Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Economics

A Drop in the Ocean?

Behavioral Spillover Effects and Travel Mode Choice

Franziska Klein

Master's thesis • 30 hec • Advanced level
Environmental Economics and Management – Master's Programme
Degree thesis No 1091 • ISSN 1401-4084
Uppsala 2017

A Drop in the Ocean – Behavioral Spillover Effects and Travel Mode Choice

Franziska Klein

Supervisor: Claes Ek, Swedish University of Agricultural Sciences, Department of Economics

Examiner: Rob Hart, Swedish University of Agricultural Sciences, Department of Economics

Credits: 30 hec
Level: A2E
Course title: Independent project in economics
Course code: EX0811
Programme/education: Environmental Economics and Management - Master's Programme

Place of publication: Uppsala
Year of publication: 2017
Part number: 1091
ISSN: 1401-4084
Online publication: http://stud.epsilon.slu.se

Keywords: behavioral spillover effects, moral licensing, travel mode choice, natural experiment

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Natural Resources and Agricultural Sciences

Abstract

This study aims to identify behavioral spillover effects related to travel mode decisions. In particular the effect of a shift towards environmentally friendly day-to-day commuting on occasional longer distance travel modes is explored. Therefore a natural experiment caused by the implementation of congestion charging in Stockholm and Gothenburg is exploited. In a fixed effects model with a control group, the average treatment effect on the treated is estimated by regressing the number of short-haul flight trips on a policy treatment dummy and the amount of the charge, respectively. The findings reveal predominantly positive estimates, confirmed and strengthened by various extensions, and thus point towards negative spillover effects.

Contents

List of Figures

List of Tables

1 Introduction

Air transportation currently contributes roughly 3% to global anthropogenic carbon emissions and is likely to increase quickly due to growing demand for air travel and decreasing returns in fuel efficiency [39]. In order to reduce anthropogenic effects on the climate and to keep global warming well below 2 degrees compared to pre-industrial levels, 195 nations have signed the *COP21* agreement in Paris in 2015. With emission reductions officially on the agenda of most of these countries, it is important to evaluate the effects of environmental policies properly to make sure they are more than a drop in the ocean. This is usually not an easy task. An increasing number of studies show the relevance of behavioral spillover effects, especially in pro-environmental contexts [45].

Spillover effects can be defined as a change in one behavior, triggered by some change in a previous behavior. When a policy is implemented, it is likely to not only affect the targeted behavior, but also to "spill over" to other actions beyond that. These effects are particularly problematic when they lead to *rebound* or *backfire* effects: a net partial offsetting of the original goal or even adverse effects. A famous example is the widely discussed *Jevon's Paradox* [41]. In order to be able to tackle climate change - and other environmental problems - efficiently, it is crucial for policy makers to understand the whole chain of impacts their decisions cause.

Regarding the increasing impact of air travel, the research question of the underlying study is whether there is evidence for spillover effects in the mobility sector. In particular, the challenge is to find out if a shift towards environmentally friendly commuting on a local level affects people's choice of travel modes for longer distances. This is studied in the context of a natural experiment in Scandinavia - the scattered implementation of congestion charging.

Exploiting the sudden and noticeable shift of many people from car commuting to public transportation triggered by the implementation of inner city congestion charging in Stockholm and Gothenburg, this study will explore the effects on short-haul air travel. In a difference-in-differences setting, the development of the number of short-haul flights in the treated cities is compared to the same outcome variable for a control group of other Scandinavian cities. These are Malmö, Copenhagen, Helsinki and Oslo. To answer the research question, three sub goals are set up: (a) find evidence for the shift in daily commuting behavior in the treated cities, (b) look for causal inference between the number of flights and congestion charging and (c) use some extensions to see if they alter the results.

The rest of the paper is structured as follows. Section 2 gives an overview of the background of this study, including a literature review, the basic conception of the problem and a summary of the effects of congestion charging. Section 3 then dives into the methodologies which are used to empirically analyze the problem at hand. The results of the main analysis and several extensions are presented and interpreted in section 4. Section 5 concludes.

2 Background of the Study

2.1 Literature Review

Over the past years, behavioral aspects departing from standard rational choice theory have gained importance. Shogren & Taylor for instance emphasize the influence of deviating behavior for environmental economics [40]. Other studies show that policy makers are becoming increasingly aware of the results from behavioral research [9, 17, 43]. Following this development, there is an increasing bulk of literature on so called behavioral spillover effects. Nilsson et al. define them as: "the effect that an initial behavior has on a subsequent behavior or the same behavior in a different time or context" [35].

Spillover effects can work in different directions [16]. They can be *promoting*, i.e. sequential behaviors have concordant signs, which often means that individuals have a strong preference for consistency [1]. Theories supporting this type of spillover effects include *cognitive dissonance theory* [4] or *foot-in-the-door* effects [20] for example. Discordant directions of sequential behaviors can be *permitting* or *purging*. They include effects like *ego depletion* [31], *moral licensing* [5] or *moral cleansing* [48]. While Cascio & Plant [11] have found evidence for prospective licensing, this study focusses on the more common assumption that people use up moral credits they already acquired through earlier behavior. Table 1 gives an overview about the directions of spillover effects.

Table 1: Spillover Effects in Pro-Environmental Behavior

Spillover Type	**Initial Behavior**	**Subsequent Behavior**	
Promoting	Pro-environmental (Polluting)	Pro-environmental (Polluting)	positive
Permitting	Pro-environmental	Polluting	negative
Purging	Polluting	Pro-environmental	negative

Dolan & Galizzi summarize that different kinds of spillovers can be facilitated in different contexts [16]. They claim that the five main streams of literature related to this focus on the cost of the initial behavior, the level of completeness of reaching a goal, the (spatial or temporal) proximity of two sequential behaviors, the trade-off between different motives and the cognitive mindset during the behaviors.

Particularly interesting to the underlying study are some findings related to costs and motives of behavior. Studies by Gneezy et al. indicate that permitting spillovers are more likely than promoting if the cost of the first behavior is low [21]. Other studies find evidence that promoting spillovers are less likely to occur when the first behavior is related to an external cause [37, 47]. According to Dhar & Simonson the type of spillover depends on the trade-off that is made. They find that trade-offs between motives and resources facilitate promoting spillovers, compared to trade-offs between different motives, which promote permitting spillovers [13].

This does not pave the way for expecting any particular outcome in the present study. The (perceived) cost of shifting towards public transportation becomes lower under congestion charging, as it becomes relatively cheaper compared to car usage. This would favor a permitting effect. The fact that a price change provides an external cause for behavioral changes lowers the probability of promoting spillover effects. The trade-off between motives (pro-environmental image) and resources (time and money) on the other hand would fuel expectations about promoting spillovers. Against this background neither positive, nor negative spillover effects can be excluded a priori.

Most of the previously mentioned results are obtained through qualitative experiments or surveys. These research methods can be desirable, especially if one is concerned with understanding the driving forces in decision making processes. However, their results are often biased due to the hypothetical nature of some experiments [34] or through *survey effects* [16] for instance. If the focus lies on finding reliable empirical evidence, data from natural experiments can offer an attractive alternative.

Behavioral spillovers have been studied empirically through actual or natural experiments before. Tiefenbeck et al. study the effect of receiving tips for saving water on households' water and electricity consumption in a controlled field experiment and find evidence for moral licensing [44]. Ek & Miliute-Plepiene find policy-driven positive spillover effects from food-waste collection in Sweden by analysing a natural experiment in a difference-in-differences approach [18]. Jacobsen et al. empirically strengthen the existence of moral licensing by showing that certain households increase their electricity consumption upon a shift to "green electricity" in a green electricity program in Tennessee [30].

While there has been much research on spillover effects, it is hard to find studies connecting this phenomenon to the demand of air travel. This might be because flying has commonly been seen as a desirable behavior afflicted with positive social norms, power and status. However, evidence is accumulating that consumers of air travel are becoming more aware of the negative environmental impacts of their behavior and even that there is a growing negative discourse around frequent flying [12].

Another reason for the gap in the literature on spillover effects and travel decisions might be that pro-environmental behavior and travel mode choices seem to be poorly linked in general. The *low-cost hypothesis*, which states that the magnitude of effects of environmental concern on pro-environmental behavior diminishes with increasing behavioral costs [14] supports that. Diekman & Preisendörfer find in a survey in Germany that car ownership and holiday air travel are only weakly related to environmental attitudes compared to "cheaper" behaviors, such as recycling, water use or shopping. In line with this are other studies, which show that most people continue holiday air travel despite their environmental concerns [12, 24, 33]. Whitmarsh & O'Neill even show that people with stronger pro-environmental attitudes travel more [46].

2.2 Conceptual Framework

The conceptual framework of this study roughly follows a scheme that Dolan & Galizzi [16] developed for the interpretation of behavioral spillovers. Similarly, there is a skeleton of three basic pillars, but some of the subordinated assumptions are adjusted. The first central assumption is that there are two different behaviors, which take place sequentially, based on the defintion of spillover effects earlier. Specifically, it is assumed that there is some intervention targeted at, and changing, the initial behavior.

Does altering Behavior Through Price Changes Count?
While Dolan & Galizzi [16] exclude all behavioral changes based on price mechanisms from their concept, the initial behavior in the underlying study is altered primarily by a change in relative prices. Imposing the charge de facto raises the price for car usage. According to the foundations of microeconomic theory, as described by Pindyck & Rubinfeld or Gravelle & Rees [23, 36], for example, this will increase the slope of the budget line and lead to shifts in demand patterns. In particular, rational choice theory predicts a decrease in demand for car travel when its price increases. It is however not as clear what happens to the consumption of the second good or service; in this case air travel. Depending on the consumers' preferences and the use of the charge revenue, its demand might either increase or decrease. Appendix A visualizes potential income and substitution effects. In

reality, it is difficult to disentangle these effects, and as mentioned earlier, attention should be paid to behavioral responses that are not covered by rational choice theory.

One might assume for example that even if people are adjusting their behavior following a change in relative prices, they might give themselves moral credit for this adaptation. This can be seen as particularly likely, if an intervention is based on aspects of environmental improvement, or if the public debate of the intervention is dominated by such effects. In addition to that, there might even be behavioral spillover effects for those people who are not adapting their behavior under the policy, because they receive a signal that their choice is bad.

The Basis for Linking Behaviors
The second pillar of the conceptual frame is that the two sequential behaviors must be linked by some kind of motive, which is a common driver underlying both actions. Assuming that people are maximizing their utility, the sources of utility need to be considered. One can argue that in the context of pro-environmental behavior, it is possible to differentiate between some form of personal and altruistic utility. People gain some utility directly from consuming a certain good or service; in this case from the travel mode they choose. Nevertheless they can receive additional altruistic utility. This can be immediate, like the often cited *warm-glow* from charitable giving [3, 29]. Moral credential hypothesis for example predicts that individuals accumulate positive feelings from taking a moral decision, in this case choosing environmentally friendly travel modes.

Apart from that, consumers might also consider some kind of long-term utility. This is particularly relevant for bridging behaviors at different points in time. This paper assumes that the two most important deeper preferences concern self-perception and external perception. Individuals can either pursue higher-order identities (life goals), or they can try to create a certain external image of themselves. While superiority of the first goal should lead to more consistent behavior and hence positive spillover effects, because inconsistent decisions cause tensions, dominance of the external perception goal is likely to lead to negative spillovers. That is because the people that surround a person daily might not ever get to know which long distance travel mode choice that person makes.

How Are Different Behaviors Linked?
The third and last pillar of the conceptual framework concerns how the initial behavior affects the subsequent behavior. Generally there could be positive, negative or no spillover effects, depending heavily on the underlying motive. Blanken et al. [5] have found evidence for the existence of a moral licensing effect, but the literature suggests mixed results about the direction of spillover effects. The fundamental idea in this paper is that people will attribute themselves credit for a shift towards more environmentally friendly travel modes in their day-to-day lives. Hence they could tend to choose more or less environmentally friendly travel modes for occasional long-distance trips afterwards. As the theory provides reasonable explanations for both behaviors, the analysis will be performed in a way that allows for both directions: a two-sided hypothesis test.

2.3 Road Pricing in Scandinavia

The basic assumption for this study is that some initial behavior is changed, which in turn causes a following behavior to change. This is supposed to happen only for the treatment, not for the control group. The intitial behavior is the decision about travel modes for daily commuting, which is equivalent to the direct effect of the congestion charge. Thus it needs to be made sure that the introduction of the policy actually caused a shift in travel mode choices (Sub goal (a)). In particular the shift should have been towards more environmentally friendly modes than cars (e.g. public transportation or cycling), if the assumption that people give themselves moral credit for the decision should hold. Once this is established, the analysis can be taken further to see if there have been any changes in the subsequent behavior (the number of flight trips).

There are various forms of road pricing, like road tolling, congestion charging, or different forms of time or distance-based fees. While road tolling for instance usually serves the purpose of financing infrastructure investments, congestion charging is rather aimed at improving environmental quality. Globally there are only a handful of cities that have established congestion charging thus far. Some examples are Singapore in 1975 [22], London in 2003 [32], Stockholm in 2006 [6], Milan in 2008 [38] and Gothenburg in 2013 [7]. Börjesson & Kristoffersson highlight that many other cities, like Edinburgh, Manchester, New York, or also Helsinki and Copenhagen have discussed the implementation of congestion charging, but have not implemented them in spite of high pollution levels [7].

This study focuses on the congestion charging schemes implemented in Stockholm and Gothenburg. Both cities have established a time-dependently priced cordon around their city center to reduce overall traffic pollution, especially during rush hours. Oslo has - as many other Norwegian cities - had road pricing through a toll ring for quite a long time (since 1990). The scheme mainly differs from the Swedish congestion charging in that fees are not time-differentiated, heavy vehicles are charged more and that the fee is only raised when entering into the city[1]. It is important to emphasize that the treatment group is not characterized by having road pricing in place, but rather by having changed travel behavior through their road pricing policy. The toll system in Oslo for instance is designed to affect travel behavior as little as possible [19,28]. Table 2 gives an overview about road pricing in the six cities which are included in the study.

Table 2: Road Pricing in Scandinavia

City	**Road Pricing Measure**	**Effects on Travel Behavior**
Stockholm	Time-differentiated cordon around the inner city (current peak price 35 SEK)	*Real* traffic reductions of up to 31%
Gothenburg	Time-differentiated cordon around the inner city (current peak price 22 SEK)	Traffic reductions around 12%
Malmö	No road pricing in place	-
Copenhagen	No road pricing in place, but widely discussed	-
Helsinki	No road pricing in place, but widely discussed	-
Oslo	Evenly priced toll belt around the city centre	Potentially minor traffic reductions[2]

Before implementing a permanent congestion charging scheme, Stockholm went through a trial phase from January 3 until July 31 2006. Parallel to that, an extension of the public

[1]The heavy vehicle category includes vehicles with a weight of more than 3,500kg. More information can be found under http://www.autopass.no/en/autopass

[2]As the toll ring in Oslo was designed in a way that affects travel behavior as little as possible, there are no detailed studies about behavioral changes. Some studies claim to find a slight increase or decrease in car usage, but the results are minor and mixed.

transportation network took place between August 31 2005 and December 31 2006. After a referendum in 2007, the charges were implemented permanently in August 2007. Prices have been phased between zero (e.g. at night) and 20SEK during peak hours. The peak charge was raised to 35SEK in 2016. The charging zone has around 300,000 inhabitants, but many people from outside this zone are affected by the charge as well. More than 210,000 people commute into Stockholm's charging zone to work and around 60,000 commute out of the zone for the same purpose [19].

Gothenburg introduced congestion charging on January 1 2013 without a trial phase and against the general public opinion [2]. Similar to Stockholm, crossing a cordon around the inner city is charged. The rates vary according to the time of the day as well and originally ranged between zero and 18SEK. In 2015 the peak charge was raised to 22SEK.

More important than technical details of the charges are the effects on travel behavior for the purpose of this study. The most important effects are summarized in the last column of table 2. During the trial phase in Stockholm, traffic volumes across the cordon decreased by 22% compared to the previous year's values according to Eliasson et al. [19], who evaluate travel time measurements and traffic flows before and during the trial. They also highlight that based on travel surveys most commuters, who changed their behavior probably switched from private cars to public transportation, resulting in an increase of 6% in public transportation trips. Börjesson et al. claim that only 1% of the commuters adapted by changing the route to work or school [6]. Discretionary trips across the cordon decreased as well, but the trend there has been more towards reduced trip frequency or alternative routes.

These trends continued in Stockholm when the charges were introduced permanently. In the first four years after the permanent introduction of the charge, traffic volumes across the cordon decreased by between 18 and 20% compared to 2005. When adjusting for other impacts, such as inflation, exemptions or fuel prices, the reductions between 2006 and 2011 were even around 27 to 31% [6]. What is interesting is that even between the trial and the final implementation of the charges in 2006/2007, the traffic volume was about 5 to 10% lower than in 2005. Börjesson et al. offer transaction costs as potential explanation. However, trying out alternative travel modes might have also convinced some people that they are better and thus represent a voluntary shift.

While the traffic reductions in Stockholm stabilized about one month after the first introduction of the charge, the adaptation process seems to have been slower in Gothenburg. Börjesson & Kristoffersson report that after eight months, the traffic reductions maintained at a magnitude of 12% across the cordon in Gothenburg [7]. The reduction in the city center that can be accounted to the congestion charge was around 3 to 5 %. Andersson & Nässén find an average reduction of 10% in traffic volume during the first year [2]. As in Stockholm, commuters in Gothenburg switched to public transport, while discretionary travellers reduced their trip frequency or switched to alternative routes [7]. These authors also claim that the increase in public transportation usage due to congestion charging can be estimated to be between 4.5 and 6.5%.

3 Methodology

3.1 Choice of Treatment and Control Group

As the behavioral shift in this study is primarily caused by a price change, not some moral conviction, it is important to have reason to assume that people give themselves moral credit for their behavior anyhow. This is generally hard to tell from the aggregated data that is used for this study. In Stockholm, however, studies have shown that the citizens felt improved air quality [19] and that the support for the charge was closely related to environmental attitudes [6]. People's attitudes towards the charging have improved tremendously and there was high compliance with and few complaints about the new regulation once it was in place. Thus it seems reasonable that people would assume their behavior adds to improved environmental goals.

The situation in Gothenburg is slightly more difficult. While the city has tried to market congestion charging as an environmental policy, the primary reason for the charge was obviously the financing of infrastructure projects. This led to low public support for the charge and most likely does not make the affected people feel like acting pro-environmentally. Andersson & Nässén for instance find that people who have not changed their travel behavior and those who have do not differ in their level of environmental concern [2]. Hence it is more likely that citizens of Gothenburg did not give themselves as much moral credit for complying with the new regulation. Gothenburg is thus a "less credible" treated entity than Stockholm and it might be wise to include an extension of the study without Gothenburg.

While there is a starting point now to assume that the treatment group experienced a change in behavior they might give themselves credit for, it is just as important to make sure the control group did not experience this. Road pricing is very common in Norway, but it is primarily used to finance infrastructure investments. Indeed the revenues from the toll ring in Oslo contributed somewhat to public transportation, but the scheme was introduced mainly to finance road enlargements and its stated purpose was to have as little traffic consequences as possible. Thus Oslo's toll ring should not be seen as an environmental measure, but rather a pure financing tool.

As mentioned above, congestion charging has also been discussed intensely in Helsinki and Copenhagen. This is really interesting, because their "almost implementation" theoretically makes them a very good control group over and above their size and spatial proximity to the treated cities. As far as I know the city of Malmö has not considered congestion charging or a similar policy thus far. It should still be included in the control group as it is the biggest Swedish city without a congestion charging scheme.

3.2 Choice of Method

This section tries to examine if there is a causal relationship between the policy introduction and short-haul flight demand. Mathematically speaking, it is the average treatment effect on the treated (ATET) which needs to be identified. The ATET, as the name suggests, is the expected effect of a treatment on some outcome variable as compared to the outcome had the unit not been treated:

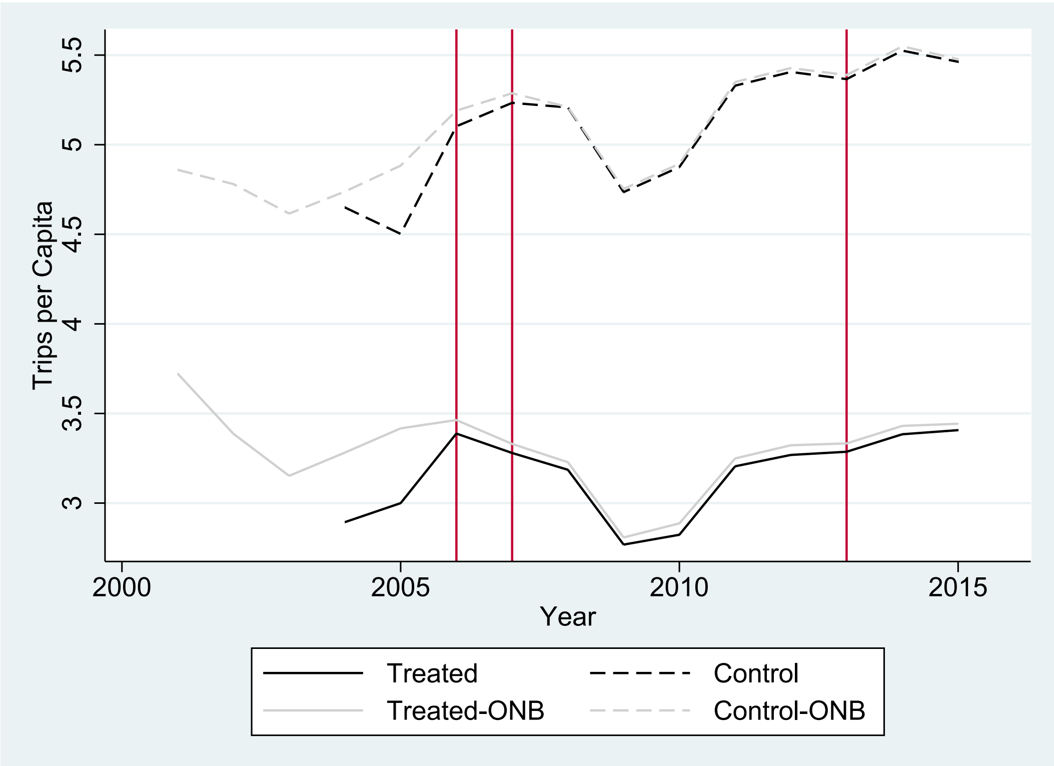
$$E[Y_{i1} - Y_{i0}|T_i = 1]$$

Y_{i1} represents the outcome variable, the number of short-haul flights per capita, in city i if this city implemented congestion charging. Y_{i0} stands for the same outcome had city i

not implemented this policy. The expected difference between those two outcomes should be conditional on that the city has actually been treated ($T_i = 1$).

It becomes obvious that the fundamental problem in this kind of estimation is that it is impossible to observe both of the two outcomes for one entity. Basically this means that before estimating the ATET, one needs to use the control group to estimate the outcome had a treated city not been treated, Y_{i0}. Therefor a standard fixed effects model, which is explained in more detail below, is used.

The simplest version of a fixed effect model is a difference-in-difference (DID) analysis. DID measures the difference in the change of the outcome variable over time for the treatment and the control group. The core of the method is to assume that the treated units would have developed in the same way as the control group, if they had not received the treatment. Essentially that requires both groups to have parallel trends in the outcome variable before the treatment. Figure 1 gives an overview of the long-term flight trends for the treatment and control group.



Red lines indicate introduction dates of congestion charge 2006 (trial) and 2007 (permanent) in Stockholm and 2013 in Gothenburg

Figure 1: Average Yearly Number of Flight Trips

The dependent variable used in this study is the number of carried passengers from or to a city's airports (black lines). As becomes clear from the graph, those values are mostly unavailable during earlier years. Helsinki and Oslo report carried passenger numbers since 1997 and 2002, respectively. For all other cities, data is only available from 2004. Unfortunately, this leaves only two pre-treatment periods for Stockholm. Therefore the trend of another variable, which moves extremely similar to the carried passengers, but is available for earlier years as well shall be considered: on board passengers (light grey lines)[3].

Apart from very small departures, the values for carried passengers move very similarly across the two groups (the black lines). The same holds for the average number of on board passengers (the grey lines)[4]. Generally the average number of flight trips is much higher in the control group than in the treatment group. That is probably largely driven by the very high number of trips per capita from and to Oslo. Although the average values are quite different in absolute terms, the groups show parallel trends throughout the observed period. The three vertical lines show the introduction of the congestion charge policy (treatment), 2006/2007 in Stockholm (first as a trial and then permanently) and 2013 in Gothenburg.

Theoretically, one would expect a disruption in the trend for the treatment group once they receive the treatment[5]. However, as congestion charging cannot be expected to have more than a marginal impact on air travel demand, especially compared to other factors, such as income for instance, it is not surprising that there are no strong blips or anything similar visible in this graph. Another problem that might weaken the graphical effect of the charge are the dissimilar introduction dates. One might be able to see a stronger effect if Stockholm and Gothenburg had introduced the charge at the same time.

Yet looking at the graph closely reveals some potential effects. There are irregular deviations from the common trend in the treated group between 2006 and 2008, right after the introduction of the Stockholm congestion charges. While the control group average continues an upwards trend, flight numbers in the treated group decline. Nevertheless, it is hard to tell if there is any effect after Gothenburg introduced its charging policy from this graph.

Observing potential effects when plotting the data of course does not allow for any causal inference. A more detailed analysis controlling for further variables is needed in order to draw any conclusions. After making sure that the parallel trend assumption holds, the analysis can be stepped up and moved on to the econometric model. The rest of section 3 explains the theoretical basis of the econometric methods, which are used to analyse the research question. It also discusses the decision making process regarding control variables, data and modeling.

3.3 The Fixed Effects Model

3.3.1 Introduction to Fixed Effects

The relationship between the implementation of a congestion charge and the number of flight-trips is estimated using a difference-in-differences (DID) approach. The "problem" with DID is that it only includes two time periods. If panel data is available for more than two periods, as is the case in this study, it is possible to extend this approach by using so called fixed effects models.

A fixed effect model is a multi-period extension of the DID model and a type of multiple regression model. It exploits panel data by controlling for variables that vary either over

[3] On board passengers are not suitable to be our dependent variable, because they include direct transit passengers. Those passengers arrive at one airport for a stop and leave it on the same plane. Hence they should not be affected by the policy at all. Still, those two variables move together closely, whenever they are both available. So for comparing long-term trends before the policy, it can be useful to consider them both. Appendix B provides a more detailed overview of these numbers for each city separately.

[4] Be aware that the two grey lines and the two black lines are supposed to be parallel.

[5] Appendix C illustrates the optimal difference-in-difference effect in theory.

time, or across entities. These variables are usually unobserved and thus hard to include in other settings. If these kind of unobserved variables exist, using a fixed effects model will reduce potential omitted variable bias. In the context of this thesis the question is whether there are city or time-specific variables, which determine the number of short-haul flights (Y). It is reasonable to assume that the cities in the sample have some very specific endemic characteristics that influence the number of flight trips per capita. These could be cultural attitudes towards flying, natural barriers like topographic conditions or sheer spatial distance to other economic centres for instance. Putting up a model with only city fixed effects could look as follows.

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 Z_i + u_{it}$$

In the study at hand, the dependent variable Y_{it} is the number of short-haul flights to and from one city per capita of the regional population. The independent variable T_{it} would be a treatment dummy equal to 1 if congestion charging is in place in city i at time t and zero otherwise. Z_i represents an unobserved city characteristic, say the topography of the region, which is inherent to that city alone and does not (quickly) change over time. Then each of the six cities will have a particular intercept $\alpha_i = \beta_0 + \beta_2 Z_i$, the city fixed effect. The slope of the regression line β_1 will be the same for all of them. Analogously, one can set up a regression model including only time fixed effects:

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_3 S_t + u_{it}$$

This will lead to year specific intersects $\lambda_t = \beta_0 + \beta_3 S_t$. S_t might include fuel price developments or an increased global integration that leads to more flight trips across all the cities. The possibility to control for fixed effects is a really nice feature. However, there might be further (un)observable variables that change both, over time and across entities. If these are determinants of Y and related to the treatment, they will lead to a biased estimator. One example could be the gross domestic product (GDP) per capita. It differs between the cities, but it also changes over time. It is very likely to determine the demand for flight trips and could potentially be related to congestion charging through car ownership or willingness to pay for environmental externalities.

The basic fixed effects model is not able to control for these kinds of confounders, so they have to be included as additional regressors. Making sure omitted variable bias is avoided whenever possible, both city and year fixed effects, as well as further confounders will be included in the analysis. This leads to the following model specification:

$$Y_{it} = \beta_1 T_{it} + \delta_j X_j + \alpha_i + \lambda_t + u_{it}$$

Where X_j stands for a number J of confounding variables, i.e. $j = 1, ..., J$. As suggested above those could be regional per capita income, but also air fares, tourism indicators or many other factors that change over time and between the different entities. The exact choice of control variables is discussed in detail and presented in tabular form in section 3.4.

3.3.2 The Model Assumptions

The fixed effects regression assumptions are basically an extension of the ordinary least squares assumptions from cross-sectional to panel data. The first assumption (A1) is that the conditional distribution of the error term u_{it}, given all confounders, has a mean of zero. Basically this means that the treatment T_{it} and the error term u_{it} may not be correlated. What is particularly important for panel data is that u_{it} cannot be correlated to either past, present or future T_{it}. In theory random treatment assignment ensures that A1 holds.

The second assumption (A2) states that all regressors must be *i.i.d.* - identically and independently distributed. If the variables are drawn from the same population, they are necessarily identically distributed. Moreover random sampling ensures independent distribution. When analysing panel data the regressors only need to be independent between the various entities. They may be - and usually are - correlated over time within one entity.

Assumption 3 (A3) demands that there are no large outliers in the data for any of the regressors. Mathematically said, they all have nonzero finite moments: $0 < Y_{it} < \infty$, $0 < T_{it} < \infty$ and $0 < X_j < \infty$. A3 is needed because OLS estimators can be responsive to these deviations outside the regular data range.

The fourth and last assumption (A4) states that there must not be multicollinearity among the regressors. In other words, none of them can be a perfect linear function of another one. While A3 and A4 have their raison d'être, A1 and A2 are usually seen as the more important assumptions. For the present study it should be kept in mind that random treatment assignment is essential for fulfilling the assumptions of the model.

3.3.3 Threats to Internal Validity

Estimates are said to be internally valid if they are consistent and unbiased. However, there are several threats to the internal validity of empirical studies. The four potentially most relevant problems to this study shall be discussed briefly: Omitted variable bias, measurement error, missing data & sample selection as well as simultaneous causality. All these cases can lead to bias through correlation between T and u and thus violate A1.

As discussed above omitted variable bias (OVB) emerges when important control variables are left out of the regression model. I.e. variables which are correlated with one or more regressors and affect the outcome of interest. If data for these variables is available, they can be included in the model as additional regressors. Yet, the choice of control variables is not always as simple as it may seem. It is usually a trade-off between bias and variance and not all the desired data might be available. To address this problem several control variables are included in this study. However, it will become clear that it is not possible to control for all the desired variables. This should be kept in mind for the interpretation of the results.

Measurement error can stem from wrong values for the dependent variable (Y), or the independent variable (*Classical Measurement Error*). Depending on the character of the measurement error, inaccuracies in the independent variable can lead to so called *errors-in-variables bias*. Errors in variables can result from wrong answers in surveys or typographical errors when administrative data is entered, for example. If there is measurement error in Y, the variance increases, but no bias is introduced. To ensure best data quality most data for this study stems from reliable official sources, such as Eurostat or national authorities.

Similarly to the measurement error, missing data only causes problems depending on why the data is missing. Stock & Watson [42] discuss three different possibilities. If data is missing independent of the values of Y, this merely reduces the sample size without introducing bias. This holds if data is missing completely at random or dependent on a regressor (which is sometimes the case in this study). However, if data is missing depending on the outcome variable, sample selection bias will be introduced. This should not be a problem in the present study, because the choice of cities was not dependent on their outcome variable (number of flights), and no city had to be discarded because of that.

Last but not least there is simultaneous causality. It basically occurs if the outcome variable can be suspected to determine the independent variable. This is an apparent problem in the present study. While the effect of a congestion charge on the number of flight trips per person is being studied, it is quite likely that more flight trips could be the cause for bad air quality and thus the introduction of a congestion charge in the first place. Unfortunately the possible solutions to this type of bias are randomized controlled experiments or instrumental variable regression, both of which are almost impossible to perform in this case. Another potential solution could be the use of lagged variables, which will be abstained from due to the limited amount of available data. Matching is used to come as close as possible to randomization, but this problem should be kept in mind when interpreting the results of the study.

Another issue for internal validity are inconsistent standard errors, as they can produce bias as well. Standard errors in the particular context of this study are thus discussed in brief in the following section.

3.3.4 Standard Errors

As mentioned earlier, most of the confounders in panel datasets are correlated over time within one entity. This arises from the fact that data is collected from the exact same entity again and again. This kind of correlation is called serial or auto-correlation. It also means that the error term u_{it} will be correlated over time. Serial correlation affects the variance of the estimator and thus the standard error. As the regular robust standard errors do not allow for serial correlation, they are not suitable for panel data analysis. They lead to downwards-biased standard errors and thus can cause "overrejection" of the null hypothesis. I.e. our estimates will be overly confident. Therefore another class of standard errors, called heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors, needs to be applied. One type of HAC standard errors that is widely used are *clustered standard errors*. Those standard errors allow for autocorrelation within clusters (here cities).

Another problem related to standard errors arises from the underlying study. Fixed effects models are often used for samples where the number of entities (n) is large. In this case, however, the sample consists of six entities at maximum. Thus, even the clustered standard errors need to be corrected theoretically. Cameron et al. suggest wild cluster bootstrapped errors as a possible solution [10]. Those are hard to combine with the weights from the matching process, however. Hence it should be kept in mind that the significance of the results might be too confident.

3.4 Control Variables

3.4.1 About Control Variables

As discussed before, there might be variables that change both across entities and over time. Those are the ones that need to be included into the model as additional regressors. However, more is not always better when it comes to control variables. Including more regressors is always a trade-off.

If they are actual determinants of Y, additional confounders decrease the bias of the main estimator, but they will also increase the variance of the other estimators. This is particularly problematic when the sample size is as small as in the underlying example. Another problem are so called "bad control variables". Angrist & Pischke [Angrist and Pischke, 2008] for example highlight that variables, which are potential outcomes of the independent variable itself should not be included as controls, because the comparison of outcome variables conditional on a bad control does not have a causal interpretation.

Stock & Watson [42] suggest 4 steps to decide on whether a control variable should be included. The first step is to identify the key coefficient of interest, that is the effect we are trying to explain (here the effect of the congestion charge). Second, the base regression and potential extensions need to be defined. This should be an a priori choice based on economic theory rather than outcome-oriented. The base specification should include few control variables, which are expected to have the most significant impact on the outcome variable. A list of more ambiguous control variables should be added for potential extensions. In the third step, the questionable control variables should be included and tested. Stock & Watson [Ibid.] recommend to keep the regressors if their estimates are statistically significant or change the main estimator (β_1) considerably. Finally all results should be summarized and made available for a transparent decision making process. The table with the main regression for all specifications can be found in the appendix C.

3.4.2 Potential Regressors

The variables which might be relevant controls in the underlying model are those that affect the number of flight trips, are correlated to the introduction of congestion charging and vary over time and across entities. Table 3 gives an overview about all variables used in the main analysis.

Table 3: Variables Used in the Fixed Effects Regression

Variable	**Description**	**Source**
Number of Flight Trips	*Main dependent variable* The number of carried passengers (domestically and on close international flights)[6] per capita of the NUTS2 region	Eurostat
Binary Policy Variable	*Main independent variable* Treatment dummy equal to one if at any time during a year congestion charging was in place	Official policy introduction
GDP	The gross domestic product per capita of the metropolitan region in purchasing power parity	Eurostat/ Statistics Norway
Air Fare Index	A country specific price index for air travel (Base year 2015)	Eurostat
Tourism	The total number of nights residents and non-residents spent at tourist accomodation establishments in the NUTS2 region per 1,000 inhabitants of the region	Eurostat
Rail Fare Index	A country specific price index for rail travel (Base year 2015)	Eurostat
Road Fare Index	A country specific price index for road travel (Base year 2015)	Eurostat

[6] Appendix D provides an overview of the close international destinations

The three most important ones - GDP, air fares and tourism - will enter the base specification. In accordance with the assumption about the importance of GDP, Brons et al. claim that there is a strong correlation between income and demand for air travel [8]. This is also in line with the environmental Kuznets curve (EKC) hypothesis, according to which higher per capita income is likely to be associated with a larger willingness to pay to prevent pollution and a more positive attitude towards pollution pricing [15].

Secondly, air fares are relevant as the demand for fight trips should depend primarily on its price, according to economic theory. Higher air fares can be expected to have a negative impact on the demand for air travel, and vice versa. Additionally, air fares might be related to congestion charging through other environmental policies for instance.

The third variable, which is the number of tourists, should naturally determine the number of flight trips in a region, too. Even though the exact interrelation is not obvious, tourism could easily be related to congestion charging. A city with many leisure tourists for example might have a special interest in keeping a recreative atmosphere - including good air quality.

There are a number of other factors which might be suspected to cause omitted variable bias, however in a more indirect or uncertain way. Two of them are road and rail fares. While the demand for a good or service is usually affected by its price, the price and availability of potential substitutes play a major role in consumption decisions, too. In the case of short-haul flight trips, the closest substitutes might be train travel, car or bus trips. Thus the price indices of road and rail travel shall be tested as potential control variables. This is far from optimal, of course, inter alia because these indices are not available per city, but on the national level. However, this is the best option available.

There are several other variables which were tested, but discarded from the analysis. Those are unemployment, political composition of local governments and air quality. Appendix E shows the base regression including the first two variables. Air quality has large potential to be a "bad control variable" and is hence included into the analysis during the another step (the matching process), instead of serving as a control variable. The functions and application of matching are explained further in the following section.

3.5 Matching

3.5.1 Functions of Matching

Some general problems naturally occur when using quasi-experiments. As mentioned above, each entity can only be observed in one state: treated or untreated. Hence one outcome needs to be estimated by comparing treatment and control group. In planned experiments, random sampling assures that both groups are the same on average. However, performing randomized experiments is expensive, time-consuming and especially within the social sciences often ethically questionable. Thus it is much more common to use observational data from quasi-experimental settings and treat them "as if" random. However, natural experiments are usually not really "as if" random. In this study, for example, the introduction of congestion charging likely depends on various observable and unobservable variables. Fortunately there are some helpful econometric techniques. This section introduces matching to bring the treatment assignment in the underlying natural experiment as close to randomization as possible.

Iacus et al. define matching as a "nonparametric method of controlling for the confounding influence of pretreatment control variables in observational data" [27]. Put differently, the idea is to pick and match those treated and control units, which are most similar before the treatment. So whereas regression models estimate the causal effect of a treatment on some outcome, matching does not produce any statistical estimators, but is merely a data pre-processing algorithm.

Controlling for every observable variable that might influence treatment reception will avoid or reduce correlation between the treatment dummy and the error term. Optimally each of these confounders would be exactly the same for a treated and a control unit (exact balance). Under this condition it would be possible to assume that treatment assignment is basically random. Matching compares those units whose empirical distributions of covariates are as similar as possible - or in practice as similar as necessary.

In general, the data can be exactly or approximately balanced. Exact balance is optimal and might be achievable, if most of the confounders are binary variables and the sample is very large. To reach exact balancing some of the cities within our sample had to be exactly equal in terms of GDP per capita, air pollution and many other potential confounders. This is highly unlikely. If the data are approximately balanced, further control for covariates is necessary after the matching in order to account for the remaining imbalance. This is done in the fixed effects model.

There are three main contributions of matching to this study. First it possesses the ability to reduce model dependence and statistical bias, even under approximate balancing [25]. Second it allows to at least partly control for variables which have to be discarded from the fixed effects model because they are subject to reverse causality or not available in panel form, for instance. Third matching allows moving closer towards “as if” random treatment assignment and thus fulfilling the assumptions of the fixed effects model.

There are various forms of matching, all of which are aiming for better balance between treatment and control group by eliminating mismatched units. One relatively new class of matching methods, “Monotonic Imbalance Bounding” (MIB) was introduced by Iacus et al. [26]. These methods actually guarantee that the imbalance will not be larger than specified ex ante. In contrast to that most common matching methods, such as propensity score or Mahalanobis matching, belong to a class of methods called equal percent bias reduction (EPBR), where some matched sample size is assured. Whereas these methods theoretically reduce imbalance on average, in practice they often increase balance only for certain covariates, while reducing it for others. As an MIB method ”Coarsened Exact Matching” (CEM) allows the researcher to specify the degree of imbalance ex ante and the different balancing choices for various variables do not affect each other, thus addressing issues of bias, rather than variance. It possesses further properties that other methods fulfill only partly, like meeting the congruence principle (data space & analysis space), demonstrably reducing model dependence, computational efficiency, etc. These advantages are the reason why CEM is the preferred method in this study.

3.5.2 Coarsened Exact Matching

As mentioned above, CEM allows the researcher to coarsen the data into reasonable strata and then match the most similar entities. Thus it is possible to determine the desired balance between the groups ex-ante. The CEM process itself consists of three main steps: (i) coarsening the variables, (ii) applying an exact matching algorithm to prune unmatched units and (iii) discarding the coarsened data. It can be performed in STATA using the *cem* command. Only one pre-treatment period is needed for the matching.

As the underlying panel data is available for more than one period before the treatment, one pre-treatment year needs to be picked. After having made this decision, the researcher needs to define the confounders she wants to use, as well as the ranges for each strata for each variable. This step involves many subjective decisions and will be discussed in more detail below. In the second step the algorithm will then perform exact matching on the coarsened data. I.e. those units that fall into the same bins, or land in the same strata for all confounders, will be matched. It is possible that more than one control unit and one treated unit fall into the same bin. Thus *cem* assigns weights to the entities:

$$w_i = \begin{cases} 1, & i \epsilon T^s \\ \frac{m_C}{m_T} \frac{m_T^s}{m_C^s}, & i \epsilon C^s \end{cases}$$

m_C and m_T stand for the overall number of matched control and treated units and m_T^s and m_C^s for the number of matched entities within strata $s \epsilon S$. Each treated unit can be matched with one or more control entities. Thus a weight of one is assigned to each treated city, while the weight for the matched control units depends on the number of matches. Those cities that cannot be matched are assigned zero weight. After the weights have been assigned, all unmatched units can be removed from the sample. The third and last step is then to drop the coarsened data and use the original (precise) numbers for further analysis.

In order to obtain causal inference further regressions are necessary. In the case of this study the same fixed effects model that is used on the full sample will be employed to the matched sample. The following section will discuss the manifold choices that must be made in order to perform the matching in this study.

3.5.3 Matching Choices

In the underlying study there are basically three main questions that need to be answered before matching can be performed: (i) Which pre-treatment year should be used? (ii) Which confounders should be included in the matching process? (iii) What range should each strata have?

The first treatment with the congestion charge started in Stockholm in 2006. Data on most of the confounders that should be included is available for at least two years before that. So rather than limiting the analysis by choosing one pre-treatment year, it is possible to match the cities separately for each pre-treatment year where the panel is more or less balanced: 2003, 2004 and 2005. This will indicate the level of robustness of the chosen strata ranges.

The second and third decision concern the choice of confounders and their strata to make sure the cities that are matched resemble each other in the most important determinants of congestion charging. Table 4 gives an overview of the five matching confounders described in this section.

It is reasonable to assume that air quality is a crucial determinant when it comes to the decision, if congestion charging should be introduced. Thus we will include long-term air pollution and short-term air pollution as confounders. The first will be measured as the average annual particulate matter 10 (PM10) concentration and the latter as the number of days per year that PM10 exceeds a threshold value of $50\mu g/m^3$. The assumption is that lower air quality will increase people's willingness to introduce pollution pricing.

Short-term air quality, measured in the number of days where the 24 hour mean exceeded $50\mu g/m^3$ will be divided into three groups: low short-term pollution (<5 days), medium short-term pollution (5 - 15 days) and high short-term pollution (>15 days). Long-term PM10 pollution will have only one break at $20\mu g/m^3$, which is the threshold value given by the World Health Organization [WHO, 2005][7].

Table 4: Matching Confounders

Variable	**Description (Range)**	**Source**
Long-term Air Quality	Average annual concentration of particulate matter of 10 microns in a diameter of air or smaller (PM10) in $\mu g/m^3$ (<5, 5-15, >15)	Eurostat
Short-term Air Quality	Number of days per year where PM10 concentration exceeds a threshold value (<$50\mu g/m^3$, >$50\mu g/m^3$)	Eurostat
Car Commuters	Share of journeys to work by car in % (<25, 25-40, >40)	Eurostat
Public Transportation Commuters	Share of journeys to work by public transport (rail, metro, bus, tram) in % (<25, 25-40, >40)	Eurostat
Commuter Share	Share of people commuting into or out of the city in % (<20, >20)	Eurostat

The extent to which people are affected by a congestion charge will always depend crucially on their travel behavior (car ownerhship, commuting across the cordon, availability of public transport, etc.). Thus some available indicators from Eurostat's urban transport indicators from before the charge introduction will be included. In particular information on the share of journeys to work by car and public transport, respectively, and the share of people commuting into or out of the city. This data is not available as a panel, but rather represents a cross section at some point in time.[8]

While it would not be possible to include this cross-sectional data in the regular fixed effects regression, using CEM allows to control for these additional variables by including them as a pre-treatment characteristic in the matching process. The share of people taking the car to work and those that take public transportation are divided into a low, medium, high scheme. The breaks are at 25% and 40%. The commuter share is only divided into two groups. That is because in general more commuters means that the policy affects more people. Here it will be assumed that the share of commuters will affect a considerable amount of people and hence public and political will when it exceeds 20%.

When matched on this basis, the pairs Stockholm and Helsinki as well as Gothenburg and Copenhagen are matched in 2005. Stockholm and Helsinki are also matched based on their 2003 and 2004 values, but no other cities. If we look at it closely, Gothenburg and Copenhagen almost match in 2003 and 2004 as well, often just being slightly left or

[7] Available from:
$http://apps.who.int/iris/bitstream/10665/69477/1/WHO_{S}DE_{P}HE_{O}EH_{0}6.02_{e}ng.pdf$ (17/05/2017)

[8] As cities do not provide yearly data on these variables, but rather occasional cross-sectional surveys, data from one city in 2003 are sometimes compared with the same variable for another city from 2004 or 2005.

right from a threshold value. So in general, based on the criteria above, these cities show a high resemblance. For that reason 2005 will be used as the pre-treatment situation for the main regressions. As there will be an extension of the study excluding Gothenburg, the regression results for that part will be similar to the results we would have gotten had the matching been based on 2003 or 2004.

4 Data

The main dependent variable is the number of short-haul flight trips per capita within one year. Short-haul flights include all domestic flights, as well as the specified international connections in appendix D. The choice of destinations was limited according to data availability. Airports located within a range of 150 km from the city centre are considered as belonging to that particular city.[9] The data for flight numbers as well as population data are taken from Eurostat. To obtain the per capita flight trips per year the sum of domestic and "close international" flights is divided by the number of inhabitants of the relevant NUTS2 region[10]. The earliest flight data is available for Helsinki from 1997. However, most cities started reporting carried passenger numbers later on, so the panel is only balanced between 2004 and 2015. Eurostat also provides monthly flight data, which is used later on as an extension of the model to the main holiday seasons.

The data for the control variables mostly stems from Eurostat as well. GDP per capita is given in purchasing power standard for the metropolitan region. Metropolitan region follows Eurostat's own definition of a city and its commuter belt[11]. The numbers for Oslo were not available at Eurostat and are thus taken from Statistics Norway. Tourism is measured as the number of overnight stays in the NUTS2 region per 1,000 inhabitants. Air, rail and road fares are country specific average annual indices from Eurostat's *Classification of individual consumption by purpose* (COICOP) dataset. It is adapted to the compilation of the harmonized consumer price index (HICP) of the European Union (EU).

The data sources for the "unused" control variables should be mentioned as well. Unemployment is measured as the total unemployment rate of over 15 year-olds for each metropolitan region. It is obtained from Eurostat and Statistics Denmark. The share of seats of the green party is given for the municipal councils of each city and is taken from the national statistical databases of Sweden, Denmark, Finland and Norway. The two different air quality measures are the annual average concentration of particulate matter of 10 microns in a diameter of air or smaller (PM10) in $\mu g/m^3$ and the number of days per year where PM10 concentration exceeds a threshold value of $50\mu g/m^3$. These numbers are taken from Eurostat as well.

5 Results

5.1 Effects of Congestion Charging

First of all the number of short-haul flights per capita is regressed on the policy treatment dummy. Table 5 presents the results of these regressions for the full sample as well as

[9] This is with the exception of Malmö and Copenhagen. For both cities only their own airports are counted.

[10] More information on Eurostat's division of NUTS regions can be found under the following link *http://ec.europa.eu/eurostat/web/nuts/overview*

[11] *http://ec.europa.eu/eurostat/web/metropolitan-regions/overview*

for the coarsened exact matched sample. Recall that this sample compares Stockholm with Helsinki and Gothenburg with Copenhagen. First of all it is apparent that none

Table 5: Main Regression Results

	Full Sample			CEM		
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.203	0.0585	0.389	0.0629	0.196	0.191
	(0.377)	(0.208)	(0.229)	(0.264)	(0.156)	(0.0955)
GDP		3.78e-05	1.01e-05		2.63e-05	2.40e-05
		(2.77e-05)	(4.95e-05)		(3.52e-05)	(3.76e-05)
Air Fare Index		0.00146	0.0153		0.000305	0.000132
		(0.00893)	(0.0105)		(0.00825)	(0.00867)
Tourism		-0.000210	-0.000457		0.000205	0.000296
		(0.000387)	(0.000384)		(0.000312)	(0.000268)
Rail Fare Index			0.0743			-0.00154
			(0.0490)			(0.0169)
Road Fare Index			-0.101			0.0233
			(0.0508)			(0.0266)
Constant	2.361***	2.905	4.085	1.513***	1.757**	-0.0899
	(0.0322)	(1.541)	(2.842)	(0.0623)	(0.537)	(1.759)
Observations	81	60	60	55	41	41
R-squared	0.532	0.460	0.621	0.853	0.750	0.753
Number of entity	6	6	6	4	4	4
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.386	0.260	0.454	0.773	0.616	0.589

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

of the results is significant on the 10% level or better. While the estimation in the first column gives a negative estimate, including more control variables increases the estimate and turns it positive. The CEM results are positive throughout and change less over the different specifications. While the estimates obtained from the full sample are actually quite high (at least the one from the extended model), the CEM estimates, which basically control for more variables, are lower. They seem to be more stable as well, having smaller standard errors than the full sample.

These first results do not indicate any causal relationship between congestion charging and the number of flight trips. Theoretically however, a positive estimate would indicate an increase in the number of flights following the implementation of a congestion charge. Neglecting other effects, this would point towards a permitting spillover effect, such as *moral licensing* for instance. In contrast to that, negative estimates would represent a consistent behavior and thus promoting spillover effects.

To test the robustness of these results and paint a somewhat clearer picture, several variations are performed as well. These are (i) using the peak charge price as an independent variable (instead of a dummy), (ii) discarding the "close international" flights and using domestic flights only, (iii) using only the holiday seasons and (iv) discarding Gothenburg.

5.2 Extensions

5.2.1 The Peak Charge

Apart from the fact whether congestion charging is in place or not, the actual amount of the charge could be crucial in determining spillover effects. This could either be due to differing psychological effects that the price might have on the consumers, or due to plain income effects. As including the peak charge as a control variable is likely to lead to multicollinearity, it is used instead as an alternative independent variable. In particular, it is the price that drivers need to pay when crossing the cordon during the most expensive hours of the day. Running the same regressions with the peak charge basically yields the same results as in the main regression, with estimates that are about the twentieth part of the original estimates. That is rather straightforward, because the peak charges are roughly around 20SEK. So while these estimates do not give any additional insights, they are still displayed in appendix F. For the following extensions only the treatment dummy is used as the independent variable.

5.2.2 Domestic Flights

The second extension concerns the dependent variable, the flight trips. While some international connections are within very short distance, one can expect that on average domestic destinations are closer. For various reasons they might also be easier to substitute with train or road travel. Some reasons could be better connections within the net of one operator, language barriers or that people are more comfortable driving in their home country than abroad. While these assumptions are not proved, they give an incentive to check whether the use of domestic flights only alters the estimates.

Table 6: Regression Results for Domestic Flights

	Full Sample			**CEM**		
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.149	0.257	0.355	0.314	0.289	0.249*
	(0.504)	(0.157)	(0.198)	(0.148)	(0.138)	(0.0962)
GDP		0.000280*	0.000272**		5.40e-05	5.12e-05
		(0.000112)	(1.00e-04)		(5.20e-05)	(5.02e-05)
Air Fare Index		-0.0293	-0.0274		-0.00333	-0.00616
		(0.0303)	(0.0307)		(0.00813)	(0.00729)
Tourism		-0.000911	-0.000981		-0.000412	-0.000335
		(0.000808)	(0.000822)		(0.000315)	(0.000291)
Rail Fare Index			0.0205			-0.0120
			(0.0337)			(0.00964)
Road Fare Index			-0.0612			0.0206
			(0.0520)			(0.0257)
Constant	2.856***	-2.339	0.454	1.899***	1.132	0.754
	(0.0239)	(2.085)	(6.155)	(0.0273)	(0.722)	(2.088)
Observations	87	67	67	57	44	44
R-squared	0.299	0.523	0.527	0.430	0.765	0.773
Number of entity	6	6	6	4	4	4
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.114	0.358	0.336	0.160	0.611	0.592

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

The results of the regression of domestic flights on the treatment dummy are depicted in table 6. The signs of the estimates are the same as in the main regression, but most of the estimates are stronger or more positive, respectively. Only the estimate from the extended model on the full sample becomes slightly weaker. Except for the specifications in column 1 and 4, the standard errors are lower than in the first results table. As above, the standard errors are lower for the CEM sample, implying more stable estimates. The estimates are not only larger in magnitude overall, but one of them also becomes slightly significant (on the 10% level). This is the extended model CEM sample estimate, which should provide the best fit. As mentioned earlier in the context of the fixed effects model, the standard errors are likely to be biased and lead to over-confident results. So while again no causal relationship is reliably supported, the larger estimates are in line with the reasoning for this extension.

5.2.3 Holiday Season

Linking back to the original idea of the study, which was that people take a decision about short-distance travel modes and afterwards they choose if and how to travel a further distance, the importance of having a choice is obvious. It is crucial that people are enabled to make a decision for the conceptual framework to hold. It is reasonable to assume that everybody is freer in making decisions about their leisure trips than about business trips. Hence this extension focusses on the main holiday seasons only, where the share of holiday trips should be comparatively high.

Table 7: Regression Results for Summer Season

	Full Sample			**CEM**		
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.0191	-0.00145	0.0201	-0.000921	0.00295	0.00205
	(0.0227)	(0.0132)	(0.0197)	(0.00861)	(0.00377)	(0.00233)
GDP		3.32e-06	1.66e-06		5.42e-07	4.16e-07
		(2.20e-06)	(2.68e-06)		(1.01e-06)	(1.02e-06)
Air Fare Index		-0.000280	0.000711*		4.33e-05	-1.43e-05
		(0.000877)	(0.000304)		(0.000230)	(0.000190)
Tourism		-4.83e-05	-6.27e-05		8.86e-06	1.33e-05
		(5.87e-05)	(5.99e-05)		(7.61e-06)	(6.68e-06)
Rail Fare Index			0.00500			-0.000269
			(0.00349)			(0.000510)
Road Fare Index			-0.00554			0.00114
			(0.00327)			(0.000702)
Constant	0.145***	0.237	0.212	0.116***	0.0482**	-0.0201
	(0.00319)	(0.241)	(0.312)	(0.00215)	(0.00830)	(0.0959)
Observations	78	60	60	52	41	41
R-squared	0.243	0.234	0.326	0.639	0.751	0.761
Number of entity	6	6	6	4	4	4
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.0446	-0.0516	0.0294	0.474	0.617	0.602

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

Tables 7 and 8 present the results for the two different main seasons: summer and winter. The dependent variable for the summer is the monthly average flight number between June and August, while the winter variable measures the average flight numbers of December and January[12]. These average values are then regressed on the congestion charge dummy again.

For both, the summer and winter season, the signs change somewhat. For the summer set, both fixed-effects-only estimates as well as the full sample base model estimate are negative. For the winter regressions even the estimate of the base specification in the CEM sample becomes negative. Overall the signs of these results are really mixed and none of the estimates is significant.

Another interesting issue is the magnitude of the effects. At first glance the effects seem much smaller than in the previous regressions. That is because the estimates have a different interpretation. While the a .01 estimate in the first regression would indicate an increase of the yearly number of flight trips per capita by .01, it would indicate the same increase, but for each month in the holiday dataset. Hence, if the effects were to be exactly the same, all holiday estimates should be one twelfth of their counter estimate in the first regressions. Comparing the estimates closely, this is roughly the case for only few of the regressions.

Table 8: Regression Results for Winter Season

		Full Sample			**CEM**	
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.0194	-0.00994	0.00446	-0.00739	-0.000784	-0.00211
	(0.0180)	(0.0146)	(0.00695)	(0.0112)	(0.00855)	(0.00850)
GDP		-1.25e-06	-2.58e-06		8.02e-07	7.19e-07
		(2.12e-06)	(4.10e-06)		(2.86e-06)	(2.81e-06)
Air Fare Index		-0.000144	0.000375		-0.000365	-0.000459
		(0.000745)	(0.000939)		(0.000696)	(0.000471)
Tourism		6.91e-06	-5.19e-06		4.51e-06	6.73e-06
		(1.06e-05)	(6.64e-06)		(1.56e-05)	(1.15e-05)
Rail Fare Index			0.00309			-0.000397
			(0.00243)			(0.00106)
Road Fare Index			-0.00526			0.000601
			(0.00291)			(0.00237)
Constant	0.323***	0.232**	0.370*	0.264***	0.251**	0.244
	(0.00267)	(0.0785)	(0.146)	(0.00285)	(0.0697)	(0.259)
Observations	77	60	60	51	41	41
R-squared	0.494	0.645	0.707	0.698	0.700	0.701
Number of entity	6	6	6	4	4	4
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.369	0.514	0.579	0.569	0.538	0.501

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

Most of the estimates for the summer season are relatively smaller than the main estimates. This would imply that the spillover effect on flights was smaller during the summer months and is contrary to the expectations formulated before. The expectation is that the share of leisure passengers is higher during summer. However, this might not necessarily hold for the short-haul flights considered in this study. The estimates for the winter season are

[12]The December value is always paired with the January value of the following year. So this is the average value for one season, rather than one year.

extremely heterogeneous and most estimates even have different signs compared to the main regression. Overall there does not seem to be a consistent deviation in the holiday estimates from the main set of regressions. What is interesting though is that there are more negative signs among the holiday estimates. In general, this would rather imply promoting spillover effects. However, none of the estimates is significant, so again it is not plausible to draw any causal inference.

5.2.4 Discarding Gothenburg

As motivated earlier, it is valid to discard Gothenburg for one further extension. Originally unloved in both cities, the attitude of the citizens of Stockholm towards congestion charging has considerably changed from the start of the trial phase and even before it was decided what was going to be done with the revenues. The people recognize the environmental impact of the charge, while the citizens of Gothenburg see the charging primarily as a financing instrument. Even though they comply as well, it is far less likely that the people of Gothenburg pat themselves on the back for pro-environmental action. Hence behavioral spillover effects should be more likely in Stockholm.

Table 9: Main Regression Results without Gothenburg

		Full Sample			**CEM**	
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.105	-0.000527	0.734*	0.656***	0.634[13]	0.549***
	(0.616)	(0.477)	(0.277)	(6.57e-08)		(5.18e-06)
GDP		3.25e-05	2.49e-05		5.85e-05	3.86e-05***
		(2.88e-05)	(4.91e-05)			(5.34e-10)
Air Fare Index		0.00520	0.0135		-0.00640	-0.00612***
		(0.0132)	(0.0120)			(2.02e-07)
Tourism		-0.000204	-0.000740		0.000339	0.000661***
		(0.000377)	(0.000447)			(3.98e-08)
Rail Fare Index			0.0836			-0.00382***
			(0.0562)			(5.42e-07)
Road Fare Index			-0.111			0.0482***
			(0.0594)			(2.90e-06)
Constant	2.948***	3.171	5.313	1.635***	1.264**	-2.349***
	(0.0519)	(1.905)	(2.971)	(1.95e-07)	(0.0377)	(0.0248)
Observations	69	50	50	31	20	20
R-squared	0.562	0.509	0.655	0.978	0.974	0.979
Number of entity	5	5	5	2	2	2
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.393	0.270	0.454	0.941	0.916	0.901

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

Table 9 displays the main regression results when Stockholm is used as the only treated unit. I.e. in the full sample regression the per capita number of flights from/to Stockholm is compared with the control group average[14] and the CEM sample only includes the direct comparison of Stockholm and Helsinki. While the first two specifications result in slightly negative, yet insignificant estimates, the extended full sample model and the three CEM sample specifications yield rather high positive estimates. An increase in flight numbers

[13]No standard errors have been generated by STATA here. This might be due to the small sample size.
[14]Gothenburg is not included in the control group, but discarded completely from the sample.

Table 10: Regression Results for Domestic Flights without Gothenburg

	Full Sample			CEM		
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.171	-0.0922	0.116	0.493***	0.465***	0.421***
	(0.553)	(0.495)	(0.296)	(2.41e-08)	(3.58e-06)	(8.85e-06)
GDP		0.000166*	0.000170*		3.22e-05***	1.78e-05***
		(7.03e-05)	(6.29e-05)		(7.62e-10)	(9.12e-10)
Air Fare Index		-0.0174	-0.0141		-0.00433***	-0.00303***
		(0.0201)	(0.0197)		(8.11e-09)	(3.44e-07)
Tourism		0.000177	6.16e-05		0.000291***	0.000573***
		(0.000499)	(0.000368)		(1.51e-08)	(6.79e-08)
Rail Fare Index			0.0276			0.00200***
			(0.0501)			(9.26e-07)
Road Fare Index			-0.0259			0.0352***
			(0.0434)			(4.96e-06)
Constant	3.248***	-5.506	-5.854	2.588***	0.739**	-2.584***
	(0.0201)	(3.589)	(6.097)	(6.96e-08)	(0.0317)	(0.0188)
Observations	72	54	54	30	20	20
R-squared	0.341	0.622	0.624	0.949	0.966	0.973
Number of entity	5	5	5	2	2	2
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.117	0.443	0.414	0.865	0.894	0.870

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

by 0.549 per year as the estimate in the last column suggests for instance, is equivalent to almost 12% of the average yearly number of flight trips[15].

The full sample estimate is significant on the 10% level and almost twice as large as the original estimate, which included Gothenburg. Two of the CEM estimates, from the model with only fixed effects and the extended model, are even significant on the 1% level. Those estimates are much larger in magnitude than the original estimates as well. Although the standard error problem that weakens the confidence of our results should still be kept in mind, this is an interesting turn.

Running the same extensions as before without Gothenburg results in further significant estimates. Tables 10 and 11 show the results for domestic flights and for the summer holiday season. The results for the winter season mostly stay insignificant, but are displayed in appendix G.

The estimates for domestic flights have the same signs as in the main regression. They are overall a bit lower, but the standard errors are similar. Again, standard errors are much smaller for the CEM sample. The estimates for domestic flights are insignificant for all specifications in the full sample, but significant on the 1% level in the CEM sample. Again, it would be wrong to jump to any conclusions, because the "true" significance level is most likely lower. The regressions for the summer season without Gothenburg have similar signs as when the full set of cities is used. The CEM estimates become significant when additional control variables are used. When the estimates are compared to the main estimates now, they all point into the same directions at least. The relative magnitudes are quite mixed, but the two significant estimates are relatively smaller than the effects in the main regression. This is in line with the summer season results for the all cities,

[15] Compared to the average of the control group and Stockholm in 2005. The exact number is 11.98%.

Table 11: Regression Results for Summer Season without Gothenburg

	Full Sample			CEM		
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.0276	-0.00751	0.0423	0.0174	0.00733***	0.00252***
	(0.0433)	(0.0264)	(0.0423)		(1.26e-07)	(1.02e-06)
GDP		3.74e-06	3.54e-06*		-3.73e-07***	-8.08e-07***
		(2.06e-06)	(1.48e-06)		(0)	(1.05e-10)
Air Fare Index		-1.16e-05	0.000695		0.000100***	-7.76e-05***
		(0.00110)	(0.000625)		(2.86e-10)	(3.95e-08)
Tourism		-4.89e-05	-8.38e-05		2.27e-05***	2.15e-05***
		(5.76e-05)	(7.68e-05)		(5.34e-10)	(7.79e-09)
Rail Fare Index			0.00597			-0.000875***
			(0.00450)			(1.06e-07)
Road Fare Index			-0.00611			0.000989***
			(0.00390)			(5.69e-07)
Constant	0.165***	0.201	0.201	0.117	0.0412***	0.0808***
	(0.00600)	(0.213)	(0.304)		(0.000194)	(0.000290)
Observations	66	50	50	28	20	20
R-squared	0.282	0.272	0.365	0.852	0.953	0.979
Number of entity	5	5	5	2	2	2
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.0480	-0.0804	-0.00417	0.637	0.851	0.899

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

which suggested a lower effect during the summer season.

To look closer at the potential effects of the policy in Gothenburg, the same analysis has been performed discarding Stockholm. The results from the main regression are presented in appendix H. While these estimates mainly point towards a negative relationship and thus explain the weaker results when both, Stockholm and Gothenburg are included, the estimates from the crucial specifications are insignificant.

5.3 Interpretation

5.3.1 Insights Gained From the Results

Although some of the extensions yield significant results, no certain conclusions can be drawn with respect to a causal effect on the outcome variable and certainly not with respect to the research question. However, there are some interesting tendencies within the various estimates which can be seen as eventual points of departure for further analysis.

Overall most of the estimates point towards a positive relationship between congestion charging and the number of flight trips. Most of the tested extensions (peak charge, domestic fights and using Stockholm only) support this finding. The domestic flights seem to be more responsive to the charge introduction, which is in line with the theory. There are most likely more substitutes available when the travel distance is shorter. The results that can be seen when Gothenburg is excluded from the sample also align with earlier expectations. The estimates for Stockholm alone are stronger in magnitude and show a higher statistical significance. This might speak in favor of behavioral shifts, because the income effects are probably similar for both cities.

Every significant estimate in the regressions has a positive sign. The significance and magnitude of these estimates is increasing for some cases where one would expect stronger spillover effects: for shorter distances and when excluding Stockholm. Against the expectations this effect is not clearly visible during holiday seasons. If one wanted to trust the estimates, they would suggest that congestion charging lead to gross substitution between car use and air travel, at least in Stockholm. This indicates potential for permitting spillover effects, such as *moral licensing.*

The results obtained here are in line with findings by Gneezy et al., Pittman and Zanna & Copper mentioned in the literature reviwe, who find that permitting effects are more likely if the perceived cost of behavior 1 is low and that promoting spillovers are less likely, if behavioral changes are related to an external cause.

5.3.2 Limitations to the Results

Various problems have been encountered throughout this study that need to make us careful about the interpretation of the results. First of all the significance levels displayed in the regression outputs are most likely overly confident, due to the small sample size. Thus even the significant results need to be handled with care. The estimates with very high significance levels might still be seen as revealing a trend. For future studies with small sample size it would be desirable to test another type of standard errors, like wild cluster bootstrapped errors for example.

Another problem mentioned earlier is potential simultaneity bias. It arises because the number of flight trips is very likely not only an outcome of congestion charging, but at the same time an underlying driver for the policy. Having more data, in particular a panel with more time periods can contribute to a solution for this problem, as lagged variables could be used. Another option would be the use of an instrumental variable or a randomized controlled experiment. Here matching was used to achieve a situation as close to randomization as possible.

Furthermore, bias due to omitted variables could be an issue. While the matching allows to include more control variables than would have been possible by only using the fixed effects model, it was not possible to control for political constellations and infrastructure investments for example. Future studies, which might compare cities on a national level might be able to include at least a political control variable to tackle potential OVB. This problem is likely to leave us with some endogeneity in the introduction of the policy. Fortunately endogeneity in the case of the present study would rather explain promoting than permitting spillover effects, because the implementation of congestion charging should be more consistent with a reduction in flight trips within one city. As the obtained results suggest permitting spillover effects, the true estimates might be even larger, if endogeneity is eliminated.

There is one more major point of criticism that might be brought forward with regard to this study. That is the fact that only a fraction of the observed passengers are actually affected by the congestion charge policy. That said, the people who are most affected by the congestion charges can on average be expected to own a car and to have an employment. Two characteristics which are likely to be correlated with demand for more (frequent) air travel. Thus their share among the passengers might actually be relatively high, though this cannot be proved here. Second, as mentioned earlier, even those people who do not adapt their behavior receive a signal about their commuting behavior, which might in turn lead to spillover effects.

6 Conclusions

This study has made a first approach to look for evidence for behavioral spillover effects related to travel mode choices by investigating the relationship between congestion charging and the demand for short-haul flight demand.

To answer the research question a literature review about the effects of congestion charging on local travel behavior in Stockholm and Gothenburg is conducted, leading to fulfillment of sub goal (a) of this study. Namely that the introduction of the policy has led to sudden shifts towards more environmentally friendly travel modes; mainly public transportation.

In search for causal inference (sub goal (b)), the yearly number of per capita flight trips is regressed on a treatment dummy using a difference-in-differences setting. The use of matching techniques, in particular coarsened exact matching, allows to control for further variables, which were impossible to control for in the fixed effects model. Matching thus adds to the study by reaching a higher balance between the treatment and the control group.

While the results do not enable to draw any certain conclusions about a causal relationship between congestion charging and the demand for short-haul air travel, the estimates mainly point towards a gross substitution between car usage and air travel. These findings are supported by most of the model extensions (Sub goal (c)). Particularly by using the peak charge as an alternative independent variable, using only domestic flights or Stockholm as the only treated unit. Extensions focusing on the holiday season lead to rather mixed and mainly insignificant results.

These findings make promoting spillover effects, which would yield decreases in air travel following reductions in car travel, seem highly unlikely. If there was to be any behavioral spillover effect, the estimates rather point towards a permitting effect. The most reliable estimate, obtained when excluding Gothenburg, suggests an increase of yearly flight trips by almost 12% following the implementation of a congestion charge in Stockholm.

A huge black hole is gaping in the literature on behavioral spillover effects in travel mode decisions and hence there is rich potential for future research. Spillover effects should be studied in different contexts, e.g. in other countries, under alternative policy scenarios or including more cities. With respect to the present study it would be desirable to exploit corresponding micro data with similar methods. While recent studies have shown that travel mode choices are often contradictory (in particular air travel), survey studies need to be complemented with empirical evidence. Furthermore, it would be interesting to disentangle the income and substitution effects triggered by the congestion charge in order to separate behavioral effects more clearly.

If in fact there are unrecognized behavioral forces at work, this would have major implications for the net changes in emissions caused by congestion charging and thus for policy making. Due to the increasing relative contribution of the travel sector to global GHG emissions, further studies should be carried out in order to obtain clarity about all the effects that pollution reduction policies trigger. Otherwise most of these policies could wind up being at most a drop in the ocean.

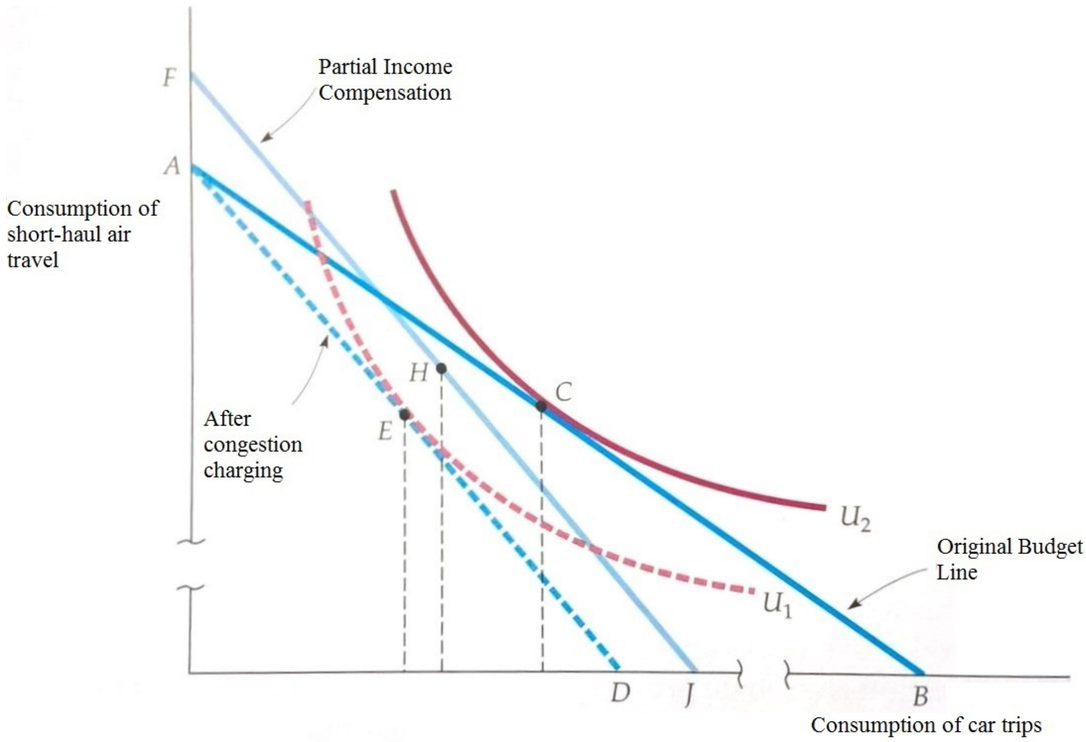
7 Acknowledgements

I would like to express my gratitude to my supervisor Claes for his patience and his sometimes puzzling, yet helpful comments. Moreover I would like to thank my family and friends for their aid and moral support. Particular credit goes to my good friend and opponent Sophie, whose valuable feedback helped me gain momentum for the final touch of this thesis and to Werner, without whom I would most certainly not be where I am today.

Appendices

A Income and Substitution Effects

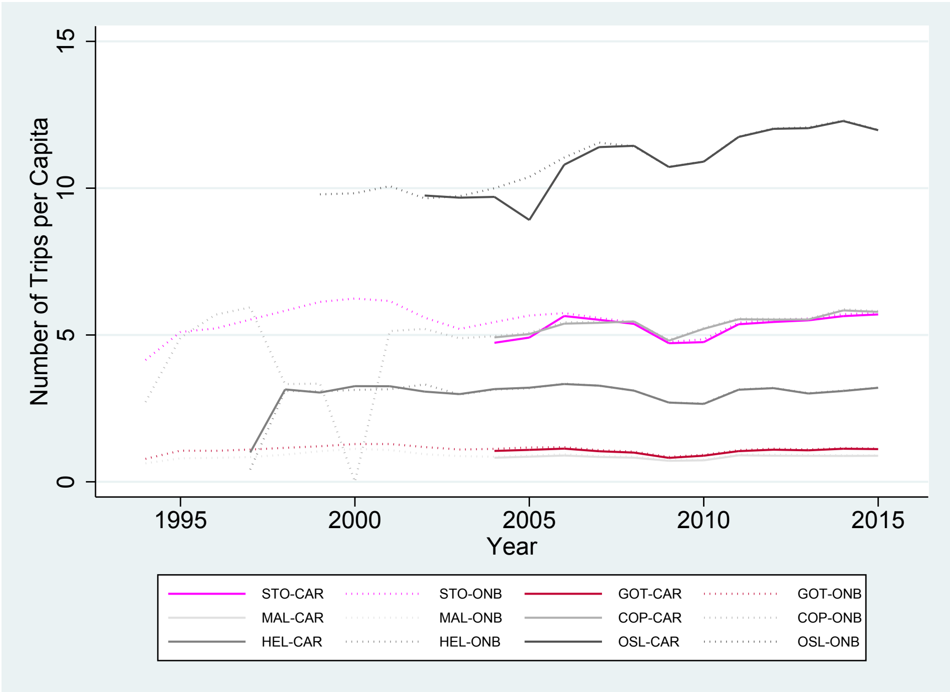
The total changes in demand depend on the magnitude of income and substitution effects. The higher price for car travel will lower people's purchasing power and should thus decrease the demand for both, car and air travel (income effect). It rotates the budget line from AB to AD. On the other hand flying will become cheaper relative to car usage (represented by the increased slope of the budget line), which would argue for increases in the demand for air travel (substitution effect). In this exemplary graph, the demnad for both goods would decrease from C to E. The budget line FJ represents a partial income compensation, which does not explicitly happen in the case of this study. However, one might argue that the improvements in public transportation might be such a partial compensation for example. From that point of view, the new consumption point H would represent an increase in demand for air travel. A change in preferences (represented by the pink indifference curves) could easily lead to an increase in air travel as compared to C as well.



Source: Pindyck & Rubinfeld (2009). This is a variation of their graph on page 124.

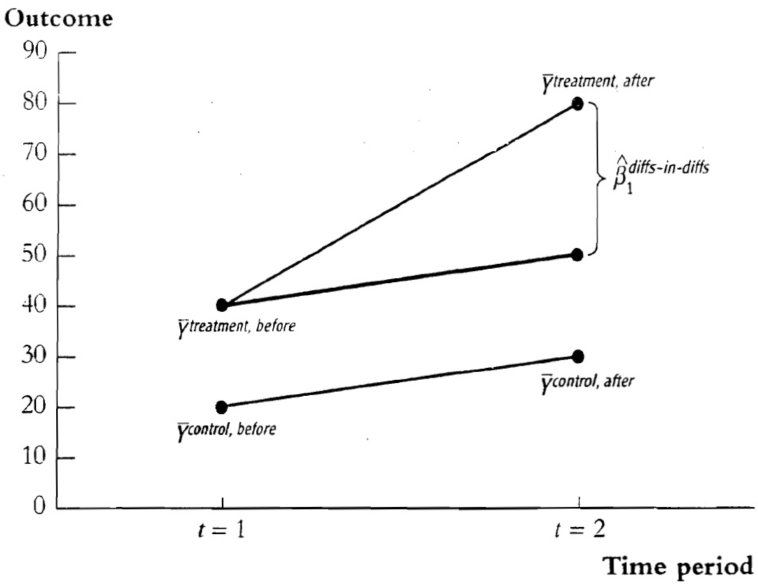
B Carried and On Board Passenger Numbers

This graph shows the number of yearly flight trips per capita for each city separately. The solid lines represent carried passenger flights, the dashed lines on board passenger numbers. While the lines for each city mostly move smoothly together, there are some exceptions. Copenhagen for instance has not reported any close international flights in 2000 and no national flights in 1998, 1999 and 2000. "Reliable" on board numbers are available from around 2001, which adds three pre-treatment years for Stockholm to the trend graph.



C Difference-In-Differences Design

As the name suggests, DID includes two types of differences. First the differences in the outcome variable of interest before and after the treatment for the treated and the control group is calculated. Secondly those two differences are subtracted from each other.



Source: Stock & Watson [42], page 492.

D Connections for Close International Flights

City, Airport	Destination
Stockholm, Arlanda	Berlin, Hamburg, Copenhagen, Tallinn, Helsinki, Tampere, Turku, Vaasa, Riga, Bergen, Statfjord, Oslo
Stockholm, Skavsta	Berlin, Bremen, Lübeck, Copenhagen, Tallinn, Riga
Stockholm, Bromma	Copenhagen, Tallinn, Helsinki
Gothenburg, Landvetter	Hamburg, Berlin, Copenhagen, Helsinki, Statfjord, Oslo
Gothenburg, Säve	Berlin
Gothenburg, Jonköping	Copenhagen
Malmö	Amsterdam, Gdansk
Copenhagen, Kastrup	Brussels, Berlin, Frankfurt, Hamburg, Köln, Düsseldorf, Hannover, Helsinki, Amsterdam, Bergen, Kristiansand, Statfjord, Oslo, Sandefjord, Trondheim, Stavanger, Göteborg, Jonköping, Stockholm, Norrköping, Karlstad
Helsinki, Vantaa	Copenhagen, Statfjord, Oslo, Göteborg, Stockholm
Oslo, Gardermoen	Billund, Copenhagen, Göteborg, Stockholm
Oslo, Sandefjord	Copenhagen, Stockholm
Oslo, Moss	Aarhus

E All Potential Control Variables

VARIABLES	(1) Carried Passengers	(2) Carried Passengers	(3) Carried Passengers	(4) Carried Passengers	(5) Carried Passengers
Congestion charge	-0.203	0.0585	0.0585	0.389	0.345
	(0.377)	(0.208)	(0.208)	(0.229)	(0.220)
GDP		3.78e-05	3.78e-05	1.01e-05	0.000107*
		(2.77e-05)	(2.77e-05)	(4.95e-05)	(4.75e-05)
Air Fare Index		0.00146	0.00146	0.0153	0.0122*
		(0.00893)	(0.00893)	(0.0105)	(0.00518)
Tourism		-0.000210	-0.000210	-0.000457	-0.000664
		(0.000387)	(0.000387)	(0.000384)	(0.000577)
Road Fare Index				-0.101	-0.133
				(0.0508)	(0.0672)
Rail Fare Index				0.0743	0.0618*
				(0.0490)	(0.0306)
Unemployment Rate					0.00122
					(0.0552)
Green Party Share					-0.00558
					(0.0247)
Constant	2.361***	2.905	2.905	4.085	5.604
	(0.0322)	(1.541)	(1.541)	(2.842)	(4.315)
Observations	81	60	60	60	57
R-squared	0.532	0.460	0.460	0.621	0.719
Number of entity	6	6	6	6	6
Entity FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Adj. R-squared	0.386	0.260	0.260	0.454	0.562

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

F Regression Results with Peak Charge

		Full Sample			**CEM**	
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Peak Charge	-0.0102	0.00339	0.0212	0.00327	0.0108	0.0106
	(0.0191)	(0.0112)	(0.0117)	(0.0134)	(0.00800)	(0.00499)
GDP		3.78e-05	9.69e-06		2.59e-05	2.37e-05
		(2.77e-05)	(4.96e-05)		(3.56e-05)	(3.81e-05)
Air Fare Index		0.00137	0.0150		9.06e-05	6.77e-05
		(0.00895)	(0.0104)		(0.00828)	(0.00868)
Tourism		-0.000214	-0.000473		0.000196	0.000282
		(0.000385)	(0.000388)		(0.000312)	(0.000265)
Rail Fare Index			0.0750			-0.000960
			(0.0489)			(0.0166)
Road Fare Index			-0.102*			0.0218
			(0.0505)			(0.0259)
Constant	2.360***	2.927	4.227	1.512***	1.830*	0.0360
	(0.0326)	(1.551)	(2.801)	(0.0630)	(0.585)	(1.768)
Observations	81	60	60	55	41	41
R-squared	0.532	0.460	0.623	0.853	0.753	0.756
Number of entity	6	6	6	4	4	4
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.386	0.260	0.457	0.773	0.620	0.593

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

G Regression Results for Winter Season without Gothenburg

	Full Sample			CEM		
VARIABLES	Only FE	Base Model	Extended Model	Only FE	Base Model	Extended Model
Congestion charge	-0.0248	-0.0238	0.00742	0.0186	0.00891	0.0185
	(0.0292)	(0.0323)	(0.0104)			
GDP		-2.13e-06	-2.60e-06		1.65e-06	1.41e-06
		(3.50e-06)	(4.93e-06)			
Air Fare Index		0.000173	0.000461		-0.000673	-1.27e-05
		(0.00105)	(0.00105)			
Tourism		1.48e-05	-8.61e-06		9.82e-05	0.000131
		(2.67e-05)	(1.69e-05)			
Rail Fare Index			0.00341			0.00278
			(0.00247)			
Road Fare Index			-0.00536			0.000790
			(0.00327)			
Constant	0.372***	0.247	0.406*	0.291	-0.0251*	-0.490***
	(0.00197)	(0.131)	(0.180)		(0.00288)	(0.000996)
Observations	65	50	50	27	20	20
R-squared	0.521	0.687	0.740	0.906	0.950	0.972
Number of entity	5	5	5	2	2	2
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.374	0.535	0.589	0.779	0.841	0.867

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

H Main Regression Results without Stockholm

VARIABLES	(1) Only FE	(2) Base Model	(3) Extended Model	(4) Only FE	(5) Base Model	(6) Extended Model
Congestion charge	-0.359	-0.0291	0.135	-0.375***	-0.213	-0.0478
	(0.327)	(0.0868)	(0.219)	(1.76e-09)		
GDP		4.27e-05	-9.26e-06		-0.000122	-9.89e-05
		(3.50e-05)	(6.18e-05)			
Air Fare Index		0.00510	0.0153		-0.0241	-0.0167
		(0.0107)	(0.00817)			
Tourism		-0.000150	-0.000651		0.000601	0.000692
		(0.000441)	(0.000534)			
Rail Fare Index			0.0873			0.0197
			(0.0509)			
Road Fare Index			-0.154*			0.0142
			(0.0582)			
Constant	2.211***	1.993	7.800	2.986*	6.395**	1.833*
	(0.0364)	(2.055)	(4.044)	(0.269)	(0.233)	(0.198)
Observations	69	50	50	24	21	21
R-squared	0.515	0.423	0.666	0.797	0.937	0.967
Number of entity	5	5	5	2	2	2
Entity FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.327	0.143	0.472	0.577	0.789	0.833

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

References

[1] Dolores Albarracin and Robert S Wyer Jr. The cognitive impact of past behavior: influences on beliefs, attitudes, and future behavioral decisions. *Journal of personality and social psychology*, 79(1):5, 2000.

[2] David Andersson and Jonas Nässén. The gothenburg congestion charge scheme: A pre–post analysis of commuting behavior and travel satisfaction. *Journal of Transport Geography*, 52:82–89, 2016.

[3] James Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477, 1990.

[4] Daryl J Bem. Self-perception theory. *Advances in experimental social psychology*, 6:1–62, 1972.

[5] Irene Blanken, Niels van de Ven, and Marcel Zeelenberg. A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, 41(4):540–558, 2015.

[6] Maria Börjesson, Jonas Eliasson, Muriel B Hugosson, and Karin Brundell-Freij. The stockholm congestion charges—5 years on. effects, acceptability and lessons learnt. *Transport Policy*, 20:1–12, 2012.

[7] Maria Börjesson and Ida Kristoffersson. The gothenburg congestion charge. effects, design and politics. *Transportation Research Part A: Policy and Practice*, 75:134–146, 2015.

[8] Martijn Brons, Eric Pels, Peter Nijkamp, and Piet Rietveld. Price elasticities of demand for passenger air travel: a meta-analysis. *Journal of Air Transport Management*, 8(3):165–175, 2002.

[9] Colin Camerer. Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences*, 96(19):10575–10577, 1999.

[10] A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

[11] Jessica Cascio and E Ashby Plant. Prospective moral licensing: Does anticipating doing good later allow you to be bad now? *Journal of Experimental Social Psychology*, 56:110–116, 2015.

[12] Scott A Cohen, James ES Higham, and Christina T Cavaliere. Binge flying: Behavioural addiction and climate change. *Annals of Tourism Research*, 38(3):1070–1089, 2011.

[13] Ravi Dhar and Itamar Simonson. Making complementary choices in consumption episodes: Highlighting versus balancing. *Journal of Marketing Research*, pages 29–44, 1999.

[14] Andreas Diekmann and Peter Preisendörfer. Green and greenback the behavioral effects of environmental attitudes in low-cost and high-cost situations. *Rationality and Society*, 15(4):441–472, 2003.

[15] Soumyananda Dinda. Environmental kuznets curve hypothesis: a survey. *Ecological economics*, 49(4):431–455, 2004.

[16] Paul Dolan and Matteo M Galizzi. Like ripples on a pond: behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47:1–16, 2015.

[17] Paul Dolan, Michael Hallsworth, David Halpern, Dominic King, Robert Metcalfe, and Ivo Vlaev. Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, 33(1):264–277, 2012.

[18] Claes Ek. Behavioral spillovers across prosocial alternatives. 2016.

[19] Jonas Eliasson, Lars Hultkrantz, Lena Nerhagen, and Lena Smidfelt Rosqvist. The stockholm congestion–charging trial 2006: Overview of effects. *Transportation Research Part A: Policy and Practice*, 43(3):240–250, 2009.

[20] Jonathan L Freedman and Scott C Fraser. Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, 4(2):195, 1966.

[21] Ayelet Gneezy, Uri Gneezy, Gerhard Riener, and Leif D Nelson. Pay-what-you-want, identity, and self-signaling in markets. *Proceedings of the National Academy of Sciences*, 109(19):7236–7240, 2012.

[22] Mark Goh. Congestion management and electronic road pricing in singapore. *Journal of Transport Geography*, 10(1):29–38, 2002.

[23] Hugh Gravelle and Ray Rees. Microeconomics. 3: e uppl. *Harlow: Prentice Hall*, 2004.

[24] Julia F Hibbert, Janet E Dickinson, Stefan Gössling, and Susanna Curtin. Identity and tourism mobility: an exploration of the attitude–behaviour gap. *Journal of Sustainable Tourism*, 21(7):999–1016, 2013.

[25] Daniel E Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236, 2007.

[26] Stefano M Iacus, Gary King, and Giuseppe Porro. Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493):345–361, 2011.

[27] Stefano M Iacus, Gary King, Giuseppe Porro, and Jonathan N Katz. Causal inference without balance checking: Coarsened exact matching. *Political analysis*, pages 1–24, 2012.

[28] Petros Ieromonachou, Stephen Potter, and James P Warren. Norway's urban toll rings: Evolving towards congestion charging? *Transport Policy*, 13(5):367–378, 2006.

[29] Alex Imas. Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics*, 114:14–18, 2014.

[30] Grant D Jacobsen, Matthew J Kotchen, and Michael P Vandenbergh. The behavioral response to voluntary provision of an environmental public good: Evidence from residential electricity demand. *European Economic Review*, 56(5):946–960, 2012.

[31] Veronika Job, Carol S Dweck, and Gregory M Walton. Ego depletion—is it all in your head? implicit theories about willpower affect self-regulation. *Psychological science*, 2010.

[32] Jonathan Leape. The london congestion charge. *The Journal of Economic Perspectives*, 20(4):157–176, 2006.

[33] Bob McKercher, Bruce Prideaux, Catherine Cheung, and Rob Law. Achieving voluntary reductions in the carbon footprint of tourism and climate change. *Journal of sustainable tourism*, 18(3):297–317, 2010.

[34] James J Murphy, P Geoffrey Allen, Thomas H Stevens, and Darryl Weatherhead. A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30(3):313–325, 2005.

[35] Andreas Nilsson, Magnus Bergquist, and Wesley P Schultz. Spillover effects in environmental behaviors, across time and context: a review and research agenda. *Environmental Education Research*, 23(4):573–589, 2017.

[36] RS Pindyck, DL Rubinfeld, Paul G Farnham, David Miles, Andrew Scott, and Francis Breedon. Microeconomics, 7th. edi—tion, 2009.

[37] Thane S Pittman. Attribution of arousal as a mediator in dissonance reduction. *Journal of Experimental Social Psychology*, 11(1):53–63, 1975.

[38] Lucia Rotaris, Romeo Danielis, Edoardo Marcucci, and Jérôme Massiani. The urban road pricing scheme to curb pollution in milan, italy: Description, impacts and preliminary cost–benefit analysis assessment. *Transportation Research Part A: Policy and Practice*, 44(5):359–375, 2010.

[39] Sgouris Sgouridis, Philippe A Bonnefoy, and R John Hansman. Air transportation in a carbon constrained world: Long-term dynamics of policies and strategies for mitigating the carbon footprint of commercial aviation. *Transportation Research Part A: Policy and Practice*, 45(10):1077–1091, 2011.

[40] Jason F Shogren and Laura O Taylor. On behavioral-environmental economics. *Review of Environmental Economics and Policy*, 2(1):26–44, 2008.

[41] Steve Sorrell. Jevons' paradox revisited: The evidence for backfire from improved energy efficiency. *Energy policy*, 37(4):1456–1469, 2009.

[42] James H Stock and Mark W Watson. *Introduction to econometrics*, volume 104. Addison Wesley Boston, 2003.

[43] Cass R Sunstein. Empirically informed regulation. *The University of Chicago Law Review*, 78(4):1349–1429, 2011.

[44] Verena Tiefenbeck, Thorsten Staake, Kurt Roth, and Olga Sachs. For better or for worse? empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57:160–171, 2013.

[45] Heather Barnes Truelove, Amanda R Carrico, Elke U Weber, Kaitlin Toner Raimi, and Michael P Vandenbergh. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, 29:127–138, 2014.

[46] Lorraine Whitmarsh and Saffron O'Neill. Green identity, green living? the role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology*, 30(3):305–314, 2010.

[47] Mark P Zanna and Joel Cooper. Dissonance and the pill: an attribution approach to studying the arousal properties of dissonance. *Journal of personality and social psychology*, 29(5):703, 1974.

[48] Chen-Bo Zhong and Katie Liljenquist. Washing away your sins: Threatened morality and physical cleansing. *Science*, 313(5792):1451–1452, 2006.