Commuting in Europe:

a dynamic macro analysis of commuting behaviour in the European Union

Name: Pieter Jonathan de Waard
Student number: 290409
Supervisor: Dr. P.A. van Reeven LLM

ABSTRACT: This thesis documents a search for influential determinants that relate to commuting streams and which might help to predict future commuting streams. It tries to add to the existing empirical work on commuting in a special way by estimating a dynamic Bårdsen error correction model using regional panel data (1999 – 2012) on the entire EU. Mainly education shows to be consistent estimator in this macro setting and reveals a significant and positive relationship with commuting outflows in both the short and long term. To a lesser extend motorisation rates possess this same consistency, but mainly on the short term. Income was expected to reveal a positive relationship on commuting, but failed to do so and did not provide significant estimates. Despite high expectations from a land use diversity index, it is not being able to explain intra-regional differences. After controlling for the choice of correct variables via a micro analysis of the same regions, the usefulness of using macro data seems to be lower compared to lower scale, and especially micro-, data.

Rotterdam, December 19th, 2013

Master's Thesis

Specialization: Urban, Port and Transport Economics
Master: Economics and Business
Erasmus School of Economics

ERASMUS UNIVERSITEIT ROTTERDAM

Contents

1 Introduction

Commuting is the main cause of particular peak hour traffic congestion and the close relation between congestion and commuting is confirmed by several studies (De Borger, 2009; Portoghese et al., 2011; Wardman and Ibáñez, 2012). Congestion is considered a serious problem in many urbanized areas throughout Europe and research in Europe shows that the average travel speed as declined dramatically over the past decades (De Borger and Proost, 2001). Among the problems caused by congestion is of course the monetary cost component such as fuel and time wasted, that is connected to these traffic delays and are estimated to be as high as hundreds of dollars per capita per year in the US (Schrank and Lomax, 1999) and comparable amounts are reported for Europe (Lindsey and Verhoef, 2000). Besides the monetary issues, evidence exists that the emissions caused by vehicle engines lead to higher premature mortality rates (Levy et al. 2010). To tackle the problem of traffic congestion and especially peak hour congestion commuting thus is an important behavioral 'habit' to investigate. Three main strategies to cope with congestion exist: First, the option of extending capacity by building new infrastructure is meant to increase supply which should relax the stress on the infrastructure system during these peak hours. Unfortunately this option sometimes even has adverse effects (Akamatsu and Heydecker, 2003). A second approach is trying to improve the current infrastructure by retiming traffic lights, adjusting the maximum speed, creating specially designated carpool lanes and so on to improve its effectiveness. A third and according to most economists the most effective approach is congestion pricing in the form of permits for a certain area, pay per mile schemes or parking restrictions and prices or any other measure that aims at increasing the price of travel in general and commuting in particular. According to the economists well-known price demand relationship (if price goes up, the demand for a good will usually goes down) demand for infrastructure use and as a result congestion will go down and so this third option is very well defendable from an economic point of view.

This research is about this third possible solution to tackle the excess demand on infrastructure in the peak hours, mainly caused by commuters. It looks into the relationship between income and commuting because pricing commuting will increase the proportion of income which is spent on commuting and as such leads to a lower disposable income. According to the standard price demand relationship mentioned above, pricing will lower commuting and peak hour traffic congestion. A subsidy for commuting of course has the opposite effect.

The main research question intends to clarify the relation and magnitude of the relationship between income and commuting on a European scale. It does not intend to propose as solution for congestion in urbanized areas but just to contribute to the existing knowledge about the influence of income or any other important determinants on commuting behavior.

First, we will look into the existing literature about commuting and review mainly the empirical work on the topic to determine what is written and concluded about the aforementioned relationship and to see what other variables are important in shaping the behavior of commuters

Next, we will check the direction and magnitude of influence of the key variables that are found in the literature review exists on a European scale. Elasticities will be calculated by developing a dynamic regression model using macro level panel data from the Eurostat database.

Finally, a chapter is dedicated to three sidesteps. First we will check if a self-constructed land use diversity index is related to commuting. Secondly, after estimating the relationships on a macro scale, we use a micro dataset that became available during this study to assess the macro analysis on a micro level scale. Final in this chapter, the findings from literature, macro and micro analysis will be compared.

The use of macro variables in commuting research is rare in itself. Using this macro data on a European scale is something not observed in a review of empirical literature on commuting. It is interesting to see whether any solid predictors for commuting behavior exist in such a large geographical scale. If so, policy makers in countries or regions for which the data on a micro scale is not available can possibly use predictions to resolve any future congestion problems in advance

Chapter two will take care of the existing body of (empirical) literature, chapter three discusses the methods and data for the main analysis and chapter four displays the results. In chapter five a land use diversity index is created and tested and a micro level analysis tests the variables used in the macro analysis and these results are compared. Chapter six concludes.

2 Literature

In the empirical literature reviewed for this paper many different variables explain commuting distance/time or frequency are used in estimation models. Ranging from the intuitive more logical socio-economic characteristics such as income or education towards the somewhat controversial variables such as the country of origin or ethnical background of the commuter in question are some examples. This chapter tries to unravel this ‘forest’ of explaining variables that have been examined in modern empirical literature concerning their influence on the practice of commuting.

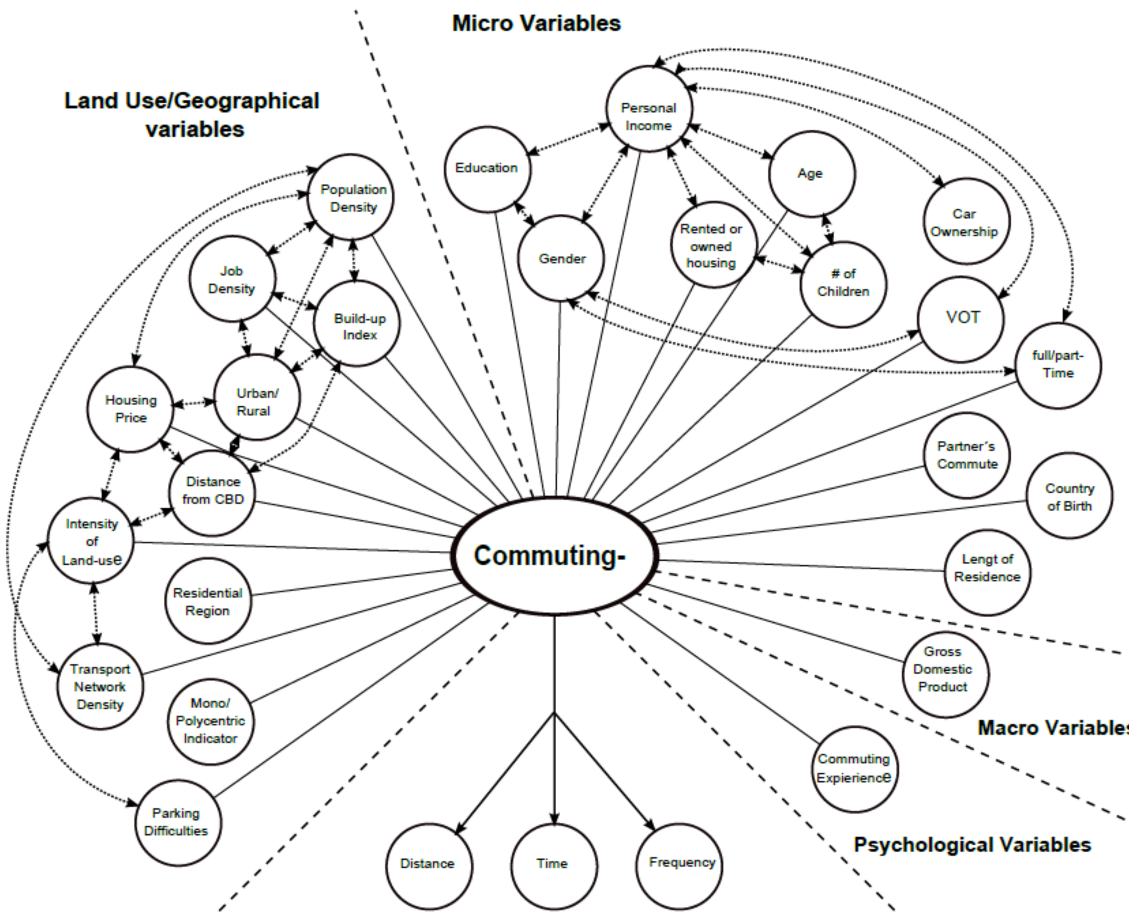
This chapter consist of three parts. First, an introduction to commuting and its possible influential factors is given. Second, the scope of existing empirical work is discussed. Sections 2.2 to 2.5 discuss the geographical coverage, the time period on which empirical work is conducted, the data sources and the level of analysis in that order. Third, in sections 2.6 to 2.9 the influencing factors on commuting behaviour are discussed. Finally a synopsis of literature findings is given in section 2.10.

2.1 Introduction to influences on commuting behaviour

As starting point the (possible) influencing components that determine the commuting behaviour are displayed in Figure 1 below.

As shown in Figure 1 the potential influences on commuting habits are numerous. All displayed characteristics can be divided into four main categories of which the micro level and land-use/geographical variables are used most often in the reviewed economical literature. Our main attention is also on these two main categories. Much more psychological variables are used in psychological literature but our focus is still mostly on (socio) economic variables and not on individual-specific psychological factors.

Figure 1: Influences on commuting behaviour



The lack of macro variables used to predict or as a control variable is limited in the literature and we came across only a few research papers (e.g. Johansson et al., 2002; Östh and Lindgren, 2012) that looked at for example (lagged) changes in GDP to predict commuting behaviour (Östh and Lindgren, 2012). The reason for the limited use of macro data in analysis is not entirely clear but a common thought among economists is that micro data provides more insights on the lowest level of decision making whereas macro data is often aggregated data per municipality, region or even country. Some authors argue that not micro data but meso- or even macro data is key to find the key determinants of commuting behaviour (Susilo and Maat, 2007; Van der Laan, 1998). An advantage that macro data entails might be that comparison among countries becomes easier as macro data is more widely available and often better comparable among different countries with different statistical definitions or methods. This issue is discussed more in-depth in the remainder of this chapter.

In Figure 1 the solid lines represent a defining variable for commuting distance, time or frequency in at least one of the studies reviewed. For a complete overview of empirical research articles reviewed for this study: See Appendix A. Important to notice is that several of these variables are interconnected in many ways. An example from micro data is income which is often related to age; the older one becomes the higher the personal income. This relationship is well known and proven to be true by several studies (e.g. Stolzenberg, 1975) and is indicated by a dashed line in the figure. Another relationship between the displayed explaining variables is the Value of Time (VOT) that is connected with both gender and personal income. VOT is said to determine commuting behaviour and is also influenced by gender as well because men seem to value time more than women (Rouwendal and Nijkamp, 2004). At the same time the VOT will be higher when personal income is higher which, in turn, also influences commuting behaviour.

Among the so-called land-use/geographical variables these connections exits as well and is again indicated by the dashed lines in Figure 1. These variables are not connected to individuals and have to be reviewed on a larger scale; for example a neighbourhood, a city or a region within a country. Among these variables are some that explain the commuting pattern very well and these will be discussed in depth later on. The connection between these land-use variables are sometimes very obvious; the relation between population density and job density is intuitive explainable and the same applies to a urban/rural setting and its connection between housing prices (Clark, 1995).

Between the different categories of variables connections exist as well but are left out of the figure because for reasons of clarity. One could imagine that education is connected to population density because of the presence of universities in cities which are in turn densely populated. Another connection could be that the presence of parking difficulties would influence the individual decision to own a car, which is mostly regarded as a micro variable. I do not claim Figure 1 to be exhaustive or complete or to display all possible connections or influences between different variables. The figure is meant to show that the influence on commuting decision is a complex task to unravel and to indicate that many influencing variables exist on different levels of decision making.

2.2 Geographical coverage

The research on commuting in Europe is mainly focused on Sweden and the UK for which very detailed micro data on commuting is available and on other (mostly northern) European counties such as Denmark, Belgium, Germany and the Netherlands for which detailed data is available as well. On these countries a lot of research is conducted in the past (Deding et al., 2008; Gutiérrez-i-Puigarnau and van Ommeren, 2012; Groot et al., 2012; Schwanen et al., 2004; Van Acker and Witlox, 2011). The rest of Europe is somewhat absent in the papers about commuting and this is either by lack of interest in those countries or simply because of non-existing data on commuting behaviour mentioned before or a combination of both. More attention to different data sources: see data sources in section 2.4. In the United States quite some research is devoted to commuting habits and shows large differences per state or even city region. Quite some research is thus focussed on separate states (e.g. Mokhtarian et al., 2003) or city regions (e.g. Rosenbloom and Burns, 1993). Because of very different attitudes towards mobility, different working attitudes, lower fuel prices and so on, the US case is completely different from the European case and beyond the scope of this study. As metioned in the introduction the empirical work on this matter is geographically rather local. Its scope is on for example a region, a city district or province. The largest geographical scale setting of commuting research in Europe seems to be national.

2.3 Time period

The time period on which empirical research commuting papers are based is relatively recent. The time period of data used in the researches reviewed spans from roughly 1990 onwards. They all use statistical software such as Eviews or Stata to assess the connection between different variables and commuting distance/time or frequency. The emerging of statistic software could be one of the reasons that attention has steadily increased from the 1990's onwards. Another explanation might be that society sees the ever growing traffic movements and its accompanying congestion and environmental problems as an important issue and expects its politicians to come up with a solution. A logical next step is for scientist to gain deeper insights in the technical determinants of commuting and thus increasing research attention to commuting emerged since the 90's.

2.4 Data sources

Data sources that contain (micro) data about commuting practices are limited to a relatively small number of mainly northern European Countries. Especially Sweden, Great Britain, Germany and the Netherlands have very detailed micro data concerning commuting behaviour. The top five most used data sources on which a majority of European based commuting research is based are briefly discussed below.

The British Household Panel Survey

The British Household Panel Survey (BHPS) data was first collected in 1991 by the University of Essex to collect information about social and economic change on a household and individual level in Great Britain. The fact that commuting time is one variable in the data makes it very relevant for research on commuting. The BHPS is a stratified sample of households in order to be representative for all Great Britain households and in 1991 contained detailed information about 5000 households. In the sample of 2001 59% of households interviewed in 1991 still remain in the sample because particular attention was paid to avoid attrition (Dargay and Van Ommeren, 2005). From 2009 onwards the BHPS is part of a larger study called Understanding Society which contains very detailed data about 40000 UK households and is one of the largest panels in the world and contains sociological, economical and health data (Laurie, 2010).

National Travel Survey

The National Travel Survey (NTS) was first conducted in 1965 and commissioned by the UK Ministry of Transport and contains information about for example travel purposes, frequencies, modes, costs and vehicles in the household. After several unregularly appearances throughout the years it became a yearly survey from 1988 onwards. It is a smaller sample compared to the BHPS (in 2000 the sample contained 5796 households and changes over time; 2011 15.048 households; Taylor et al., 2012) but is truly devoted to travel behaviour. One part of the NTS is the travel diary where participants are asked to hand in about a week of their traveling behaviour. The NTS is a cross-section with a stratified sampling procedure to ensure that the sample is representative for British households (Kershaw et al., 2001).

ASTRID Database

The ASTRID database is a database which contains data about all Swedish residents between 1985 and 2003 on a micro level (income, family status, education, occupation, residential location, workplace location etc.). The beauty of the database is that next to the micro individual data it contains micro-geographical data using Geographical Information Systems (GIS) which makes it possible to calculate the commuting distance between home- and workplace very accurate (100 m2 resolution). This means the ASTRID database cannot say anything about the commuting time of the individual but everything about the Euclidian or “as the crow flies” distance between home and workplace and combined with the numerous other variables in the database is a very powerful tool to review commuting practices. Another plus is that it contained all Swedish residents (approx. 9 million) between 1985 and 2003 and thus no sampling errors could have occurred (Holm and Timpka, 2007; Sandow, 2008). On the other hand the lack of information about commuting time or mode is a disadvantage of the ASTRID database (Sandow and Westin, 2010).

German Socio-Economic Panel

The German Socio-Economic Panel (SOEP or GSOEP) is a panel of around 25.000 individuals in Germany and started in 1984. In 1990 East Germany was included as well and the sample has been refreshed several times during its lifetime (Wagner et al., 2007). It contains micro level individual specific data of the panel subjects comparable to the BHPS mentioned above and is interesting for commuting researchers because, among other things, it contains information about the usual commuting distance (Gutiérrez-i-Puigarnau and Van Ommeren, 2012). Unfortunately the commuting relevant questions are left out of the survey during some years (i.e. between 1990 and 2010 four years of commuting data is missing). A plus of the SOEP is that it contains data on subjective well-being of the subjects which possibly allows us to draw conclusions on the happiness of commuters. This is exactly what Stutzer and Frey (2008) did in their very entertaining article Stress that Doesn't Pay: The Commuting Paradox*.

Dutch National Travel Survey

The Dutch National Travel Survey is carried out by Statistics Netherlands (CBS) since 1978. It is a yearly cross-section which at the start contained 10.000 households and was increased up to

60.000 households (ca. 600.000 trips made by 150.000 individuals) in 1995. Due to several adjustments in survey methods and the merger of some municipalities in 2005 the surveys are unfortunately not entirely comparable anymore. The Dutch travel survey contains not only data about commuting lengths and distances and the usual micro variables (income, family status, age, gender, education etc.) but also detailed geographical variables such as distance from motorway ramps, metro stations, transport network density and urbanization degree. This makes research interesting on the relationship between commuting and the built environment (Susilo and Maat, 2007) or metropolitan structure (Schwanen en al., 2004) possible. The disadvantage of this NTS is that it is not a panel but a yearly cross-section although it is large enough to produce very credible results.

Other useful data sources

The only other European country with a long lasting history of travel surveys is Denmark with the Danish National Travel Survey that runs from 1975. In many other European countries (besides Sweden, Germany, UK and the Netherlands) similar databases containing travel behaviour data does exist but these are set up more recently (Austria; 2000), or appeared irregular, a few times or only once (Spain, France, Luxemburg, Norway) and contain no panel data (Violland, 2011).

Connecting the geographical scope, the time period reviewed and the data sources availability of existing empirical work it appears to be that the different scopes of research on commuting can be traced back to the data availability. From the 90's the use of statistical software became more common and data availability improved in some countries and in turn resulted in empirical research on commuting on the wider available data and a larger geographical scope. It would not be surprising if new empirical work would emerge on a more diverse set of regions/countries in the near future.

2.5 Level of analysis and type of data

A majority of the studies reviewed for this paper use highly detailed micro data on the individual level. Very few use mainly macro data to predict commuting (Östh and Lindgren, 2012) and some studies look at regional variables next to the usual micro data to draw conclusion on commuting patterns (Groot et al., 2012). Schwanen et al. (2004) argue that micro individual data is preferred because the (financial) resources to facilitate commuting at all, differ substantially among individuals. Although this seems to be a plausible argument in favour of using micro data, it might be a difficult practice to continue because of data availability issues mentioned before. The geographical coverage of research on commuting is constrained even more than it already is by only conducting research on micro data. At the same time the use of larger scale variables to explain commuting also seems relevant. Some authors claim that regional or municipally scale variables help to explain and predict commuting patterns (Groot et al., 2012) and even the earlier mentioned proponents (Schwanen et al; 2004) of the use of micro data use larger than individual scale data in their regression models. Examples are land rent in a certain area, population density in a city or region or even the nationwide tax regulations on transportation. At least the first two aforementioned variables seem to have a significant impact on commuting patterns (Kwon, 2005; Rouwendal and Nijkamp, 2004) and tax regulations might be an interesting subject of future commuting research. Examples of the use of true macro data to make statements about commuting are scarce but do produce significant results. An example from Östh and Lindgren (2012) detects that GDP change is significant correlated with commuting distance and the strength and direction of correlation are influenced by other geographical, demographical and socio-economic variables. The authors also state “the study is justified by the fact that there is a lack of knowledge concerning the longitudinal impact of economic cycles on workers’ commuting behaviour” (Östh and Lindgren, 2012) and in this way justify the use of macro variables in commuting research. Another reason in favour of the use of macro data not mentioned before or in the existing literature lies in the wider availability of macro data. With the use of (or at least in combination with) macro data a wider geographical scope can possibly be added to the existing empirical research on commuting behaviour which is now relatively limited to areas for which detailed micro data is available.

Usually the commuting researcher has little choice between different types of data because it is simply unavailable. Panel data is the most interesting tool because it tracks changing commuting decisions from a single individual under changing circumstances over time. Unfortunately commuting related panel data sets are scarce (GSOEP, BHPS, ASTRID database) and relatively few authors use panel data in their research (Benito and Oswald 2000; Dargay and Ommeren, 2005; Gutiérrez-i-Puigarnau and van Ommeren, 2012; Johansson et al., 2002; Östh and Lindgren, 2012; Stutzer and Frey, 2008). All of these studies use the large British (BHSP), German (GSOEP) or Swedish datasets. More widely available are cross-sectional surveys as with a European example the Dutch National Travel Survey. Several other non-European travel surveys such as some Australian travel surveys in different states as well as the US travel survey exist (Australian Time Users Research Group, 2013) and numerous studies use this type of datasets (Ben-David and Sharabi, 2009; Deding et al., 2008; Groot et al., 2012; Pucher and Renne, 2003; Schwanen et al., 2004; Stutzer and Frey, 2008; Susilo and Maat, 2007). Sometimes these cross-sections are grouped into household cohorts based on several grouping variables to create so called pseudo-panel data which result in reliable estimation if the cross-section is large enough and thus the household cohorts are large enough to simulate a trustworthy panel (Dargay, 2007). Time series are logically only used in combination with cross-sections (combined also called TSCS and is another name for a panel data). Consensus in literature exists on the matter and states that panel data is most desired when conducting research on commuting, followed by pseudo panel data if the cross-sections are large enough and the grouping variables are carefully selected.

2.6 Income and commuting

The relationship between income and commuting time/distance is discussed frequently in literature. Appendix A shows a summation of literature discussing the relationship between commuting and income. The urban economic theory basically states that richer people live further away from their workplace and therefore commute longer. The need for larger (affordable) residences for this higher income group forces this group to commute longer to avoid the densely populated areas where land rent is usually higher (Colwell and Munneke, 1997). In addition to this 'standard' urban economic theory there are two additional labour market explanations to deal with the positive income commuting relationship (Gutiérrez-i-Puigarnau and van Ommeren 2012). First, firms located far from residential areas will offer relative higher wages compared to relative closer located firms to become equally attractive for workers. Second, because of imperfect labour markets (e.g. search frictions or incomplete information) employers will offer higher wages and by doing so enlarge their attraction radius to compete for (scarce) employees. From these labour market explanations it is hard to draw conclusions about the causal direction of influence. Causal direction of influence is discussed later on.

From the consulted literature a consensus seems to exist because most authors show that richer people do live further away from their place of work. A very diverse range of different statistical modelling techniques have been used to research this income commuting relationship are not discussed in this thesis, but are briefly discussed as shown in Appendix A. We came across only five studies that actually calculated income elasticities related to commuting distance or time and their results are displayed in table I. As displayed in table I, most estimation results are positive elasticities which indicate that richer individuals (Benito and Oswald, 2000; Dargay and Van Ommeren, 2005; Groot et al., 2012) or households (Dargay and Clark, 2012; Gutiérrez-i-Puigarnau and Van Ommeren, 2012) live further away from their workplace than the relative poorer people do. From these five articles only one produced negative income elasticities which would mean that with rising income people would live closer to their place of work (Benito and Oswald, 2000).

This IV estimation model uses union membership as instrumental variable and assumes that this is related to higher income. Union membership might be positively correlated with income in Great Britain but doubtful for many other European countries. Benito and Oswald (2000) also estimate another model without this IV approach and estimate positive elasticities in this latter estimation.

The authors also differentiate between men and women and show that elasticity is higher for women compared to men, which would mean that an income change for women affects their commuting time (and likely distance) more than income changes for men. A possible explanation could be the still traditional division of child care responsibilities (if present in a household) between men and women and thus women stop working (and at the same time stop commuting and earning income) completely when children are born. If women do start to work again the commuting distance makes a jump from zero to a positive value which is reflected by the larger elasticity. Unfortunately Benito and Oswald (2000) estimate different models for men and women but do not try to explain these differences.

Dargay and Van Ommeren (2005) estimate a very simple model which results in a rather small elasticity of 0.04 possibly caused by the lack of any control variables. The most recent work on elasticities show only neutral or positive results ranging from 0.00 until 1.34.

Table I

Income Elasticities

Authors	Income elasticity	Model	Time Period	Data Source and Country
Benito and Oswald (2000)	≈0.15-0.2 for male ≈0.30 for female	OLS estimation with log(commute time as dependent	1991 - 1998	BHPS (Panel) Great Britain
	≈ -0.44 for male ≈ -0.36 for female	IV (2SLS) estimation with log(commute time) as dependent and union membership as IV	1991 - 1998	BHPS (Panel) Great Britain
Dargay and Van Ommeren (2005)	≈ 0.04	Fixed effects model with log(commute time) as dependent	1991 - 2001	BHPS (Panel) Great Britain
Groot et al. (2012)	≈0.29 ≈0.27	OLS estimates with log(commute distance) as dependent OLS estimates with log(commute distance) as dependent and job and population density added	2000 - 2008	EBB (9 cross-sections) The Netherlands
	≈0.16 ≈0.15	OLS estimates with log(commute time) as dependent OLS estimates with log(commute time) as dependent and job and population density added	2000 - 2008	EBB (9 cross-sections) The Netherlands
Gutiérrez-i-Puigarnau and van Ommeren (2012)	≈ 0.15	OLS and IV estimates with gross household income as IV and both with Log(commute distance) as dependent	1991 - 2003	GSOEP (Panel) Germany
Dargay and Clark (2012)	=0.57	Weighted least squares estimation with log(distance) as dependent	1995 - 2006	NTS (11 cross-sections) Great Britain
	=0.31 for car =1.34 for rail =0.00 for coach	Total of 47 models of which 7 on commuting	1995 - 2006	NTS (11 cross-sections) Great Britain

Notes: ≈: rounded to two decimals. All models estimate short term interactions (static)

Groot et al. (2012) estimate income elasticity for both commuting times and distances in two models with major difference with added variables for population and job density which does not result in large estimation changes. Their results show that the income elasticity is higher for distance compared to time estimates and this could partially be explained by the more extended modal options of higher income groups. An example could be: a highly skilled worker taking the train to work every day and decides at some point to switch to car commuting and as a result shortens his commuting time but not his distance. A low skilled worker taking the bus to work could probably not afford to take the shorter (in time terms) car trip to his work. From this example one could conclude that it is easier to shorten time than distance of a commuting trip, hence a lower income elasticity regarding commute time. In the research of Dargay and Clark (2012) a similar pattern can be detected because they estimated elasticity for every separate mode except for air commuting, because observations in this category were too few. They estimate an average income elasticity of 0.57 and their lowest estimates is close to zero for coach commuting. Assuming that bus commuting is an inconvenient way of traveling for most people it is understandable that a rising income does not result in a higher demand for bus commuting. Completely the opposite is the case of rail commuters, for which the income elasticity is the largest (1.34; Dargay and Clark, 2012). An explanation could be in the often long commute that is undertaken by train and an accompanying high salary. A second explanation could be in a value of time approach (VOT) which states that time is valued more by high income earners (Rouwendal and Nijkamp, 2004) who can work in a train in contrast to working while commuting by car. A third possible explanation could be that travel to highly urbanized areas where most high income earners tend to work (Fielding, 1989) is likely undertaken by train (or another form of public transport) because congestion problems play a large role in choosing the mode of transport.

Summarizing the different income elasticities estimated we can conclude that the consensus among several different authors is that higher income results in a longer commute both in terms of distance and time. Women show larger responses in this respect compared to men and bus commuting is showing low response on income change, followed by car commuting and finally train commuting shows the highest income change sensitivity because of reasons mentioned above.

Causality issue

Some authors ask questions regarding the direction of influence between income and commuting and some do not even care to address this issue while it is a quite important one. The question whether higher income earners commute longer or longer commutes result in a higher salary is a difficult one to tackle. Gutiérrez-i-Puigarnau and van Ommeren (2012) are among the authors that do look into this causality issue and state: “due to the presence of wage gradients, job search imperfections, tax systems and unobserved variables, it appears that based on a standard regression of commuting distance on household income (with controls), the effect of income on the commute is difficult to interpret as a causal effect of household income”. The authors are tackling the causality issue by removing people that changed jobs from their sample (the so-called “workplace spell”) and still come up with a positive income elasticity (see Table I) and conclude that richer people do live further away from their workplace as a result of income. In other words they conclude that a longer commute is *caused* by a higher income. Unfortunately they are among the few authors who actually addressed this causality issue. In this thesis we are not diving into the causal direction of influence but merely into the direction of influence between variables.

2.7 Education and commuting

Education level is used as a control in a large number of studies but Groot et al. (2012) treat education as the most important factor influencing commuting behaviour. They control for wages and conclude that higher educated workers commute (ceteris paribus) 26 per cent more by bike and 27 per cent less by car in the Netherlands. Overall the higher educated commute longer (and further. The effect of wages on the commute distance of low educated workers is highest and decreases with education level. The effect of wages on commuting distance of university graduates is very low and still they commute further. This leads Groot et al. (2012) to conclude that education is a more important determinant for commuting compared to wages. Multiple other authors have concluded that a higher education levels are usually accompanied by longer or further commutes (Sandow 2008; Benito and Oswald, 2000; Susilo and Maat, 2007; Östh and Lindgren, 2012). Of course education is often closely correlated with income (Griliches and Mason, 1972) so drawing simple conclusions about the education commuting relationship should be done with caution. Despite this caution it seems that education plays a large role in the individual decision on commuting practices; the higher educated commute further and longer independent of income differences.

2.8 Other socio-economical (micro) determinants for commuting

Gender is mentioned in many studies as a factor of importance influencing commuting. In general men commute longer, further and more frequent compared to women (Groot et al., 2012; Östh and Lindgren, 2012; Schwanen et al, 2004; Susilo and Maat, 2007; Van Acker and Witlox, 2011;) and this difference is greater when children are present in a household because of the still traditional larger role in child care by women (Deding et al., 2009; Van Ancker and Witlox, 2011). Age seems to be another determinant for commuting behaviour as younger people tend to commute longer and further then older people (Dargay and Clark, 2012; Sandow, 2008; Simonsohn, 2006). Whether housing is rented or owned seems to have an effect on commuting behavior as well because renters seem to be more mobile on the housing market and are more likely to adapt their commuting behavior by switching homes (Deding, 2009). Even the housing type seems to be relevant for the commute as Dargay and Clark (2012) estimate that individuals living in detached houses travel further compared to owners of semi-detached houses. This might be due to a larger investment which makes the homeowner less mobile and not triggered to move closer to work. The sector of employment is influencing the commute as well. Several authors look into this matter (Sandow and Westin, 2010; Deding et al., 2009; Groot et al., 2012) and found that private sector employees commute further compared to public sector employees. Very low commute times are found in the agricultural industry. Size of the workplace does also matter in the sense that larger plants or companies attract workers from further away and this leads to a longer commute for its employees (Benito and Oswald 2000). Car availability is also positively influencing the commuting behavior but before drawing conclusions one should be aware of the causal direction of the relationship between commuting and car availability (Dargay and Clark, 2012; Schwanen et al., 2004; Van Acker and Witlox, 2011; Susilo and Maat, 2007; Pucher and Renne 2003).

2.9 Non micro economical determinants for commuting behaviour

A different and relatively ‘younger’ approach towards commuting behaviour is concerned with less personal characteristics of the commute but with the physical environment the commuter lives and/or works in. For example in remote and sparsely populated areas the commute of workers seems to be longer regardless of income and education (Sandow, 2008). Also population density, residential density and employment density are all affecting the commute (Schwanen et al., 2004). Job density lowers the commuting distance as well as population density and residential density does because of interrelatedness. It is likely that commuting distances are affected more than commuting times by these indicators because of slower commuting times in densely populated congested areas. Schwanen et al. (2004) also checked whether poly or mono centric agglomerations show significant differences in commuting times and distances but do not find any striking results for the Netherlands. Job growth ratios are expected to influence the commuting times/distances negatively but this seems not to be the case probably because newly created employment is currently often situated outside congested residential areas and thus does not lead to shorter distances. In addition to the research on the connection between density and commuting Östh and Lindgren (2012) check whether a rural or urban home municipality has an influence on commuting distance. Their conclusion is that in response to income rise both urban and rural areas start to show an increase in commuting distance but the urban commuters' daily trip shows a faster response. Johansson et al. (2002) approached commuting from a rather original perspective and constructed an accessibility measure for municipalities and related this to commuting behaviour. Their significant results show that accessibility is an important determinant for commuting behaviour and is good news for policy makers because commuting habits can be altered via altering the accessibility policies in municipalities. Van Acker and Witlox (2011) try to determine if the type of land use is influencing commuting distances in Belgium with their own constructed variable called the "land use diversity index". This index quantifies the degree of balance between several land use purposes such as residences, services, nature, agriculture, industry etc. They estimate that this diversity lowers commuting distances because of a mix of working and living purposes in the same region lowers distances between them. This, again, is good news for policy makers as it seems they can influence the type of land use, especially within municipalities, at least partially.

Non economical motives are also important do determine future commuting behaviour. Simonsohn (2006) checked if previous experience with long commutes (so-called "contrast effects" in psychology) have an influence on commuting discussions. He concluded that people with previous long commutes tend to choose long commutes again in the future but do adapt their commuting behavior to the standards of their new geographical and social environment eventually. Deding et al. (2008) and Sandow and Westin (2010) also found a positive relationship between 'commuting experience' and the commuting behavior.

2.10 Synopsis of empirical literature

Table II shows a summary of research consensus about the relationship between the listed variables and commuting.

Table II

Variables influencing commuting decisions

Variable	**Influence on commuting time/distance**	**Papers**
Income	Income elasticity between 0.00 – 1.34	Benito and Oswald (2000); Dargay and Van Ommeren (2005); Dargay and Clark (2012); Gutiérrez-i-Puigarnau and Van Ommeren (2012); Östh and Lindgren (2012); Pucher and Renne (2003); Rouwendal and Nijkamp (2004); Sandow (2008); Sandow and Westin (2010); Schwanen et al. (2004); Simonsohn (2006); Susilo and Maat (2007)
Education	Positive	Benito and Oswald (2000); Östh and Lindgren (2012); Sandow (2008); Susilo and Maat (2007)
Male	31 – 37% longer commuting time	Dargay and Clark (2012); Gutiérrez-i-Puigarnau and Van Ommeren (2012); Groot et al. (2012)Sandow (2008); Sandow and Westin (2010); Simonsohn (2006)
Age	negative	Dargay and Clark (2012); Östh and Lindgren (2012); Sandow and Westin (2010); Susilo and Maat (2007); Van Acker and Witlox (2011)
Children	negative	Gutiérrez-i-Puigarnau and Van Ommeren (2012); Sandow (2008); Simonsohn (2006)
Home ownership	positive	Deding et al. (2008); Groot et al. (2012)
Car ownership	positive	Dargay and Clark (2012); Pucher and Renne (2003); Schwanen et al. (2004)
Length of residence Time (social cohesion)	positive	Dargay and Clark (2012); Rouwendal and Nijkamp (2004)
Previous commuting "experience"	positive	Deding et al. (2008); Sandow and Westin (2010); Simonsohn (2006)
Private sector	positive	Deding et al. (2009); Groot et al. (2012)
Population/residential density	negative	Dargay and Clark (2012); Sandow (2008); Schwanen et al. (2004); Susilo and Maat (2007).
Job density	negative	Johansson et al. (2002); Rouwendal and Nijkamp (2004).

Notes: 'positive' or 'negative' influence on commute time *or* distance.

It does not list elasticity ranges for all variables because the results and methods vary widely and many are not estimated in the form elasticities. If the direction of the relationship between a variable and commuting distance and/or time is confirmed by at least two papers it is included in the table in which 'positive' indicates a positive and 'negative' a negative relationship between the displayed variable and commuting time *or* distance. As far as possible or comparable a range of results is displayed in table II.

From table II above it seems that a young, highly educated male without children, who lives in a more rural area in a rented house, owns a car and works in the private sector has a high probability of a longer commute compared to a woman with children, who is poorly educated, on a low income and lives in the city.

Research on commuting seemed to focus only on socio economic micro characteristics of individuals or households for a long time but is now shifting more towards a combined approach in which micro as well as geographical variables are combined. A combined approach seems to explain and predict commuting most accurate and future research on commuting should focus on both (Van Acker and Witlox, 2011). Furthermore it is notable that most empirical studies on commuting behaviour use static modelling and estimations of long run relationships or dynamic models are rare. These types of models are quite important as predicting tools for future commuting behaviour. Again, this depends to a large extent on the data availability of panel data in which the different variables that have an influence on commuting are included. Another characteristic of current empirical work is the rather local approach in geographical sense. It is mainly focused on regions or cities and European wide studies are rare (Schwanen, 2002). In the future added insights from psychology and behavioural economics might be used to construct new models that reflect reality more accurate than the current models do.

3 Data and Methods

In this chapter the data and methods used for constructing a dynamic macro level error correction model will be discussed. First, the nature of the data described. Second, the used variables are discussed and finally the main regression model is presented, after solving some statistical issues.

3.1 Data

The data used in this research is mostly obtained from Eurostat, the official statistics bureau of the European Union in Luxemburg. A panel data set is composed of eleven different Eurostat dataset to finally define a regression model to assess what is influencing commuting behavior on a European scale. The panel is limited to the amount of data available in the Eurostat database and is explained in more detail below.

The NUTS (Nomenclature des Unités Territoriales Statistiques*)* classification is developed and regulated by the European Union and is developed to ensure a uniform approach to statistics within Europe. Several non-European Union countries have also adopted the NUTS classifications. The classification system consist of three NUTS “sizes” (1,2,3) and are mostly based on population size. Nuts 1 is the largest size and nuts 3 is the smallest and for this reason it might be so that an entire country is designated as one NUTS region. The population size criteria is not applied very rigidly by the NUTS classification.

For the composed panel the NUTS 2 region is the smallest geographical statistical region for which data on commuting is available and for some of the important determining variables identified in the existing empirical work no data at all is available on the NUTS 2 scale. Examples of NUTS 2 regions are: the entire country of Lithuania or the city of Berlin which both have been classified as NUTS 2 regions because of population sizes but differ significantly in geographical size. A map of the statistical NUTS 2 regions is included in Appendix B, to have a clear idea about the size differences

These nuts 2 regions are the units of the cross-sections in the panel used for this research. The time series dimension of the panel stretches from 1999 until 2012 but not for all cross-sections all time units are available or the other way around.

3.2 Variables Used

In literature commuting time or distance is most often used as dependent variable in regression models. Unfortunately commuting time or distance is not available in the Eurostat database so a balance of the outflow of commuters compared to the total population of a NUTS 2 region is used as dependent variable. The model uses the number of commuters working outside the region of residence and not the more common measures of commute length or distance because this is simply not available.

To control for the differences in workforce or population between the regions which of course influences the number of commuters, two options exist. The first is to insert the total population of a region as a control in the independent variables but this number does not control for any unemployment rate or demographical differences between regions. In a region where for example the share of people above retirement age is significant the share of commuters compared to economically active population is likely to be underestimated. It is intuitively sounder then to include the economically active population as a control variable. Unfortunately this also has some disadvantages because now other population demographics such as the number of children in a region, is likely to be overestimated because it is compared to active population instead of total population. In the estimation models active population is used instead of total population because the commuting outflow is the important dependent variable in our regressions.

On the effects of income on commuting behavior consensus exist that this will positively influence the commuting distance and/or time in micro data analysis. It is also expected that on a larger data scale income will have an effect on commuting behavior and in our case the commuters balance between regions. The income variable is an aggregated number of the income in euros per person per year over the entire population of the NUTS 2 regions.

The next most important issue in explaining commuting behavior discussed in empirical literature is education level. Several data strings on education are available from Eurostat. Eurostat measures education levels among countries based on the International Standard Classification of Education (ISCED) scale of UNESCO which is meant to make across border education levels comparable. The ISCED (1997) framework is valued from 0 to 6 in which level 0 stands for post primary education and level 6 stands for "the second stage of tertiary education" such as an advanced research qualification (e.g. a Ph.D. degree).

Eurostat contains data on numbers of students enrolled in all levels of the ISCED education framework as well as numbers on education levels actually finished. The latter number seems more relevant because higher educated people that actually finished their education are more likely to be active on the labor market and engage in commuting behavior. For the proposed model the number of people that have finished tertiary education (ISCED 1997 levels 5 and 6) is taken as a measurement of the height of the education level of a NUTS 2 region. The reason for this it that several studies show that the highest educated employees commute the farthest, regardless of income.

Several studies include car availability in their models on commuting behavior. Based on the NUTS 2 region data from Eurostat the next best thing to use seems to be the motorization rate of the different regions. The motorization rate that will be included is the number of passenger vehicles per 1000 inhabitants and this is expected to have a positive relation with the commuting outflow of a region.

The presence of children in a household is estimated to bring down the commuting distance or time in several studies. On a regional (NUTS 2) aggregate level data is only available on number of people under 5, under 10 or under 15. Especially the presence of very young children (<5) has been estimated to be significantly related to the commuting decisions of households individuals. The question is if this also holds on an aggregated scale and it seems interesting to include the ratio of children under 5 in the model. Besides the very small children, other age groups will be tested as well in separate models. On the complete opposite of the demographical scope are the oldest age groups who are said to commute less compared to the younger part of the population. Using both the share of young children and e.g. the share of people over 65 could lead to problems in terms of multicollinearity, which will be checked later on.

Besides the personal economical characteristics that relate to commuting behavior literature showed that geographical variables show a sometimes strong relation to commuting behavior. Some of the geographical variables are available in the Eurostat database as well.

Population or job density showed a negative relationship with commuting time or distance because these ‘job centers’ attract workers from different regions. Unfortunately ‘job density’ is not available from Eurostat but population density shows high correlations with population density in some studies. This is why population density is included in the model and is expected to show a negative relationship with commuting outflows because of a concentration of labor demand.

The density of the transport network is a variable that would be very interesting to include in the model because some studies show its relation with commuting practices (e.g. Van Acker and Witlox, 2011). Rail and road density are available from Eurostat but the data is missing for many years and regions and jumps up and down between years. Because of these unreliable features and the large number of missing cases this variable is not included in the analysis but might be interesting when data quality improves.

A last possible influence on commuting for which data is available is the price of the actual commute. The price of travelling has multiple components such the tax regime on motor vehicles, insurance costs, toll costs, petrol taxes and so on. Unfortunately this data is not available from Eurostat at all. Only a nationwide consumer petrol price could be obtained from the monthly oil bulletin from the European Commission[1]. This nationwide data is included in the panel in the sense that it is assumed the same for every NUTS region within a country. Although the petrol costs are included, not too much is expected from the relationship between commuting and these aggregated prices because they might be too general.

Al the claims about the expected direction of influence of the variables and the accompanying empirical papers are displayed in table II. An overview of all the variables used, a more detailed description their official Eurostat names and their code names used in regression analysis can be found in appendix C.

3.3 Model

This empirical part uses the Stata 11 statistical software package to analyze the data. From both the dependent and independent variables the natural logarithm is taken to end up with a log-log

[1] A publication of the European Commission's Directorate-General for Energy. http://ec.europa.eu/energy/observatory/oil/bulletin_en.htm, may 2013

model which makes interpretation in terms of elasticities possible. The variables are first tested for non-stationarity because neglecting this data feature could produce biased coefficients. Because the panel is unbalanced the more regular unit root test such as the Levin–Lin–Chu, Harris–Tzavalis, Breitung, Im–Pesaran–Shin do not produce (trustworthy) results. The one used on the data is a Fisher type Dickey-Fuller test for unit roots. All the variables are tested separately and the analysis shows that at least some of the variables contain unit roots and for some mixed results are shown. Test statistics can be found in Appendix D.

A dynamic model seems to be a good approach to also asses the long term effects of a change in certain explanatory variables on commuting behaviour. For this reason an Error Correction Model (ECM) is used which also allows for stationary and non-stationary data to be used in one model so the stationarity problem described above is tackled as well. The Bårdsen Transformation of the Error Correction model is used because it shows both short- and long run effects of a change in the explanatory variables. The Bårdsen Transformation of the ECM is given by (Bårdsen, 1989):

(3.1)
$$\Delta y_t = \alpha_0 + \sum_{i=1}^{m-1} \alpha_i^* \Delta y_{t-i} + \sum_{j=1}^{p} \sum_{i=0}^{n-1} \beta_{ji}^* \Delta x_{jt-i} + \alpha_m^* y_{t-m} + \sum_{j=1}^{p} \beta_{jn}^* x_{jt-i} + \varepsilon_t$$

In equation (3.1), *m* is the number of lags for the dependent variable, *n* the number of lags for the independent variable and p is the total number of exogenous variables. This transformation does directly show the short run effects in the differenced terms of the explanatory variables x. The long run dynamics have to be revealed by equation (3.2):

(3.2)
$$\theta_j = \frac{-\beta_{jn}^*}{\alpha_m^*}$$

In which θ_j is the long run coefficient and α_m^* is called the adjustment coefficient and can be interpreted as the speed of adjustment towards equilibrium. The standard errors for the long run coefficients can be calculated from the calculated Bårdsen Transformation estimation results as in equation (3.3):

$$var(\theta_j) = (\alpha_m^*)^{-2}\left[var(\beta_{jn}^*) + (\theta_j)^2 var(\alpha_m^*) + 2\theta_j cov(\beta_{jn}^*, \alpha_m^*)\right] \quad \textbf{(3.3)}$$

After the model is chosen the appropriate number of lags for the dependent and independent variables have to be determined. According to Pesaran and Shin (1999) the maximum number of lags included is two when analysing annual data so four models have to be reviewed to determine te appropriate number of lags: ECM(1.1), ECM(1.2) ECM (2.1) and ECM (2.2). The numbers between brackets is the number of lags applied to the dependent/independent variable respectively. The model is tested with the Akaike Information Criterion (AIC) and reveals that the model with both the dependent and independent variables lagged twice fits the data best (Appendix E).

After deciding on the number of lags (ECM 2.2) the fixed effects and random effects form of the model is assessed. It is intuitively plausible to use fixed effects estimation because of the nature of the data. The cross-sections (regions) are expected to contain various different characteristics that are not included in the model e.g. cultural or taxation differences between regions. This expectation is formally tested by conducting a Hauseman test which does not produce results because of the unbalance nature of the panel. A Sargan-Hansen test formally confirms that a fixed effects model better fits the data compared to random effects estimation (APPENDIX E).

Controlling for hetroskedasticity is performed by estimating the model with and without robust standard errors. The large differences between the estimation results show that hetroskedasticity is present and robust standard errors should be used in the estimation to obtain trustworthy results.

The intuitive interrelatedness of some of the explanatory variables could potentially result in multicolinearity problems and is thus tested. Correlation table III is constructed and shows that the number of children in certain age categories is heavily and significantly correlated and poses a possible multicolinearity problem and resulting biased estimates. To resolve the high correlation shown in table III the two ages groups (age between 10 and 15 and from 15 up to 20 are removed from the regression. This is theoretically justified as well by the findings of several authors (e.g.Sandow, 2008; Susilo and Maat, 2007) that especially the youngest children affect the commuting behaviour of the parents most.

Table III

Correlation Coefficients Between Different Age Groups

	Under 5 years	5 to 10 years	10 to 15 years	15 to 20 years
Under 5 years	1.0000			
5 to 10 years	**0.9922**	1.0000		
10 to 15 years	**0.9760**	**0.9909**	1.0000	
15 to 20 years	**0.9617**	**0.9754**	**0.9900**	1.0000

Notes: **bold numbers** are significant at the p<0.001 level

The two remaining age groups are merged into one variable containing children aged 0 until 9 years old because significantly large correlation between these two age groups is present as well. Correlation between the economically active population and some other population variables is quite high as well but not dropped from the regression because the dependent variable is a number of commuters compared to this active population. Not controlling for the size of the workforce would lead to serious omitted variable bias and leave the regression analysis meaningless. An example is: if active population is left out, suddenly number of children, people over 65 years of age and number of cars show significant estimation in the regression analysis. Fortunately the model proposed used lagged and differenced variables that are not so sensitive to the correlation in the original variables. A full correlation table of all the original variables used is presented in Appendix E.

After the necessary adaptations from initial expectation due to the different statistical issues discussed above, the final model estimated looks like:

$$\begin{aligned}(3.5)\qquad \Delta lncommute =\ & \Delta lncommute_{t-1} + \Delta lnprimary + \Delta lnactive + \Delta lnpopdens + \Delta lnchild \\ & + \Delta lneduhifi + \Delta lncars + \Delta lnover65 + \Delta lnpetrol + \Delta lnprimary_{t-1} \\ & + \Delta lnactive_{t-1} + \Delta lnpopdens_{t-1} + \Delta lnchild_{t-1} + \Delta lneduhifi_{t-1} + \Delta lncars_{t-1} \\ & + \Delta lnover65_{t-1} + \Delta lnpetrol_{t-1} + lncommute_{t-2} + lnprimary_{t-2} \\ & + lnactive_{t-2} + lnpopdens_{t-2} + lnchild_{t-2} + lneduhifi_{t-2} + lncars_{t-2} \\ & + lnover65_{t-2} + lnpetrol_{t-2} + \varepsilon_t\end{aligned}$$

In equation (3.5) delta stand for the first difference method used and the subscripts for the number of lags applied explained in detail in equation 3.2. The terms containing $lnchild$are replaced in the different models with different age groups Finally year dummies are created to check for trends in time sense. Adding time dummies does not reveal any clear patterns and does not lead to any notable changes in the estimations results.

4. Results

This chapter displays the results of the model proposed in chapter 3. Section 4.1 will show the quantitative outcomes of the regression analyses. Section 4.2 discusses the outcomes of section 4.1

4.1 Estimation results

Descriptive statistics of the original variables used can be found in Appendix F but as these are numbers which depend partially on the population size which, in turn, could depend on the absolute size of the NUTS 2 region, they do not give a useful insight in relative differences. The adapted descriptive statistics shown in table IV show a more useful insight in the numbers (and their differences between regions) used in the regression equations and are meant to understand the differences between them.

Table IV

Descriptive Statistics of the Data Adapted

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
Commute	60,220.00	87,330.00	0.00	1,076,300.00
Primary (in €)	15,855.59	10,284.94	716.4	147,844.8
Active (thousands)	840.0656	715.6931	12.1	5,712.9
Popdens (per km²)	341.627	852.5626	1.9	9,673.7
Percentage < 5y	5.36%	1.20%	2.99%	13.78%
Percentage <10y	10.92%	2.35%	6.33%	28.43%
Percentage <15y	16.86%	3.42%	10.00%	42.04%
Perc. Eduhifi	22.55%	8.94%	3.20%	63.00%
Cars (p. 1000 inh.)	431.36	140.14	14	1,101
Percentage > 65y	16.39%	3.52%	3.06%	26.70%
Petrol (€ per liter)	1.155	0.200	0.649	1.667

Notes: percentages compared to total population of NUTS 2 region *Source:* Eurostat

Table IV reveals that large differences exist between regions in many ways. The differences between sizes of the different age groups are quite obvious and might predict differences in commuting behaviour. An example is the economically active population which ranges from an economically active population of just over 12.000 (Åland, Sweden) up to figures well over 5 million people (Île de France) which is over 400 times as large.

Before starting to analyse the regression estimation results it is wise to bear in mind that the dependent variable in the regression models is the number of commuters that work outside their region of residence. It is not the number of kilometres or travel time used to travel to work. All variables are included in natural logarithm scale and can be interpreted as elasticities if needed. The long run coefficients θ are calculated using equation (3.2). An example helps to clarify its working: the adjustment α_m^* factor is -0.71 and the eduhifi coefficient lagged twice is 0.21. Applying formula (3.2) the long term coefficient θ for education as calculated as in (4.1)

(4.1) $$\theta(0.30) \approx \frac{-(0.21)}{-0.71}$$

Table V below shows three different models in which the width of the age groups of children is chosen differently in each model. Model I includes children under the age of 5 (child5), model II children under 10 (child10) and finally model III with children under 15 (child15). Purpose of doing so is to check whether a broader age group alters the estimated coefficients and the influence of children and whether it changes other estimated coefficients. This appears not to be the case because no coefficients changed signs or turned out to be suddenly (in)significant if the age groups were altered.

Table V

Regression Coefficients Estimates

	Model I		**Model II**		**Model III**	
	Coefficient	(SE)	Coefficient	SE	Coefficient	(SE)
$\Delta primary$	0.24	(0.18)	0.20	(0.18)	0.19	(0.18)
$\Delta active$	**2.02**	**(0.27)**	**2.00**	**(0.27)**	**1.97**	**(0.27)**
$\Delta popdens$	**3.79**	**(1.26)**	**3.75**	**(1.34)**	**3.78**	**(1.21)**
$\Delta child5$	-0.62	(0.78)				
$\Delta child10$			0.31	(1.48)		
$\Delta child15$					0.69	(1.48)
$\Delta eduhifi$	**0.23**	**(0.11)**	**0.24**	**(0.12)**	**0.24**	**(0.11)**
$\Delta cars$	**0.28**	**(0.08)**	**0.26**	**(0.08)**	**0.24**	**(0.09)**
$\Delta over65$	-1.10	(1.21)	-1.35	(1.20)	-1.41	(1.19)
$\Delta petrol$	-0.09	(0.14)	-0.09	(0.14)	-0.09	(0.14)
$\Delta commute_{t-1}$	**-0.70**	**(0.05)**	**-0.70**	**(0.05)**	**-0.70**	**(0.05)**
$\Delta primary_{t-1}$	0.23	(0.17)	0.18	(0.17)	0.18	(0.17)
$\Delta active_{t-1}$	**0.94**	**(0.30)**	**0.94**	**(0.30)**	**0.92**	**(0.30)**
$\Delta popdens_{t-1}$	**-4.52**	**(1.50)**	**-4.00**	**(1.54)**	**-3.13**	**(1.37)**
$\Delta child5_{t-1}$	1.17	(0.62)				
$\Delta child10_{t-1}$			-0.07	(0.91)		
$\Delta child15_{t-1}$					-1.70	(1.40)
$\Delta eduhifi_{t-1}$	0.11	(0.12)	0.11	(0.13)	0.12	(0.12)
$\Delta cars_{t-1}$	**0.29**	**(0.14)**	0.26	(0.13)	0.22	(0.13)
$\Delta over65_{t-1}$	0.07	(1.20)	0.07	(1.21)	0.02	(1.22)
$\Delta petrol_{t-1}$	-0.12	(0.14)	-0.14	(0.15)	-0.17	(0.15)
$commute_{t-2}$	-0.71	(0.06)	-0.71	(0.06)	-0.72	(0.06)
$primary_{t-2}$	-0.07	(0.11)	-0.08	(0.12)	-0.07	(0.12)
$active_{t-2}$	1.59	(0.35)	1.56	(0.34)	1.50	(0.34)

Table V (continued)

Regression Coefficients Estimates

	Model I		**Model II**		**Model III**	
	Coefficient	(SE)	Coefficient	(SE)	Coefficient	(SE)
$popdens_{t-2}$	-2.33	(0.82)	-2.11	(0.77)	-1.62	(0.93)
$child5_{t-2}$	-0.03	(0.18)				
$child10_{t-2}$			-0.11	(0.21)		
$child15_{t-2}$					-0.32	(0.37)
$eduhifi_{t-2}$	0.21	(0.13)	0.21	(0.14)	0.23	(0.14)
$cars_{t-2}$	0.25	(0.13)	0.23	(0.13)	0.18	(0.13)
$over65_{t-2}$	0.50	(0.28)	0.41	(0.26)	0.26	(0.27)
$petrol_{t-2}$	-0.05	(0.17)	-0.04	(0.17)	-0.07	(0.17)
Long run parameters						
$\theta_{primary}$	-0.10	(0.17)	-0.12	(0.17)	-0.09	(0.17)
θ_{active}	**2.23**	**(0.43)**	**2.18**	**(0.42)**	**2.09**	**(0.42)**
$\theta_{popdens}$	**-3.28**	**(1.08)**	**-2.96**	**(1.05)**	**-2.27**	**(1.29)**
θ_{child5}	-0.05	(0.25)				
$\theta_{child10}$			-0.16	(0.30)		
$\theta_{child15}$					-0.44	(0.52)
$\theta_{eduhifi}$	**0.30**	**(0.18)**	**0.30**	**(0.19)**	**0.32**	**(0.18)**
θ_{cars}	0.35	(0.18)	0.32	(0.18)	0.26	(0.19)
θ_{over65}	0.70	(0.37)	0.58	(0.34)	0.37	(0.37)
θ_{petrol}	-0.06	(0.23)	-0.06	(0.24)	-0.10	(0.24)
Observations	1341		1341		1341	
R^2	0.4397		0.4376		0.4391	

Notes: **bold number** are significant at the $p<0.05$ level

4.2 Discussion of results

The next step is to discuss the short and long term estimation results in the same order displayed in table V. Income does not show significant estimation results in any of the models estimated in long or short term. Striking is the change of sign between short run (positive) and long run (negative) values of income in all three models. Economically intuitively this could be explained by the (on micro level) multiple times estimated relation between rising income and longer commutes. A rise in income on the short term is related to more commuting. More income means more commuters to another region on the short term (table II). After rising income the outflow of commuters seems to halt after two years (which are the long run parameters Θ in table V) and the region could become an attraction region for commuters (hence the negative long run parameters) because employment in this richer area (compared to two years before) is created. Despite of the intuitive economical defendable reasoning hard evidence is missing because the estimations are not significant. An explanation for the lack of significance could be the aggregated nature of the variable primary income. All differences in income and their accompanying different commuting behaviour is levelled out by this feature of the data. Income is proven to be a significant predictor of commuting behaviour but on this macro scale it turns out to prevent us from drawing strong conclusions.

The economically active population is expected to be a significant predictor of numbers of out commuters both in the long and short term. This proved to be true as p-values are smaller than 0.05. Notable is that the coefficients are all close to two which means (since all variables are in natural logarithm form) that a one per cent rise in the active populations size leads to a twice as large increase in the number of people commuting to another NUTS region in the short term (*ceteris paribus*). The long term coefficients are even larger which means that with a delay of two years the reaction on a rising economically active population is even larger. Explanation for this rather logical relationship is that if the number of people working rises, the number of people working outside their "home region" rises as well. The magnitude of the coefficients is somewhat surprising but could be partially explained by the fact that a rapid rise in employees and a lagging employment demand growth forces people to look for work further away from home. The increasing coefficient in the long term could be caused by adjustment or frictions problems in finding new jobs and housing.

Population density appears to be a significant estimator for out commuting as well. Both the short and long term estimations show significant estimation coefficients but differ in the fact that short term coefficient's are highly positive and long term coefficients show negative signs. The short term population density growth (which we used as proxy for job density) results in higher commuting outflows in the estimation model. This is explainable in economic terms because a rapid increase in people per square km does not lead to immediate job growth in the region and people are forced to search for employment elsewhere. The change indicates that the long term effect of a density growth results in a decline of out commuters *ceteris paribus*. An explanation might be that employers are locating employment centres in or near denser populated areas but need time to relocate facilities to adapt to changing geographies. This rather rigid adaptation of employment supply is shown in the negative sign of the estimated long-term coefficient ϴ. Another explanation could be the merger of several NUTS 2 regions causes the estimates to be unreliable so caution is advised while interpreting.

The relationship between particularly young children and their parents commuting behaviour is negative on a micro level in several studies (see literature review, table II) on a micro level. On this macro scale it is harder to prove which is reflected by the non-significant nature of the estimations. There is however the expected negative sign in estimation one which could indicate that having young children (in this case under 5 years) reduces the number of commuters to another NUTS2 region. Widening the age groups leaves us with an unclear picture of the short term effects of the presence of children in a certain age group. Striking nonetheless is that whatever the width of the age group, the long term coefficients for every age group are negative. Again, the results are not even close to any standard significance levels so conclusions are hard to draw. An explanation for the significant findings on the micro level and the lack of those on the macro level is that having children is a very personal matter on which decisions are taken on the family scale. The diverse nature of people or families included in one number for a whole region appears not to be statistically and intuitively sound.

Education level is found to be a significant estimator in the regression analysis on both long and short term. A short term coefficient of 0.23 is found in model I and this means that a region in which the number of people with at least a finished tertiary education is 10 per cent higher, the number of commuters is 2,3 per cent higher. Model II and III show similar short term coefficients. The long term coefficient for education shows that the future effect of finished high education is even higher compared to short term effects. The estimations are all 0.3 or higher which means that a 10 per cent increase in people with finished tertiary education relates to a 3 per cent increase of "out commuters" two years later. Education is the only significant estimation coefficient that shows similar results with the micro studies discussed in the literature review section. The notion that the long term relationship of rising education levels is higher compared to the short run relationship could be caused by the search time that the market needs to match employment supply and demand when entering the labour market after finishing education.

The number of cars per 1000 people shows the final significant estimation relationship in table V and is positively related to commuting practices. In previous research this relationship was found as well but the causal direction is hard to determine. More cars mean more commuters in the short term *ceteris paribus*. The long term coefficients are very close to the significance level of 5% but are not considered significant. The car-commuting relationship is probably a very tight one and does not appear in lagged form. To be able to commute one often needs a car and owning a car makes it possible to commute further away. The present state (short term effect) of car availability seems to be more important than the past car availability does.

The size of the age group over 65 years shows no significant estimation results and also shows conflicting signs. As with children the data might be too aggregated to show an effect on commuting behaviour. Added thereto, the retirement age in some countries is above and more often below 65 so the sudden behavioural change with regards to commuting does not appear at this age and the effects in the regression might be even more marginalized.

The petrol price coefficient shows negative signs in short and long term but is not significant. The coefficients are rather small which could indicate that petrol price rises do not have a very large effect on commuting behaviour. This rather weak relationship between gasoline price and demand has been demonstrated before in literature (Brons et al., 2008), but is not proven in this study because this data might again be too aggregated. Another possible cause of not showing significant estimation results could be the rigid demand for gasoline for private users which have to commute to work by car and do not have very close alternatives to choose from.

5 Land use diversity, micro level check and comparison of results

In this chapter we will make three separate 'sidesteps' from the main macro analysis in this thesis. First, an attempt is made to explain the differences between regions by using a land use diversity index. Next, a check will be performed to assess whether the same variables used in the dynamic macro analysis of chapter three and four show comparable results in a micro level static regression model. Finally a comparison will be made between the existing literature, the macro analysis and the micro analysis from this chapter.

5.1 The land use diversity index and commuting

Van Acker and Witlox (2011) estimate that the diversity of land use is related to commuting behavior already discussed in chapter two. The authors show that the more diverse land use purposes are, the shorter commuting time seems to be. In this section a measurement for this land use diversity is constructed to test whether it shows a connection with the commuting outflow out of the NUTS 2 regions under research in chapter three and four. A dynamic model is not possible due to the lack of a time series on the land use diversity of the regions. Only for 2009 the Eurostat database contains the percentages of land use in different categories in a 'basic' and an 'extended' distribution of land use patterns explained below.

Eurostat has data from 2009 on the diversity of land use on a NUTS 2 scale. For this study we have calculated two different land use diversity indexes based on the methods used by van Acker and Witlox (2011). A 'basic' one that uses only the main categories of land use that are used by Eurostat (Agriculture, Forestry, Hunting and Fishing, Heavy and Environmental, Services and Residential and No Visible Use) and is calculated as shown in equation 5.1.

(5.1) $$Land\ use\ diversity = 1 - \left[\frac{\left|\frac{a}{T}-\frac{1}{6}\right| + \left|\frac{f}{T}-\frac{1}{6}\right| + \left|\frac{hu}{T}-\frac{1}{6}\right| + \left|\frac{he}{T}-\frac{1}{6}\right| + \left|\frac{s}{T}-\frac{1}{6}\right| + \left|\frac{n}{T}-\frac{1}{6}\right|}{\frac{10}{6}}\right]$$

in which *T* stands for total land area, *a* stands for agriculture, *f* for forestry, *hu* for Hunting and Fishing, *he* for Heavy and Environmental, *s* for Services and Residential and *n* for No Visible Use. The diversity number can only be positive or zero between 0 and 1 because of the use of absolute values. A value of 1 means that the land use is equally distributed between all the different categories and a value of 0 means that the land use diversity consists of only one category. Another somewhat more complex index splits up agriculture, heavy environmental and services and residences categories up into several subcategories. The calculation basically remains the same and is explained in detail in Appendix G. For reasons of clearness we now refer to the more complex division measure as ‘extended’.

Model

Because only a cross-section of the land use diversity indexes is available, a model will be estimated using region dummies from the dynamic Bårdsen model and apply a simple linear regression model to estimate the influence of different land use patterns on different regional commuting patterns. The dependent variable will be the region dummies from the dynamic model estimates and the independent variable will be the ‘basic’ and ‘extended’ land use diversity indexes respectively. What the model basically does is try to explain the regional differences based on the land use diversity index constructed, after controlling for the variables used in the dynamic regression approach.

Results

The results from the basic and extended estimations are displayed in table VI below. Both indexes do not show significant or consistent results. The ‘basic’ land use diversity index shows a negative sign which would mean that more diversified regions show a lower commuting outflow but the ‘extended’ version shows an exactly opposite direction of influence. Besides the different signs, the estimations are not significant at all and a relationship based on this geographical scale of measurement is not showing a pattern at all. Acker and Witlox (2011) already said that the land use diversity at the work location might be just as important as the residential conditions and those are not included in this analysis due to a lack of data. This issue, combined with the huge scale difference in the current analysis (NUTS 2 regions) and the authors’ (neighbourhoods within Ghent), might cause this analysis to fail to produce a coherent picture of the influence of land use mix on commuting behaviour.

Table VI

Land Use Diversity and Commuting

	Coefficient (Standard Error)	R^2
Basic land use diversity index	-2.15 (2.12)	0.0064
Extended land use diversity index	0,85 (2.71)	0.0006

Notes: Both coefficients are not significant

5.2 Micro level data and commuting

During the study period a new dataset was published based on the European Working Conditions Survey from 2010. It is based on personal interviews in all EU countries and some outside it and is conducted every five years. It is a cross-section of individuals and no time dimension is included in the dataset which makes a dynamic approach impossible. To detect whether the relationships found in literature review exist in the countries under consideration in the macro analysis of chapters three and four, a micro analysis is conducted using the same variables and countries used in the original regression model.

Opposed to the original dynamic approach any possible long term relations are not detectable in this analysis. It is particularly meant to clarify the results of the panel data analysis on the Nuts 2 level and prove that the independent variables chosen are relevant in the geographical area under research, at least on the short term. The analysis is constrained to the variables available in the Eurostat database to allow for a sound comparison between the two analyses as far as possible.

Table VII

Descriptive Statistics of Micro Dataset

	Observations	Mean	Standard Deviation	Min	Max
Commute (minutes)	39,408	41.29	33.99	1.00	360
Income	29,617	1,135.31	1,053.55	0.00	30.231
Cars (per 1000 inhabitants)	23,270	419.71	138.80	19.00	779.00
Density	29,110	487.365	1,108.54	3.30	9,673.70
Petrol	24,902	1,328.04	101.49	1,109.72	1,492.98

Model

The EWSC survey of 2010 is used to conduct a relatively simple linear regression analysis and the summary statistics of the quantitative data is shown in table VII. All the data from the EWCS is gathered in 2010 and is individual specific micro level data. The motorization rate and population density is transferred from the NUTS 2 level panel data set used in chapter three and four and connected to the subjects with the help of the geographical information attached to each observation in the EWCS. The motorization and density are thus on a larger than individual scale. The petrol prices of 2010 are also transferred from the original panel and are true macro numbers of the petrol price per country. Interpreting the estimation based on this variable should be done with caution because of the large geographical scale difference in the variable measurement.

The variables from the original EWCS are the commute time (in total minutes per day), income in euros per month, a dummy for children in the household, the possession of a tertiary education degree and a last dummy for being over 65 years of age. Testing for different assumptions of linear regressions does not lead to large problems with the variable or model properties. Multicollinearity is not a problem according to an analysis based on variance inflation factors because they are all under 10 and 5 which are safe thresholds according to Myers (1990) and Menard (1995) respectively. Appendix H contains a table which displays the VIF's for all the used variables. Any problems with heteroscedasticity of the quantitative variables are resolved by taking the natural logarithm of these variables. The variable income is logged only after adding one to the observations as there is no natural logarithm of 0 and zero income individuals are observed (table VII). This change in approach does not lead to very different results in terms of estimated coefficients, standard errors or significance levels. The presence of heteroskedasticity is not clear after visual inspection (Appendix H), but just in case robust standard errors are used. The use of using robust standard errors does hardly change any of the aforementioned estimation results The approximate normality of the standard errors is assessed using a plot of the fitted against the observed values and does not show a very large deviation from normality and is shown in Appendix H.

The final regression equation used for the micro analysis looks like:

(5.1)

$$lncommute = lnincome + lncars + lndensity + lnpetrol + D_1.child + D_2.over65 + D_3.eduhifi + \varepsilon$$

In (5.1) the ln variables in natural logarithms and D_1, D_2 and D_3 are the dummy variables in the child, over 65 years of age and finished tertiary education categories.

Results

As shown table VIII the regression results are all highly significant. The expected directions of coefficients based on the literature review are visible from the estimation coefficients. The coefficients in logarithm form are directly interpretable as percentage change. A ten percent rise in income will lead to 1.4 percent rise in commuting time in minutes per day *ceteris paribus*. A higher motorization rate will lead to a lower commuting time, probably due to the fact that car commuting still is the fastest feasible way of traveling to work. Higher residential densities lead to a longer commute which is most likely the result of slower commuting speeds (and modes) in urbanized areas (Levtnson and Kumar, 1997). The petrol price has, according to this model, hardly any influence on the commuting time, probably due to the availability of modes and regulations which make a direct connection between petrol price and commuting cost marginal. Nonetheless this coefficients should be reviewed with caution because the petrol price is assessed as a nationwide average because a lack of local data of this variable.

Table VIII

Regression Analysis of Micro Data

Variable	**Coefficient (SE)**
constant	**4.603 (0.27)**
Lnincome	**0.146 (0.01)**
Lndensity	**0.113 (0.01)**
Lncars	**-0.182 (0.05)**
Lnpetrol	**-0.001 (0.00)**
Over65 (dummy)	**-0.203 (0.06)**
Child (dummy)	**-0.052 (0.02)**
Eduhifi (dummy)	**0.125 (0.02)**
R^2	0.05
Observations	

Notes: all **bold** coefficients are significant at the $p<0.01$ level

The dummy variables are showing the same signs expected based on the literature review as well. The coefficients cannot be interpreted directly because the dependent variable in this regression is in natural logarithm form and should be transformed 'back' using Euler's number *e*.

Being over 65 years of age leads (after transformation) to a reduction in commuting time per day of 22.5 % and is the most influential dummy in this regression. This intuitively makes sense because in many European countries under consideration here the retirement age is reached at 65. If children present in the household the commuting time will be on average 5% lower and a finished tertiary education program will on average indicate a 13.3 % longer commuting time. All statements above are in line with the consulted literature in chapter two and based on the *ceteris paribus* assumption. The negative sign of car availability seems surprising but when realizing commuting time is the dependent variable it seems more logical. Car commuting is still one of the fastest modes and owning a car will likely reduce the commuting *time.* The value of R^2 is not really high but is usually lower in cross-sectional analysis compared to panel analysis and the same applies for large datasets compared to smaller ones (Reisinger, 1997). Adding predictors to the regression equation above to increase R^2 is not done in this case because the current variables represent the same relationships assessed in the panel analysis of chapter three and four.

This regression analysis shows highly significant results in all chosen variables. It could be viewed as a confirmation of the right choice of macro variables which was made before obtaining the EWCS 2010 dataset.

5.3 Comparison of results from different analyses

Table IX below shows a comparison of the results from the literature review (chapter 2), the macro analysis (chapter 3 and 4) and the micro analysis (section 5.2 above). Column one is a reproduction from the literature findings in table II and is mostly indicated with positive or negative as explained in section 2.10. Column two displays the results from table V and column three shows the results from table VIII. Some numbers in column three do not exactly match the numbers in table VIII because they had to be transformed as explained before (section 5.2). In this section we will mainly try to explain contrasting findings from the three analyses.

Table IX

Comparison of Results of Direction and Magnitude of Influence

Variable	**Literature *(commuting time or distance)***	**Macro Analysis *(commuting outflow)***	**Micro Analysis *(commuting time)***
Income	0.00 - 1.34	No significant results	+0.15
Higher Educated	positive	0.23 (ST) - 0.32 (LT)	+0.13
Children present in household	negative	No significant results	-0.05
Car availability	positive	0.24 – 0.29 (ST)	-0.18
Population density	negative	3.75 (ST) -3.28 (LT)	+0.11
>65 years	negative	No significant results	-0.22
petrol price	unknown	No significant results	≈0.00

Notes: All numbers can be interpreted as elasticities
(dependent variable used between brackets)
LT: Long term; ST: short term

Table IX shows that income is a positive indicator for commuting time or distance in empirical literature. In the micro analysis in section 5.2 the effect of income on commuting time is estimated to be positive as well. The macro analysis does not yield significant results because of the aggregated nature of the data discussed before. Nonetheless, it is consistent that income is a positive estimator for commuting distance and time in micro analysis.

Education is a very consistent estimator. Literature, micro and macro analysis al shows that higher education is significantly and positively related to commuting distance, time and streams. The macro analysis shows that rising education levels in a region have an even stronger positive impact on commuting behaviour in the long term.

The presence of children in a household can only be estimated on a micro scale with households or individuals as subjects. In literature review and our own micro analysis children have a negative influence on the commuting time. In the macro analysis no significant coefficients are estimated and again, the cause is the aggregated nature of the data (number of children belonging to a certain age group in a region).

Investigating car availability results in a mixed view. The literature and macro analysis show a positive relationship with commuting distances but in our macro analysis the coefficient is negative. The explanation lies in the observation that most reviewed literature uses distance as measurement for commuting. The macro analysis uses commuters working in another region (not the residential region) as a measurement and this is at least related to distance. On the contrary, the micro analysis uses commuting *time* as measurement for commuting. Because car commuting is still the fastest commuting mode (Van Ommeren and Dargay, 2006), is makes sense that when cars become widely available the commuting speed increases but commuting *time* decreases at the same time.

Density also shows mixed results. In literature the negative relationship of density and commuting distance is explained by the attraction of both employers and employees to the denser areas. In these concentrations (mostly cities) the commuting distances are lower. In our macro analysis the sign of the coefficient switches from positive (short term) to negative (long term). This discrepancy is already explained in section 4.2. Again, caution is advised because region mergers or splits might bias the estimates. The negative result from the micro analysis can be explained by slower commuting speeds in urban areas and the accompanying longer commute times (Levtnson and Kumar, 1997), hence the negative sign of the coefficient.

Being of age (over 65 in the micro analysis) shows a negative relationship with commuting in existing literature and our micro analysis. The consistency between these two is caused by the use of mostly micro data in literature. The macro analysis does not show significant estimates regarding higher age. Again, this is caused by the aggregated nature of the data.

Finally, petrol price is not showing any clear patterns of influence on commuting behaviour. Either it is not tested, not significant or close to zero. One cause could be the existence of other modes of transport, besides cars, for which commuting costs do not directly relate to petrol prices. Another reasonable explanation is that petrol costs are not the main cost component in commuting costs and other costs play a bigger role. Again the aggregated nature of the data is also a possible cause for not showing significant results in the analyses, because it is the average yearly and nationwide petrol price.

6 Conclusions

The dynamic approach with macro level panel data and the European scope of the study is what distinguishes this study from the majority of empirical studies on commuting. The majority of empirical studies is rather local in geographical sense and use a static micro level research approach. The use of (regional) macro data has an advantage in the sense that it is available for more countries and time periods compared to micro data. This makes dynamic modelling possible for more geographical areas and in this case on almost the entire European Union.

Income was initially expected to show a positive relationship with commuting but in the dynamic macro analysis this was not confirmed. In a separate micro analysis the connection between income and commuting behavior is confirmed. The aggregated nature of macro data is the main cause for the absence of significant estimations in the main analysis.

On the contrary, education is a consistent and significant estimator in all analyses. The dynamic macro model reveals that long term effects of higher education levers are higher compared to short term effects on commuting outflows. The expected direction of this relationship was also confirmed in the micro analysis, although less strong compared to the main model.

Motorization rate and population density are showing significant results in the different analyses and should be carefully interpreted. The way in which commuting is measured can determine the sign of the estimates coefficients to be positive or negative.

Some other well-known variables with a relationship to commuting behavior known from the literature review are not showing significant results in our main estimations. Added to the absent income-commuting relationship mentioned above, the number of children, the size of the group of elderly and petrol prices are showing no significant estimation results in macro analysis as well. The aggregated nature macro data is most likely to be the cause of absents of significant estimation results for these variables as well.

Checking the same variables on the same countries assessed in the dynamic macro level analysis in a static micro level analysis yields, in contrast to the macro model, all the results that are expected from the literature review. Although not dynamic, the short term indicators for income, belonging to the elderly, children in a household and the indicators already significant in macro analysis show significant relationship with commuting time.

The tests using the land use diversity indexes was expected to produce negative estimates of the relationship between diversity of land use and commuting practices. On the regional (NUTS 2) scale this is not the case. The reason is probably the large geographical scale of the regions for which just one diversity index could be constructed.

The main advantage of the macro analysis applied in this thesis is wider data availability. More countries and areas can be examined compared to research solely based on micro data which is not so widely available. The obvious disadvantage of the macro approach is the fact that many well-known micro level indicators which could contribute to predicting or explaining commuting behavior are too aggregated and do not reveal the patterns they are expected to show.

This research shows that, although rare, some useful indicators exist on a macro scale on which policy makers can base their (future) policy. On the macro level, rising education levels and rising motorization rates are the most important and consistent indicators for rising commuting streams. Observing these changes should warn policymakers when trying to counteract commuting caused congestion problems. This result mainly applies to regions where education levels are lower and large growth is still possible, compared to relatively high and steady education levels in countries in mainly northern Europe. On the 'lower educated' regions highly detailed high quality quantitative micro data is rare and if interested in commuting behavior, the use of macro data seems to be the only option. Luckily this data seems to contain characteristics that allow predicting commuting behavior.

Further research has to reveal whether research on commuting behavior yields different results when based on other continents or separate countries. Research on commuting is important in the sense of acting on time. Many developing countries face huge congestion problems which in the future could be counteracted on time when more predictors are available. It also seems necessary to use dynamic models to estimate relationships to prevent over- or underestimating short and long run effects.

If micro data is not available, the next smallest data scale should be used in a research approach, but it is still micro data that contain the largest predicting power when commuting behavior is investigated.

References

Akamatsu, T. and Heydecker, B. (2003). Detecting dynamic traffic assignment capacity paradoxes in saturated networks. *Transportation Science*, *37*(2), 123-138.

Bårdsen, G. (1989). Estimation of long run coefficients in error correction models. *Oxford Bulletin of Economics and Statistics*, *51*(3), 345-350.

Ben-David, N., Sharabi, M. (2009). Commuting and Its Effect on Work Decisions. *International Journal of Economic Perspectives*, *3*(3), 183-187

Benito, A. and Oswald, A. J. (2000). Commuting in Great Britain in the 1990s.

Brons, M., Nijkamp, P., Pels, E., and Rietveld, P. (2008). A meta-analysis of the price elasticity of gasoline demand. A SUR approach. *Energy Economics*,*30*(5), 2105-2122.

Clark, E. (1995). The rent gap re-examined. *Urban Studies*, *32*(9), 1489-1503.

Colwell, P. F. and Munneke, H. J. (1997). The structure of urban land prices.*Journal of Urban Economics*, *41*(3), 321-336.

Dargay, J. (2007). The effect of prices and income on car travel in the UK.*Transportation Research Part A: Policy and Practice*, *41*(10), 949-960.

Dargay, J. M. and Clark, S. (2012). The determinants of long distance travel in Great Britain. *Transportation Research Part A: Policy and Practice*, *46*(3), 576-587.

Dargay, J. M. and Ommeren, J. (2005, August). The effect of income on commuting time using panel data.
In *45th Conference of the European Regional Science Association at the Vrije Universiteit Amsterdam, Amsterdam*.

De Borger, B. (2009). Commuting, congestion tolls and the structure of the labour market: Optimal congestion pricing in a wage bargaining model. *Regional Science and Urban Economics*, *39*(4), 434-448.

De Borger, B., Proost, S. (Eds.), (2001). Reforming Transport Pricing in the European Union. Edward Elgar Publishing Company, Cheltenham, UK

Deding, M., Filges, T., and Van Ommeren, J. (2008). SPATIAL MOBILITY AND COMMUTING: THE CASE OF TWO-EARNER HOUSEHOLDS*. *Journal of Regional Science*, *49*(1), 113-147.

Fielding, A. J. (1989). Migration and urbanization in Western Europe since 1950. *The Geographical Journal*, *155*(1), 60-69.

Gutiérrez-i-Puigarnau, E., van Ommeren, J.N. (2012). Do richer households live further away from their workplace? *Manuscript Title. Unpublished manuscript, Free University, Amsterdam, NL.*

Griliches, Z. and Mason, W. M. (1972). Education, income, and ability. *The Journal of Political Economy, 80(3), S74-S103.*

Groot, S. P., de Groot, H. L., and Veneri, P. (2012). The Educational Bias in Commuting Patterns: Micro-Evidence for the Netherlands (No. 12-080/3). *Tinbergen Institute*

Holm, E. and Timpka, T. (2007). A discrete time-space geography for epidemiology: from mixing groups to pockets of local order in pandemic simulations. *Studies in health technology and informatics, 129(1), 464.*

Johansson, B., Klaesson, J., and Olsson, M. (2002). Time distances and labor market integration. *Papers in Regional Science*, *81*(3), 305-327.

Kershaw, A., Barton, J., Noble, B., Williams, D., and Hird, D. (2001). *National travel survey*. London:: Technical Report 2000. Office for National Statistics.

Kwon, Y. (2005). Urban comparative statics when commuting cost depends on income. *Journal of Housing Economics*, *14*(1), 48-56.

Laurie, H. (2010). Continuity and innovation in the design of Understanding Society: the UK Household Longitudinal Study (No. 2010-02). *Understanding Society at the Institute for Social and Economic Research.*

Levtnson, D. M. and Kumar, A. (1997). Density and the journey to work. *Growth and Change*, *28*(2), 147-172.

Levy, J. I., Buonocore, J. J., and Von Stackelberg, K. (2010). Evaluation of the public health impacts of traffic congestion: a health risk assessment.*Environmental Health*, *9*(65), 1-12.

Lindsey, C. R. and Verhoef, E. T. (2000). *Traffic congestion and congestion pricing* (No. 00-101/3). Tinbergen Institute.

Menard, S. (1995). Applied logistic regression analysis: Sage university series on quantitative applications in the social sciences.

Mokhtarian, P. L., Collantes, G. O., and Gertz, C. (2003). Telecommuting, residential location, and commute distance traveled: Evidence from State of California employees

Myers, R. H. (1990). *Classical and modern regression with applications* (Vol. 2). Belmont, CA: Duxbury Press.

Östh, J. and Lindgren, U. (2012). Do Changes In Gdp Influence Commuting Distances? A Study Of Swedish Commuting Patterns Between 1990 And 2006. *Tijdschrift voor economische en sociale geografie.*

Portoghese, A., Spissu, E., Bhat, C. R., Eluru, N., and Meloni, I. (2011). A copula-based joint model of commute mode choice and number of non-work stops during the commute. *Rivista Internazionale di Economia dei Trasporti*,*38*(3), 337.

Pucher, J., and Renne, J. L. (2003). Socioeconomics of urban travel: evidence from the 2001 NHTS. *Transportation Quarterly*, *57*(3), 49-77.

Reisinger, H (1997) "The impact of research designs on R2 in linear regression models: an exploratory meta-analysis", *Journal of Empirical Generalisations in Marketing Science, Vol 2*, No.1

Rosenbloom, S. and Burns, E. (1993). Gender differences in commuter travel in Tucson: implications for travel demand management programs.

Rouwendal, J. and Nijkamp, P. (2004). Living in Two Worlds: A Review of Home-to-Work Decisions. *Growth and Change, 35*(3), 287-303

Sandow, E. (2008). Commuting behaviour in sparsely populated areas: evidence from northern Sweden. *Journal of Transport Geography*, *16*(1), 14-27.

Sandow, E. and Westin, K. (2010). The persevering commuter–Duration of long-distance commuting. *Transportation Research Part A: Policy and Practice, 44*(6), 433-445.

Schrank D, and T. Lomax (1999). The 1999 Annual Mobility Report Information For Urban America, Texas Transportation Institute, Texas A and M University System, College Station, Texas

Schwanen, T. (2002). Urban form and commuting behaviour: a cross-European perspective. *Tijdschrift voor economische en sociale geografie, 93*(3), 336-343.

Schwanen, T., Dieleman, F. M., and Dijst, M. (2004). The impact of metropolitan structure on commute behavior in the Netherlands: a multilevel approach. *Growth and Change, 35*(3), 304-333

Simonsohn, U. (2006). New Yorkers commute more everywhere: contrast effects in the field. *Review of Economics and Statistics*, *88*(1), 1-9.

Stolzenberg, R. M. (1975). Occupations, labor markets and the process of wage attainment. *American Sociological Review*, 645-665.

Stutzer, A. and Frey, B. S. (2008). Stress that Doesn't Pay: The Commuting Paradox*. *The Scandinavian Journal of Economics*, *110*(2), 339-366.

Susilo, Y. O. and Maat, K. (2007). The influence of built environment to the trends in commuting journeys in the Netherlands. *Transportation, 34*(5), 589-609.

Taylor, E., Humphrey, A., Pickering, K., Williams, D., and Tipping, S. (2012). *National travel survey*. London:: Technical Report 2011. The Department for Transport.

Van Acker, V. and Witlox, F. (2011). Commuting trips within tours: how is commuting related to land use?. Transportation, 38(3), 465-486.

Van der Laan, L. (1998). Changing urban systems: an empirical analysis at two spatial levels. *Regional studies*, *32*(3), 235-247.

Van Ommeren, J. and Dargay, J. (2006). The optimal choice of commuting speed: consequences for commuting time, distance and costs. *Journal of Transport Economics and Policy*, 279-296.

Violland, M. (2011). Travel/mobility surveys: some key findings.

Wagner, G., Frick, J., and Schupp, J. (2007). The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements.

Wardman, M., and Nicolás Ibáñez, J. (2012). The congestion multiplier: Variations in motorists' valuations of travel time with traffic conditions.*Transportation Research Part A: Policy and Practice*, *46*(1), 213-225

Online resources:

Australian Time Users Research Group:
retrieved from: http://australian-time-users-group.org/travel-time/au/ accessed July 2013

The Oil Bulletin: A publication of the European Commission's Directorate-General for Energy. Retrieved from: http://ec.europa.eu/energy/observatory/oil/bulletin_en.htm, accessed May 2013

Eurostat Statistics Database:
Accessed via http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database, Accessed throughout 2013

EWCS 2010 (European Working Conditions Survey)
Accessed via: http://www.data-archive.ac.uk/, accessed august 2013

Appendix A

#	APA Reference	Aims, Model and dependent variable	Explanatory Variables, DATA, geographical area of the research and time coverage	Results/Conclusions	Comments
1	Dargay, J. M., and Ommeren, J. (2005, August). The effect of income on commuting time using panel data. In *45th Conference of the European Regional Science Association at the Vrije Universiteit Amsterdam, Amsterdam*.	Fixed effects model specified: $\ln c_{it} \;\; = a_i + y \ln w_{it} + \beta x_{it} + u_{it}$ Ln commute dependent variable in minutes from door to door trip to work)	1) lnincome: real income in constant 2000 pounds. Source: BHPS. Area of coverage: United Kingdom 1991- 2001 Panel Data	Very small positive effect: income elasticity of 0.04 or 0.03	Unreliable Limited Dataset, very simple regression model, real income makes foreign comparison impossible.
2	Goodwin, P., Dargay, J., and Hanly, M. (2004). Elasticity's of road traffic and fuel consumption with respect to price and income: a review. *Transport Reviews, 24*(3), 275-292.	Meta Data analysis of 175 studies using cross-section, time series and panel data.	1) fuel consumption 2) vehicle km 3) vehicles 4) fuel efficiency 5) 4 other. Sources: vary widely. Area of coverage: Europe, OECD, USA, Australia, Canada, Japan and other 1929 - 1998 Panel, cross-section, time series data	Effects on fuel price change: almost all negative depending on the use of static/dynamic model and type of data. General: real price 10% up, traffic 1%(ST) or 3%(LT) down and volume consumed 2.5%(ST) and 6%(LT) down. Effect of income change: 10% up, # of vehicles and fuel consumption 4%(ST) and 10%(LT) up. Traffic volume 2%(ST) and 5%(LT) up.	Wide range of studies and sometimes hard to compare. Panel data shows the smallest elasticity's in static modeling. Not a very close connection to commuting but more overall car use.
3	Sandow, E., and Westin, K. (2010). The persevering commuter–Duration of long-distance commuting. *Transportation Research Part A: Policy and Practice, 44*(6), 433-445.	Multivariate regression analysis with duration of long distance commuting for dual households as dependent and as base year the 2000 data.	1)Long distance commute 2)age 3)gender 4)education level 5)income 6)employment sector 7)residential region 8)previous mobility experience 9)partner long distance commuting Source: Statistics Sweden (Astrid Database contains Micro data on individual level). Area of coverage: Sweden 1995 – 2005 Panel data	Biggest impact on the dependent variable (long distance commuting) is the coefficient of the variable years of LDC before. Income for both men and women are positively related to long distance commuting. Several conclusions on education level, age and the presents of children are drawn (page 439)	Might me reversed causality in the connection between long distance commuting and income (commuting has a positive effect on income) The independent variable 'years of experience' in commuting might be more psychological than it is an economical reason for commuting. Education, age and income are all independent and are likely to be related.
4	Blauwens, G., De Baere, P. and Van de Voorde, E. (2008). Transport Economics, *3nd edn, Antwerp: De Boeck*	Just a review on existing literature in a student's text book.	Population size, employment, income, car ownership, distance to various centers of activity, household composition.	No actual conclusion (just description of influential variables)	Trip generation as main topic, no specific attention to commuting
5	Sandow, E. (2008). Commuting behaviour in sparsely populated areas: evidence from northern Sweden. *Journal of Transport Geography*, *16*(1), 14-27.	Study commuting behavior and factors influencing individuals' propensities to commute longer distances. Analyze commuting behavior in a sparsely populated area with an emphasis on gender based differences Model: binary logistic regression with probability of long distance commute as dependent	1)age 2)education 3)income 4)employment sector 5)family status 6)presence of children 7)gender 8)employment opportunities 9)residential density Source: Statistics Sweden (Astrid Database contains Micro data on individual level) Area of coverage: Northern/Middle Sweden 1991 – 2003	Commutes are longer in low density areas for both men and women. Men and women working in the private sector tend to have longer commutes. High income and education men and women are longer commuters. Higher age groups decrees their commuting distances. Women with spouse and or children show shorter commutes. Men's commutes are always longer than women's'	Beautiful dataset with lot of variables (even 100m accurate distance information) on the individual level. Narrow in in space (only northern Sweden.) Why binary logistic regression if numbers on distance are available? Geo data possibly incorrect as only the location of headquarters is known.
6	Dargay, J. M., and Clark, S. (2012). The determinants of long distance travel in Great Britain. *Transportation Research Part A: Policy and Practice, 46*(3), 576-587.	Determine the effects of socio-economic, demographic and geographic factors on long distance travel. Model: $D_i = \sum_{k=1}^{k} \beta_k X_{ki} + \Upsilon t + \varepsilon_i$ Mean Distance traveled when over 50 miles as dependent	1)age 2)gender 3)employment status 4)household composition 5)household income 6)housing type 7)length of time of residence 8)size of municipality (population) 9)car ownership Source: National Travel Surveys. Area of coverage: Great Britain, 1995 – 2006 11 sets of cross-section Data	Income elasticity's fall between short and long run. Conclusions on commuting: Income elasticity of commuting travel all modes 0.57. Highest for rail commuting. Negative influence on commuting distance: women and 'complex household'(3+ adults). Positive: Employed, 1 adult, rural area. Urban areas travel less (London).	Mostly on long distance travel instead of commuting, but attention os paid to commuting on its own. Short/long run difference is explained and calculated (p583). All results per mode and travel distance on which this article is based are found in Dargay, J. (2010). http://www.theitc.org.uk/docs/7.pdf

Appendix A (continued)

#	APA Reference	Aims, Model and dependent variable	Explanatory Variables, DATA, geographical area of the research and time coverage	Results/Conclusions	Comments
7	Rouwendal, J., and Nijkamp, P. (2004). Living in Two Worlds: A Review of Home-to-Work Decisions. *Growth and Change, 35*(3), 287-303.	Discussing the various aspects of the economic analysis commuting behavior. No empirics. Monocentric model Value of time analysis	Possible drivers for commuting: 1)VOT (and wage rate) 2)housing prices 3)gender 4)density of job vacancies 5)real transport costs 6)population density 7)intensity of land use 8)distance from CBD 9)Income 10) history of residence (social cohesion within areas) 11)scheduling costs	Critical of the homocentric model and its unrealistic assumptions. VOT is mostly positive but sometimes even negative (people actually like commuting). Lot of research done and still lots to do.	No real strong conclusions. Article covers many areas of theory on commuting but not in depth. Somewhat disappointing. References to panel studies on the relation commuting and
8	McCann, P. (2001). *Urban and regional economics* (Vol. 15). Oxford: Oxford University Press, chapter 3	Von Thunen Model: monocentric, linear, land an non land input fixed, transport costs, homogenous land and produce Bid-rent model: monocentric, convex, land and non-land input substitutable, distance, homogenous land and produce	1)distance from market 2)land rent 3)transport costs 4)market price of products 5)non-land inputs	In the bid rent model, it is assumed that an individual tries to maximize his utility as he chooses between land and non-land inputs. Further from the CBD the land is cheaper but the transport costs higher. The individual maximizes this tradeoff.	First notion of 'commuting' in transport costs to the market place (products or people) and one of the first theories with attention to the special structure of a 'market place' (M) and the movement between them. Unrealistic assumptions in both models.
9	Gutiérrez-i-Puigarnau, E., van Ommeren, J.N. (2012). Do richer households live further away from their workplace? *Manuscript Title. Unpublished manuscript, Free University, Amsterdam, NL.*	To examine the long-run causal effect of income on the workers' commute. Static panel model which only includes people that stayed in de same workplace and moved during the survey period to control for reversed causality. Commuting Distance as dependent	1)gross household income 2)gender 3)number of children Source: German socio-economic panel Area of coverage: Germany, 1990 – 2010 Panel Data	Long run income elasticity of 0.15. Higher for 2 earner household than ones. Children have a negative effect. Likely underestimated because of amenity based choices of higher incomes.	Not published. Interesting approach and only empirical work that tries to test for (reversed) causality between income and commuting I know of. The limitations of panel data are discussed!
10	Simonsohn, U. (2006). New Yorkers commute more everywhere: contrast effects in the field. *Review of Economics and Statistics*, *88*(1), 1-9.	Research on whether a person's commuting decisions are influenced by experiences from the past. Regression model with commuting time as dependent variable.	1)previous commuting length (t-1 and t-2) 2)family income 3)number of children 4)age of head 5)gender Source: PSID and Census Data combined (1972 – 1986) Area of coverage: United States, Panel Data	People moving from one city to another within the US show a positive relation between previous commuting time in their former city and their new city. After a while they tend to adopt to the new commuting time standards.	The paper combines insights from psychology (background contrast effects) with economy. Interesting. Paper is careful with the interpretation of coefficients which is not always the case! (Reverse causation etc.)
11	Groot, S. P., de Groot, H. L., and Veneri, P. (2012). The Educational Bias in Commuting Patterns: Micro-Evidence for the Netherlands (No. 12-080/3). Tinbergen Institute.	To understand the role of education as a determinant of differences in travel behavior across individuals in the Netherlands. Regression model with fe and re. Commuting distance, duration and balance of commuters as dependent variables.	1)Education level 2)age 3)municipality 4)country of birth 5)gender 6)full/part time 7)industry 8)married 9)land rent 10)wage premium. home ownership, willingness to travel, transaction costs in residential mobility, search imperfections Source: EBB, CBS 2000–2008, The Netherlands 9*cross-sectional data	Higher education commute more (in time and distance) than lower educated do, even after controlling for their wage. Among lower educated workers the relation between wage and commuting distance is strong. For higher educated this relationship declines and for university graduates this relation seems to be non-existing. Higher educated use more public transport.	Good description of commuting patterns in the Netherlands (section 3). And in Europe OECD (2010) Attention to modes of transport as well. Nice combination of meso and micro data.
12	Deding, M., Filges, T., and Van Ommeren, J. (2008). SPATIAL MOBILITY AND COMMUTING: THE CASE OF TWO-EARNER HOUSEHOLDS*. *Journal of Regional Science, 49*(1), 113-147.	Examine the effect of the spatial configuration of workers' residence and workplace location on intraregional residential and job moving decisions of workers belonging to two-earner households. Search theoretical model Logit regression: job mobility and residential mobility as dependent	1)male commuting distance 2)female commuting distance 3)distance between workplace 4)education 5)residence size, duration 6)owned rented housing 7)age 8)children 9)sector 10)experience 11 Source: Statistics Denmark 1999 and 2000, age group 25-40 ,Denmark Cross-section of '99	Residential mobility is positively affected by distance of both spouses and negatively with distance between their workplaces. Job mobility depends positively on commuting distance, positively on the distance between workplaces and negatively on the spouse's commuting distance.	References to neoclassical view on commuting (p114). Short time period and Denmark as research country, which is an abnormality in commuting behavior compared to the rest of the European Union. Original approach! Only two earner households. Residential dummy's explanation.
13	Benito, A., and Oswald, A. J. (2000). Commuting in Great Britain in the 1990s.	Considers how the burden of commuting falls disproportionally on certain types of workers and the characteristics these individuals possess. OLS Regression model with log home to work time as dependent variable	1)wage 2)owned or rented 3)age 4)education 5)workplace size 6)job change in last year 7)part/full time 8)gender 9)sector Source: British Household Panel Survey 1991-1998. United Kingdom: Panel Data	Long commutes depend positively on education, homeownership and working in large plants/offices. Commuting distance depends negatively on the wage rate.(!)	Earning and commuting are negatively related whereas most papers find positive relations between wage rate (or income) and the commuting distance or time. Talks about the externalities of commuting as well, not in detail.

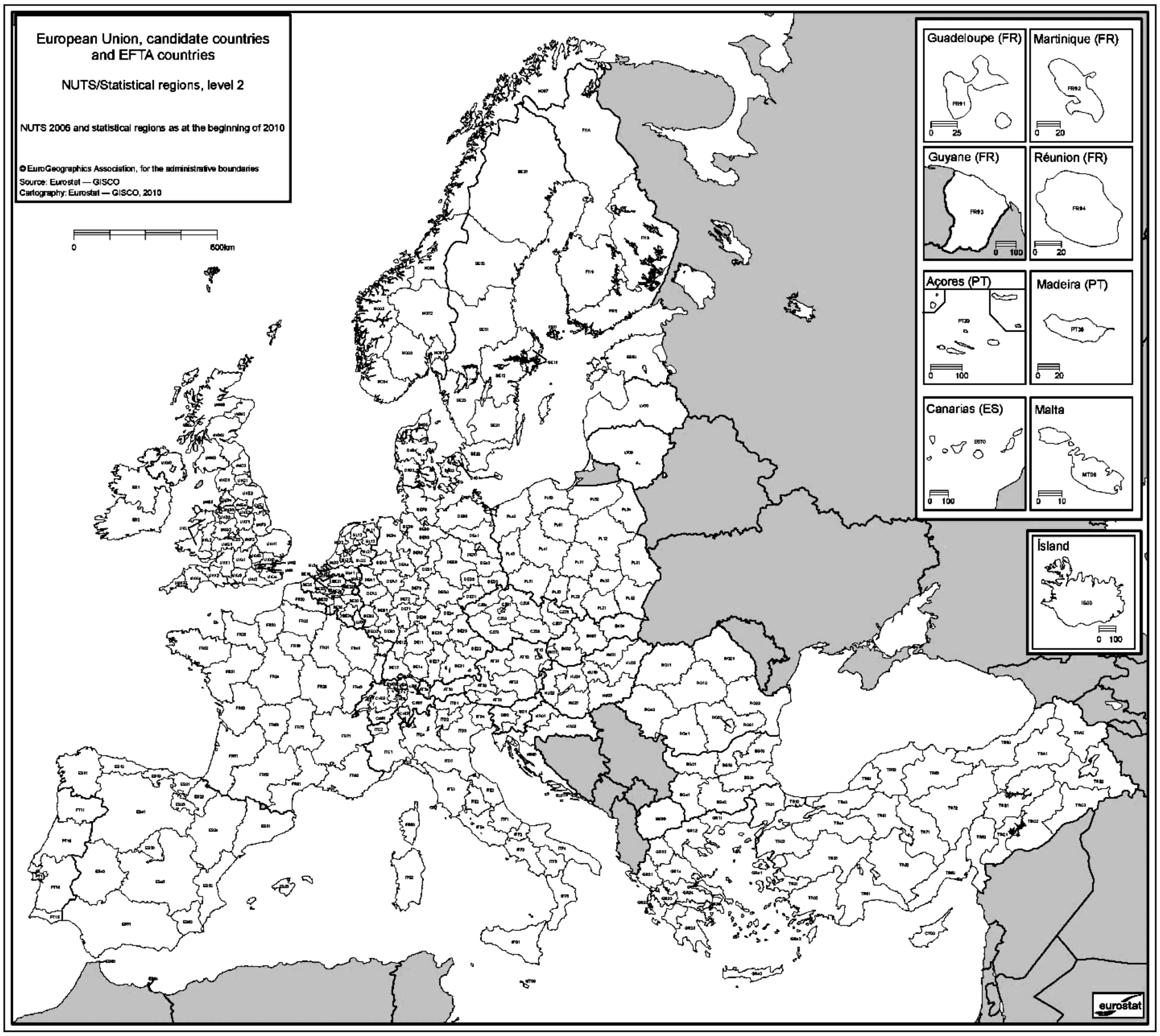
Appendix A (continued)

#	APA Reference	Aims, Model and dependent variable	Explanatory Variables, DATA, geographical area of the research and time coverage	Results/Conclusions	Comments
14	Schwanen, T., Dieleman, F. M., and Dijst, M. (2004). The impact of metropolitan structure on commute behavior in the Netherlands: a multilevel approach. *Growth and Change, 35*(3), 304-333.	Investigation of the impact of metropolitan structure on the commute behavior of urban residents in the Netherlands. Multilevel regression model with commute time and commute distance as independent	1)auto availability 2)personal income 3)education 4)age 5)gender 6)household type 7)pop. Density 8)residential density 9)employment density 10)municipality size 11)mono/polycentric- indicator (urban structure 12)job growth Source: Dutch national travel survey 1998, The Netherlands: Cross-sectional data	Most differences are explained at the individual worker level. Socioeconomic status (income, car ownership, etc.) and gender are important explanatory factors. Density and job growth lowers car commuting. Polycentrism does not seem to boost car usage in the Netherlands, contradictory to US research. This might be explained by regulated markets and compactness of the Netherlands compared to the US.	Macro/micro data references. Multilevel regressions are recent and referenced (p309). Only auto drivers are considered. Nice model and well explained. Only one year is covered.
15	Östh, J., and Lindgren, U. (2012). Do Changes In Gdp Influence Commuting Distances? A Study Of Swedish Commuting Patterns Between 1990 And 2006. *Tijdschrift voor economische en sociale geografie.*	Explore the long term relationships between changes in the economic cycle and the effects on individual commuting distances. RE GLS Regression models: all with annual change in commuting distance between t and t-1 as dependent.	1)GDP (and lagged 1 and 2 years) 2)urban/rural 3)labor market centrality 4)age 5)gender 6)poor/rich country of origin (race) 7)job type by sector 8)education 9)personal income group 10)unemployed Source: PLACE (micro) and Statistics Sweden (macro) 1990-2006, Sweden: Panel Data	There are significant effects of GDP changes on the commuting distance. Different variables determine the strength and direction of the change. Urban areas are more positively related then rural areas are. Non metropolitan areas increase commuting faster; Recently unemployed and young commuters are highly responsive to changes in GDP.	Many well-known variables explaining commuting are explained and referenced.
16	Glenn, P., Thorsen, I., and Ubøe, J. (2004). Wage payoffs and distance deterrence in the journey to work. *Transportation Research Part B: Methodological, 38*(9), 853-867.	Suggestion of a micro-economic model for how commuting flows relate to travelling distance in a two region system Theoretical	Variables in the model: Number of work centers, Number of jobs in a region, number of workers, Job types, distance deterrence function (with costs in terms of utility and distance embedded)	The game theoretical approach to commuting and commuting costs shows that there is a concave distance deterrence function (aversion to distance rises but with falling aversion towards distance)	Highly theoretical and simplified as with most theoretical models.
17	Van Ommeren, J., and Rietveld, P. (2007). Compensation for commuting in imperfect urban markets*. *Papers in Regional Science*, *86*(2), 241-259.	Development of a monocentric urban equilibrium job search model with imperfect labor market to predict compensation for commuting costs Theoretical	Commuting costs, residential location, wage rate, employed/unemployed, labor market power (bargaining), residential moving costs, job location, search costs	The model predicts: workers are partially compensated for their commuting costs; higher labor market power predicts higher wages but less compensation for commuting costs; rent gradients are less steep than expected; compensation derived from the labor/housing markets is less than 100%. Moving costs and labor market imperfections are key to understand the relation between urban labor/housing markets and the implications for commuting compensation.	Highly theoretical and simplified as with most theoretical models.
18	Kwon, Y. (2005). Urban comparative statics when commuting cost depends on income. *Journal of Housing Economics*, *14*(1), 48-56.	Investigation on the comparative static properties of an urban model where commuting cost is a function of income Model based on the standard urban model Theoretical	Utility function, budget constraint, commuting cost (operating and time costs), land market with distance from CBD.	Standard urban model commuting costs depends only on distance. In this model commuting costs per mile is an increasing function of income. Land rent at CBD rises as income grows if the time cost of commuting is greater than the operating costs.	Highly theoretical and simplified as with most theoretical models.
19	Johansson, B., Klaesson, J., and Olsson, M. (2002). Time distances and labor market integration. *Papers in Regional Science*, *81*(3), 305-327.	To illustrate the usefulness of the purpose-specific accessibility measure (with internal and external labor market characteristics). Regression analysis (static) with out commuting and in commuting (into a municipality) as dependent variable	Self-constructed measures of: 1)internal/external job accessibility 2)internal/external accessibility to realized labor supply Statics Sweden (macro) and Swedish National Road Administration (macro). Panel Data	The macro indicators of job accessibility and accessibility to realized labor supply are positive and significant. These measures can thus be used to predict in and outflow of commuters. Maximum commute time in Sweden is about 45 minutes.	Nice description of commuting. Creative approach to construct own accessibility measures. Only macro data used and highly significant!
20	Dargay, J., and Gately, D. (1999). Income's effect on car and vehicle ownership, worldwide: 1960–2015. *Transportation Research Part A: Policy and Practice,33*(2), 101-138.	Making projections of the growth in the car and total vehicle stock to the year 2015, for OECD countries and a number of developing countries. Gompertz function with car ownership as dependent variable	1)per capita GDP World motor vehicle data, Penn world tables, United Nations, International road federation. Panel Data	The largest growth in ownership will be in low income countries with high levels of income growth. Relative low growth will happen in rich OECD countries which are close to a level of 'car ownership saturation'. Strong historical relationship between income and ownership.	Lot of possible influential variables omitted. Examples of the authors: costs, demographics,, population density, road density

Appendix A (continued)

#	APA Reference	Aims, Model and dependent variable	Explanatory Variables, DATA, geographical area of the research and time coverage	Results/Conclusions	Comments
21	Pucher, J., and Renne, J. L. (2003). Socioeconomics of urban travel: evidence from the 2001 NHTS. *Transportation Quarterly, 57*(3), 49-77.	Look into the socioeconomics of urban travel behavior. No model, just data description	Variables discussed: age, gender, income, city size, ethnicity, mode of transport, region within US, peak/off-peak travel, household/car ratio National Household Travel survey 2001, US Cross-section	Private car use dominates urban travel. Work and work related trips rely for 92.1% on car use. Public transport trips declined in compared to 1995. As transit is also small among poor, funding public transport is not the main strategy. In cities with 3m+ inhabitants funding is more justified. Poor tend to walk most but walking is neglected. The NHTS reveals the importance for the first time. Ethnic minorities are correlated with poverty.	Good to see the differences between US/Europe in travel behavior. Not really specific about commuting! Comparison between Europe and US in reference 10 p75
22	Paumgarten, N.,(2007, April). There and back again, the soul of the commuter. *The New Yorker* (p?-?)	Description of real life stories of some extreme commuters with some scientific background information	No variables	One of the 'conclusions' the autor states: "People may endure miserable commutes out of an inability to weigh their general well-being against quantifiable material gains".	The history of the word commuting and a humoristic approach to extreme commuting
23	Van Acker, V., and Witlox, F. (2011). Commuting trips within tours: how is commuting related to land use?. Transportation, 38(3), 465-486.	Contribute to the existing research debate on the relationship between land use and commuting and differentiate between work only trips and more complex tours. Structural equation model (variable can act as dependent in one and independent in another equation)	1)job density 2)built up index 3)land use diversity index (interesting) 4)distance to nearest bus/train stop 5)job accessibility by car (15/30 min.) 6)parking difficulties 7)gender 8)age 9)marital status 10)car needed during work 11)household size 12)number of children <6 13)household income 14)number of cars per abled driver 15)income 16) full/part-time Ghent Travel Behavior Survey, 2000 and 2001, Ghent area, Belgium Panel data (only two years)	Land use aspects have a larger effect on commuting compared to more complex tours. Land use characteristics of the workplace have significant effects on commuting (distances and time) and are often neglected. Land use characteristics effect commuting but not as direct as of the thought because of intervening variables. Land use policy can reduce commuting times as long as the workplace characteristics are also accounted for, and if the land use effects on car use are. (public transport raises time, not distance!)	Many land use variables in the lit. review with good references (entropy index, mixed land use, commercial/residential land use ratio, etc. p467-469). Original self-constructed variables that proof to be significant as well
24	Ben-David, N., Sharabi, M. (2009). Commuting and Its Effect on Work Decisions. *International Journal of Economic Perspectives, 3*(3), 183-187	Investigate the effect of commuting on reservation wage and hours worked, booth theoretical and empirical. Difference between predicted and actual working hours as independent variable	1)difference actual hourly wage and predicted hourly wage 2)predicted hourly wage*hours worked (income effect) 3) commute time needed to drive to and from work Sample of 680 Israeli > 17 years, 2006, Israel. Cross-section data	Substitution effect of wage is positively related to hours worked, the income effect of hours worked is also positively related to hours worked. Commuting time has a negative effect on hours worked.	Use of a model to predict variables that ere used in a model to predict working hours. Not really a model which uses real world data. For this practice 680 agents is a bit small. Might be reversed causality. Not very convincing research.
25	Susilo, Y. O., and Maat, K. (2007). The influence of built environment to the trends in commuting journeys in the Netherlands. *Transportation, 34*(5), 589-609.	Identify trends in commuting journeys in the Netherlands in the last decade (1995-2005) and examine the influence of urban form and travel accessibility on commuting journeys over time. Binominal logit regression with workplace in/outside home municipality as dependent Multinomial logit regression with preferred commuting mode as dependent Regression model with commuting time as dependent variable	1)gender 2)age 3)children<12 4)#household members 5)income 6)(high) education 7)car availability 8)# inhabitants per municipality 9)urbanization degree 10)located/not located in RMA (randstad) 11)network density 12)distance from train/metro/motorway 13)mode of transport Dutch National Travel Survey Three sets of cross-sectional data, 1995, 2000, 2005, the Netherlands	Somewhat sad conclusion for urban policy makers: There is no single conclusion to be drawn about the influence of urban form variables on commuting. Some have become less/more significant over time (1995, 2000, 2005) and some have even changed sign in the same regressions in different years. Strong clues that socio-economic or socio demographic variables have more influence on commuting decisions or patterns than urban form/land use does.	Nice introduction and history (of research). Socioeconomic characteristics have more effect on commuting then urban structure does. Us/Europe comparison (p591). Pro macro data p599. Binominal logit model could be interesting for Eurostat commuting Data! Nice article, useful references
26	Stutzer, A., and Frey, B. S. (2008). Stress that Doesn't Pay: The Commuting Paradox*. *The Scandinavian Journal of Economics, 110*(2), 339-366.	Study whether commuters are compensated for their commuting time, as predicted by rational choice urban location theory. Different regression models; all with self-reported well-being as dependent variable.	1)commuting time 2)commuting time squared 3)commuting distance 4)change of residence 5)change of job German Socio-economic Panel Study, 1985 – 2003, Germany, Panel Data	In all variants of the models there is a significant and large negative effect of commuting time on self-reported well-being. This cannot be explained by the wrong unit of measurement (individuals vs. households) and can possibly explained by the inability to weight the full cost of commuting. A rational choice explanation is thus not yet available. Behavioral explanations could possibly explain the paradox trough a lack of will power or loss aversion.	Other fields discussed (mainly psychology) and the classical urban location theory is explained well and referenced. References to data are mentioned.

Appendix B



Source: Eurostat http://epp.eurostat.ec.europa.eu/cache/GISCO/yearbook2009/RYB-Full-NUTS2-2009-EN.pdf, retrieved august 2013

Appendix C

Table C

Eurostat Variable Description

Variable	Name	Unit	Eurostat Dataset Code
Number of commuters	commute	# of people working outside their 'home' NUTS region	[lfst_r_lfe2ecomm]
Primary Income	primary	Average euro per capita per year	[nama_r_ehh2inc]
Economical active population	active	# of people marked as economically active in a NUTS region	[lfst_r_lfp2act]
Population density	popdens	Population density per km²	[demo_r_d3dens]
Finished tertiary education	Eduhifi	# of people with a finished university or higher education	[educ_renrlrg1]
Children under 5 years	child5	# of children under 5 years of age in the NUTS region	[demo_r_pjangroup]
Children under 10 years	child10	# of children under 10 years of age in the NUTS region	[demo_r_pjangroup]
Children under 15 years	child15	# of children under 15 years of age in the NUTS region	[demo_r_pjangroup]
Motorisation rate	cars	# of cars per 1000 inhabitants	[tran_r_vehst]
People over 65 years	Over65	# of people over 65 years of age	[demo_r_d2jan]
Petrol price	petrol	Average yearly petrol price per country in euros	Not applicable
Land use diversity	LUbasic and LUextended	Between 0 and 1. Explained in 3.2 and appendix G	Based on [lan_lu_ovw], [lan_lu_agr] and [lan_lu_inf]

Appendix D

Table D

Four Types of Unit Root Tests

Variable	***P***	***Z***	***L****	***Pm***
Lncommute (1 lag)	**1073.29**	**-10.77**	**-14.17**	**18.73**
Lncommute (2 lags)	**898.84**	**-4.86**	**-9.76**	**14.96**
Lnprimary (1 lag)	307.37	7.07	6.82	-6.98
Lnprimary (2 lags)	209.63	14.61	14.87	-9.93
Lnactive (1 lag)	**1234.68**	**-8.92**	**-12.36**	**19.41**
Lnactive (2 lag)	**1003.65**	**-2.24**	**-5.96**	**14.02**
Lnpopdens (1 lag)	322.37	16.34	16.93	-8.5
Lnpopdens (2 lags)	357.86	19.31	18.89	-7.49
Lneduhifi (1 lag)	**964.76**	**-8.14**	**-9.58**	**10.76**
Lneduhifi (2 lags)	**972.74**	**-4.4**	**-7.42**	**11.37**
lnchild5 (1 lag)	376.36	7.65	7.68	-6.96
lnchild5 (2 lags)	387.67	6.32	6.22	-6.64
lnchild10 (1 lag)	331.92	9.2	9.19	-8.22
Lnchild10 (2 lags)	326.82	8.36	8.13	-8.37
Lncars (1 lag)	422.87	7.13	5.4	-3.99
Lncars (2 lags)	**624.25**	7.69	1.99	**2.37**
lnover65 (1 lag)	528.29	4.8	4.48	-2.66
Lnover65 (2 lags)	476.61	6.48	6.49	-4.12

Notes: **bold numbers** are significant at the $p<0.05$ level; ***P*** = Inverse chi-squared test; ***Z***= Inverse normal; ***L**** = Inverse logit t test; ***Pm*** = Modified inverse chi-squared test.

Appendix E

Table E.1

Akaike Information Criteria Tests for Lags

	ECM (1.1)	**ECM (1.2)**	**ECM (2.1)**	**ECM (2.2)**
AIC statistic	-1517.98	-1531.29	-1561.72	-1565.49

Table E.2

Test for fixed effects vs. Random Effects

Sargan-Hansen statistic	-660.975 (0.000)

Table E.3

Correlation Coefficients of Used Variables

	primary	**active**	**popdens**	**child5**	**child59**	**eduhifi**	**cars**	**over65**	**petrol**
primary	1.00								
active	0.02	1.00							
popdens	0.15	0.12	1.00						
child5	0.02	0.96	0.14	1.00					
Child10	0.01	0.96	0.11	0.99	1.00				
eduhifi	0.16	0.89	0.23	0.83	0.80	1.00			
cars	0.09	0.93	0.07	0.79	0.77	0.79	1.00		
over65	0.05	0.95	0.06	0.80	0.80	0.78	0.96	1.00	
petrol	0.30	0.04	0.05	0.04	0.02	0.11	0.08	0.06	1.00

Appendix F

Table F

Descriptive Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
commute	2,982	60,220.00	87,330.00	0.00	1,076,300.00
primary	2,848	15,855.59	10,284.94	716.40	147,844.80
active	3,764	840,070.00	715,690.00	12,100.00	5,712,900.00
popdens	3,768	341.63	852.56	1.90	9,673.70
child5	4,080	98,397.82	93,934.82	1,361.00	1,084,589.00
child10	4,080	199,338.80	186,132.70	2,909.00	2,118,157.00
eduhifi	3,537	423,722.00	435,965.10	5,732.44	4,932,732.00
Cars	2,949	431.36	140.14	14	1,101
over65	3,810	295,259.80	252,109.80	3,397.00	2,042,476.00
petrol	3,186	1.16	0.20	0.65	1.67
lusimple	3,360	0.44	0.10	0.12	0.76
lumax	3,360	0.35	0.08	0.21	0.62

Notes: popdens in people per km^2; cars per 1000 inhanitants; petrol in euro per liter

Appendix G

Calculation of the 'extended' land use diversity index is done along the same principles on which the 'basic' index is calculated. The categories agriculture and service and residential are split up further. Agricultural use is split up in three categories: agriculture (excluding fallow land, kitchen garden and personal consumption areas), fallow land and abandoned land and kitchen garden. The services and residential land use is spit up in five categories: commerce, finance, business; community services; recreation, leisure and sport; residential and a category for nature reserves. The 'extended equation is comparable in composition to (3.1) but contains more categories.

(G.1)

$$extended = 1 - \left[\frac{\left|\frac{a}{T}-\frac{1}{12}\right| + \left|\frac{af}{T}-\frac{1}{12}\right| + \left|\frac{ak}{T}-\frac{1}{12}\right| + \left|\frac{f}{T}-\frac{1}{12}\right| + \left|\frac{hu}{T}-\frac{1}{12}\right| + \left|\frac{he}{T}-\frac{1}{12}\right| + \left|\frac{\mathrm{c}}{T}-\frac{1}{12}\right| + \left|\frac{\mathrm{cs}}{T}-\frac{1}{12}\right| + \left|\frac{\mathrm{rl}}{T}-\frac{1}{12}\right| + \left|\frac{\mathrm{r}}{T}-\frac{1}{12}\right| + \left|\frac{n}{T}-\frac{1}{12}\right|}{\frac{20}{11}}\right]$$

In (D.1) *T* stands for total land area, *a* stands for agriculture (general), *af* for fallow or abandoned la*nd, ak* for kitchen garden, *f* for forestry, *hu* for Hunting and Fishing, *he* for Heavy and Environmental. The category services and residential is spit up as well where *c* stands for commerce, finance and business; *cs* for community services; *rl* for recreation, leisure and sport; *r* for residential and *n* for nature reserves.

Appendix H

Table H.1

Variance Inflation Facors

Variable	VIF
Lnincome	1.52
Lndensity	1.25
Lncars	1.14
Lnpetrol	1.53
Over65 (dummy)	1.03
Child (dummy)	1.01
Eduhifi (dummy)	1.10

Figure H.1

P-P plot of Regression Residuals