theoretical framework, one could argue that the operator must receive a subsidy to ensure coverage of costs, such that the socially optimal production can be obtained (price = marginal costs) (Preston 2008, p. 189). However, whether the rail industry exhibits significant economics of scale effects have also been questioned (Graham et al. 2003, p. 456). Furthermore, the bus industry is known to have a constant returns to scale (Preston 2008, p. 189). An argument in favor of subsidizing buses is user economics of scale, which is often referred to as the Mohring effect (Mohring 1972). Increased demand for public transport leads to more frequent transit, which benefit consumers through reductions in passenger waiting times. Mohring argues that the reductions in passenger waiting time is a reduction in passengers generalized costs of using public transport. A market solution does not take these costs into account, and subsidies are therefore necessary to reduce fares and increase output such that the benefits of user economics of scales can be obtained. Studies from the UK suggests that the optimal transport system would require subsidies to cover 50 percent to 75 percent of costs (Preston 2008, p. 190). In comparison, Ruter receives subsidies just below these numbers (Ruter 2017, p. 22).

Merit goods

The public sector promotes consumption of merit goods because they produce positive externalities. Positive externalities are benefits that a third party receives without consuming or purchasing the good that creates the positive externality. The public sector fear that these goods will be under-consumed because positive externalities are not taken into account in individuals decision-making processes.

Public transport is an example of a merit good that can be under-consumed. People with low income might not be able to travel because they cannot afford the fares. That is a problem for two reasons. First, because public transport creates positive externalities and other amenities by providing easy access to public health services and education. These externalities are not realized in a market solution. Second, because it is politically desirable, from an equity and fairness standpoint, that everyone in a society has access to transport services. The notion is that everyone should have the right to access to transport, such that basic needs such as education facilities and health care services can be provided to all citizens. (Button 2010, p. 97). These objectives are not met in a free market, and as a consequence, public transport subsidies are necessary to accomplish them.

The effectiveness of public transport subsidies

A market solution does not take into account the full cost of car use, as costs related to pollution, infrastructure, congestion, road safety risks and other environmental externalities[2] are not included in the market price for cars (Serebrisky et al. 2009). Hence, a market without regulation will allow cars to be under-priced (Becker et al. 2012).

Taxes are one of the financial tools that are applied to reduce external costs related to cars use. By taxing car use, the marginal costs increase and usage will be reduced. If the socially optimal tax is enforced, the marginal costs will be equal to the social marginal costs. Hence, all the external costs will be internalized in the price of cars and the socially optimal production of the good will be produced.

Another way to reduce car use is through subsidizing substitutes, such as public transport. If an increase in subsidy leads to a reduction in car use, it can be viewed as an effective policy tool to reduce car externalities. The validity of the argument for subsidizing public transport depend on these three factors:

1. The price elasticity of demand for transit modes [3]
2. The cross-elasticity of demand between transit and private cars [4]
3. The social marginal costs of congestion, pollution, wear and tear, accidents and noise associated with private cars.

The price elasticity of demand for metro and bus is estimated to be around -0.3 (de Grange et al. 2013, p. 179). This means that a 1% reduction in the ticket price for bus and metro increase demand for bus and metro with 0.3%. An elasticity of -0.3 shows that subsidies used to reduce ticket prices have an effect on public transport demand, but that the effect is limited.

Fearnley et al. (2017, p. 67) examined a number of reports on the cross-elasticity of demand between transits and private cars, and found the average cross-elasticity to be 0.055. This suggests that a 1% reduction in public transport fares on average reduces car use by 0.055 %. Thus, demand for cars is not sensitive to fare changes. This weakens the argument for subsidizing public transport because even very low fares does not necessarily influence car users such that they substitute cars for public transport.

Furthermore, there is no need for subsidies if the social marginal cost of car use is equal to the individual's marginal cost. That can be achieved by taxes on car use, such that externalized costs

[2]" Externalities exist when the activities of one group affect the welfare of another group without any payment or compensation being made."

[3]The change in frequency of ridership with transit modes as a result of changed price.

[4]The change in demand for transit modes when the price of private ridership changes

become internalized. Such a solution is also likely to be more efficient than using subsidies.

We conclude that public transport subsidies have a limited effect on reducing car externalizes, as both price elasticity and cross-elasticities are fairly low. Anyhow, there are efficiency argument that support subsidizing public transport such as economics of scale and the mohring effect. It is also desirable from a social equity standpoint to provide public transport access for all.

2.2 Literature review

The progressiveness of public transport subsidies and the redistributive effects that follow is not a heavily researched topic in transport economics. However, some empirical research does exist. Frankena (1973) came to the conclusion that transit subsidies in Canada had regressive distributional effects, while Pucher (1983) found the opposite in six American metropolitan areas. Of the more recent articles, Asensio et al. (2003) and Fearnley (2006) conclude that subsidies for public transport are progressive for the most part, but that it can vary between transport modes. The latter articles have influenced our choice of method and model. Their methods and results are therefore briefly presented before they are summarized and discussed at the end of this section.

"Redistributive Effects of Subsidies to Urban Public Transport in Spain"

Asensio et al. (2003) analyzed distributive effects generated by subsidization of urban public transport services in Spain. The analysis was based on Spanish microdata from the Spanish Household Budget Survey (Encuesta de Presupuestos Familiares, EPF) of 1990 to 1991 with households as the cross sectional unit. The biggest cities in Spain where included in the analysis (Spain, Barcelona, Valencia, Sevilla, Malaga and Zaragoza).

By applying a two-stage model, they first estimated the probability of using public transport, before regressing a log-log model in order to find the elasticity of income on urban public transport expenditures. The last model is conditional on the first, which means that if the probability of taking public transport is zero, the expenditures in the last model will not be observed. Asensio et al. (2003) also estimated the probabilities of owning of 1, 2 or 3 cars, as car ownership is a major factor in the choice of transport mode. Another aim of including these variables in the two-stage model is to interpret the income elasticity as a long-run elasticity. Increased income influences the probability of owning a car, which in turn affects the probability of taking public transport and therefore public transport expenditures. The other variables included are household specific, such as the number of members or the amount of employed in the household, with the exception of a variable that provides a measure of road quality.

The article define subsidies as the difference between annual revenue and annual costs. By assuming that the passengers are the sole beneficiary of these subsidies, they can divide this amount by the passengers. They start with the groups with discounted tickets (such as students, disabled or retirees) then assign the rest of the subsidies to the groups with ordinary tickets. The result is subsidy per trip per household, given the assumption that these groups have the same travel pattern. When each groups total amount of subsidies are calculated, subsidy as a percentage of costs can be calculated by dividing total subsidies on the total amount of costs in a group (Asensio et al. 2003, p. 436). Subsidies are progressive if the subsidy (as a percentage of household income) decreases when income rises. By looking at different groups and cities in Spain, Asensio et al. (2003) found that subsidies on public transport are progressive, but that the redistributive effects are relatively small.

"Public Transport Subsidies in the UK: Evidence of Distributional Effects"

Fearnley (2006) analyzes the difference in dependency on public transport between income groups, using a national travel survey (NTS) from 1995-1997 in the UK. Like Asensio et al. (2003), this article aims to find the progressivity of the subsidies granted for public transport. It also relies on the assumption that passengers are the main beneficiary of subsidies, but as subsidies are not calculated in this article, some leakage is allowed as long as the passengers receive most of the subsidies. The article also presents a discussion toward the use of expenditures as measurement of dependency and why the number of trips is the favoured measure. We elaborated on this in Section 4.1. In order to estimate progressivity, Fearnley (2006) applies a propensity approach to measure the dependency on public transport. However, the analysis cannot estimate the magnitude of the distributive effects. The article also analyzes how these propensities vary between gender and age groups.

Fearnley (2006) finds that bus subsidies in the UK are quite progressive. As income increases, dependency on bus transport decreases markedly. On the other hand, train usage has the opposite pattern and therefore suggests a largely regressive subsidy system. Gender propensities revealed that females are substantially more dependent than men on transport by bus while train is equal at 1 for both genders. The results based on age groups indicated that young adults are by far the most dependent group on both transport modes. They are followed by the oldest population (retirees among other) with regards to bus, while the second highest dependency on trains belongs to adults between the age 25-59.

Summary and Discussion

Both articles estimate a measure of public transport subsidy progressiveness using different approaches. While the results in Spain was based on public transport as a whole, Fearnley (2006) analyzed bus and train transport separately. Asensio et al. (2003) findings where progressive, Fearnley (2006) found clear opposites in progressiveness between bus and train. This motivates us to focus on public transport as a whole, but still examine different transport modes to see if dependency differs substantially between them.

Similar to our analysis, both articles apply cross sectional data collected through national surveys. In order to analyze the progressivity and calculate the redistributive effects similar to Asensio et al. (2003), the articles are conditional on several assumptions. Some of these assumptions are incorporated in this analysis as well and are presented in Chapter 4.

As both articles examine how income influences the dependency of public transport and the income groups that benefit most from public transport subsidies, we can compare methods and results at the end of this analysis. However, the findings by Asensio et al. (2003) and Fearnley (2006) are not necessarily comparable to the Norwegian population and our sample from 2013-2014. Not only are the results based on data from 1990-1991 and 1995-1997, which is at most 24 years ago, but there are also some cultural and demographic differences to consider. Spain is probably the least comparable of the two countries, due to large differences in GDP per capita[5] and income distribution[6]. While the UK is relatively closer to Norway in GDP per capita, it has a relatively higher degree of income inequality than Spain prior to 2008. Fearnley (2006) also looked at the UK as a whole and did not delineate the area of analysis to major cities like Asensio et al. (2003).

Nonetheless, the articles have provided insight into methods and assumptions that guide and shape this analysis. Even if the magnitude of progressivity and distributional effects are difficult to compare between countries, the signs and patterns of the results are still relevant.

[5]GDP per capita can be used as a measure of welfare. Spain's GDP per capita was 14800 USD in 1991, while Norway was close to twice this amount in the same year (The World Bank 2018)

[6]The Gini coefficients, which is a measurement of income equality, presents a more unequal income distribution in Spain, compared to Norway in 1990. In addition, the difference in income equality increases over time (Atkinson et al. 2017).

3 | Data

This section provides a summary of the data set used in our empirical analysis, as well as a discussion of the benefits and disadvantages. We then present our data adjustments and the variables we intend to use in our econometric models. At last, we present some relevant descriptive statistics.

3.1 The National Travel Survey 2013/2014

Our empirical analysis uses data from The National Travel Survey 2013/2014 (RVU)[1], which is the largest national travel habit survey in Norway and contains information from approximately 61 400 respondents all around the country, all above the age of thirteen. The data consists of a base sample of 10 000 interviewees, which are distributed geographically close to proportionally with the population in Norway (Hjorthol & Uteng 2014, p. 4). This means that if 2 percent of the base sample lives in Finnmark[2], then approximately 2 percent of the Norwegian population should reside in this county. The remaining observations have been gathered through additional regional samples[3].

The information is divided into three data files; personal data, daily journeys and information regarding long journeys. The first file contains information on individual characteristics, such as age, gender, income and residence, as well as household characteristics. Additionally, it covers detailed information on respondents access to both public and private transport. The second file is made up of individual‘s travel diaries. The travel diary contains information on the number of journeys, purpose of journeys and more. Within these journeys, the respondents also specify the different transport modes used to reach their travel destination. The last file consists of information about domestic journeys with a length of minimum 100 kilometers, and journeys abroad.

Benefits and Disadvantages with the RVU Data

To ensure representativeness, the base sample and the regional samples have been randomly drawn from the National registry in Norway [4]. In addition, the interview process for each

[1] Conducted by the Institute of Transport Economics on behalf of Avinor, The Ministry of Transport and Communications, The Norwegian Coastal Administration, The Norwegian National Rail Administration and The Norwegian Public Roads Administration (Hjorthol & Uteng 2014).

[2] The most northern county in Norway.

[3] On behalf of The Norwegian Public Roads Administration and regional authorities, who aim to apply these samples in regional analysis.

[4] "The national registry contain information of everyone that resides or have resided in Norway" (The Norwegian Tax Administration 2018).

person is identical. The chosen individuals first receive a letter describing the process and the aim of the survey. This letter also contains a travel diary to record their journeys on the pre-selected registration day. Ahead of this day, the individuals receive a phone call with the aim of motivating and helping the potential respondents. A few days after the registration day the actual interview takes place via telephone (Hjorthol & Uteng 2014, p. 3). In addition to the travel activity, they collect personal and relevant information from the respondents, but also from other members of the household. To observe daily and seasonal variation in travel activity, the data collection is continuous over a whole year, including every day of the week except special holidays. The total sample therefore provides a reliable presentation of the travel activity and travel habits of the Norwegian population within a year.

The response rate is 20 percent, and has been declining since the first RVU in 1985 (Hjorthol & Uteng 2014, p. 5). Difficulties in making contact with individuals and technical issues caused two thirds of the non-participation. Declining response propensity in the population explains the last third. The low response rate can make it more difficult to determine whether a sample is random or not. However, as long as the selection is random and the non-response is not based on clear systemic characteristics, a low response rate alone should not be enough to condemn data material (Hellevik 2015, p. 226).

To avoid a loss of representativeness due to additional regional samples, the RVU have been weighted to correct for geographical differences in the probability of being drawn (Hjorthol & Uteng 2014, p. 6). It means that regions which are over represented have a lower weighting than regions which are under represented. The weighting is also adjusted for age, weekday and season within the regional zones. This means that if you look at the age distribution in Oslo for instance, it coincides with the actual population distribution in this area based on data from Statistics Norway (Gregersen 2017, p. 9-11). The weights will therefore, in some cases, contribute to unbiased estimates, while in other cases, uncritical use of the weights may lead to the opposite.

One of the weaknesses with the RUV-data is the lack of weighting for educational levels. Urbanet Analyse (2015) discovered that individuals with higher education are over-represented in the national sample as well as in the Oslo area, compared to population data from SSB (Urbanet Analyse 2015, p. 5). This is illustrated in Figure 3.1[5]. As a result, RVU is not representative of the target population if there are substantial differences in travel activity between education levels. This can lead to econometric problems in the model, which will be further discussed in Chapter 4.

[5]Borrowed with permission from the report made by (Urbanet Analyse 2015, p. 5). A comparison between the RVU 13/14 and the Statistic Norway's educational statistics during the same period.

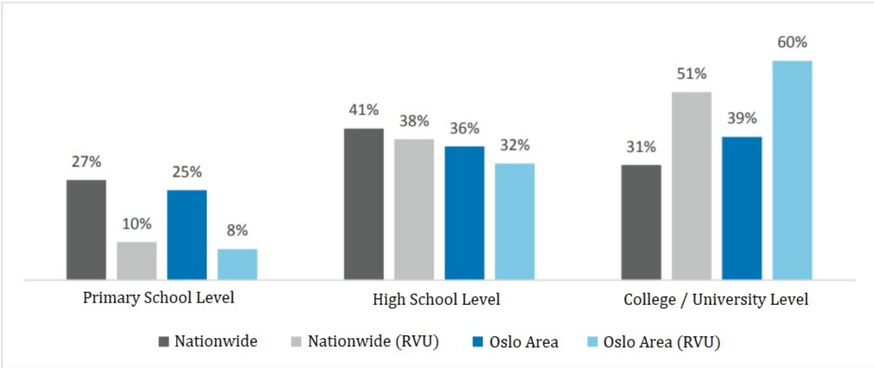

Figure 3.1: The distribution of education levels in the population versus the RVU sample. For individuals above the age of sixteen (Urbanet Analyse 2015, p. 5).

Figure 3.1 clearly display that individuals with higher education are over represented in the RVU relative to the true population, both nationwide and in the Oslo area. It can also be an indicator for additional systematic biases. As education and income tend to be heavily correlated there is reason to believe that there may be a bias in income as well. Are the individuals with a low education level not participating in surveys due to opinions regarding the survey or are there difficulties in making contact with this group? One explanation for the lack of participation from the low and middle education levels can be the contribution from immigrants in the survey. In comparison with the rest of the population, they are over-represented in having no education, or having only primary school as their highest education based on data from 2016 (Directorate of Integration and Diversity 2018). The Directorate of Integration and Diversity argues that immigrants tend to have a high degree of mobility, which in turn provide challenges when it comes to contacting these individuals (Directorate of Integration and Diversity 2010).

3.2 Data Adjustments

As the scope of this thesis is shorter travels on public transport within a certain area, we have chosen to exclude the last file of the RVU in our analysis. The file containing long journeys of 100 kilometers or more. One reason is that the pattern of subsidies for these types of journeys are likely to be very different. So is the frequency of these journeys. Another point is that tram and subway do not cover such distances and therefore have zero observations.

The two remaining files were therefore merged. The file containing personal information was structured in a wide format, as expected from cross-sectional data, but the daily travel file was shaped in a long format, somewhat similar to a panel data set. In the sense that each individual had multiple observations based on how many journeys were registered. This posed a challenge when merging these files into one data set, as the detailed data from each journey had to be extracted into new variables before the merging, to not lose any valuable information. While all the variables in the personal data were unaffected, most of the journey specific variables from the daily travel file had to be counted or adjusted. The main reason for merging the data into a wide format rather than a long one, was so we could have one observation per individual. The latter would be optimal when analyzing changes over time, but seeing as the personal information and all the individual journeys were made within one day, there are little to none variation in most of the variables.

After structuring the new variables, the duplicates for each individual are removed before combining it with observations from the personal information file. The 5 888 respondents who did not merge, had no registered journeys in the daily travel file. Determining the reason for their information absence proves quite the challenge. Lack of transportation options or income-related choices toward public and private transport would clearly be relevant. However, they could simply have decided not to answer, forgot it, or not left their house due to their own will or potential illness. As any of these arguments are not certain, the unmatched individuals are removed from the data set.

Due to the geographical area of our research question, the next step was to narrow it down to individuals residing in Oslo and Akershus. In order to perform an income adjustment and grouping, we exclude the respondents who did not answer, wanted to specify or did not know their household income at the time. Additionally, all other observations with missing data or unspecified answers[6] are removed. At last, we remove 422 individuals who are registered with tickets that are free[7], as their decision-making process to use private or public transport is not comparable to others. This brings the number of observations down to 8 861 respondents, which will serve as the main sample for this empirical analysis.

3.3 Model Variables

This section provides a presentation of the variables in our data set which we intend to use in our econometric demand model. We start with the dependent variable before the explanatory variables and summary statistics are presented at the end.

The Dependent Variable

Our main measure of demand for public transport is the total number of trips with public transport per day per person, which is represented by the dependent variable $Trips$. Trips by

[6] Answers such as "I do not know" and "I do not want to specify."

[7] Ticket types denoted as "Free card", School card (no cost) and TT-card (which targets disabled individuals who need alternative transportation methods than Ruter can offer)

bus[8], metro and tram are defined as public transport trips, while train is excluded. The justification for this exclusion is explained later in Chapter 4. The same chapter also contains a discussion to why number of trips is a preferred measure of dependency on public transport.

Independent Variables

The following variables will be used in this empirical analysis to explain the progressivity of public transport subsidies. First, our main variable of interest, namely income, is presented. We then continue to display variables which will serve as demographic controls to the effect of income on public transit usage. Also, note that all variables in this subsection are binary. It means they take the state of "Yes" or "No", given by the values 1 and 0.

Income

We have access to both individual and household income in our data. By adjusting the household income we obtain a better unit of comparison than individual income as it takes into account the size and composition of the household. As a result, adjusted household income is more comparable and representative of actual purchasing power per individual than individual income. In order to perform the adjustment, an equivalence scale is applied. This provides an expression of how much income a household need in order to have the same economic welfare as a single individual with a given income. By dividing the household income on the square root of the number of persons in the household, the income is adjusted toward the respondent. This method is appropriately named the Square Root Scale, and provides more weight to economies of scale than similar equivalence scales[9].

The relationship between income and public transport is not necessarily non-linear, as usage can differ between income levels. To find a more precise effect of income on public transit usage, we divide the individuals into five groups based on adjusted household income. Why five groups? We see no wrong in choosing fewer or more groups, but the income levels and its effects can be harder to interpret when the size of the intervals are very small or large. It is therefore preferable to find a middle ground, especially when our data has limited variation in income[10]. As Le Grand (1982, p. 116) and Fearnley (2006, p. 34) also used five groups in their research, we can easier compare our results. However, four or six groups can be just as adequate. That is why we test whether the number of groups have an effect on our results as a part of our robustness analysis in Chapter 6. Each income group represent close to 20% of the population. Summary statistics for each income group is presented below in Table 3.1 .

[8]We count both city and regional bus trips as $Trips$, as our data set cannot separate between these.

[9]Alternative equivalence scale measures such as the EU-scale and the OECD-scale are reviewed in Section 6.3.

[10]The limited variation is visualized graphically in the appendix, figure A.1, with a histogram of adjusted income.

Income Group	Income Interval	Number of People	Share of Population
1	25 000 - 350 000	1 656	0.19
2	353 553 - 491 935	1 708	0.19
3	494 974 - 519 615	1 649	0.19
4	550 000 - 635 085	1 622	0.18
5	636 396 - 1 100 000	2 226	0.25
Sum	-	8 861	1

Table 3.1: Income intervals for income group 1-5. The limited variation in income, results in an unequal number of people in the income groups.

The poorest group is defined as $Incomegroup1$, and is set as the reference category. The other groups have similar variable names corresponding to their group number.

There is a large literature on the effect of income on demand for public transport. Using data from the UK, Paulley et al. (2006, p. 303) found that the effect of income on bus demand is significant, with a long run income elasticity in the range -0.5 to -1 percent. This means that a 1% increase in income lowers bus demand by 0.5 to 1 percent. Given that tram and metro have similar income elasticities, we expect individuals in low income groups to use more public transport than the other groups. Note that these estimates include the indirect effect of car ownership on public transport demand. Further discussion on the link between car ownership and income and public transport demand will be presented in Chapter 6.

Oslo

The variable takes on the value 1 if the person lives in Oslo County , 0 if the person lives in Akershus county. As Oslo is geographically smaller and more densely populated, it provides better access to public transport than the municipalities in Akershus County. (Christiansen et al. 1972, p. 20). For this reason, the proportion of the population in Oslo that use public transport is higher than in Akershus (Hjorthol & Uteng 2014, p. 25)

Figure 3.2 display the share of environmental friendly transport in central Oslo versus share of car based transport. Environmental friendly transport includes bike, walking and public transport, while car based transport represents car drivers and passengers. The figure illustrate that public transport use, cycling and walking is more common in central Oslo than less urban areas. 35% of the our sample population lives in Oslo, and we expect that these individuals use more public transport than those who live in Akershus.

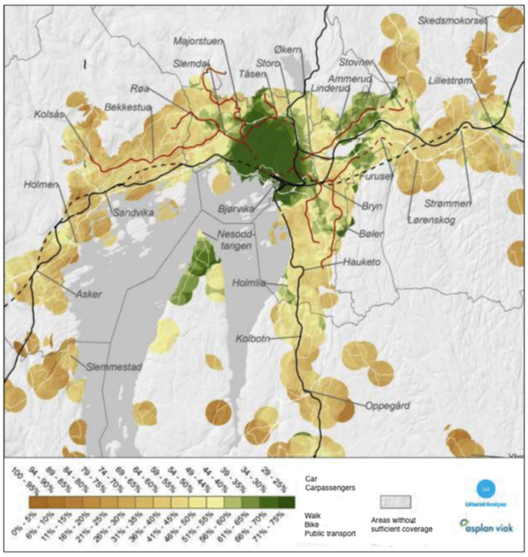

Figure 3.2: Share of environment friendly transport versus share of car based transport, based on base statistical units (Steinsland et al. 2016)

Age

The relationship between age and demand for public transport is likely to be non-linear, as children, adults and retirees have different travel habits. Reflecting this, rather than introducing age as a linear term in our analysis, we create a number of age category dummy variables. This allows for a more flexible relationship between age and transportation usage. The age dummies are created to capture these differences. They represent people within the age intervals 13-17, 18-25, 26-60 and 61-92.

Male

Our data set consists of 51.8% males, which is slightly higher then the national average of 50.2% (Statistics Norway 2014). Hjorthol & Uteng (2014, p. 26) found that women in Norway travel more frequently on public transport than men. This leads us to expect that the the male dummy will have negative effect on public transport use.

Children

The variable takes on the value of 1 if one or more of the household members are children[11]. Hjorthol & Uteng (2014, p. 14) found that couples with children travel longer distances than people without children and that the most frequent users of public transport are singles without children. Strathman et al. (1994) also found that household composition and children does effect travel behavior, as children are dependant on their parents to travel. This increase the demand

[11] Household members under the age of 17 are defined as Children

for transport for families since children attend day care, school or other activities. Hjorthol (2006, p. 10) suggests that parents also view driving their children around as a aspect of being a good parent. We therefore expect individuals with children to use less public transport than people without.

Other Demographic Variables

In addition to the variables already described, we will briefly introduce the remaining ones. $Higher\ education$ represent individuals who have a university degree. $Worktrip$ and $Free\ parking$ are both related to employment and describe if a person has one or more trips to work during a day and if a person has access to free parking at work. Free parking incentivize car use as it makes it cheaper and more convenient, and hence makes public transport relatively less preferable. Vibe et al. (2005) studied the effect of free parking and found that access to free parking reduce the probability of using public transport. 41% in our sample have free parking at work. At last, our weekend and seasonal dummies, which are used to control for seasonal and daily differences between income groups. $Weekend$ denotes if the day of registration is either Saturday or Sunday, while season is divided into $Fall$, $Winter$, $Spring$ and $Summer$[12]. The latter season is set as the reference category as travel activity is expected to be lower due to vacation from work.

Table 3.2 displays the summary statistics for all these variables.

[12]The division is based on the Northern Meteorological Seasons.

Table 3.2: Summary statistics for relevant variables.

Variable	Mean	Std. Dev.	Min.	Max.	N
Trips	0.529	1.101	0	9	8861
Income group 1	0.187	0.39	0	1	8861
Income group 2	0.193	0.394	0	1	8861
Income group 3	0.186	0.389	0	1	8861
Income group 4	0.183	0.387	0	1	8861
Income group 5	0.251	0.434	0	1	8861
Oslo	0.348	0.476	0	1	8861
Age 13-17	0.029	0.168	0	1	8861
Age 18-25	0.075	0.263	0	1	8861
Age 26-60	0.62	0.485	0	1	8861
Age 61+	0.276	0.447	0	1	8861
Male	0.508	0.5	0	1	8861
Children	0.299	0.458	0	1	8861
High education	0.621	0.485	0	1	8861
Work trip	0.449	0.497	0	1	8861
Free parking	0.414	0.493	0	1	8861
Summer	0.238	0.426	0	1	8861
Weekend	0.226	0.418	0	1	8861
Spring	0.268	0.443	0	1	8861
Fall	0.229	0.42	0	1	8861
Winter	0.264	0.441	0	1	8861

3.4 Descriptive Statistics for Income Groups

The progressivity of public transport subsidies are based on the effect of certain income levels on public transit usage. How the groups differ in composition can be a key element in the interpretation of these effects. Therefore, this section presents descriptive statistics for each group in regards to demographic characteristics and travel activity.

We start by illustrating the occupational status and age for different income groups, given as proportions, in Table 3.3.

	OCCUPATION				AGE			
Income Group	Employed	Retiree	Student	Other[13]	13-17	18-25	26-60	61-92
1	0.43	0.23	0.22	0.12	0.04	0.21	0.47	0.28
2	0.62	0.11	0.20	0.07	0.06	0.07	0.64	0.23
3	0.61	0.30	0.03	0.07	0.01	0.04	0.55	0.40
4	0.88	0.01	0.08	0.03	0.05	0.05	0.86	0.03
5	0.77	0.19	0.01	0.03	0.001	0.02	0.59	0.39

Table 3.3: Presenting the share of different occupations and age groups within the income groups.

Age group 13-17 represent teenagers who are too young to have a drivers licence. Many within this age group live with their parents, which means that they must rely on their parents, public transport or other modes of transport in order to travel. Age group 18-25 represents young adults. The average salary in this group is low reflecting that a large proportion of the group are students or employees without little to work experience. Age group 26-60 is defined as the work force group because the majority in this group are long term employed. Many within this group are married and have children. Age group 61-92 represent retirees, as a large fraction within this group are retired, or are close to retirement. People within this group are likely to live with their spouse or alone [14]. Table 3.4 show demographic statistics about sex, education level, car ownership and residence. All the numbers in this table represent proportions.

Income Group	Male	Oslo	Higher Education	Car Ownership
1	0.44	0.42	0.44	0.63
2	0.51	0.26	0.55	0.92
3	0.48	0.38	0.58	0.82
4	0.54	0.29	0.76	0.97
5	0.56	0.38	0.74	0.90

Table 3.4: Share of individuals being male, living in Oslo, has a higher education and who owns a at least one car, within the different income groups.

The distribution of men and women is close to equal, as none of the income groups consists of more than 56% or less than 44% males. Only income group 4 and 5 are slightly over-represented by men, which can be explained by the income gap between men and woman in Norway (Hirsch et al. 2010, p. 14).

The share of people living in Oslo vary across income groups. We expect income to increase with the share of people living in Oslo because rent and housing prices are higher in Oslo than

[13] Individuals that are unemployed, homemakers, in military service , on maternity leave, have a long term illness or are disabled.

[14] Distribution of household members for age group 61-92 shown in appendix A.2.

Akershus (Statistics Norway 2017). Surprisingly, the highest share of people living in Oslo can be found in income group 1. A possible explanation is the relatively high proportion of young people (Table 3.3), that are likely to have a preference to live in urban areas.

Individuals with higher education have higher incomes, and the difference is substantial. Only 44% of the ones in the lowest income group have higher education, while 74 % of those in the highest income group have higher education. People in income group 1 own far less cars than people in all the other groups. Only 63% own cars in group 1, while the share of car ownership in all the other groups are above 80%.

Table 3.5 present the total number of trips in each income group and the share of trips made with different transport modes. What is interesting to observe is that the share of total trips per income group is surprisingly similar to the share of individuals within the same groups displayed in Table 3.1. It means that the groups approximately travel the same amount per day.

	TOTAL TRIPS		CAR[15]	PUBLIC TRANSPORT	BUS	TRAM	SUBWAY
Income Group	Number	Share	Share	Share	Share	Share	Share
1	6980	0.19	0.38	0.18	0.09	0.03	0.06
2	7116	0.19	0.55	0.10	0.06	0.01	0.03
3	6684	0.18	0.47	0.13	0.6	0.02	0.04
4	7370	0.20	0.54	0.10	0.05	0.01	0.03
5	9169	0.25	0.49	0.12	0.06	0.02	0.04

Table 3.5: Transport statistics: The share of trips made with different transport modes compared to total trips.

The most used public transport mode in every group is bus, followed by subway and tram respectively. This makes sense seeing as the bus covers larger parts of Oslo and Akershus than the other two transport modes. Table 3.5 also illustrate the fact that that group 1 travel relatively more with public transport. These transportation patterns are elaborated on when the results from our descriptive propensity approach are presented in Chapter 5.

[15] As driver and as passenger

4 | Method

To study the distributive effects of public transport subsidies, one would ideally prefer detailed information regarding the subsidies, such as subsidy per route per passenger per kilometer. Unfortunately, this kind of detailed information is nonexistent. Furthermore, one would need very specific travel information for individuals and households to complement this ideal data scenario. This level of personal travel details, such as the actual route, distance and time used per transport mode are not available in our data.

Similar to Frankena (1973), Pucher (1983), Asensio et al. (2003), and Fearnley (2006), we have a cross-sectional data set at our disposal, with limited information on individual subsidies, travel history, and demographics. Our analysis therefore depends on certain definitions and assumptions to ensure that the chosen methods can analyze the distributional effects we seek to find. This chapter starts with a presentation of these before we elaborate on our methods of analysis, namely the propensity approach and our econometric models. The latter receives extra focus due to the challenges that follow with econometric estimation.

4.1 Definitions

Before the assumptions and methods are presented, we first discuss some key definitions.

Public Transport

Public transport is a common term for transportation modes that are available for everyone, unlike private transport modes. In Oslo and Akershus, the alternative modes are bus, metro, tram, train, and ferry. As trains and especially ferries are substantially different in travel length and departure frequency compared to the other modes, we have chosen to exclude them from our main analysis. However, trains are included in the descriptive part of our analysis in Chapter 3 and as a part of our robustness check in Chapter 6. This means that when public transport is mentioned, we are referring to transport by bus, tram and metro in Oslo and Akershus unless otherwise specified.

Subsidy

Public transport subsidies can be defined in a number of different ways and there is no consensus or single definition of transport subsidies. OECD (2005, p. 16) defines a subsidy as "a result of a government action that confers an advantage on consumers or producers, in order to supplement their income or lower their costs". Following Asensio et al. (2003, p. 435) we have chosen to define total transport subsidies as the difference between operators revenue and operating costs:

$$\text{Public transport subsidies} = \text{operating costs} - \text{revenue} \tag{4.1}$$

This definition does not take externalities into account, such as environmental effects or road deterioration. For example, subsidies to public transport can reduce air pollution in urban areas through reduced car use and increased public transport usage. This benefits urban citizens more than rural citizens. Our approach to calculate subsidies and progressiveness will be unable to capture these external effects. However, we argue that Equation 4.1 provides a sufficient indicator of total public transport subsidies, as these effects are difficult to measure accurately.

Public Transport Dependency

We define dependency as the degree of public transport usage for individuals and demographic groups. It can be measured in a number of ways, for example by user expenditures, total distance travelled and number of public transport trips (Fearnley 2006, p. 33).

User expenditures, as a dependency measure, is defined by the monetary cost of using public transport in a given time frame for an individual or a household. It means that the higher a person's public transport expenditures, the more dependent a person is. Asensio et al. (2003) applied this dependency measure in their analysis, but there are a range of issues to consider. Fearnley (2006, p. 34) argues that expenditures may capture valuable information, as it is likely related to distance[1]. Nevertheless, he emphasizes that the measure cannot determine whether low expenditure is a result of discounts or minor dependency. Given that public transport is a normal good (i.e. price goes down, demand goes up), expenditure data will not accurately measure usage (and hence dependency) for groups that receive discounted or free tickets.

Another measure of public transport usage relies on estimates of distance travelled. It means that the longer a person travels by public transport, the more dependent the individual is. Mileage is normally related to welfare, as people with higher incomes tend to travel further than those with lower incomes (Stokes & Lucas 2011, p. 2-3). However, as the scope of this analysis is on shorter travels in Oslo and Akershus this relation is not necessarily apparent. Also, an exact measure of the distance travelled on each transport mode can be difficult to provide due to memory or calculative challenges. A lack of detailed data therefore restricts the opportunities of using travel distance as a measure of public transport use.

The last measurement, number of trips on public transport, is stated by Balcombe et al. (2004, p. 32) to be the most common measure of aggregate usage. One reason is that it is relatively easier

[1]The price on Ruter's tickets increase when passing through additional zones.

for people to remember and keep track of, which makes it a more reliable measure of transport mode usage than former methods discussed. However, there are some issues to consider as well. Trip length may vary substantially between passengers and possibly income groups. Also, people can have a non-optimal route nearby, which means they need relatively more trips than others to reach their destination. This situation can be related to income, as public transport access within close proximity likely has a positive effect on housing prices. Nonetheless, similar to Fearnley (2006, p. 34) we conclude that number of trips is the preferred measure of public transport usage. In that context, we must define what a trip is, how it is counted and how it differs from another travel definition in this thesis - namely journeys.

A journey is defined by the purpose of the final travel destination. This definition differs from how the public might perceive a journey. For example, if a person travels with the intention of going to work, but has to drop off his children at school on the way, the public would perceive this as one journey. By our approach, it will be defined as two journeys. This is because the person has two purposes: to take their children to school and go to work. When a single travel destination is reached, it is defined as the end a journey.

Hjorthol & Uteng (2014, p. 1) defines a trip as "a movement from outside your home, school, work or cabin, regardless of length, duration, purpose, or transport mode used". A person with the intention of travelling to work can illustrate this. Walking to a bus station, taking the bus, and at last walking from the bus stop to the workplace. This is defined as three trips within one journey, but only one counts as a trip on public transport.

As journeys can contain multiple trips with different transportation modes, both private and public, the number of journeys on public transport can be an inaccurate measure of public transport dependency compared to the number of trips.

4.2 Assumptions

The methods described in this chapter are based on four assumptions. The preciseness and validity of the analysis will be reduced if any of the assumptions below are violated.

Equal ticket price

The average purchase price for tickets must be equal across income groups. If one group receive more discounted ticket prices than the others, but their public transport use is identical, our measure of public transport use will predict that they receive an equal amount of subsidies.

Equal operating costs

The operating costs must be, on average, equal across income groups. This assumption neglects the fact that operating costs vary a lot between traveling routes. Knowing the exact operating costs would be very useful to measure subsidies for each individual and income group, but this kind of data is difficult to obtain. As a result, we are forced to assume that operating costs are equal across income groups.

No subsidy leakage

The full amount of subsidies must be passed on to travelers. This implies that subsidies will only reduce ticket prices or increase service levels, and not leak into higher costs. It therefore follows that operating costs are not influenced by subsidies.

Subsidies depend on frequency of trips

As the usage of public transport is measured by the number of trips by public transport, we assume that very frequent passengers receive a larger share of subsidies relative to people who travel less frequently.

Discussion

The first assumption is the crudest one, as it is not very likely that individuals in Oslo and Akershus pay the same price for their tickets. A quick look at the fares for Ruter's transportation modes (such as tram, bus, and metro) reveals that there are several prices for the same tickets. Some are discounted for elderly or children, while some discounts are based on the duration of the ticket. As the population consists of groups that are eligible for age or occupational discounts and individuals that are inclined to minimize their cost per trip, there is little support for this assumption. Nevertheless, to conclude that the most dependent income group receives a relatively larger share of subsidies than the less dependent groups is a necessary condition.

There are many ways of defining the costs of operating public transport routes. A total sum is not always comparable, as traffic, road quality and distance are not identical within routes. Therefore, the cost per route may be high, even if the cost per passenger or cost per kilometer is low. Also, since passengers tend to hop on and off the transport mode at different stops, the cost per kilometer per passenger will vary, which means the subsidy per individual also varies. However, this detailed information is not available in our data set and can be difficult to obtain from external sources. Hence, we consider equal operational costs, regardless of the route and length traveled, as a reasonable assumption.

If 100 percent of a granted subsidy benefits the intended targets, there is no leakage. With data from 15 countries, Bly et al. (1980, p.327) found that a 1% increase in transit subsidy, gained passengers by 0.4 to 0.8% through reduced fares and higher service levels. Five years later Bly & Oldfield (1985, p. 19) found that, on average, roughly half of the subsidies to public transport leaked into higher costs when they examined 16 countries. Based on these findings, Pucher (1983, p. 40) and Fearnley (2006, p. 30-31) both argue that existing passengers are the primary beneficiaries from public transport subsidies, even in the presence of leakages. Additionally, the structure of Ruter's bidding process and gross cost contracts to operate buses in Oslo and Akershus minimizes the cost and the likelihood of subsidy leakage into higher costs. The competitive bidding process ensures that Ruter can pick the company that provides the lowest operating cost, which mean that subsidies do not necessarily cover excessive costs. Gross cost contracts means that Ruter pays the operating bus company a fixed amount for their services. The contract implies that Ruter receives the full revenue from tickets, and administer the marketing, planning of routes and determine the price of tickets. By doing this themselves, the operating bus companies have less chance of directing a leakage into higher costs. Thus, we conclude that no subsidy leakage is a fair assumption. It is a condition Asensio et al. (2003, p. 434) took for granted in order to calculate the distributive effects in their analysis.

The last condition comes as a result of using trips by public transport as a measure of transport dependency. It is also conditional on assumption one, that everyone pay the same ticket price. We can therefore conclude that the income group with the highest frequency receives relatively more subsidies than the rest. However, as trips may differ in length and time spent traveling, the individual subsidy per trip can vary. Time spent is not necessarily positively correlated with welfare, but the length of trips can vary significantly between income groups (Le Grand 1982, p. 116-117). It has also been found that individuals with higher incomes travel further than those with lower incomes (Stokes & Lucas 2011, p. 2-3). The latter receives support from our own descriptive statistics, as income groups 5 and 4 are shown to travel the longest[2] on an average day compared to the other income groups. The arguments against are strong, but Asensio et al. (2003, p. 436) applied a similar condition by assuming that concessionary fare passengers have the same travel pattern. We are thus concerned by a lack of support, but it is a necessary condition to conclude on which income group that receives most subsidies.

If these assumptions hold, we are able to analyze the progressiveness of public transport with the methods presented in the following sections.

[2]The average distances traveled per day (not including journeys above 100 km) can be found in appendix A.3.

4.3 Propensity Approach

The first method we use to measure the progressiveness of public transport is of a descriptive nature. The propensity approach can be used to determine whether income groups are over- or underrepresented in public transport usage.

It can be calculated by using the following formula

$$\text{Propensity}_i = \frac{\text{Share of public transport use}_i}{\text{Share of population}_i} \tag{4.2}$$

where i denotes a given income group. If the propensity score is 1, the proportion of usage is identical with the proportion of the population. It means that a group is neither under- or over-represented in the use of public transport. If the propensity value goes below 1 it means the group is relatively underrepresented and therefore less dependent on public transport. Likewise, a group is relatively overrepresented and dependent on public transport if the value is above 1. For instance, if a group containing 20% of the sample has made 25% of the total trips by public transport, it results in a propensity score of 1.25. This indicates above average public transport use for that income group. If the income group with the lowest income has the highest propensity score at the end of these calculations, it indicates, given the assumptions stated in Section 4.2, that public transport subsidies in Oslo and Akershus are progressive. The approach is therefore a simple, but effective measure of dependency, which Fearnley (2006) also applied in his analysis. The final propensity results are presented at the beginning of Chapter 5.

4.4 Main Model

By using econometric models in our analysis we intend to complement the propensity approach and further analyze the relationship between income and dependency on public transport. Econometric models have several advantages compared to the descriptive approach, such as the possibility to test and control for other variables. More specifically, these models enable us to test whether public transport usage is significantly different between groups. Additionally, it allows us to control for other variables that also affect public transport demand, such as demographic factors. We can then explore whether specific characteristics in each group explain the differences in usage or if income-related factors are the main cause.

We apply a demand model as our main model with the following model specification,

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \tag{4.3}$$

where y_i represent the total number of trips with public transport for individual i. The vector X_1 represents the income groups, while X_2 is a vector denoting all control variables. At last comes the error-term u_i , which captures all unexplained variation in the model.

Ordinarily least squares (OLS) is used as the estimation approach for the demand model. A number of statistical properties for OLS-estimation must hold in order to achieve precise and unbiased estimates[3]. These will be elaborated on in the next paragraphs. At last, an alternative estimation approach with another measure of dependency will presented and discussed.

4.4.1 Econometric Challenges

When the assumption of exogenous independent variables are violated, the estimators produced by OLS will be biased and inconsistent. This is often referred to as the endogeneity problem, and is typically a consequence of omitted variables, measurement error or simultaneity (Woolridge 2010, p. 54-55). In addition, non-random sampling can also lead to endogeneity if the pattern in which the sample has been drawn is correlated with the error term. (Wooldridge 2016, p. 303)

The presence of heteroscedasticity and multicollinearity can also influence our results. Our application of robust standard errors is a common solution to the former case, and in the robustness analysis we apply a different estimation technique that assumes heteroscedasticity. Moreover, a low Variance Inflation Factor (VIF) and low correlation values suggests that multicollinearity is not a substantial problem in our model [4]. This section therefore focus on the issues that are most relevant to our situation, namely measurement error and omitted variables.

Measurement Error

Measurement error (ME) occurs when the observed value deviate from the true value of a variable (Brewer & Picus n.d., p. 456-457). A measurement error can be systematic or random. A systematic error means that respondents would systematically register an imprecise answer due to the method of obtaining the information. This may lead to biased and

[3]The classical linear properties for OLS are described by Wooldridge (2016, p.139-140).
[4]The VIF values and a correlation matrix are presented in Table A.4 and A.1 in the appendix.

inconsistent estimates, and the ability to draw conclusions toward a target population will be severely limited. A random error can be caused by misreading a question, accidental mistakes in the registering or challenges with recalling correctly. The effect is normally more noise as standard deviations increase, but a considerable amount of random errors may lead to less precise estimates. In econometrics, it all depends on which variable contains the measurement error and the assumptions of the error in question.

If there is Measurement error in the dependent variable, Ordinary Least Squares (OLS) will produce consistent estimates, as long as the the ME is not correlated with any of the explanatory variables. The downside is that ME will increase the variance in the model because it is correlated with the model's error term (Woolridge 2010, p. 76-78). The number of trips with public transport may contain some error, as individuals may have forgotten the number of trips or the specific mode they used during their travels. However, by notifying the respondents in advance, specifying the day of registration and supplying a travel diary, the RVU survey tries to avoid these random errors. Denstadli & Lian (2003, p. 130-131) found that these actions improved the reporting, especially the trips performed by foot.

Measurement error in the explanatory variables tend to have more serious effects on the estimates. If the ME correlates with the true value, the error is random and the effect is the same as in the previous example. If the ME correlates with the observed value, instead of the true value, there is a case of systematic error. The following result is that the estimated coefficient will be biased towards zero. This is referred to as the attenuation bias (Wooldridge 2016, p. 291). In a model with multiple variables, ME causes inconsistency in all the estimators as long as the variable with ME is correlated with the other explanatory variables - which it generally is.

The income variable is an important part of this empirical analysis, so precise values are much desired. In surveys, it is common that reported gross income is undervalued (Moore 2000, p. 332). Aside from job earnings, there are many potential income sources such as dividends on stocks, interests on savings account, government transfers and earnings from smaller engagements. These income sources are not always obvious at the time for the individuals reporting their income. By including the earnings of other members in the household it amplifies this challenge. However, compared to other methods of gathering income information, such as regional averages, the RVU provide a more precise measure. The reason is that the individuals themselves report the income category they belong to.

As the income in RVU is registered in categories and not absolute income, there is already an imprecise measure at hand. By transforming these categories into the average of each category, the measures are worsened. For instance, if an individual registers a household income of 210 000 in the category 200 000 - 400 000, the transformed income for the individual is 300 000. However, in order to adjust the income for household size, this was a necessary action. If a systematic error is present, then the effect of income on public transport usage will be underestimated.

Omitted Variables

The omitted variable problem (OVB) occurs if a relevant variable, that correlates with one or more of the model variables, is left out of the model. Since the error term will capture the effect of the omitted variable, it may lead to correlation between the error term and included model variables. In such a case, the assumption of exogenous explanatory variables is violated and OLS will produce biased estimates for variables that correlates with the error term (Wooldridge 2016, p. 79). The models used in this empirical analysis is based on cross sectional data. This makes it possible to exploit variation between cross sectional units, but not over time. As a consequence, time varying variables are unavailable and omitted from the models. Time varying variables such as gas-prices (Mattson 2008), parking fees and population growth are likely to effect public transport demand, and since we cannot include these variables in the models, the models will have an omitted variable problem. Furthermore, the variable for car ownership is omitted for these reasons, even though it is one of the main determinants of demand for public transport.

Still, it is neither realistic or necessarily desirable to include all relevant variables in a model, as it might cause estimation problems such as multicollinearity or simultaneity. The goal of this analysis is not to find the causal effect of income on public transport demand. Instead, we seek the effect of income on public transport dependency after controlling for demographic factors, in order to determine if such factors drive progressivity. Therefore, choosing the right number of variables in the model specification involves a trade-off. An under-specified model can lead to a biased coefficients and over-specification can cause high standard errors and insignificant estimates. Having an omitted variable problem is common in applied econometrics, at least to some extent. Nevertheless, it is still important to know which variables that are omitted and how this will affect the results. Car ownership is therefore included as a part of the robustness analysis in Chapter 6.

.

5 | Results

In this chapter, we present and discuss the results from our chosen methods of analysis, namely the propensity approach and our econometric models. The outcomes from the propensity approach are first explored and discussed in Section 5.1. We then motivate the need for econometric models and present and examine these results in Section 5.2. The chapter is then finalized with a summary and discussion of the interpretation and robustness of these results. The econometric regressions and tests are all produced in the statistical software Stata.

5.1 Travel Propensity

The propensity score provides a measure of how dependent individuals are on public transport conditional on their income groups. Table 5.1 shows propensity scores for total number of trips, public transport and car use, while Figure A.2 in the appendix illustrate the results graphically.

Income Group	(1) Total Trips	(2) Public Transport (excluding train)	(3) Public Transport (including train)	(4) Car[1]
1	1.00	1.45	1.30	0.78
2	0.99	0.75	0.79	1.12
3	0.96	1.01	0.97	0.94
4	1.08	0.84	0.97	1.19
5	0.98	0.96	0.98	0.98

Table 5.1: Propensity scores for total trips, public transport and cars. Total trips propensities are calculated based on frequency of trips by car, bus, metro, tram and train. Public transport (excluding train) propensities are based on trips by bus, metro, tram. Public transport (including train) propensities are based on the same modes, in addition to trips by train. Car propensities are based on trips by car.

A propensity score equal to 1 implies that the groups share of the population is equal to the groups share of total public transport usage. Propensity scores above 1 indicate over-representation in use, and scores below 1 represent the opposite. Detailed description of the propensity approach was provided in Section 4.3.

Table 5.1 shows that propensity scores differs between income groups. Column (1) demonstrate that travel propensities does not vary substantially across income groups, although income group 4 is slightly over-represented in the frequency of total trips.

[1]Trips by car, as driver and as passenger.

Column (2) displays the public transport propensities and indicate that income group 1 is relatively more dependent on public transport than the other groups. The low propensity score can be related to area of residence, as the two income groups with the lowest propensity score also have the fewest people living in Oslo (Table 3.4). The underlying factor can be that residents in Oslo have relatively better access to bus, tram and metro than the average resident in the county of Akershus. Another explanation is that these income groups have the highest propensity for car usage, which indicates that they are more dependent on cars rather than public transport.

In Column (3) we include trains to examine the sensitivity of these measures to this form of transportation. The inclusion of trains leads to transport propensity scores settling closer to 1, where the difference in public transport propensities between groups is reduced. This suggests that high income groups are over-represented among train passengers.

Column (4) shows that car propensities does not increase with income. This is unexpected, as the literature suggests that income and car ownership is positively correlated (Dargay & Gately 1999) and that car ownership has a positive effect on car travel (Dargay 2007). The relationship between income and car propensities appears to follow a zigzag pattern.

Income Group	Bus	Tram	Subway	Train
1	1.37	1.59	1.50	0.78
2	0.88	0.53	0.67	0.92
3	0.96	1.11	1.06	0.80
4	0.86	0.73	0.85	1.44
5	0.95	1.04	0.95	1.05

Table 5.2: Propensity score for different transport modes

Table 5.2 presents the propensity scores disaggregated by different public transport modes. It shows that income group 1 is particularly dependent on bus, tram and subway transport, but not train transport. The top income groups, 4 and 5, are more dependent than the other groups on transport by train. These measures fits with the results from Column (4) in Table 5.1. This indicates that any subsidies attached to train are to some extent regressive, simply because the highest income groups are most dependent on train transport. This fits with evidence from the UK (Fearnley 2006).

The main findings from the propensity approach is that individuals in income group 1 are more dependent on public transport than individuals in the other groups, both including and excluding transport by train.

5.2 Demand Model

Table 5.3 displays the estimates from our demand model, where we use a specific-to-general approach[2]. All models are estimated with ordinary least squares estimation and robust standard errors. First, we present the effect of income on public transport use. Second, we discuss how each control variable affects public transport use.

The Effect of Income on Public Transport Demand

Table 5.3 shows that income group 2-5 have a lower demand for public transport than income group 1 (the poorest group). Column (1) only includes the main variable of interest, namely income. The estimates reveal that individuals in income groups 2-5 take between 0.367 and 0.229 fewer trips with public transport per day than those in income group 1. Column (2) adds a control variable for Oslo and the difference between income group 1 and all the other groups gets smaller. Moreover, there is no longer any noticeable differences in public transport use across income groups 2-5. This suggest that some of the income differences in public transport usage found before (Column 1 and propensity score) reflect differences in income and public transport usage across Oslo and Akershus.

Column (3), (4) and (5) show that the inclusion of demographic control variables tend to further decrease the difference between income group 1 and all the other groups. This implies that some of the differences in public transport usage found in Column (2) express differences in income and public transport usage between age, sex, educational level and employer-related factors.

Column (6) demonstrates that the difference in public transport usage between income groups is not driven by season and weekend variation between income groups. We conclude that the estimates in Table 5.3 indicate that income group 2-5 use less public transport than income group 1, before and after control variables are included.

[2]The specific-to-general approach is a model building method. The first step is to only include the main variable of interest in the model. The proceeding steps is to include one variable at the time, until the preferred model specification is created.

Table 5.3: Demand model
Dependent variable: Number of trips with public transport

	(1)	(2)	(3)	(4)	(5)	(6)
	Trips	-	-	-	-	-
Income group 2	-0.367***	-0.258***	-0.204***	-0.186***	-0.200***	-0.191***
	(0.0404)	(0.0381)	(0.0375)	(0.0377)	(0.0373)	(0.0372)
Income group 3	-0.229***	-0.201***	-0.0816**	-0.112***	-0.141***	-0.128***
	(0.0432)	(0.0405)	(0.0396)	(0.0405)	(0.0400)	(0.0397)
Income group 4	-0.322***	-0.234***	-0.205***	-0.194***	-0.237***	-0.226***
	(0.0411)	(0.0391)	(0.0398)	(0.0414)	(0.0410)	(0.0409)
Income group 5	-0.255***	-0.230***	-0.0988***	-0.155***	-0.207***	-0.192***
	(0.0401)	(0.0381)	(0.0374)	(0.0399)	(0.0398)	(0.0395)
Oslo		0.672***	0.607***	0.579***	0.556***	0.555***
		(0.0273)	(0.0271)	(0.0276)	(0.0273)	(0.0272)
Age 13-17			0.226***	0.271***	0.390***	0.358***
			(0.0684)	(0.0722)	(0.0742)	(0.0740)
Age 18-25			0.573***	0.579***	0.604***	0.602***
			(0.0606)	(0.0626)	(0.0626)	(0.0621)
Age 61+			-0.184***	-0.197***	-0.101***	-0.118***
			(0.0241)	(0.0272)	(0.0294)	(0.0298)
Male				-0.119***	-0.132***	-0.128***
				(0.0220)	(0.0217)	(0.0216)
Children				-0.118***	-0.113***	-0.109***
				(0.0329)	(0.0323)	(0.0322)
Higher education				0.154***	0.147***	0.149***
				(0.0241)	(0.0237)	(0.0237)
Work trip					0.420***	0.353***
					(0.0250)	(0.0279)
Free parking					-0.195***	-0.184***
					(0.0251)	(0.0252)
Constant	0.766***	0.483***	0.436***	0.462***	0.369***	0.342***
	(0.0333)	(0.0305)	(0.0325)	(0.0377)	(0.0392)	(0.0440)
Seasonal dummies						Yes
Weekend dummies						Yes
Observations	8861	8861	8861	8861	8861	8861
Adjusted R^2	0.012	0.095	0.121	0.129	0.160	0.164

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All independent variables in the model are binary. Income group 1 and Age 26-60 are set as reference categories.

The high standard errors of the income dummies cause imprecise coefficient estimates. A plot of the coefficient estimates in column (1) and (6) with corresponding 95% confidence intervals are shown in Figure 5.1.

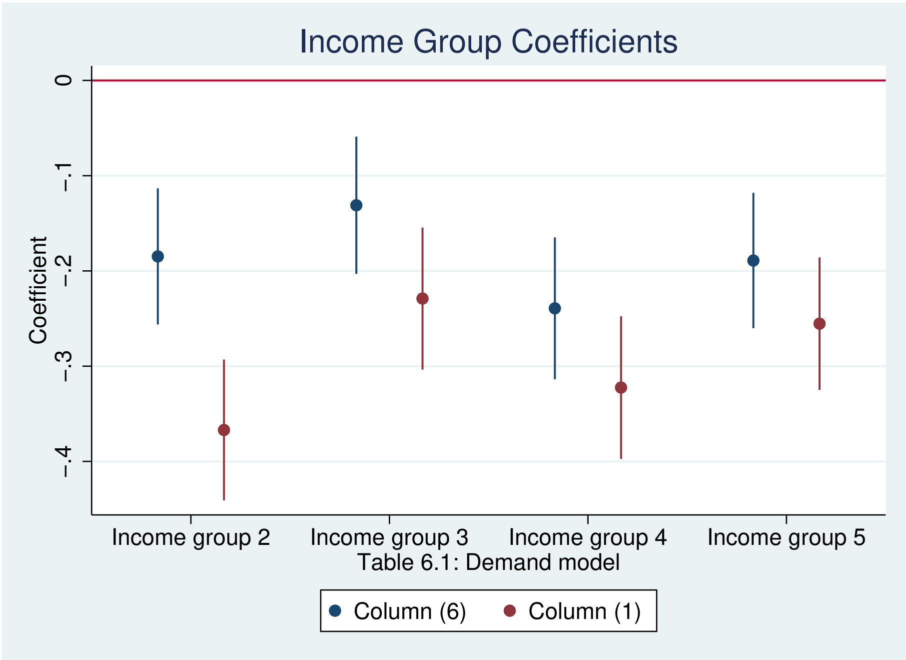

Figure 5.1: Coefficients for income groups 2-5 with corresponding 95% coefficient intervals.

The question is whether public transport use is significantly different between income groups 2-5. This is formally tested by enforcing restrictions in Column (1) and (6). The null- and alternative hypothesis is:

$$H_0 : income\ group_i = income\ group_j$$
$$H_1 : income\ group_i \neq income\ group_j$$

where i and j denote the income group. In other words, whether income group$_i$'s coefficient is equal to income group$_j$'s coefficient. We set a 5% significance level for the F-tests, which means that the null hypothesis is rejected for p-values above 0.05. Test results are reported in Table 5.4.

Table 5.4: Test of Restrictions of Column (1) and (6) in Table 5.3.

Restriction	Prob > F	
	(1)	(6)
Income group 2 = Income group 3	0.00	0.09
Income group 2 = Income group 4	0.24	0.33
Income group 2 = Income group 5	0.00	0.98
Income group 3 = Income group 4	0.01	0.01
Income group 3 = Income group 5	0.46	0.06
Income group 5 = Income group 4	0.06	0.39

Prob > F represents p-values for the F-statistics.

Table 5.4 displays whether or not public transport use is equal between income groups 2 to 5, before and after control variables are added to the model. Column (1) represents the demand model without control variables. It shows that public transport use is statistically equal between income groups 2 and 4, and between income groups 3 and 5. In other words, we cannot reject the null hypothesis in two of the tests, income groups 2 = 4 and income 3 = 5, as the p-values are above 0.05. In column (6), the six restrictions are enforced after all of the control variables are added to the model. The tests show that public transport use is equal between income groups 2 to 5. However, the test is inconclusive regarding income groups 3 and 4. It suggests that income group 2‘s coefficient is statistically equal to income groups 3 and 4‘s coefficient, which implies that income groups 3 and 4‘s coefficient are equal. However, the test between groups 3 and 4 evidences otherwise. The F-statistics P-value of 0.01 means that we must reject the null hypothesis of equal coefficients. As a consequence, we cannot determine if the coefficients are equal or not. Except for groups 3 and 4, the tests in Table 5.4 indicate that public transport use is equal between income groups 2 to 5 after control variables are added to the model. So while there is statistically a difference between group 1 and the others, there is no clear income gradient in public transport usage.

Effect of Additional Explanatory Variables

The control variables used in Column (2)-(6) have statistically significant effects on public transport demand and on all of the income coefficients. The following section discusses these effects based on the preferred model specification in Column (6), which from now on will be referred to as the baseline model. All the control variables are significant at a 0.01% significance level.

The model estimates that people who live in Oslo take 0.6 more trips by public transport per day than those who live in Akershus. This is expected, as the proportion of people who use public transport in Oslo is the highest in Norway (Hjorthol & Uteng 2014, p. 40).

The relationship between age and demand for public transport is represented by an inverted U-shape. Age-group 13-17 and 18-25 use relatively more public transport than Age-group 26-60 (the reference group), while age-group 61+ use less. In comparison, Balcombe et al. (2004, p. 8) found that bus use in Britain is highest at each end of the age spectrum, and that the age variation among rail passengers is small.

The additional demographic controls such as male, children and higher education have the expected signs. The male dummy is negative, indicating that men use less public transport than woman. The children dummy is negative, meaning that households with one child or more use less public transport than households without children. This is expected because households with children are more likely to own a car (Hjorthol & Uteng 2014, p. 17). The dummy variable for higher education shows that a university degrees increases public transport use by 0.15 trips per day. One possible reason is that higher education raises awareness of environmental issues, which increases the propensity to use public transport. (Meyer 2015).

A dummy for education is included in the model to avoid, or at least reduce, possible bias in the income coefficients. If $highereducation$ was omitted from the model and the variable correlates with both income and public transport demand, it would cause an omitted variable problem and hence biased income coefficients. There is a large literature on the causal effect of education on earnings (Card 1999). We include the dummy to control for the effect education has on demand for public transport. The isolated effect of education on the income group coefficients can be studied in the appendix, Column (4) and (5), Table A.7. All the income coefficients are close to unchanged after education level is controlled for, implying that the difference in public transport use between income groups is not explained by differences in education level.

People with a one or more work related trips have 0.42 more trips with public transport than people without work related trips. This is expected as people with work related trips are employed, and employment often increases the need for transportation. The model estimates that free parking at work, or close to work, reduces daily public transport trips by -0.184. The fact that free parking lowers public transport demand is not surprising. It lowers the marginal cost of car use, and hence incentivizes the use of it relative to public transport. Metz (2010) partly confirms this, by claiming that a lack of parking options reduce car use.

The estimates reported in Table 5.3 demonstrate that the individuals in the poorest group are the most dependent on public transport, before and after the inclusion of demographic controls. This suggests that the differences in public transport usage by income groups do not simply reflect demographic differences. These estimates can then be used as an input to predict how progressive public transport subsidies are.

6 | Robustness checks

In this chapter, we perform several checks to evaluate the robustness of the results in Chapter 5. Most of the tests are conducted on the baseline model, primarily to test if the effect of income on public transport use is sensitive to changes in the number of income groups, equivalence scales, explanatory variables and change of dependent variable. First, we check for sensitivity of changing the dependent variable.

6.1 Trips vs. Usage

In this section, we test the robustness of our results by changing the measure of dependency on public transport. A weakness of our demand models is the use of number of trips as the dependent variable. Many $Trips$ can be an indicator of non-optimal travel routes, not high usage. Hence, our model might over-estimate some peoples dependency on public transport because of frequent transit changes. This problem can be avoided by applying a binary variable, representing *probability of using public transport*, as regressand. We can then examine if the main results are affected by the change of dependent variable.

We estimate the *probability of using public transport* by using $Trip$ as the dependent variable. It takes on the value of 1 if the individual uses public transport during the day, 0 otherwise. The appropriate estimation strategy is using a probit model because the dependent variable is binary. As the model applies Maximum Likelihood, the robustness check does not only test a change in dependent variable, but also in estimation technique.

The magnitude of the estimated probit coefficients are difficult to interpret and compare with other OLS coefficients, but the sign of the coefficients are still correct (Wooldridge 2016, p. 530). By calculating the marginal effects of each coefficient, we can interpret both the magnitude and sign of the estimated effect of the different variables. Specifically, we take the average of the individuals marginal effect, also known as APE[1], as it is more suitable for models with dummy variables than other common alternatives.

By looking at the average marginal effects from Table A.10 in the appendix, we find that income groups 2-5 are between 3.56 to 6.05 percentage points less likely to use public transport than income group 1. This implies that even if we estimate the probability of using public transport instead of the number of trips, the results are largely the same. Income group 1 is more dependent on public transport than all the other income groups.

[1] Average Partial Effect.

6.2 Different Income Groups

The income coefficients in the baseline model (Table 5.3) are estimated using five income groups. The size of the income intervals and number of people in each income interval may have an impact on the estimated coefficients. In the following check, we estimate the baseline model with four and six income groups, in addition to the initial model with five income groups. The purpose is to check whether the main results from the baseline model are affected by a change in the number of income groups. The proportion of the population that each income group represents is shown below.

Table 6.1: Three Different Divisions of Income Groups.

	Proportions of the Population		
Income Group	4 Income Groups	5 Income Groups	6 Income Groups
1	0.19	0.19	0.17
2	0.28	0.19	0.18
3	0.28	0.19	0.19
4	0.25	0.18	0.14
5	-	0.25	0.18
6	-	-	0.14
Sum	1	1	1

Table 6.2 shows the estimated coefficients in the baseline model using three, four and five income groups. All of the income coefficients in Column (1)-(3) are significant at a 0.01% level and are in the interval of -0.226 to -0.128. This means that the previous findings that all income groups use less public transport than income group 1 holds when we change the number of income groups.

The income coefficients does not seem to by particularly affected by the number of groups. Individuals in income group 1 use more public transport than the others, while the differences between income groups 2-5 are not substantial. We conclude that the main findings from the baseline model holds, even if four or six income groups are used instead of five.

Table 6.2: Demand Model with Different Income Groups

	(1) Trips	(2) Trips	(3) Trips
incomegroup 2	-0.184***	-0.191***	-0.182***
	(0.0349)	(0.0372)	(0.0398)
incomegroup 3	-0.166***	-0.128**	-0.165***
	(0.0382)	(0.0397)	(0.0407)
incomegroup 4	-0.202***	-0.226***	-0.218***
	(0.0393)	(0.0409)	(0.4433)
Incomegroup 5		-0.192***	-0.216***
		(0.0395)	(0.0416)
incomegroup 6			-0.211***
			(0.0454)
Constant	0.354***	0.342***	0.357***
	(0.0437)	(0.0440)	(0.0447)
Control variables	Yes	Yes	Yes
Seasonal dummies	Yes	Yes	Yes
Weekend dummies	Yes	Yes	Yes
Observations	8861	8861	8861
Adjusted R^2	0.163	0.164	0.1652

Robust standard errors in parentheses

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

Column (2) is equivalent to column (6) in model 5.3. Column (1) and Column (3) have the same model specification, but with three and five income groups.

6.3 Equivalence scales

As mentioned in the previous section, limited income variation can be a reason for the uneven groups, which in turn can affect the results. Additionally, the lack in variation can cause less significant results and imprecise estimates. In an optimal situation, our data would contain the actual disposable household income. Unfortunately, this is not the case. Our data set contains gross household income, reported in categories, so to approximate the disposable income we must apply an equivalence scale. As a part of our robustness analysis, this section seeks to find the effect of using different equivalence scales. How does it influence the variation in income and does this affect the final results? First, we present alternatives to our current equivalence scale, the square root scale, which we first introduced in Chapter 3. The new and original scales are then compared, before we summarize and come to a conclusion on the effect of using different equivalence scales.

There are a number of equivalence scales used in the litterature, but there is little to no consensus on which scale is considered the most accurate as they all value economies of scale differently (OECD et al. 2012). The OECD[2] scale is the most modest in this respect, assigning the next adult and child in a household, 0.7 and 0.5 respectively. The same scale has been modified to a more moderate weight of 0.5 for an adult, and 0.3 for a child, which is known as the EU-scale (or modified OECD scale). Children above 13 years of age are considered adults in both of these cases. The last scale is the square root scale. It weigh economies of scale slightly more than the EU scale and is applied by dividing the household income by the square root of individuals in that household. Even though the OECD's standard equivalence scale is the square root scale, OECD (2013) states that the EU scale is the preferred choice by Statistics Norway and Eurostat[3]. For this reason we have chosen to compare these two scales. In the case of increased income variation, the EU scale separates children and adults[4], while the square root scale only considers the number of household members. This distinction can result in a higher variation in income after the adjustment compared to our original approach. Figure 6.3 clearly displays such a result. The amount of unique values after adjusting the household income variable is close to tripled with the new EU scale than the original square root scale.

[2]The Organisation for Economic Co-operation and Development (OECD)

[3]Eurostat is the statistical office of the European Union.

[4]So does the original OECD scale, but as the approach is the same, the variation in income will be identical to the application of the EU Scale.

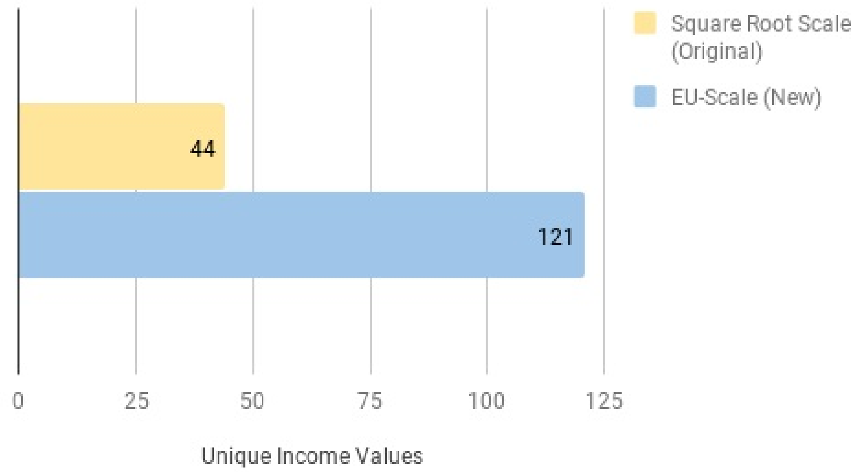

Figure 6.1: The difference in unique income values when applying different equivalence scales.

A larger amount of unique values does not necessarily mean that the income is distributed more evenly or if the distribution is more suited for this analysis. We therefore compare the distribution of adjusted household incomes with overlapping histograms in Figure 6.2. It illustrates that the new scale seemingly provides a more even household adjusted income distribution than the original scale.

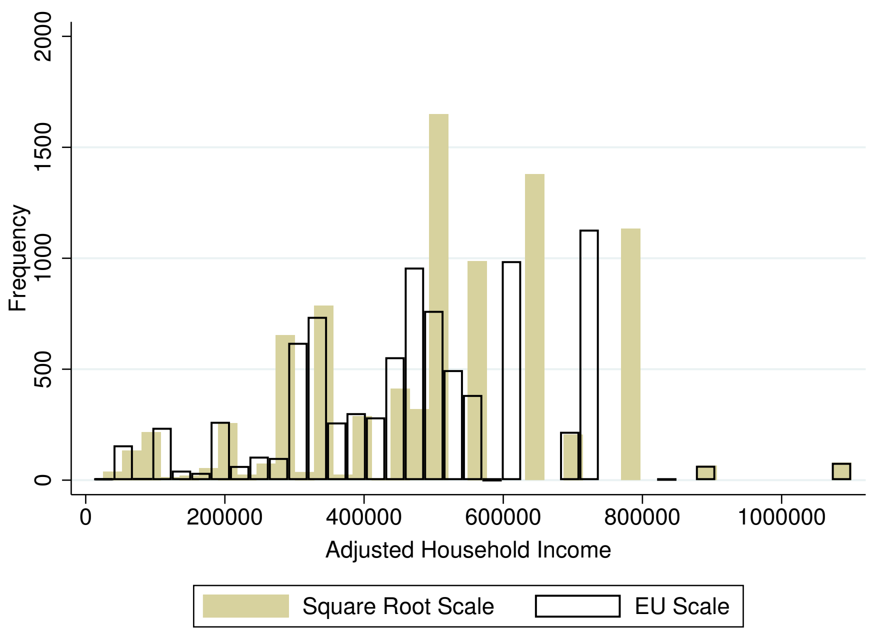

Figure 6.2: A histogram comparison of adjusted household income using two different equivalence scales.

These arguments indicate that the new scale has increased the variation in income. It is therefore interesting to see if the size of the income groups become more proportionate as result of this. Table 6.3 displays this comparison. We first notice that group 2 and 3 increase their proportion by 0.02 and 0.03 respectively, while group 5 ends up with the optimal share (0.20) of the population. In other words, the groups receive a more even distribution with the new scale compared to the previous. Another point is that the intervals have been lowered in sum, as a result of using a scale that put less weight on economies of scale. We can now compare how these income groups will differ in effect with the application of our econometric models.

	Square Root Scale (Original)			EU Scale (New)		
Income Group	Number	Share	Interval	Number	Share	Interval
1	1656	0.19	25 000 - 350 000	1660	0.19	12 048 - 321 429
2	1708	0.19	353 533 - 491 935	1865	0.21	323 529 - 440 000
3	1649	0.19	494 974 - 519 615	1970	0.22	450 000 - 500 000
4	1622	0.18	550 000- 635 085	1620	0.18	523 810 - 600 000
5	2226	0.25	636 396 - 1 100 000	1746	0.20	611 111 - 1 100 000

Table 6.3: Comparison of how different equivalence scales influence the segmentation of income groups.

The results are visualized in Table 6.4. From the Table and Column (4) we observe that the effects per income group have all become slightly more negative, except for group 4 which has a become less negative compared to the reference group. These small changes in effect make sense as the individuals travel habits are unaltered, even if the variation in income is higher and the new method values economies of scale less than the previous method.

Summary

The application of the EU scale creates more unique values of income than the square root scale as it distinguish between adults and children in the household. The new scale therefore increases the variation in income and contributes to distributing it relatively more even. The division of groups also becomes slightly more balanced than before, but the final results were only marginally affected by all of these improvements. Furthermore, by dividing the newly adjusted household income into four and six income groups[5], we get very similar results to Section 6.2. It indicates that low income variation is not necessarily a good explanation as to why the effect of income changes between various divisions, and that there are other forces behind this. Overall, the main conclusion is consistent, as income group 1 still is substantially more dependent than the other groups.

[5]The division of groups and the results from the econometric estimation are displayed in Table A.5 and A.9 in the Appendix .

Based on these results, we conclude that (in this context) a change of scale will not substantially alter the main results or the main conclusion, but since the EU-scale improves the distribution and variation in income, the scale is marginally more appealing than the square root scale.

Table 6.4: Demand Model with Different Equivalence Scales

	Square Root Scale (1) Trips	Square Root Scale (2) -	EU Scale (3) -	EU Scale (4) -
Incomegroup 2	-0.367***	-0.191***	-0.390***	-0.218***
	(0.0377)	(0.0361)	(0.0369)	(0.0352)
Incomegroup 3	-0.229***	-0.128***	-0.270***	-0.157***
	(0.0381)	(0.0366)	(0.0364)	(0.0348)
Incomegroup 4	-0.322***	-0.226***	-0.326***	-0.193***
	(0.0382)	(0.0387)	(0.0382)	(0.0370)
Incomegroup 5	-0.255***	-0.192***	-0.243***	-0.202***
	(0.0355)	(0.0366)	(0.0375)	(0.0375)
Constant	0.766***	0.342***	0.779***	0.355***
	(0.0269)	(0.0437)	(0.0268)	(0.0435)
Control variables	No	Yes	No	Yes
Seasonal dummies	No	Yes	No	Yes
Weekend dummies	No	Yes	No	Yes
Observations	8861	8861	8861	8861
Adjusted R^2	0.012	0.164	0.014	0.164

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.4 Car ownership

A variable for car ownership is intentionally left out of our models despite the fact that it is a large determinant of public transport use. The justification for this exclusion, and an analysis of how car ownership affects the results from the demand model, will be presented in this section. The car ownership variable in use takes on the value of 1 if the household owns or has access to a car, 0 if not.

Table 6.6 shows that including a car ownership variable in the baseline model dramatically lowers the effect of income on demand for public transport. Column (1) represents the baseline model and shows that income group 2-5 use less public transport than income group 1. Column (2) adds the car ownership variable, and all the income coefficients become small and not statistically significant. They are also insignificant at a 10% significant level. The estimates show that car owning households take 0.7 trips fewer trips with public transport per day than non car owning households. Column (3) shows that the car variable‘s coefficient and significance level only changes slightly if income groups are left out of the model.

The results from Table 6.6 indicate that the baseline model may suffer from an omitted variable problem. The income dummies could be significant mainly because they pick up the effect of car ownership on public transport demand. Whether car ownership is the underlying reason for significant and negative income coefficients can be examined by looking at table 6.5.

Income Group	Share of Car Ownership	Coefficient from the baseline model
1	0.63	-
2	0.92	- 0.191
3	0.82	- 0.128
4	0.97	- 0.226
5	0.90	- 0.192

Table 6.5: Share of households who owns at least one car in income group 1-5 and estimated coefficients from the baseline model.

Table 6.5 shows the share of households that own a car in income group 1-5 and estimated coefficients from the baseline model. The magnitude of the income coefficients seems to be highly correlated with car ownership, as income group 4 has the highest car ownership share and the most negative coefficient. Additionally, income group 3 has the lowest car ownership share and the least negative coefficient. This indicates that car ownership and income is interlinked, despite the low correlation between the variables[6]. The income coefficients may suffer from a negative bias because they capture the negative effect of car ownership on public transport demand.

This robustness check has shown that households in income group 1 owns far less cars than any other income group. This makes the income group relatively more reliant on public transport. If car ownership is taken into account, income does not seem to effect public transport dependency.

The interlinked relationship between car ownership, income, and demand for public transport has been discussed since the first book on public transport demand (Paulley et al. 2006, p. 23). It is hard to distinguish the effect of income from the effect of car ownership on demand for public transport. The preferred model specification depends on the purpose of the analysis. As we measure progressiveness of subsidies to public transport, the primary variable of interest is income. The effect of car ownership on demand for public transport is of secondary interest, which leads us to exclude the variable from the model.

[6]Correlation illustrated in the correlation matrix, A1 in appendix.

Table 6.6: Models with and without car variables

	(1) Trips	(2) Trips	(3) Trips
Incomegroup 2	-0.191***	-0.0527	
	(0.0372)	(0.0365)	
Incomegroup 3	-0.128***	-0.0136	
	(0.0397)	(0.0387)	
Incomegroup 4	-0.226***	-0.0463	
	(0.0409)	(0.0405)	
Incomegroup 5	-0.192***	-0.00968	
	(0.0395)	(0.0385)	
Car		-0.704***	-0.712***
		(0.0458)	(0.0450)
Constant	0.342***	0.821***	0.811***
	(0.0440)	(0.0575)	(0.0547)
Control variables	Yes	Yes	Yes
Seasonal dummies	Yes	Yes	Yes
Weekend dummies	Yes	Yes	Yes
Observations	8861	8861	8861
Adjusted R^2	0.164	0.203	0.203

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The car ownership variable takes on the value of 1 if the individual owns a car or have a car at disposal, and 0 if not

6.5 Train Use as a Dependent Variable

$Trips$ is currently the dependent variable in the baseline model and represents a person's number of trips by bus, metro, and tram. Trips by train are presently left out, even though train use is commonly conceived as a public transport mode. How do the results change if trips by train are included in the $Trips$-variable? This will be examined by estimating the baseline model with two new dependent variables; $Trips\ including\ train$ and $Trips\ by\ train$. The former denotes trips by public transport including train, while the latter expresses only trips by train.

Table 6.7 shows that the inclusion of train trips in the dependent variable changes the estimated effect of income slightly in both directions. Column (1) is the baseline model, with $Trips$ as the dependent variable. By including trips by train in column (2), the estimated income coefficients are close to unchanged for income groups 2 and 3, but reduced for groups 4 and 5. The difference in public transport use seems to be reduced when trips by train are included as a measure of public transport use. The dependent variable in column (3) is solely the number of trips by train. The estimated coefficients for income group 2 and 3 are below 0.001, which demonstrates a negligible effect. Income groups 4 and 5 have positive coefficients, which implies that they have more trips by train than income group 1. However, all of the coefficients are insignificant at a 5% significance level. Consequently, the model predicts that income does not significantly impact demand for train transport.

The findings in Table 6.7 are similar to the results from the propensity approach, which showed that the richest groups use trains more than the poorest and that the difference in public transport use between income groups reduces when trips by train are included in the $Trips$ variable. The main takeaway from this robustness check is that the exclusion of trains does not dramatically affect our main estimates of interest.

Table 6.7: Baseline Model with Different Dependent Variables

	(1) Trips	(2) Trips including Train	(3) Trips by Train
Incomegroup 2	-0.191***	-0.193***	-0.00205
	(0.0372)	(0.0425)	(0.0162)
Incomegroup 3	-0.128***	-0.132***	-0.00430
	(0.0397)	(0.0446)	(0.0158)
Incomegroup 4	-0.226***	-0.189***	0.0372*
	(0.0409)	(0.0475)	(0.0196)
Incomegroup 5	-0.192***	-0.178***	0.0138
	(0.0395)	(0.0451)	(0.0168)
Constant	0.342***	0.452***	0.111***
	(0.0440)	(0.0505)	(0.0195)
Control Variables	Yes	Yes	Yes
Seasonal Dummies	Yes	Yes	Yes
Weekend Dummies	Yes	Yes	Yes
Observations	8861	8861	8861
Adjusted R^2	0.164	0.147	0.061

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Control variables represent all the control variables in the baseline model: Oslo, Age 13-17, Age 18-25, Age 61+, Male, Children, Higheducation, Worktrip, Freeparking.

6.6 Summary

The robustness tests show that the results from the baseline is not sensitive to changes in the number of income groups, equivalence scales or to modifications of the dependent variable. However, it has demonstrated that car ownership is a significant factor in the decision on whether or not to use public transport. The problem is that income is a determinant of car ownership, and car ownership is a predictor of income level, which makes the relationship between the variables endogenous. This forces the exclusion of the car variable from the demand model.

We conclude that the baseline model provides a good estimate of income groups demand for public transport. These estimates can be used as an input to calculate subsidies, which again is used to analyze the distributional effects.

7 | Subsidy Calculation

Our econometric models show that the group with the lowest income is the most dependent on public transport relative to the other groups. This can be interpreted as an indication of progressive subsidies to public transport, given our assumptions in Chapter 4. By calculating the actual subsidies received by each income group we want to test whether we can draw such a conclusion. As the process depends on certain assumptions, we apply four different approaches to examine if the distributive effects vary with different assumptions and methods of calculation.

This chapter first presents a simple approach before we expand and alter the approaches by relaxing certain assumptions. Finally, we summarize and formulate a conclusion based on our calculative results and the chosen methods.

In order to calculate distributive effects, we require information on subsidies. Fortunately, the cost, revenue and total amount of subsidy per mode is made available to the public by Ruter's annual reports. We have used numbers from the annual report for 2016[1] as a base for our calculations (Ruter 2017, p. 23-27). Table 7.1 displays these numbers from the years that coincide with our observations in the RVU data set.

Year	Transport Mode	Cost per Trip	Revenue per Trip	Subsidy per Trip	Proportion of Cost
2013	Bus[2]	17.27	9.7	7.57	0.44
	Tram	16.73	8.67	8.06	0.48
	Metro	18.67	8.7	9.97	0.53
2014	Bus	17.37	9.91	7.46	0.43
	Tram	15.98	8.68	7.3	0.46
	Metro	19.11	8.74	10.37	0.54

Table 7.1: All numbers are in NOK and have been adjusted according to the Norwegian consumer price index for 2017.

As the table presents, there are differences between transport modes and the amount of subsidy they receive. In 2013 and 2014 the metro was the most subsidized mode per trip compared to bus and tram, as a result of higher costs. Since there are no differences in fares between these modes, it indicates that metro users receive more subsidies than bus users who have the same travel pattern, given the following assumptions from Chapter 4:

[1] The calculations were made prior to the publishing of the 2017 report.

[2] Our data set cannot separate between trips on city and regional buses.

1. Equal Price
2. Equal Operating Costs
3. No Subsidy Leakage
4. Subsidy Depends on Trips

As our data set contain observations over 2 years we apply the average subsidy for these years in our calculation. These averages are presented in Table 7.2.

	Bus	Tram	Metro	Public Transport
Average Subsidy per Trip 2013/2014	7.5	7.7	10.1	8.5

Table 7.2: The average subsidies per passenger trip in 2013/2014. All numbers are in NOK.

We also need a predicted average of trips per income group. We base this on the estimates from Table 5.3. The predicted trips for group 1 in the descriptive model is given by the constant in Column (1) while the trips in the Baseline model are found by holding all control variables at their sample means. This means multiplying each demographic variable with the sample mean, and then adding these together with the constant. By subtracting the coefficient from the other income groups, we find each groups predicted average trips per day per person.

Alternative 1

The first approach calculates subsidies on the condition that all four assumptions hold. It takes the average subsidy per public transport trip and divides it by the predicted average number of trips per income group. This method provides results that express the amount each group receive on average in subsidy per trip made with public transport. It therefore follows an assumption which states that there is no difference in subsidies between modes. Figure 7.1 illustrates the difference in average subsidy per day after controlling for demographic factors.

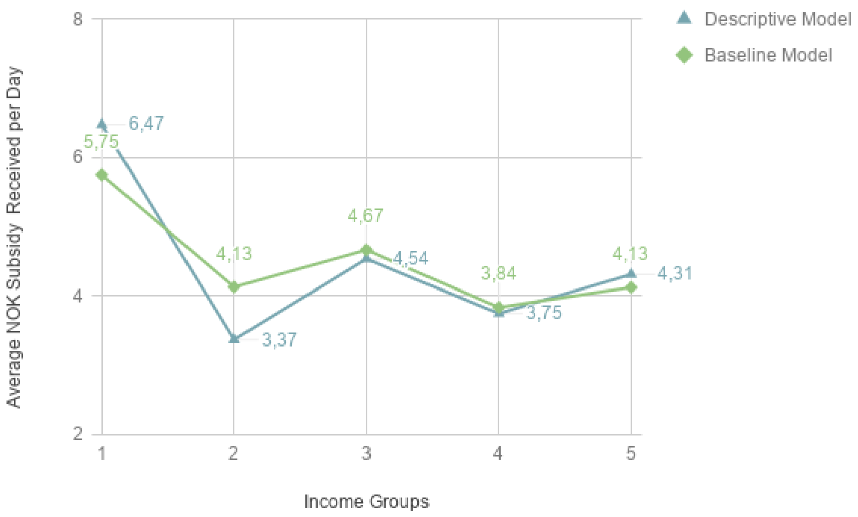

Figure 7.1: Comparison of the calculated subsidies between the descriptive and econometric model. All numbers are given in NOK.

The figure illustrates that demographic characteristics in each group does not solely drive the progressivity of public transport in Oslo and Akershus. The control variables in our baseline model even out the effect of income on public transport dependency, which results in less subsidies to income group 1 compared to the effect in the descriptive model from Column (1) in Table 5.3.

Alternative 2

In the previous alternative we assumed that there is no difference in subsidy between transit modes, even if Table 7.1 and 7.2 presents such a distinction. This alternative method takes this difference into account, which means we need information regarding the distribution of travel on different modes in each group. Table 7.3 present the proportion of trips per mode per group.

Income Group	1		2		3		4		5	
	Trips	Share	Trips	Share	Trips	Share	Trips	Share	Trips	Share
Bus	615	0.49	408	0.60	429	0.48	381	0.53	576	0.51
Tram	227	0.18	78	0.11	157	0.18	102	0.14	199	0.18
Metro	426	0.34	195	0.29	299	0.34	236	0.33	361	0.32
Total	1268	1	681	1	885	1	719	1	1136	1

Table 7.3: The share of transport modes used within each income group.

Since we already know the average amount of subsidy per trip per mode from Table 7.2, we use these numbers to calculate a new average subsidy per trip for public transport by multiplying them with the proportions in Table 7.3. The new averages are presented in Table 7.4.

Income Group	1	2	3	4	5
Bus	3.64	4.50	3.64	3.98	3.81
Tram	1.37	0.88	1.36	1.09	1.35
Metro	3.42	2.91	3.44	3.34	3.23
Total	8.44	8.29	8.44	8.41	8.39

Table 7.4: The new group specific subsidy averages per trip, when taking the difference in subsidy per mode into account. All numbers are in NOK.

The first impression is that the new averages per group are very similar to each other and the average subsidy we used in Alternative 1. It means that taking the difference in subsidy per mode into account has little effect on the subsidies received. The main reason for this is that the groups distribute their travels proportionately, with a small exception of group 2. Another point is that the subsidies per trip per mode are not substantially different. It results in group 2 only having a marginally lower subsidy per trip than the other groups. Figure 7.2 clearly demonstrate the small effect.

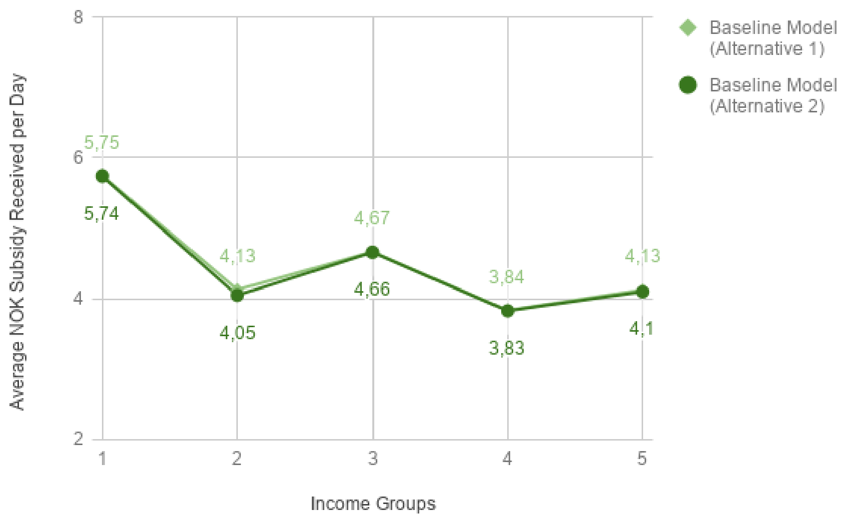

Figure 7.2: Comparison of the calculated subsidies between the Alternative 1 and 2. All numbers are given in NOK.

These results demonstrate that the condition in Alternative 1, which assumes no difference in subsidy between modes, holds, given the travel pattern in our sample. We will therefore apply this assumption in the following approach as well.

Alternative 3

We have so far applied the assumption that prices are equal, but this approach relaxes the condition. As Ruter offer discounts to certain groups in society based on age and occupational status, we seek to implement these discounts into our calculation.

Which ticket was bought? Was it bought with a discount? How many zones? All these factors will influence the price paid and the discounts received, but our data does not contain such information. This means we either need to make multiple crude assumptions or find a different way of measuring the subsidization of prices. Fortunately, even if prices differ between tickets and zones, they have something in common. The discounts based on occupation and age have a similar proportion of the cost no matter the ticket type and number of zones[3], compared to the most expensive ticket. In order to calculate these numbers we must assume that individuals are rational and would minimize their cost of transit and therefore buy the cheapest tickets available. The fare subsidies are calculated using the difference in price between the discounted groups. These subsidies are displayed as proportions in Table 7.5. The higher the share, the higher the degree of subsidization.

Discounted Groups	Adults	Child/Youth	Seniors	Students[4]
Subsidization of tickets	0	0.6	0.5	0.4

Table 7.5: The discounted share of a full price ticket per group.

Adults denotes all individuals not eligible to buy concessionary fares. We assume that their ticket is not subsidized in any way and therefore they pay the full price. This is denoted by the number 0. Child/Youth are children and youth up to 19 years old. Ruter separates these tickets, but as their prices (and hence proportion of subsidies) are the same, we have chosen to merge them. Seniors are defined as elderly above the age of 67 and individuals who are either disabled, blind or a registered partner to someone in these categories. Students are individuals below the age of 30 and enrolled in school or college. Table 7.5 clearly shows that children and youth receive the largest share of price subsidization, followed by Seniors and then Students. Based on these definitions we designate these values to the travelling individuals in our data set and calculate the average per income group, as displayed in Table 7.6

[3]The discounts are constant except for children/youth which increases slightly when adding multiple zones. We therefore apply the average discounts between zones for this age group.

[4]Students only receive discounts on 7 or 30 day tickets, so according to the assumption of minimizing costs we assume that they would buy these tickets.

Income Group	1	2	3	4	5
Average Subsidization of Tickets	0.25	0.16	0.10	0.09	0.07

Table 7.6: The calculated average discount given per ticket.

From Table 7.6 we see that group 1 receives the largest amount of fare discounts, and that the effect is negatively correlated with earnings. This implies that there are more discount-eligible persons in the low income groups. This is expected, as these discounts are directed explicitly towards individuals with relatively less income. If we disregard that subsidies also depend on the number of trips, the proportions demonstrate a progressive subsidy system. However, in this alternative method, we seek to implement trips and discounts into the estimation. By taking the average trip subsidies per day, which we found via Alternative 1, and the average discounts in Table 7.6 we can incorporate both in our subsidy calculation:

$$Average\ Subsidy\ Per\ Day_i \cdot (1 + Average\ Ticket\ Discount_i)$$

It means that the average subsidy per day increases by the average percentage of discounts for each group, which is denoted by i. For instance, since income group 1 has an average ticket discount of 0.25, the average subsidy per day will increase by 25%. The new subsidy per day is presented in Figure 7.3

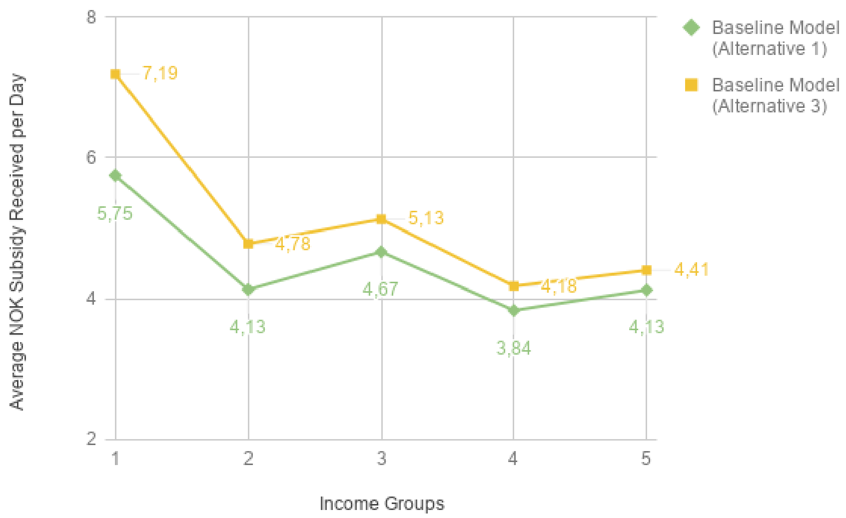

Figure 7.3: A comparison of using discounts in subsidy calculations versus not. All numbers are in NOK.

It illustrates that when fare discounts are considered, the calculated amount of subsidies for the lower income groups increase relatively more than those with higher income due to a higher share of discount-eligible individuals.

Summary

In this chapter, we have estimated subsidies received by individuals in each income group. By applying three different methods we demonstrate that income group 1 receives the most subsidies, regardless of our chosen methods. We also observe that discounted tickets have a substantial impact on the subsidies received per group. This result argues against the validity of assumption 1, equal prices, and rather that prices and discounts should be included in the calculation of subsidies. Another point is that even if subsidies vary between transport modes, it only has a minor impact on the overall distributional effects. The probable reason for this is that our groups have very similar travel patterns and that the difference in subsidy per mode is not substantial.

What is the magnitude of the distributive effects? By multiplying our estimated daily subsidies found via Alternative 3, with the number of days in a year (365) we can more easily compare the distributive effects of subsidies with income. With this method we find that a person in income group 1 receives on average 2 624 NOK in a year in public transport subsidies. Compared to the average income in group 1, which is 228 000 NOK, the average effect is 1.15% per year.

In comparison, a person in income group 5 receives on average 1 610 NOK. It means that the poorest receive 38% more public transit subsidies on average relative to the richest group. As these numbers are based on averages they are likely to be substantially lower than the distributional effect of someone who is a frequent user of public transport. However, our results indicate that the poorest in society receive the most subsidies, the system is progressive and that the distributional effects of public transport subsidies in Oslo and Akershus are noticeable.

8 | Discussion and Conclusion

The goal of this thesis has been to answer the question; Are subsidies to public transport progressive? There are no standard guidelines, no *golden rule* on how to calculate the distributive effects of public transport subsides. We have drawn inspiration from the few relevant articles on this particular topic, in order to create our own approach of analysis.

Taking inspiration from Fearnley (2006) we calculate travel propensities to measure relative dependency on public transport between different income groups. We find that the poorest group is by far the most dependent group on public transport in Oslo and Akershus with a propensity of 1.45, well above the dependency threshold of 1. By analyzing each specific transport mode we find similar results. The interpretation of the remaining groups dependency on public transport prove a tougher challenge (0.75-1.01) and emphasizes one of the weaknesses of this approach. Specifically, the ability to determine the difference in dependency, and if the differences are at all significant, especially when controlling for demographic factors. This motivates our use of econometric modelling.

The econometric analysis demonstrates that the effect of income on public transport is reduced for all groups when controlling for demographic factors. Moreover, the results were consistent with the propensity approach as the poorest group remained the most dependent on public transport in terms of number of trips. The demand model demonstrates that the other groups have significantly lower dependency both before and after the inclusion of demographic control variables. This suggests that the differences in public transport usage by income groups do not simply reflect demographic differences.

While there is statistically a difference in public transport use between the poorest group and the others, there is no clear income gradient in public transport usage. We perform a range of tests to evaluate the robustness of our results. These showed that our main results were robust to changes in the number of income groups, equivalence scales and change of dependent variable.

Together, the results from the propensity approach and the econometric models indicate that public transport subsidies are progressive, insofar as individuals with the lowest income have the highest usage (and therefore dependency). We calculate subsidies to provide estimates of the magnitude of the distributional effects. Our different calculations demonstrate several interesting aspects, including an increase in progressivity when ticket discounts where taken into consideration. It shows that the poorest group receives on average 2 624 NOK in public transit subsidies per year, which is 38% more than the richest group. The yearly amount of subsidies received for group 1 corresponds to 1.15% of their average income. Compared to other ways of distributing income, public transport subsidies may not be as effective. However, considering that the main focus of these subsidies are on mobility and environment, the distributional effects are noticeable.

Even if our results exhibits progressiveness, one should be careful with drawing conclusions based on our results because of several weaknesses with our data, the econometric models and some of the underlying assumptions.

First, the fact that our sample contains an over-representation of individuals with higher education compared to actual statistics from Statistics Norway can mean that our sample is non-random and thus cause biased coefficients. This over-representation implies that we are missing individuals from low education backgrounds , who consequently could belong to low income groups as income and education tend to be heavily correlated. If these missing observations have substantial differences in travel activity than the ones already analyzed, the progressivity could fall or increase based on this information.

Second, even if the number of trips can be a measure for public transport dependency, it can also be an indicator of non-optimal travel routes. As areas with less optimal travel routes are likely to have a higher proportion of low-income households, which could lead to biased coefficients. Especially the poorest group‘s public transport use could have an upward bias due to our current measure of public transport use. Other assumptions used in the analysis, such as equal operating costs and no subsidy leakage, are similarly not guaranteed to hold. Nonetheless, this approach is the most accurate given our data set, and the time and resources available.

Therefore, our results should be viewed more as an indicator of progressive subsidies to public transport in Oslo and Akershus than as actual proof. One should also be careful with generalizing these results to other cities or Norway as a whole since city and population characteristics, travel habits, income distribution and public transport networks can differ substantially. Although these results are not representative of other cities, it would still be interesting to compare them with similar cities and analyze the potential differences.

We introduced this thesis by implying that the distributive effects of public transit subsidies are not a first or even a second priority when subsidies and policies are determined by politicians. Instead, the spotlight has been placed on environmental concerns and the switch from private to public transport. If the results were regressive, the Norwegian government should take an active interest in the mechanisms behind such a result. This knowledge can help create new and improved actions to redistribute income and increase mobility to the relatively poorest in society.

Bibliography

Asensio, J., Matas, A. & Raymond, J.-L. (2003), 'Redistributive Effects of Subsidies to Urban Public Transport in Spain', *Transport Reviews* **23**(4), 433–452.

Atkinson, A. B., Hasell, J., Morelli, S. & Roser, M. (2017), *Inequality by Country*. Data retrieved from The Chartbook of Economic Inequality's web page., `https://www.chartbookofeconomicinequality.com/inequality-by-country/`[Accessed 4th of May, 2018].

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M. & White, P. (2004), 'The Demand for Public Transport: A Practical Guide', *TRL Report TLR 593* .

Baujard, A. (2013), '*Welfare Economics*', `https://halshs.archives-ouvertes.fr/halshs-00906907/document` [Accessed: 3th of June, 2018]. Working paper GATE 2013-33.

Becker, U. J., Becker, T. & Gerlach, J. (2012), The True Costs of Automobility: External Costs of Cars Overview on Existing Estimates in EU-27, Technical report, Technische Universität Dresden, Institute of Transport Planning and Road Traffic.

Bly, P. H., Webster, F. V. & Pounds, S. (1980), 'Effects of Subsidies on Urban Public Transport', *Transportation* **9**(4), 311–331.

Bly, P. & Oldfield, R. (1985), 'Relationships Between Public Transport Subsidies and Fares, Service, Costs and Productivity', *TRRL research report 24* .

Brewer, D. J. & Picus, L. (n.d.), *Encyclopedia of Education Economics and Finance.*, 1. Edition, SAGE Publications, Inc.

Button (2010), *Transport Economics, 3rd Edition*, EE.

Card, D. (1999), Chapter 30 - The Causal Effect of Education on Earnings, *in* O. C. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3, Elsevier, pp. 1801 – 1863.

Christiansen, P., Gundersen, F. & Gregersen, F. A. (1972), 'Kompakte byer og lite bilbruk? Reisemønster og arealbruk [Dense Cities and Low Car Use? Travel Behavior and Land Use Development in Norwegian Cities]', *TØI report I1505/2016* .

Dargay, J. (2007), 'The Effect of Prices and Income on Car Travel in the UK', *Transportation Research Part A Policy and Practice* **41**(10), 949 – 960.

Dargay, J. & Gately, D. (1999), ‘Income’s Effect on Car and Vehicle Ownership, Worldwide: 1960–2015’, *Transportation Research Part A: Policy and Practice* **33**(2), 101 – 138.

de Grange, L., González, F., Muñoz, J. C. & Troncoso, R. (2013), ‘Aggregate Estimation of the Price Elasticity of Demand for Public Transport in Integrated Fare Systems: The Case of Transantiago’, *Transport Policy* **29**, 178 – 185.

Denstadli, J. M. & Lian, J. I. (2003), Combining Surveys of Daily and Long-Distance Mobility: An Empirical Investigation of Non-Response Problems, *in* K. W. Axhausen, J. L. Madre, Polak & P. L. Toint, eds, ‘Capturing Long-Distance Travel’, Research Studies Press, Baldock, pp. 121–132.

Directorate of Integration and Diversity (2010), ‘*Tiltak for å heve svarprosenten*’. `https://www.tolkeportalen.no/no/brukerundersokelser/Kapittel-3/34-Tiltak-for-a-heve-svarprosenten/` [Accessed: 19th of April, 2018].

Directorate of Integration and Diversity (2018), ‘*Barnehage, skole og utdanning*’. `https://www.imdi.no/om-integrering-i-norge/innvandrere-og-integrering/barnehage-og-utdanning/#ssb_utdanning2` [Accessed: 19th of April, 2018].

Elgar, I. & Kennedy, C. (2005), ‘Review of Optimal Transit Subsidies: Comparison Between Models’, *Journal of urban planning and development* **131**(2), 71–78.

Fearnley, N. (2006), ‘Public Transport Subsidies in the UK: Evidence of Distributional Effects’, *World Transport Policy and Practice* **12**(1), 30–39.

Fearnley, N., Flügel, S., Killi, M., Gregersen, F. A., Wardman, M., Caspersen, E. & Toner, J. P. (2017), ‘Triggers of Urban Passenger Mode Shift–State of the Art and Model Evidence’, *Transportation Research Procedia* **26**, 62–80.

Frankena, M. (1973), ‘Income Distributional Effects of Urban Transit Subsidies’, *Department of Economics Research Reports* **7311**.

Graham, D. J., Couto, A., Adeney, W. E. & Glaister, S. (2003), ‘Economies of Scale and Density in Urban Rail Transport: Effects on Productivity’, *Transportation Research Part E: Logistics and Transportation Review* **39**(6), 443–458.

Gregersen, F. A. (2017), ‘Weights in the norwegian national travel survey’, *TØI Report nr. 1548/2017)* .

Hellevik, O. (2015), *The Implications of Nonresponse in Survey Research*, Norwegian Journal of Social Research 02/2015 (Volume 56), Universitetsforlaget.

Hirsch, A. A., Flatebø, G., Modig, I., Sandnes, T., Aalandslid, V., Thorsen, L. R. & Stene, R. J. (2010), 'Women and Men in Norway, What the Figures Say. Revised edition', p. 14.

Hjorthol, R. (2006), 'Bilens betydning for barns og foreldres bevegelsesmønstre', *TØI Report 834/2006* .

Hjorthol, R., E. Ø. & Uteng, T. (2014), '2013/14 National Travel Survey - Key Results', *TØI Report 1383/2014* .

Langeland, T. (2009), Language and Change: An Inter-Organisational Study of the Zero Vision in the Road Safety Campaign, PhD thesis.

Le Grand, J. (1982), *The Strategy of Equality: Redistribution and the Social Services.*, Gearoge Allen & Unwin Ltd, London.

Mattson, J. (2008), 'The Effects of Gasoline Prices on Bus Ridership for Different Types of Transit Systems ', *Journal of the Transportation Research Forum* **47**(3), 5–21.

Metz, D. (2010), 'Saturation of demand for daily travel', *Transport Reviews* **30**(5), 659–674.

Meyer, A. (2015), 'Does Education Increase Pro-environmental Behavior? Evidence from Europe', *Ecological economics* **116**, 108–121.

Ministry of Finance (2001), 'St.meld. nr. 30'.

Mohring, H. (1972), 'Optimization and Scale Economies in Urban Bus Transportation', *The American Economic Review* **62**(4), 591–604.

Moore, Jeffrey C. & Welniak, E. J. (2000), 'Income Measurement Error in Surveys: A Review', *Journal of Official Statistics* **16**(4), 331–361.

Norwegian Ministry of Transport and Communications (2016), 'St.meld.33 nasjonal transportplan 2018–2029'.

OECD (2005), *Environmentally Harmful Subsidies*, OECD Publishing, Paris.

OECD (2013), '*Income Distribution Data Review - Norway*'. `https://www.oecd.org/els/soc/OECDIncomeDistributionDataReview-Norway.pdf` [Accessed: 6th of May, 2018].

OECD, Directorate for Employment, L., Affairs, S. & the Statistics Directorate (2012), Quality review of the OECD database on household incomes and poverty and the OECD earnings database, Part I, Technical report, OECD.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J. & White, P. (2006), 'The Demand for Public Transport: The Effects of Fares, Quality of Service, Income and Car Ownership', *Transport Policy* **13**(4), 295 – 306.

Preston, J. (2008), Public Transport Subsidisation, *in* S. Ison & T. Rye, eds, 'The Implementation and Effectiveness of Transport Demand Management Measures: An international perspective', Ashgate Publishing Limited, pp. 189–209.

Pucher, J. (1983), 'Who Benefits from Transit Subsidies? Recent Evidence from Six Metropolitan Areas', *Transportation Research Part A: General* **17**(1), 39 – 50.

Ruter (2017), '2016 Annual Report', `https://ruter.no/globalassets/dokumenter/aarsrapporter/ruter-arsrapport-20161.pdf`[Accessed: 27th of March, 2018].

Serebrisky, T., Gómez-Lobo, A., Estupiñán, N. & Muñoz-Raskin, R. (2009), 'Affordability and Subsidies in Public Urban Transport: What Do We Mean, What Can Be Done?', *Transport Reviews* **29**(6), 715–739.

Statistics Norway (2014), '*Population and Population Changes*'. `https://www.ssb.no/statbank/table/07459/tableViewLayout1/?rxid=a66369cc-ad34-4e36-a940-ebfdd3d0985d` [Accessed: 15th of May, 2018].

Statistics Norway (2017), '*Highest rents in Oslo*'. `https://www.ssb.no/en/priser-og-prisindekser/artikler-og-publikasjoner/highest-rents-in-oslo--332565` [Accessed: 15th of May, 2018].

Steinsland, C., Østli, V. & Fridstrøm, L. (2016), 'Fordelingseffekter av endret bilbeskatning / distributional effects of changed car taxation', *TØI rapport 1463* p. 3.

Stokes, G. & Lucas, K. (2011), 'National travel survey analysis', *TSU Working Paper Series* **Ref. 1053**.

Strathman, J. G., Dueker, K. J. & Davis, J. S. (1994), 'Effects of Household Structure and Selected Travel Characteristics on Trip Chaining', *Transportation* **21**(1), 23–45.

The Norwegian Tax Administration (2018), '*National Registry*'. `https://www.skatteetaten.no/en/person/national-registry/` [Accessed: 27th of April, 2018].

The World Bank (2018), '*GDP per Capita, in USD*'. Data retrieved from World Bank national accounts data, and OECD national accounts data files, `https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=NO-ES-GB`[Accessed 4th of May, 2018].

Urbanet Analyse (2015), ‘Reisevaner i osloområdet. en analyse av den nasjonale reisevaneundersøkelsen 2013/14’, *PROSAM Report nr. 218* .

Vibe, N., ystein Engebretsen, O. & Fearnley, N. (2005), ‘Determinant of Urban Transport Development in Norway’, *TØI report 761/2005* .

Wooldridge, J. M. (2016), *Introductory Econometrics - A modern approach*, sixth edition, Cengage Learning.

Woolridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2. Edition, The MIT Press.

Appendices

A Appendix

A.1 Descriptive statistics

Table A.2: Distribution of Household Members for Age Group 61+

Size of household	Frequency	Percent	Cumulative Percent
1	652	26.68	26.68
2	1686	68.99	95.66
3	79	3.23	98.90
4	18	0.74	99.63
5	4	0.16	99.80
6	3	0.12	99.92
7	2	0.08	100.00
Total	2444	100.00	

Table A.3: Average Travel Length per Income Group

Income Group	Average Distance Traveled (km)
1	38.14
2	50.17
3	48.98
4	57.91
5	60.57

Table A.1: Correlation matrix of the variables used in our preferred model specification.

	(1)																				
	Trips	Income group 1	Income group 2	Income group 3	Income group 4	Income group 5	Oslo	Age1	Age2	Age3	Age4	Male	Children	Higher education	Worktrip	Free parking	Weekend	Summer	Spring	Fall	Winter
Trips	1																				
Income group 1	0.103	1																			
Income group 2	-0.057	-0.234	1																		
Income group 3	0.003	-0.229	-0.234	1																	
Income group 4	-0.036	-0.227	-0.231	-0.226	1																
Income group 5	-0.00992	-0.278	-0.283	-0.277	-0.274	1															
Oslo	0.298	0.073	-0.092	0.031	-0.058	0.042	1														
Age 13-17	0.022	0.017	0.086	-0.060	0.063	-0.096	-0.032	1													
Age 18-25	0.196	0.240	-0.011	-0.06	-0.037	-0.119	0.144	-0.049	1												
Age 26-60	-0.004	-0.144	0.022	-0.069	0.235	-0.038	0.037	-0.221	-0.364	1											
Age 60+	-0.120	0.009	-0.049	0.133	-0.257	0.147	-0.113	-0.107	-0.176	-0.789	1										
Male	-0.058	-0.067	0.003	-0.030	0.029	0.058	0.012	0.017	0.006	-0.048	0.043	1									
Children	-0.073	-0.056	0.228	-0.188	0.424	-0.367	-0.108	-0.094	-0.171	0.479	-0.384	-0.026	1								
Higher education	0.057	-0.177	-0.076	-0.036	0.137	0.138	0.135	-0.221	-0.167	0.244	-0.083	-0.023	0.193	1							
Worktrip	0.161	-0.136	-0.040	-0.037	0.124	0.081	0.039	-0.140	-0.018	0.353	-0.321	0.029	0.157	0.127	1						
Free parking	-0.065	-0.126	0.004	-0.044	0.110	0.052	-0.085	-0.124	-0.034	0.323	-0.285	0.004	0.196	0.09	0.308	1					
Weekend	-0.123	0.002	0.0003	0.012	-0.011	-0.003	-0.006	-0.008	-0.0007	-0.012	0.016	0.013	0.008	0.0002	-0.396	0.001	1				
Summer	-0.055	-0.028	0.013	0.02	-0.003	-0.002	-0.02	-0.032	-0.0001	-0.012	0.025	0.003	-0.002	0.003	-0.043	-0.005	-0.01	1			
Spring	-0.006	-0.011	0.009	-0.013	0.009	0.005	0.006	0.029	-0.01	-0.001	-0.004	-0.009	-0.006	0.01	-0.003	-0.016	0.006	-0.339	1		
Fall	0.03	0.03	-0.015	0.008	-0.02	0.002	0.012	-0.02	0.01	-0.006	0.007	0.012	-0.008	0.009	0.008	-0.02	0.02	-0.305	-0.33	1	
Winter	0.032	0.01	-0.007	-0.014	0.016	-0.005	0.002	0.019	-0.0001	0.018	-0.027	-0.004	0.015	-0.022	0.037	0.038	-0.017	-0.335	-0.363	-0.327	1

Table A.4: The corresponding Variance inflation factors (VIF) connected to the preferred model specification. A VIF below 10 is normally considered

Variable	VIF	1/VIF
Income Group 2	1.78	0.56
Income Group 3	1.78	0.56
Income Group 4	1.96	0.51
Income Group 5	2.20	0.45
Oslo	1.11	0.90
Age 13 - 17	1.22	0.82
Age 18 - 25	1.28	0.78
Age 61 +	1.59	0.63
Male	1.02	0.98
Children	1.98	0.51
Higher Education	1.22	0.82
Worktrip	1.54	0.65
Free Parking	1.23	0.81
Weekend	1.24	0.81
Spring	1.56	0.64
Fall	1.52	0.66
Winter	1.56	0.64
Mean VIF	1.52	

Table A.5: Three different divisions of income groups based on EU-scale adjusted household income.

	Proportions of the Population		
Income Group	4 Income Groups	5 Income Groups	6 Income Groups
1	0.26	0.19	0.18
2	0.25	0.21	0.16
3	0.29	0.22	0.17
4	0.20	0.18	0.16
5	-	0.20	0.18
6	-	-	0.14
Sum	1	1	1

Table A.6: Correlation Between Car Ownership and Trips

	Car Ownership	Trips
Cars Ownership	1.00	
Trips	-0.3366	1.00

A.2 Figures

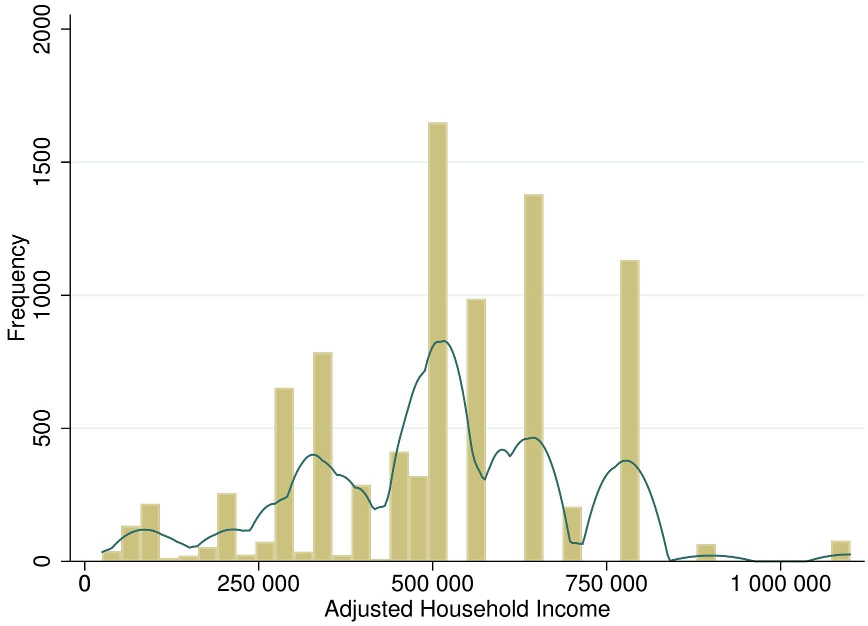

Figure A.1: Histogram of Adjusted Household Income

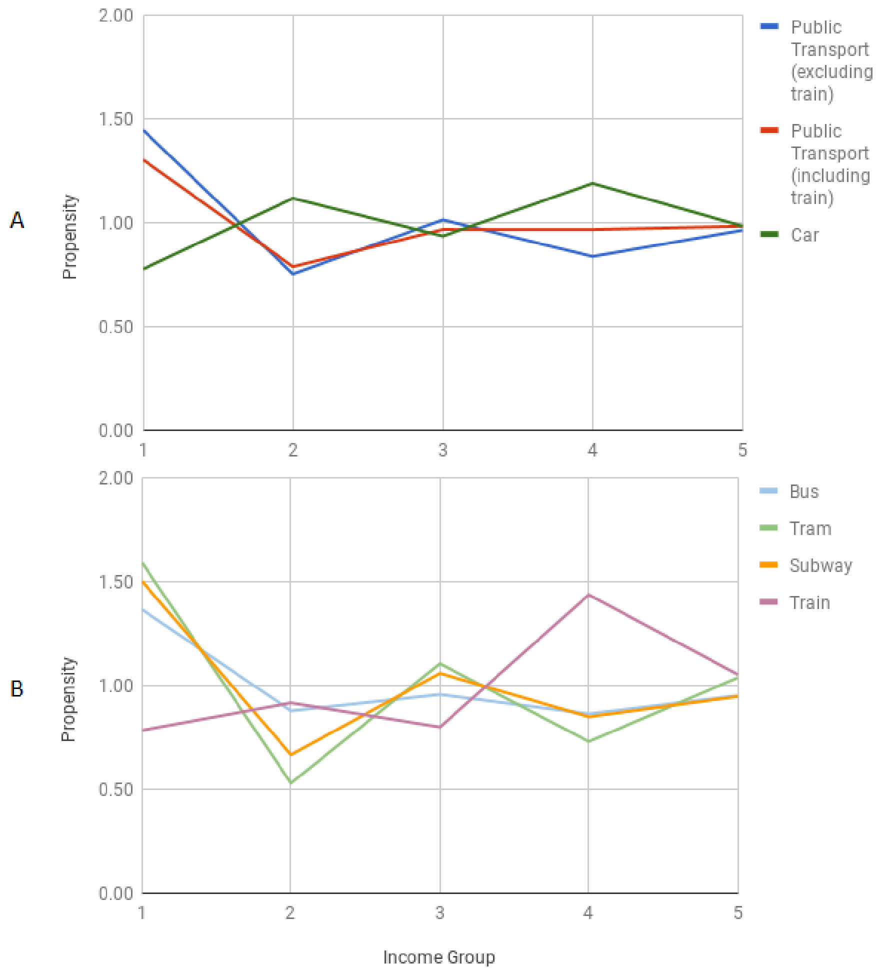

Figure A.2: Graphic visualization of the propensity scores from tables 5.1(A) and 5.2(B).

A.3 Demand Model

Table A.7: Demand model - Part 1

	(1) Trips	(2) -	(3) -	(4) -	(5) -	(6) -
Incomegroup2	-0.367^{***}	-0.359^{***}	-0.313^{***}	-0.217^{***}	-0.222^{***}	-0.186^{***}
	(0.0404)	(0.0404)	(0.0404)	(0.0385)	(0.0385)	(0.0377)
Incomegroup3	-0.229^{***}	-0.225^{***}	-0.246^{***}	-0.211^{***}	-0.231^{***}	-0.112^{***}
	(0.0432)	(0.0433)	(0.0436)	(0.0408)	(0.0414)	(0.0405)
Incomegroup4	-0.322^{***}	-0.311^{***}	-0.232^{***}	-0.167^{***}	-0.188^{***}	-0.194^{***}
	(0.0411)	(0.0411)	(0.0427)	(0.0409)	(0.0416)	(0.0414)
Incomegroup5	-0.255^{***}	-0.242^{***}	-0.282^{***}	-0.243^{***}	-0.281^{***}	-0.155^{***}
	(0.0401)	(0.0402)	(0.0413)	(0.0391)	(0.0404)	(0.0399)
Male		-0.113^{***}	-0.117^{***}	-0.126^{***}	-0.122^{***}	-0.119^{***}
		(0.0233)	(0.0233)	(0.0224)	(0.0224)	(0.0220)
Children			-0.171^{***}	-0.119^{***}	-0.150^{***}	-0.118^{***}
			(0.0297)	(0.0282)	(0.0289)	(0.0329)
Oslo				0.669^{***}	0.652^{***}	0.579^{***}
				(0.0272)	(0.0277)	(0.0276)
Higheducation					0.101^{***}	0.154^{***}
					(0.0244)	(0.0241)
Age1						0.271^{***}
						(0.0722)
Age2						0.579^{***}
						(0.0626)
Age4						-0.197^{***}
						(0.0272)
Constant	0.766^{***}	0.815^{***}	0.859^{***}	0.569^{***}	0.537^{***}	0.462^{***}
	(0.0333)	(0.0351)	(0.0367)	(0.0332)	(0.0335)	(0.0377)
Observations	8861	8861	8861	8861	8861	8861
Adjusted R^2	0.012	0.015	0.018	0.100	0.102	0.129

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Demand model - Part 2

	(1)	(2)	(3)	(4)
	Trips	-	-	-
Incomegroup2	-0.211^{***}	-0.200^{***}	-0.198^{***}	-0.191^{***}
	(0.0374)	(0.0373)	(0.0372)	(0.0372)
Incomegroup3	-0.157^{***}	-0.141^{***}	-0.134^{***}	-0.128^{***}
	(0.0400)	(0.0400)	(0.0398)	(0.0397)
Incomegroup4	-0.251^{***}	-0.237^{***}	-0.232^{***}	-0.226^{***}
	(0.0411)	(0.0410)	(0.0409)	(0.0409)
Incomegroup5	-0.237^{***}	-0.207^{***}	-0.196^{***}	-0.192^{***}
	(0.0395)	(0.0398)	(0.0396)	(0.0395)
Male	-0.133^{***}	-0.132^{***}	-0.128^{***}	-0.128^{***}
	(0.0217)	(0.0217)	(0.0216)	(0.0216)
Children	-0.129^{***}	-0.113^{***}	-0.109^{***}	-0.109^{***}
	(0.0324)	(0.0323)	(0.0322)	(0.0322)
Oslo	0.578^{***}	0.556^{***}	0.557^{***}	0.555^{***}
	(0.0271)	(0.0273)	(0.0272)	(0.0272)
Higheducation	0.148^{***}	0.147^{***}	0.148^{***}	0.149^{***}
	(0.0238)	(0.0237)	(0.0237)	(0.0237)
Age1	0.454^{***}	0.390^{***}	0.367^{***}	0.358^{***}
	(0.0733)	(0.0742)	(0.0741)	(0.0740)
Age2	0.613^{***}	0.604^{***}	0.601^{***}	0.602^{***}
	(0.0626)	(0.0626)	(0.0623)	(0.0621)
Age4	-0.0511^{*}	-0.101^{***}	-0.117^{***}	-0.118^{***}
	(0.0285)	(0.0294)	(0.0298)	(0.0298)
Worktrip	0.382^{***}	0.420^{***}	0.360^{***}	0.353^{***}
	(0.0244)	(0.0250)	(0.0279)	(0.0279)
Freeparking		-0.195^{***}	-0.183^{***}	-0.184^{***}
		(0.0251)	(0.0252)	(0.0252)
Weekend			-0.146^{***}	-0.150^{***}
			(0.0240)	(0.0240)
Spring				0.0636^{**}
				(0.0287)
Fall				0.120^{***}
				(0.0306)
Winter				0.128^{***}
				(0.0296)
Constant	0.302^{***}	0.369^{***}	0.420^{***}	0.342^{***}
	(0.0382)	(0.0392)	(0.0411)	(0.0440)
Observations	8861	8861	8861	8861
Adjusted R^2	0.153	0.160	0.162	0.164

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Demand model with different income groups based on household income adjusted with the EU-Scale.

	(1) Trips	(2) -	(3) -
Incomegroup 2	-0.165^{***}	-0.218^{***}	-0.202^{***}
	(0.0311)	(0.0365)	(0.0392)
Incomegroup 3	-0.108^{***}	-0.157^{***}	-0.228^{***}
	(0.0324)	(0.0379)	(0.0388)
Incomegroup 4	-0.144^{***}	-0.193^{***}	-0.132^{***}
	(0.0359)	(0.0393)	(0.0433)
Incomegroup 5		-0.202^{***}	-0.213^{***}
		(0.0407)	(0.0409)
Incomegroup 6			-0.217^{***}
			(0.0447)
Constant	0.302^{***}	0.355^{***}	0.368^{***}
	(0.0412)	(0.0435)	(0.0437)
Control Variables	$Y.es$	$Y.es$	$Y.es$
Seasonal Dummies	$Y.es$	$Y.es$	$Y.es$
Weekend Dummies	$Y.es$	$Y.es$	$Y.es$
Observations	8861	8861	8861
Adjusted R^2	0.163	0.164	0.164

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.4 Probit Model

Table A.10: Marginal Effects

	(1) Marginal Effects Probit Trip
Income group 2	-0.0595***
	(-4.75)
Income group 3	-0.0356**
	(-2.75)
Income group 4	-0.0605***
	(-4.65)
Income group 5	-0.0470***
	(-3.67)
Male	-0.0482***
	(-5.88)
Children	-0.0345**
	(-2.85)
Oslo	0.202***
	(20.34)
Higher education	0.0626***
	(6.81)
Age 13-17	0.217***
	(6.88)
Age 18-26	0.199***
	(9.65)
Age 61+	-0.0554***
	(-5.00)
Worktrip	0.144***
	(13.87)
Free parking	-0.0806***
	(-9.24)
Weekend	-0.0635***
	(-5.93)
Spring	0.0269*
	(2.24)
Fall	0.0511***
	(3.99)
Winter	0.0430***
	(3.52)
Observations	8861

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$