Does improving Public Transport decrease Car Ownership?
Evidence from the Copenhagen Metropolitan Area

Ismir Mulalic[1]
Ninette Pilegaard[1]
Jan Rouwendal[2]

[1] Technical University Denmark, Denmark;
[2] Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.

Does improving public transport decrease car ownership?

Evidence from the Copenhagen metropolitan area

Ismir Mulalic[1], Ninette Pilegaard[1] and Jan Rouwendal[2]

Abstract

Car ownership is lower in urban areas, which is probably related to the availability of better public transport. Better public transport thus may offer the possibility to relieve the many problems (congestion, health, and parking) associated with the presence of cars in urban areas. To investigate this issue, we develop and estimate a model for the simultaneous choice of a residential area and car ownership. The model is estimated on Danish register data for single-earner and dual-earners households in the greater Copenhagen metropolitan area. We pay special attention to accessibility of the metro network which offers particularly high quality public transport. Simulations based on the estimated model show that for the greater Copenhagen area a planned extension of the metro network decreases car ownership by 2-3%. Our results suggest also a substantial increase in the interest for living in areas close to the metro network, that affects the demographic composition of neighbourhoods.

Keywords: car ownership, public transport, residential sorting.

JEL codes: R4, R1, D1.

[1]Department of Transport, Technical University of Denmark, Denmark
[2] Department of Spatial Economics, VU University, and Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands

1. Introduction

Public transport is, potentially, an important substitute for the car and recent research suggests that its presence may have an important impact on urban congestion (Anderson, 2014). Good public transport may make car ownership less attractive. If parking spaces are difficult to find, or if parking is expensive, as is the case in many city centers the benefits of owning a car may still be lower (Van Ommeren et al., 2011). Moreover, the availability of many amenities at walking distance in residential areas decreases the value of owning a car even further. It is therefore no surprise that the share of car-owners is lower in urban than in rural areas (see, for instance, Dargay (2002) and Pyddoke and Creutzer (2014)).

In this paper we study car ownership in relation to the availability of public transport. We look at the choice of car ownership and the residential location as a simultaneous decision, taking into account that households may want to live in a particular location especially because of the availability of public transport. The interaction between car ownership and public transport has been addressed in an older literature (see for instance Goodwin (1993)), but appears to have been neglected in recent decades. It is, nevertheless, of considerable interest because road congestion is still an important problem in many urban areas, pollution by cars is associated with health problems and global warming is perhaps the most important environmental problem of our age. Cities can be relatively green places (see e.g. Kahn (2006)) and the lower share of car owners contributes to that.

Despite the – at least potential – importance of the relationship between cities and car ownership, the topic has received little attention in economics. There is an older literature in economics looking at car ownership (see, for instance, Mannering and Winston (1985)) that pays marginal attention to it. For instance De Jong (1998) develops a binomial model in which car ownership and use are modelled simultaneously and reports that living in a rural area increases the probability of owning a car.

There exists a small geographic literature on the impact of urban form and urban amenities on car ownership. See for instance Dieleman et al. (2002) or Potoglou and Kanaroglou (2006). In this literature car ownership is usually estimated as a binomial choice, conditional on the characteristics of the residential area. For instance, Potoglou and Kanaroglou (2006) find that mixed land use is associated with a lower share of car owners. It is perhaps more surprising that

1

even in transportation the impact of urban form and urban amenities on car ownership decision does not seem to be an intensively studied topic. Matas et al. (2009) is an exception.

In this paper we develop a simultaneous structural model for residential location and car ownership. That is, we assume that households looking for a residential location contemplate to live in a particular area while owning a car or not. Our model extends a logit-based 'horizontal' residential equilibrium sorting model (see Kuminoff et al. (2013)) with car ownership. The methodology employed in this type of model was developed by Berry (1994) and Berry et al. (1995) who studied the market for new cars. Bayer et al. (2007) pioneered the application of this approach to housing market analysis.

The choice alternatives we consider in our model are combinations of residential areas and car ownership. Interactions between characteristics of the residential areas and car ownership are the focus of interest. The residential area characteristics include public transport related as well as more traditional urban amenities. Our model can alternatively be viewed as one explaining car ownership while paying special attention to its relationship with residential area characteristics (urban amenities).

We use the estimated version of the model to simulate the impact of an extension of the Copenhagen metro network that is currently under construction. The model predicts house prices, demographic composition of neighbourhoods and car ownership in this counterfactual situation. We also compute the impact on welfare of this improvement in public transport.

The paper is organized as follows: in the next section we briefly describe the most relevant characteristics of the data and the study area (the Great Copenhagen Area (GCA)). In section 3 we present and discuss the theoretical model and the specification we use in our empirical work. In section 4 we discuss household heterogeneity and urban amenities. Section 5 reports the estimation results and presents some robustness checks. In Section 6 we use the estimated model to simulate the response of households to a metro extension in the city of Copenhagen. Finally, section 7 concludes.

2. Data and descriptives

2.1. The greater Copenhagen area (GCA)

The Greater Copenhagen Area (GCA) is part of the Danish island Zealand (see map 1). Copenhagen (the capital city of Denmark) is its centre. The GCA is the political, administrative, and educational core region of Denmark and accounts for more than 40 % of Denmark's GDP, 1.6 million people (app. one third of Danish population), and 1 million workplaces (app. one half of workplaces). The GCA is divided into 166 areas, which are designed for the purpose of detailed traffic modelling. The geographical area of GCA is rather small (615.7 km^2).[1]

Map 1: The Greater Copenhagen Area (GCA)

---
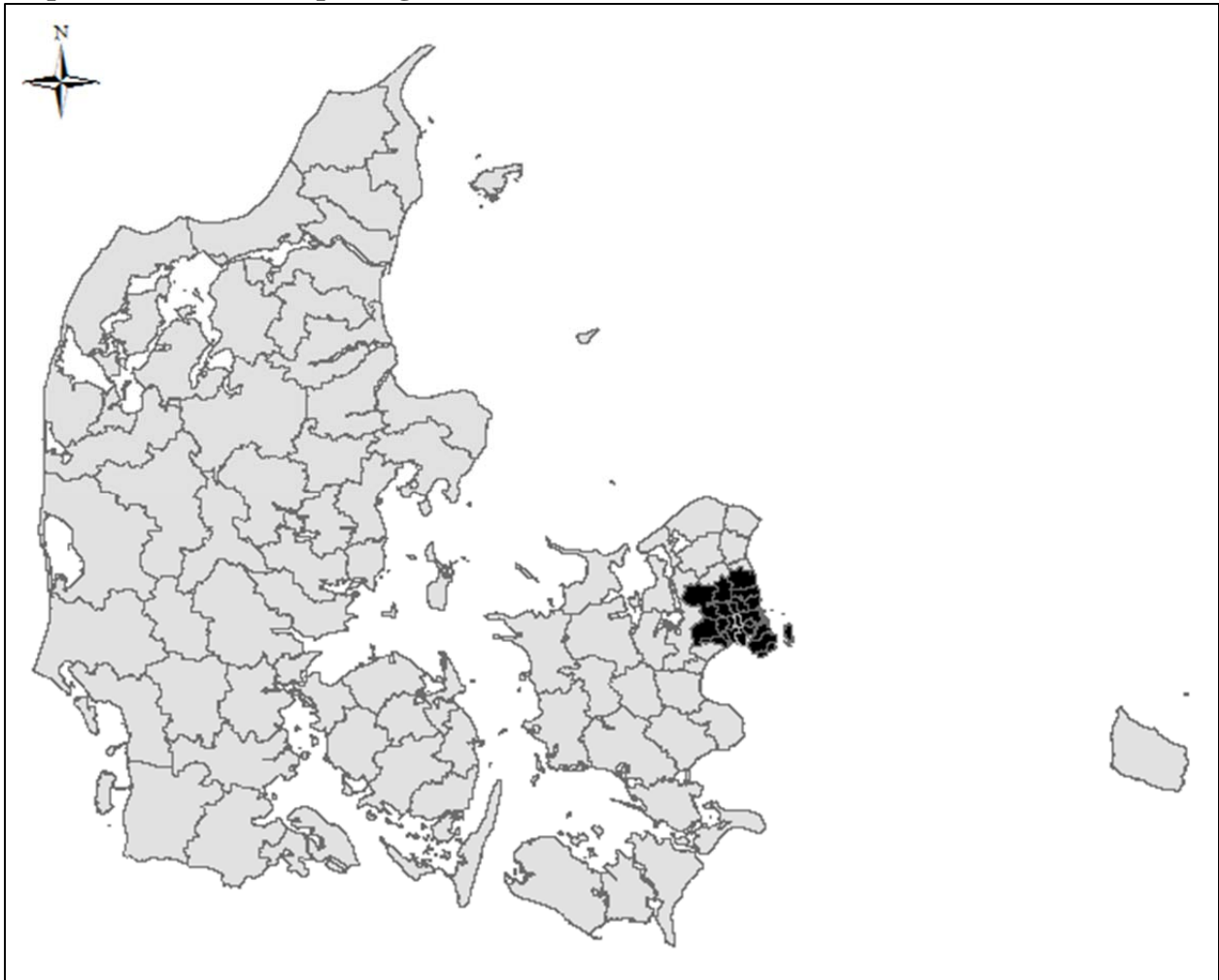
[1] The average mean travel time with car within areas in the GCA is about 17 min., and the maximum is less than 1 hour (51.8 min.). The average mean travel time with public transport is about 48 min., and the maximum is almost 2 hours (112.7 min.).

It is a fair simplification to claim that the GCA constitutes a single spatial labour market. This implies that the estimated effects on location choices from our model are not disturbed by labour market effects. Commuting from GCA to other parts of Zealand is negligible, whereas commuting flows inside GCA are relatively large. There is a tendency to have most commuting towards the centre of Copenhagen although the flowsare not just one-way. This suggests that workers in the GCA consider the whole area when looking for a job and that wage differences within the GCA can be ignored.[2]

In our model we consider the household location decision to be related to the decision of car ownership. This is especially relevant in Danish context. Car ownership in Denmark is extremely expensive compared to international standards due to taxation. The purchase-tax of a

Map 2: Car ownership in the GCA (number of cars per household)

_____
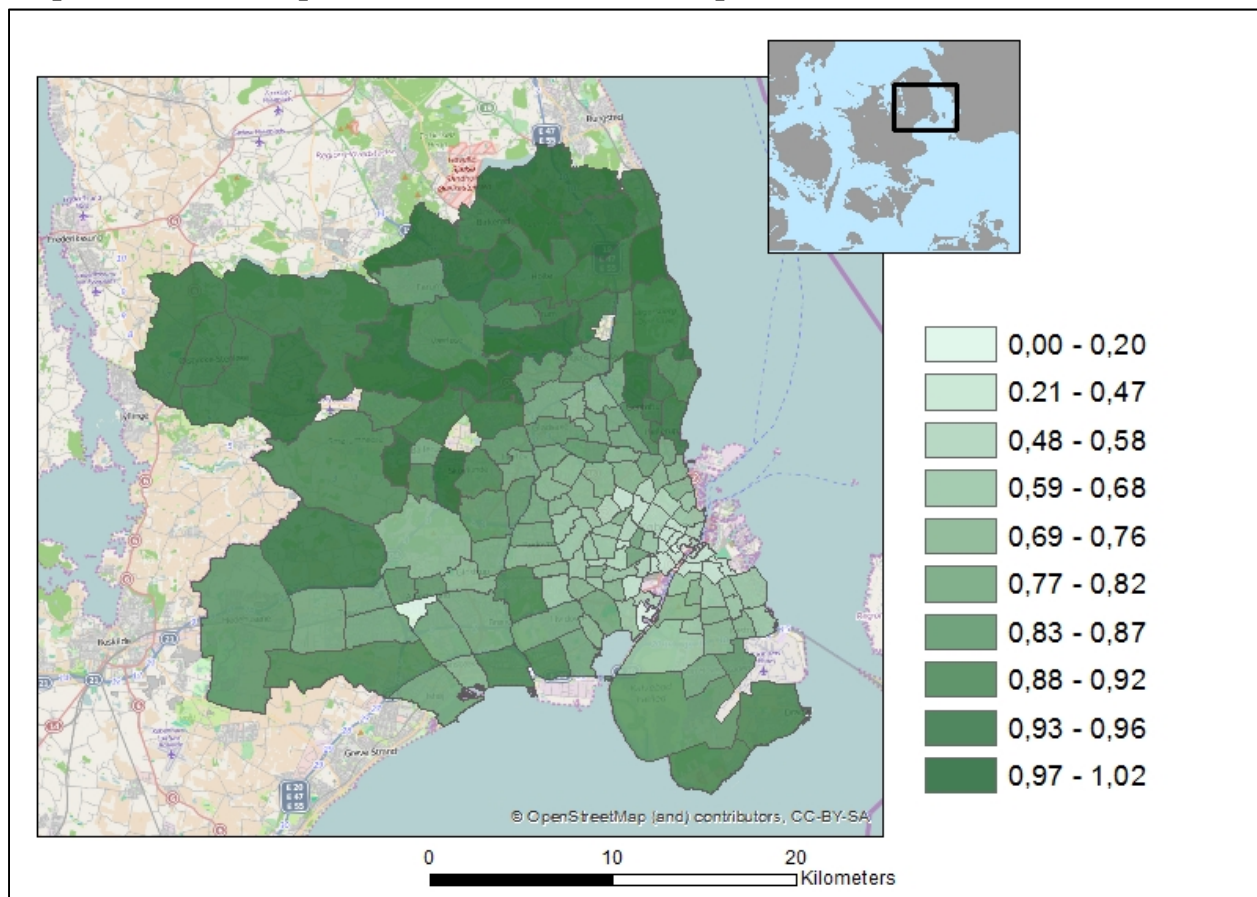
[2] In all probability the same is true for the prices of consumer goods (except housing).

car is 105% for the value of the car below app. €10.500 and 180% of the value of the car above. In addition there is an annual ownership tax of app. € 500 (300-900) depending on the characteristics of the car. Consequently, car ownership is relatively low in Denmark relative to other comparable countries (0.81 cars per household in Denmark, 0.71 cars per household in GCA). For many low income households car ownership is hardly affordable and even many medium income households choose not to own a car. The number of households with two cars is also quite low (8.2% of households in Denmark). The alternative travel mode to car is of course public transport but a bike is also a common mode of transport, especially in Greater Copenhagen and other bigger cities, and among younger people.

The high cost of car ownership also implies that car ownership is often reconsidered when households change residence. Households thus experience an active trade-off between car ownership and housing expenditures. Many young Danish families, even with relatively high incomes and high income expectations for the future, choose to prioritise housing over car ownership in their first years as house-owners and then use future income increases to buy a car in later years. This has an impact on their location choices as not owning a car typically implies additional considerations about accessibility to the labour market, shopping and other urban amenities by other modes of transport.

2.2. Selection of sample

The equilibrium sorting model is estimated on data derived from administrative registers for all Danish households with residence in the GCA for the year 2008. We use a 20% sample of the GCA population living in *owner-occupied housing*. The model focuses strictly on the location choices of households active on the labour and housing market. We only consider owner-occupiers. Our model can be considered as part of a broader nested logit model in which the housing tenure choice is on the top of the utility tree and the choice of the combination of housing type (apartment or other) and the geographical area refers to the lower level.[3] The share

---

[3] The market for rented housing in Denmark is strictly regulated in many ways. Because of the regulation, the price of a rented residence does not equilibrate demand and supply. Hence, the price of the rented residences does not reveal marginal willingness to pay. Only in the market for owned residences households have a free choice, given their budget constraint, to choose residence with respect to e.g. type and location.

of owner-occupied residences constitutes just over 50% of the housing stock (see Figure 1). The price in this segment represents a market equilibrium that is conditional upon the state of the rental market.

The households in our sample are distributed over 166 areas in which they can choose to live either in a multi-family house (apartment or flat) or a single house (covering detached villas and terraced houses).

We assume that the supply of owner-occupied residences is fixed. This implies that the housing stock does not react to the housing price. This is not very restrictive because changing the housing stock takes time, implying that it cannot react immediately to changing market circumstances. Moreover, in the CGA area year-to-year changes in the housing stock are small. This is particularly true for owner-occupied residences in the GCA, see Figure 1.

Figure 1: Occupied dwellings by tenure
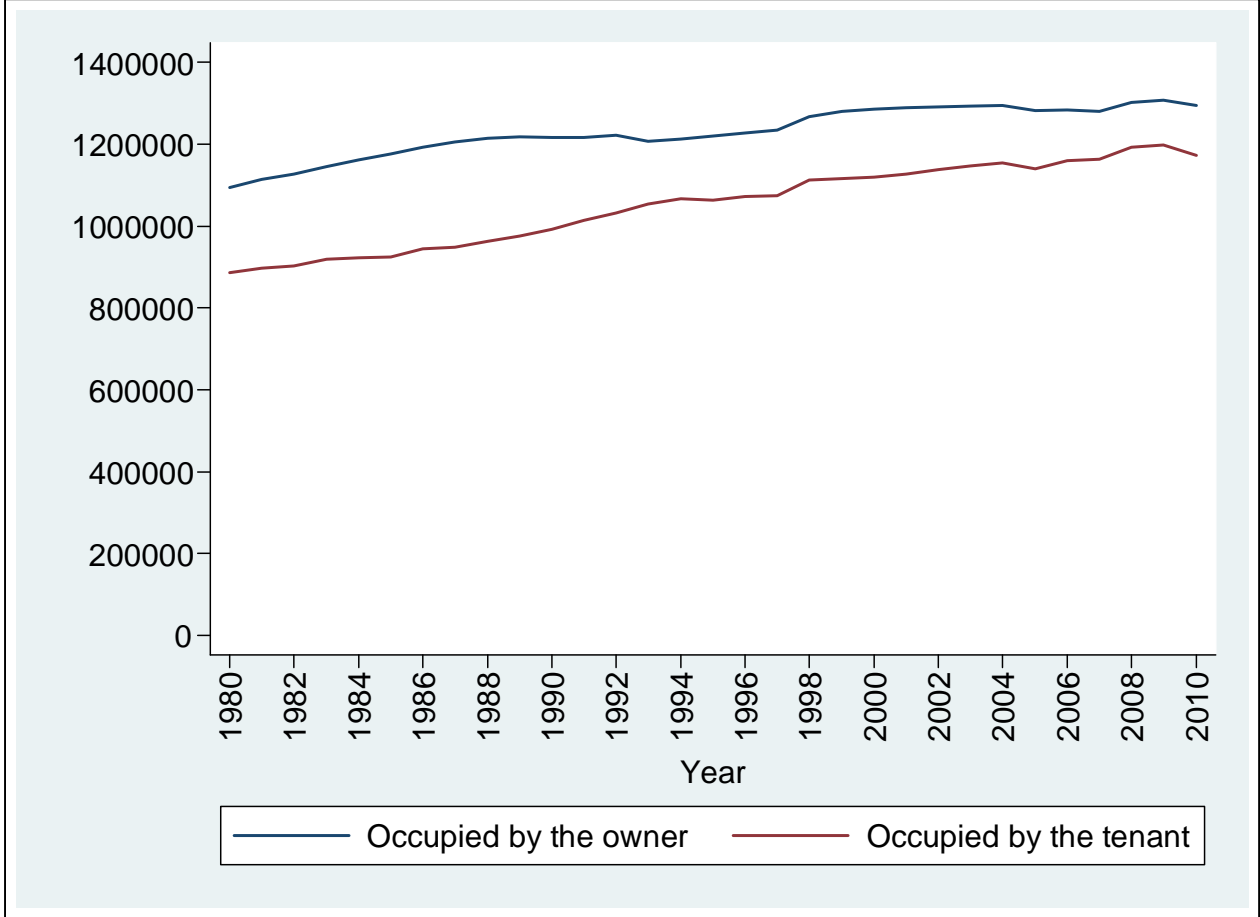
The choice to focus on the location choices of households active on the labour and housing market also implies, that we exclude households where both (or the one if single) members are either student, unemployed, retired or otherwise inactive on the labour market (23.8%).[4] There are two reasons for this choice. First of all, we want to examine how households value and potentially trade-off urban amenities and labour market accessibility when choosing a residential location. This is only relevant for households active in the labour market. Second, households where both (or one if single) adults are unemployed will rarely be active on the housing market – even if this was to improve the accessibility to jobs. However, when at least one of the two persons in a retired couple is still active, they are included in our sample (59.2%). We also include households in which one household member is employed and the other is studying. In summary, we think that excluding households without a worker from the estimated model allows us to focus on the most relevant group of household and does not imply substantial biases.

We distinguish between single earner households (66,012) and dual earners households (87,330) and estimate separate models for these two groups. The reason for this is that these household types are quite different in many respects. In Denmark the norm for families is that both adults in the household are active in the labour market. Families where one of the adults is inactive on the labour market for a longer period of time by choice are uncommon.

In some areas we observe only single family housing types (mainly at the outskirts of the study area) and in some areas only multi-family housing (apartments, in the centre of the GCA). There are also areas in which we do not observe single family households with a car, or without a car. Similarly not all areas have two-earner households without a car, with one car and with two cars. In case we have no observations of a particular choice alternative we assumed it was not in the choice set of the relevant household type. We also model car ownership, i.e. car ownership for the single earner households and car 1-2 for the dual earners households. The total choice set includes 538 and 636 elements for single earner households and dual earners households, respectively.

---

[4] The majority of these inactive households are pensioners (89.93%).

3. The model

This section presents the theoretical model that underlies the empirical analyses. We introduce the model for single earner households, but most of the analysis remains unchanged when the model is extended to dual earners households. The model for single earner households can be extended to model for dual earners household by including the choice of 2 cars.

3.1 A discrete choice model and its implications for car ownership

The model we estimate in this paper considers car ownership and residential location as a joint decision. Households choosing a residential area know about the availability of public transport in that area, about the parking possibilities and the presence of other amenities. These characteristics of the area determine the value of having car and it is plausible that the decision to own a car is closely related to the choice of the residential location.

Following this reasoning we develop a discrete choice model in which combinations of car ownership and residential areas are the choice alternatives. A household thus considers living in a residential area with and without having a car and chooses the alternative that offers the highest utility.

We consider households who derive utility from housing, owning a car, local amenities and a composite that represents all other consumption goods. Car ownership is included as a simple indicator that takes on the dichotomous values of 0 and 1. This implies that we do not distinguish between car brands, new or second hand cars or any other car characteristics. We thus ignore the heterogeneity of cars in the interest of focusing on the interaction between the availability of public transport and car ownership.[5]

We assume that housing services are available at a given price per unit that is specific for the residential area. The number of units consumed is determined by choosing from the stock or adjusting an existing house. This approach follows Muth (1969) and was further developed by Rouwendal (1998) and Epple and Platt (1998) who used it to study location

---

[5] We also ignore car sharing and carpooling. Moreover, we do not model car usage.

choices within urban areas. It is very convenient since it allows the researchers to abstract from heterogeneity in the housing stock.[6]

However, the neglect of the durable aspects of housing may be problematic if quality differences are substantial. In particular the distinction between single and multifamily housing seems to be a fundamental one. We have therefore decided to distinguish between these two types of houses, while maintaining the "Muth-framework" for the stocks of these two types of housing.[7]

This means that for each residential area in the GCA there are in principle four choice alternatives in our model: single and multifamily housing, both with and without having a car. Choice alternatives are therefore defined by three dimensions: area ($a = 1 \dots n$), house type ($h = s, m$), and car ownership ($c = 0,1$) and we denote the utility of a choice alternative for household i as $u_{a,h,c}^i$. We specify the utilities associated with each choice alternative as the sum of a deterministic and a random term (McFadden (1973)):

$$u_{a,h,c}^i = v_{a,h,c}^i + \varepsilon_{a,h,c}^i \tag{1}$$

and assume that the $\varepsilon_{a,h,c}^i$'s are multivariate extreme value (MEV) distributed.[8] A MEV distribution is characterized by a generator function $G\left(e^{v_{a,h,c}^i}\right)$ where $e^{v_{a,h,c}^i}$ is the vector of the exponentiated deterministic parts of the utilities. Choice probabilities can be written as:

$$\pi_{a,h,c}^i = \frac{e^{v_{a,h,c}^i} * G_{a,h,c}\left(e^{v^i}\right)}{G\left(e^{v_{a,h,c}^i}\right)} \tag{2}$$

where $G_{a,h,c}$ denotes the first derivative of G with respect to the argument that corresponds to choice alternative $\{a, h, c\}$. In this paper we consider only the special case in which $G\left(e^{v_{a,h,c}^i}\right) =$

---

[6] Moreover, it overcomes a problem associated with treating individual houses as choice alternatives, viz. that not every household can afford to live in every house. We assume here that every household can find affordable (single- or multi-family) housing in every area. This allows for the possibility that a (large) part of the housing stock that is available in an area may not be affordable for specific households.

[7] A single family housing represent a house or a villa, typically with private garden, while a multifamily housing typically represent a flat in an apartment building with no or shared outdoor facilities.

[8] Often also referred to as GEV distributed.

$\sum_a \sum_h \sum_c e^{v_{a,h,c}^i}$, which implies that the choice probabilities are given by the multinomial logit model (MNL).[9]

We are interested in what the model can tell us about the decision to own a car and the impact public transport has on that decision. The consumer will own a car if the maximum utility of the alternatives in which a car is owned exceeds the maximum utility of the alternatives in which no car is owned. The former maximum utility, which we denote as U_1^i $(= max\{u_{a,h,c}^i | c = 1\})$ is:

$$U_1^i = \ln\left(\sum_a \sum_h e^{v_{a,h,1}^i}\right) + \varepsilon_{c=1}^i. \tag{3}$$

For the utility U_0^i of not having a car we have:

$$U_0^i = \ln\left(\sum_a \sum_h e^{v_{a,h,0}^i}\right) + \varepsilon_{c=0}^i. \tag{4}$$

The first terms on the right-hand side of (3) and (4) are known as the logsums. The random terms $\varepsilon_{c=1}^i$ and $\varepsilon_{c=0}^i$ are independent and Extreme Value Type I distributed. The choice whether or not to own a car can therefore be described as a binomial logit model in which the logsums are the deterministic parts of the utilities. Denoting the probability of car ownership as $\pi_{c=1}^i$ we thus have:

$$\pi_{c=1}^i = \frac{e^{\ln\left(\sum_a \sum_h e^{v_{a,h,1}^i}\right)}}{e^{\ln\left(\sum_a \sum_h e^{v_{a,h,1}^i}\right)} + e^{\ln\left(\sum_a \sum_h e^{v_{a,h,0}^i}\right)}} \tag{5}$$

It should be noted that this model differs from one in which we estimate car ownership conditional on the choice of a residential area and housing type, as is the case in the literature on the impact of urban form on car ownership cited in the introduction. In that literature binomial models of the type:

$$\pi_{c=1|a,h}^i = \frac{e^{v_{a,h,1}^i}}{e^{v_{a,h,1}^i} + e^{v_{a,h,0}^i}} \tag{6}$$

---

[9] That is: $\pi_{a,h,c}^i = \dfrac{e^{v_{a,h,c}^i}}{\sum_{a\prime} \sum_{h\prime} \sum_{c\prime} e^{v_{a\prime,h\prime,c\prime}^i}}$.

are estimated. Model (5) allows the consumer to choose a different neighbourhood and housing type depending on whether a car will be owned, whereas (6) compares the utility a household would be able to reach with and without owning a car in a given neighbourhood.

Our welfare analysis follows De Palma and Kilani (2003). The average value of car ownership can be measured as the compensating differential of the utility that can be reached with and without a car. To discuss how it can be computed we observe that utility depends on household income y^i. We write $v_{a,h,c}^i = v_{a,h,c}^i(y^i)$ and define the conditional compensating variation Δy_{ccv}^i as the change in income that makes consumer i indifferent between having and not having a car conditional when residential area and housing type remain unchanged. Δy_{ccv}^i is defined implicitly by the equation:

$$v_{a,h,1}^i(y^i - \Delta y_{ccv}^i) - v_{a,h,0}^i(y^i) = 0 \ . \tag{7}$$

The conditional compensating variation ignores the possibility that a consumer may decide to live in a different residential area or housing type, depending on car ownership. This *is* taken into account by model (5) and we define the unconditional compensating variation Δy_{ucv}^i on the basis of that model analogously as:

$$\ln\left(\sum_a \sum_h e^{v_{a,h,1}^i(y^i - \Delta y_{ucv}^i)}\right) = \ln\left(\sum_a \sum_h e^{v_{a,h,0}^i(y^i)}\right) \tag{8}$$

Equation (8) can be rewritten as:

$$\sum_a \sum_h \left(e^{v_{a,h,1}^i(y^i - \Delta y_{ucv}^i)} - e^{v_{a,h,0}^i(y^i)}\right) = 0$$

To analyse the impact of public transport, we must specify how it enters the utility of the consumers. In our empirical model we use two variables: accessibility of jobs through public transport (apt) and accessibility of the metro network (amt). We introduce them now explicitly into the utility function and write: $v_{a,h,c}^i = v_{a,h,c}^i(apt, amt; y^i)$. Both variables will be described in more detail in the next section together with the other urban amenities. We expect both variables to have a nonnegative impact on the utility of all choice alternatives and we expect that the impact on the utility of a given residential area and housing type without a car is at least as large as that on utility with a car. This means that the conditional compensating variation of car ownership will never increase when public transport improves. We conjecture that the same

holds for the unconditional compensating variation. According to our model (see (5)), improving public transport will have a nonpositive impact on car ownership.

For a two-worker household we extend the model by including the choice of 2 cars. This implies that the number of choice alternatives increases as every pair of residential area and housing type can be combined with 0,1 or 2 cars.

In our empirical work we measure the impact of car ownership and public transport on the utility of the choice alternatives without imposing a priori restrictions on the signs or relative magnitudes. In the next subsection we discuss our empirical specification.

3.2 Model specification

In this subsection we specify the utility function and discuss some estimation issues. We estimate separate models for single earner households and dual earners households. We discuss the specification for the single earner households first.

Utility depends on the characteristics of the choice alternative and of the household. The former set includes accessibility of public transport and the metro system apt_a and amt_a, car ownership for which we use a dummy d_c, the housing type for which we use a second dummy d_h representing a single house, the housing price, which depends on the housing type as well as the area and will be denoted as $P_{h,a}$ and other area characteristics X_a (e.g. distance from the CBD, number of protected/conserved buildings, etc.). Household characteristics include the (natural) log of income y^i and other characteristics Z^i (e.g. age and education of the head of the household, the number of children in the households, etc.). All household characteristics are used in demeaned form.

The deterministic part of the utility of a choice alternative is:

$$v_{a,h,c}^i(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i, Z^i) = \alpha_1^i apt_a + \alpha_2^i amt_a + \alpha_3^i d_c +$$

$$\beta_1^i d_h + \beta_2^i P_{h,a} + \beta_3^i X_a + \left(\gamma_1^i apt_a + \gamma_2^i amt_a + \gamma_3^i d_h + \gamma_4^i X_a\right)d_c + \xi_{a,h,c} . \qquad (9)$$

Utility is the sum of three parts, indicated by coefficients α, β and γ, respectively and an alternative-specific variable ξ that reflects unobserved (by the researcher) characteristics of the alternative. The superfix indicates that they are functions of household characteristics, as will be

discussed below. Equation (9) gives the most extensive specification considered, in the empirical work we decided to leave some variables out.

The first part of the utility refers to transport variables: availability of public transport and car ownership; the second part refers to area characteristics as they are included in equilibrium sorting models as used by Bayer et al. (2007); the third part refers to interactions of car ownership with the availability of public transport and with other neighbourhood characteristics. These interactions are key in our model that focuses on the interaction between residential location choice and car ownership. We indicated in the previous subsection that we expect car ownership to be less valuable for a household if there is better public transport. Hence we expect γ_1^i and γ_2^i to be positive. Since single family houses often have more parking space either on their own plot or on the street (density is usually lower in areas with single family housing) one may expect γ_3^i to be positive. The signs of the elements of γ_4^i depend on the nature of the area characteristic. For instance, if it is an indicator for the presence of parking charges, we expect the sign to be negative as this makes car ownership more expensive.[10] The final term was originally proposed in Berry, Levinsohn and Pakes (1995) in the context of discrete choice models for car type choice. Bayer at al. (2007) used it in the context of neighbourhood sorting and we follow them here. Incorporating this term is helpful in fitting the model and in the analysis of potential endogeneity problems associated with the housing price and other potentially endogenous variables.

The coefficients α, β and γ all depend on household characteristics and we specify them further as:

$$\alpha_j^i = \tilde{\alpha}_j^0 + \tilde{\alpha}_j^1 \ln y^i + \sum_{l=1}^{L} \tilde{\alpha}_j^{l+1} Z_l^i \qquad (10)$$

and analogous expressions for the βs and γs. Note that for the coefficients with a tilde, the superfix refers to the associated household characteristic. Since we have demeaned the household characteristics, $\tilde{\alpha}_j^0$ is the average value of the coefficients α_j^i in the population.

To estimate the model we use a two-step procedure introduced in Berry et al. (1995). We substitute (10) and the analogous expressions for the βs and γs into (9) and write the result as the

---

[10] Although one could perhaps argue that the presence of such charges makes parking space less scarce, which makes car ownership more valuable. Moreover, parking charges may reduce cruising for parking (Van Ommeren et al., 2011).

sum of the average utility of the alternative (that only includes the coefficients $\tilde{\alpha}_j^0$ $\tilde{\beta}_j^0$ $\tilde{\gamma}_j^0$ and $\xi_{a,h,c}$) and a household-specific deviation from that average. The average is then viewed as a single alternative specific constant which is, in the first step estimated as a single coefficient, jointly with the remaining parameters. This first step thus involves estimation of a MNL model.

In the second step the alternative-specific constants are written out again as a function of the coefficients $\tilde{\alpha}_j^0$ $\tilde{\beta}_j^0$ $\tilde{\gamma}_j^0$:

$$v_{a,h,c}^0\big(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a\big) = \tilde{\alpha}_1^0 apt_a + \tilde{\alpha}_2^0 amt_a + \tilde{\alpha}_3^0 d_c +$$

$$\tilde{\beta}_1^0 d_h + \tilde{\beta}_2^0 P_{h,a} + \tilde{\beta}_3^0 X_a + \quad (\tilde{\gamma}_1^0 apt_a + \tilde{\gamma}_2^0 amt_a + \tilde{\gamma}_3^0 d_h + \tilde{\gamma}_4^0 X_a)d_c + \xi_{a,h,c}. \qquad (11)$$

In this equation $\xi_{a,h,c}$ is the error term. (11) can be estimated using methods for linear equations. In the context of the present paper OLS is not appropriate, since the housing price should be expected to reflect the impact of the unobserved neighborhood characteristics $\xi_{a,h,c}$. We therefore use an instrumental variables approach.

3.3 Endogeneity

Several variables in our models can be considered as endogenous. That is, it may be argued that the values of these variables are correlated with the error term $\xi_{a,h,c}$ in the second stage regression (11). In this subsection we discuss these variables as well as the instruments we use to deal with these endogeneity concerns.

Since the unobserved characteristics $\xi_{a,h,c}$ affect the attractiveness of a choice alternatives directly, it must be expected that they have an impact on the equilibrium price of housing. This problem was observed by Berry et al. (1995) in their study of the automobile market and they proposed the use of the sums of car characteristics as instruments.[11] In the context of residential sorting the use characteristics of alternatives that are geographically close have been used as instruments by some researchers (see Klaiber and Phaneuf, 2010). A potential drawback of this practice is that characteristics of residential areas that are physically close may well have a direct impact on the utility of the choice alternative considered as residents may easily cross the borders of their area of residence to visit areas in the vicinity that have attractive

---

[11] They use sums over all car makes as well as over the makes offered by a given producer. This choice was inspired by the literature on optimal instruments (see Chamberlain, 1987).

amenities.[12] Bayer et al. (2004, 2007) adopted a different approach. They construct an instrument that intends to summarize the relative position of a choice alternative on the housing market on the basis of all available exogenous information. Their proposed instrument is the counterfactual equilibrium price predicted by the model when the term $\xi_{a,h,c}$ that reflects the unobserved characteristics is absent. This instrument is by construction independent of the unobserved heterogeneity terms ξ and most likely strongly correlated with the observed housing prices.[13] We follow Bayer et al. (2007) here.

A second variable that may be considered endogenous is the share of higher educated. To instrument for this variable we use information about the location of private schools before 1890 in the GCA. At that time only the rich could afford to send their children to such schools and the location of these schools was related to the preferred residential locations of the upper class at that time. In 1890 there were 12 such schools, only a few of them located in – what is now – the centre of Copenhagen. The idea behind this instrument is that unobserved characteristics that make a location currently (un)attractive for the average Danish household are unrelated to those that determined the location of the private schools more than a century ago, while the clustering of high income people in the early 21st century is correlated with that in the 19th century. Our instrument is the distance to the private school that is closest to the area of the choice alternative.

Thirdly it can be argued that accessibility to employment could also be endogenous as many firms nowadays are 'footloose' with respect to inputs and outputs, and may tend to locate close to where their potential workers live, while other firms – for instance shops – want locate close to the households to which they sell their goods. The instrument we use for this variable is the train stations that were founded before World War II. Many of these stations were constructed in the 1930s for the purpose of serving local industries and incidental trips from rural areas to the capital and vice versa. At the time commuting by train was exceptional, but when it became more common in the 1960s the lines connecting these stations served as the starting point for the extensive rail network constructed later on. For this reason the distance to the nearest of these older stations (which we use as our instrument) must be expected to be still correlated with accessibility to employment by public transport. Moreover, the unobserved

---

[12] Van Duijn and Rouwendal (2013) develop a model in which this is explicitly taken into account.

[13] This instrument is thus a function of all exogenous areacharacteristics (urban amenities). It may be observed that this requires area characteristics to be excluded from the equation for the average utility (11).

characteristics that make an area attractive as a place of residence for the average Danish household are unrelated to the factors that determined the location of these stations.

4. Household heterogeneity and urban amenities

Now that we have introduced the econometric model, we describe in detail the socioeconomic variables we use to control for household heterogeneity and then the variables for the chosen amenities.

We include following socioeconomic variables to account for the household heterogeneity: i) age (and square of age) of the head of the household, ii) three dummy variables indicating the highest education level obtained by the head of the household, iii) the number of children in the household, and iv) household income. Moreover, for single earner households we also include a dichotomous variable indicating a one-person household (single). For dual-earner households we also include socioeconomic variables for partners. First of all we control for age of the head of the households to account for the life-cycle preferences of households. Because we do not expect the age effect to be linear, we also include the square of the age of the head of the households. It has been argued in the literature that the highly educated households are attracted by the better access to labour market and urban amenities. We control for the highest education obtained by the head of the household, because we expect e.g. that the highly educated dual earners to have different preferences for accessibility to transport facilities than the highly educated single earner households (in some cases singles). We also expect the presence and the number of children in household to affect household preferences, e.g. because of the provision of public goods (Fernandez and Rogerson (1996); Nechyba (2000)). Examples of the public goods are e.g. childcare, schools and recreational facilities. Finally we also control for the household income.[14] In general, the demand for urban amenities depends positively on household income (e.g., Van Duijn and Rouwendal, 2013). Moreover the household income represents the budget constraint.

---

[14] Information about households income is based on third-party reporting (includes both reporting from firms who tax wages and banks, mortgage institutions, brokers, etc.) and is considered highly reliable. Kleven et al. (2011) show that the tax evasion rate is close to zero for income subject to third-party reporting.

Table 1. Household characteristics

	Single earner households		Dual earners households	
	Mean	Std. dev.	Mean	Std. dev.
Household's income (1000 DKK)	393.574	470.470	630.634	435.415
Number of children in household	0.379	0.779	1.220	1.039
Non-apartment owner (share)	0.530	0.499	0.819	0.385
Car ownership, one or two cars (share)	0.600	0.490		
One car (share)			0.753	0.431
Two cars (share)			0.108	0.310
Age, head of the household	47.006	13.382	46.109	10.015
Low education (share), head of the household	0.565	0.496	0.501	0.500
Medium education (share), head of the household	0.244	0.429	0.242	0.429
High education (share), head of the household	0.192	0.394	0.257	0.437
Age, partner			42.733	9.626
Low education (share), partner			0.487	0.500
Medium education (share), partner			0.281	0.450
High education (share), partner			0.233	0.422
Singles	0.648	0.478		
Number of observations	66,012		87,330	

Notes: low education obtained includes: basic school, general upper secondary school, vocational upper secondary school and vocational education; medium education obtained includes: short-cycle higher education and medium-cycle higher education; and high education includes: bachelor, long-cycle higher education and PhD-degree.

Table 1 shows the summary statistics of the household characteristics. It is worth to mention some interesting differences between single earner households and dual earner households. Not surprisingly, household income for dual earner households largely exceeds income for single earner households. Dual earner households also have more children and more of them live in single family houses. Note that single family houses are typically larger than multifamily houses as well as they often have a private garden and parking space whereas outdoor facilities are typically limited or shared in multi-family houses. Car-ownership is also higher for dual earner households and they hold a larger share of higher educated.

Table 2: Descriptive statistics area characteristics

	Mean	Std. dev.	Min.	Max.
Employment access with public transport / 1000	235.088	41.075	99.261	288.520
Proximity to the nearest metro station (km)	0.234	0.331	0.000	0.902
Standardized house price (DKK million)	2.409	0.484	1.309	3.519
Share of higher educated	0.249	0.128	0.046	0.500
Number of conserved/protected buildings per sq. m.	0.0004	0.0003	6.36E-6	0.0011
Distance to the CBD. (km)	10.607	7.161	0.000	32.570
Parking charging (share)	0.133	0.340	0.000	1.000
Social housing (share)	0.243	0.235	0.000	0.950

Notes: number of observations is 166.

We expect the different types of households to have different preferences for urban amenities. We focus on 8 local urban amenities where the two first are transport and employment related: i) employment access by public transport, ii) proximity to the nearest metro station and

the last six are characterizing the neighbourhood in other ways: iii) standardized house price, iv) share of higher educated population, v) cultural heritage i.e. conserved buildings in the neighbourhood, vi) distance to the CBD, vii) parking charging, and viii) social housing. Table 2 shows the summary statistics for the considered urban amenities.

In order to account for the labour market attractiveness of each area we include the measure of the employment access. The employment access measure has been compiled using the number of the full time job equivalents (J) for each area applying the following equation

$$EA_a = \sum_{a'} J_{a'} * e^{-\delta d_{a,a'}}$$

where EA is employment access, a is the area index, d is the travel time by public transport from area a to area a', and δ is a parameter. We set δ to 0.05. This implies, for instance, that one additional job at the distance of 120 min (the max) has weight 0.0025 in the EA calculation. If the distance is 48 min (the mean) the weight is =0.09 and if the distance is 60 min it equals 0.05. Jobs 'around the corner' have a weight 1.

Accessibility to transport facilities is of main interest. Denmark has a highly developed transport infrastructure. The accessibility to public transport is particularly highly developed in the GCA. Therefore it is not surprising that there is excellent access to e.g. a bus stop in all the considered areas in the GCA. Consequently there is no variation in this or similar variables in our sample, so they are not useful in the model estimation. For the metro, however, this is different. The proximity to the nearest metro station is particularly important because it represents the high quality public transport with frequent services and attractive stations (physical environment). Proximity to the nearest metro is compiled for an area as the average distance from each address in the area to the nearest metro station:[15]

$$\text{Proximity} = \begin{array}{ll} \frac{3-distance}{3} & if \quad distance \ to \ metro \ st. \leq 3 \ km \\ 0 & if \quad distance \ to \ metro \ st. > 3 \ km \end{array}$$

As also mentioned earlier, we include a dummy for housing type (single family house vs. multi-family housing). We capture the housing market by including an area price index for a standard house, which we interpret as the price of housing services. Standardized house price has been compiled from two separate hedonic models with area fixed effects, one for single family

---

[15] The majority of bike-and-ride users travel up to 3 km to a public transport stop (Martens, 2004).

houses and the other for apartments. Estimation results of these hedonic price functions are reported in the Appendix A.1. The standard house has the average size and other characteristics (for the whole GCA) that are used in the hedonic price equation.

The average standard house price (3.2 million DKK) is almost two times higher than the average standardised apartment price (1.7 million DKK). Map 3 shows the weighted average standardised house and apartment price in the GCA. The map shows the expected pattern of high house prices in the northern part of the GCA that is considered as highly attractive by Copenhagen households.

Map 3. Weighted average standardised housing price in the GCA (1000 DKK)
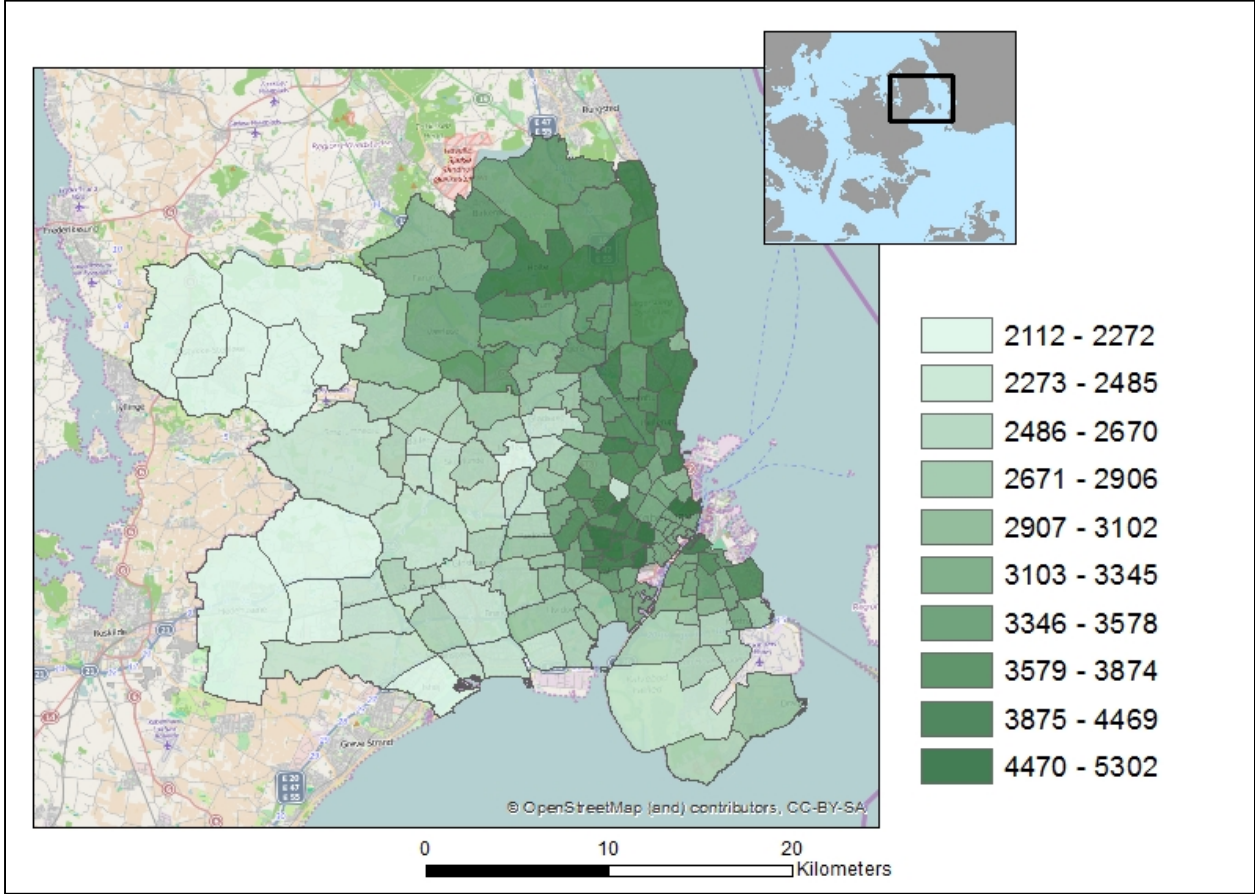
Notes: standardized house price has been compiled from the two separated hedonic models with area fixed effect, i.e. one for the houses and one for the apartments.

We now turn to the other amenities characterizing a neighbourhood. We include the share of higher educated people from the population as an indicator of endogenous amenities. It is often argued in the literature that the attractiveness of living in a particular area is partly

determined by the demographic composition of that neighbourhood. For instance, in sociology the phenomenon of homophily which holds that households interact preferably with other households that are similar, is well-known. In the urban economics literature, the importance of this factor for location choice within the San Francisco Bay area was documented by Bayer et al. (2007). Map 4 shows the share of higher educated in the sample distributed over the considered areas. It is interesting to notice the similarity over maps 3 and 4. Map 4 shows a higher share of the higher educated in the northern part of the GCA, the same part that is considered as highly attractive by Danish households. This variable is also highly correlated with house prices. It may also be noted that it is not necessarily the share of higher educated households *per se* that is important. It may well be the case that the presence of such households has an impact on the attractiveness through shops, restaurants and other facilities that are offered in the vicinity.

Map 4. The share of higher educated in the sample
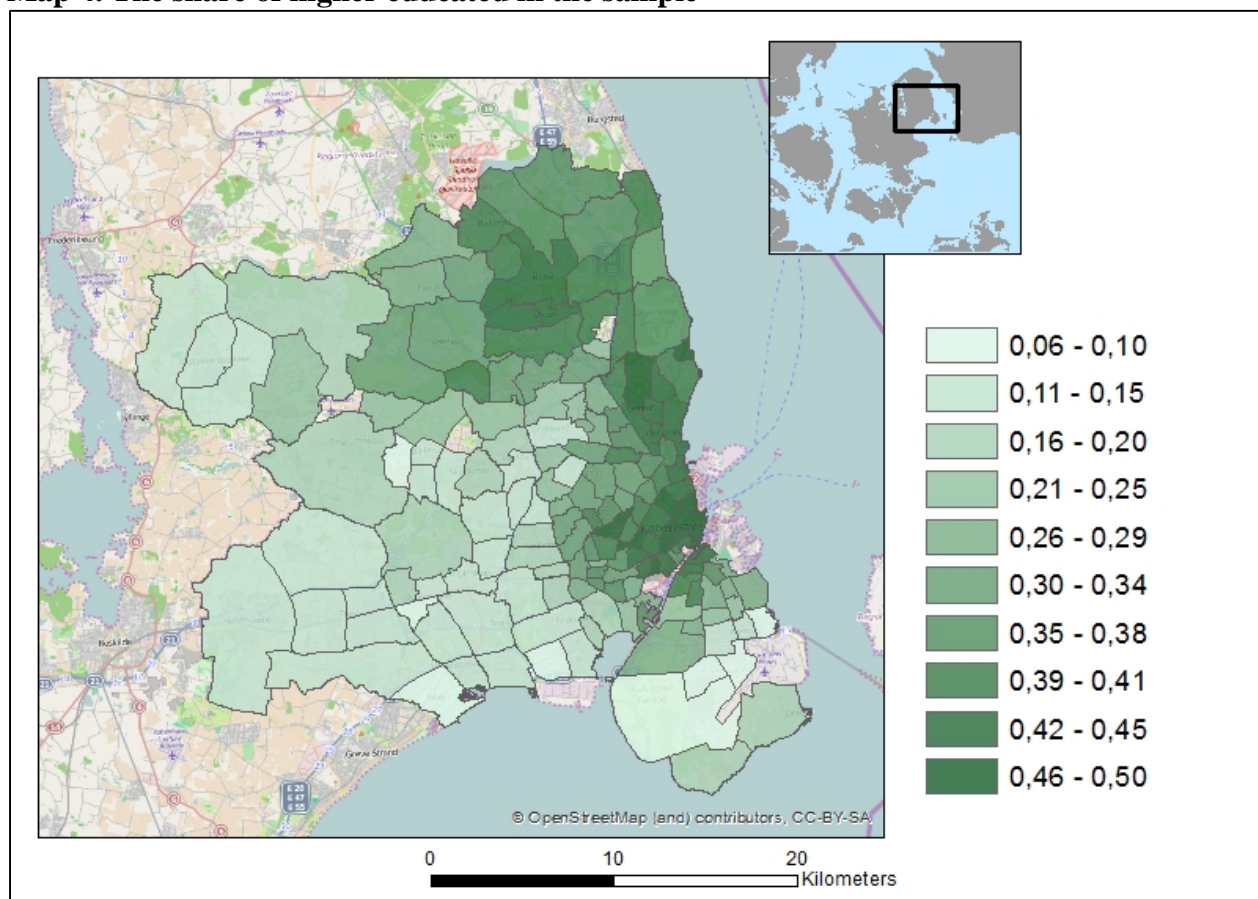
It has been shown in the literature that the concentration of historical buildings is important for household location choice (Van Duijn and Rouwendal, 2013) either because this cultural heritage is appreciated itself or because it helps to attract shops, restaurants, cinema's and other endogenous amenities. Since there is not a generally accepted measure of cultural heritage that reflects differences in its quality, we use the number of conserved and/or protected buildings per sq.km. as an indicator for it. We also include the distance to the CBD. Distance to the CBD has been compiled as the distance from an area to the area representing the city centre in Copenhagen (the city hall).

We include a dichotomous variable indicating whether curb side parking in the area is subject to charges or not. Curb side parking charges are especially found in the centre of Copenhagen and gets less frequent the further you get from the centre. Areas with parking charging are typically also areas where parking spaces are scarce and where a lot of cruising for parking potentially takes place. In neighbourhoods with parking charges it is typically possible for residents to buy a yearly parking permit at a low cost.

Finally, we include a parameter indicating the share of social housing relative to the total number of houses in the neighbourhood. The reason is similar to that for including the share of higher educated households: social housing is more or less only accessible to households with low incomes and other households may have preferences for (or against) living in the proximity of such households.

5. Estimation results

We estimate two models: one for single-earner households and another for dual-earner households. For both samples we first estimated a logit model in which the deterministic part of the utility of choice alternative (a, h, c) is specified as:

$$v_{a,h,c}^i(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i, Z^i) = \delta_{a,h,c} + \Delta\alpha_1^i apt_a + \Delta\alpha_2^i amt_a + \Delta\alpha_3^i d_c + \quad (12)$$

$$\Delta\beta_1^i d_h + \Delta\beta_2^i P_{h,a} + \Delta\beta_3^i X_a + \quad \left(\Delta\gamma_1^i apt_a + \Delta\gamma_2^i amt_a + \Delta\gamma_3^i d_h + \Delta\gamma_4^i X_a\right)d_c.$$

In this equation $\delta_{a,h,c}$ is the utility attached to (a, h, c) by the average Danish household[16] and $\Delta\alpha_j^i$ is the difference between α_j^i and $\tilde{\alpha}_j^0$, $j = 1,2,3$ [17] and similar for the $\Delta\beta$'s and $\Delta\gamma$'s. In the

---

[16] That is $\delta_{a,h,c} = v_{a,h,c}^0(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a)$ in (11).

second stage we put the estimated values of the alternative-specific constants on the left-hand side of (12) and estimate its coefficients using OLS and IV.

5.1 The average household

Tables 3.a and 3.b show the results of the second stage, which refers to the utility attached by the average Danish household to the various choice alternatives. Table 3a refers to the single-earner households and 3b to the dual-earner households, respectively. Tables 4.a and 4.b report the coefficients that show deviations from the average utilities are related to household characteristics, for the same groups of households.

Table 3.a Second step estimation results for single earner households: decomposition of the household's mean indirect utilities

		[1] OLS	[2] IV (2SLS)
α's	Employment access with public transport / 1000 * dummy variable indicating no car	0.008*** (0.003)	0.007* (0.004)
	Proximity to the nearest metro station (km) * dummy variable indicating no car	0.454** (0.207)	0.547** (0.230)
	Dummy variable indicating one car	0.960*** (0.227)	0.889*** (0.304)
β's	Dummy variable indicating non-apartment	1.432*** (0.235)	1.980*** (0.353)
	Log (standardized house,apartment price)	-2.178*** (0.324)	-3.032*** (0.517)
	Share of higher educated	1.874*** (0.532)	3.130*** (1.043)
	Number of conserved/protected buildings per sq.m.	0.937*** (0.167)	0.903*** (0.167)
	Distance to the CBD.	0.020** (0.008)	0.016* (0.009)
	Social housing (share)	-0.418** (0.206)	-0.410* (0.219)
γ's	Dummy variable indicating non-apartment * dummy variable indicating one car	0.128 (0.151)	0.126 (0.152)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.168 (0.194)	-0.179 (0.196)
	Constant	-1.189*** (0.324)	-0.937** (0.392)
	R-squared	0.214	
	No. of observations	538	538

Notes: standard errors in parentheses; standardized house/apartment price, share of higher educated and employment access with public transport are instrumented; see Table A.2.1 in the Appendix A.2 for first-stage regression estimates of the 2SLS; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

Table 3.b Second step estimation results for dual earners households: decomposition of the household's mean indirect utilities

	[1]	[2]

---

[17] That is $\alpha_j^i = \tilde{\alpha}_j^1 y^i + \sum_{l=1}^{L} \tilde{\alpha}_j^{l+1} Z_l^i$, see (10).

		OLS	IV (2SLS)
α's	Employment access with public transport / 1000 * dummy variable indicating no car	0.012***	0.010*
		(0.003)	(0.005)
	Proximity to the nearest metro station (km) * dummy variable indicating no car	0.712***	0.800***
		(0.215)	(0.236)
	Dummy variable indicating one car	1.728***	1.770***
		(0.298)	(0.392)
	Dummy variable indicating two cars	1.033***	0.912**
		(0.327)	(0.444)
β's	Dummy variable indicating non-apartment	2.743***	3.428***
		(0.277)	(0.463)
	Log (standardized house, apartment price)	-2.321***	-3.357***
		(0.361)	(0.651)
	Share of higher educated	2.644***	3.880***
		(0.586)	(1.255)
	Number of conserved/protected buildings per sq.m.	0.897***	0.848***
		(0.159)	(0.161)
	Distance to the CBD.	0.039***	0.027**
		(0.009)	(0.012)
	Social housing (share)	-0.370*	-0.443**
		(0.199)	(0.215)
γ's	Employment access with public transport / 1000 * dummy variable indicating one car	0.004	0.002
		(0.003)	(0.005)
	Proximity to the nearest Metro station (km) * dummy variable indicating one car	0.243	0.300
		(0.217)	(0.235)
	Dummy variable indicating non-apartment * dummy variable indicating one car	0.495***	0.471***
		(0.168)	(0.174)
	Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.147	-0.142
		(0.236)	(0.245)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.130	-0.122
		(0.212)	(0.214)
	Dummy variable indicating parking charging * dummy variable indicating two cars	-0.072	-0.143
		(0.424)	(0.431)
	Constant	-2.854***	-2.370***
		(0.368)	(0.498)
	R-squared	0.570	
	No. of observations	636	636

Notes: standard errors in parentheses; standardized house/apartment price, share of higher educated and employment access with public transport are instrumented; see Table A.2.2 in the Appendix A.2 for first-stage regression estimates of the 2SLS; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

Tables 3.a and 3.b show the results of the second step of the estimation procedure based on (11). The dependent variable is the vector of mean indirect utilities that were estimated as alternative specific constants in the first (logit) step of the estimation procedure. These $\delta_{a,h,c}$'s represent the part of the utility that is equal for all one earner or two-earner households. Table 3.a gives the results for single earner households. For the alternatives in which no car is owned, accessibility to employment by public transport and proximity to a metro station are important. Ownership of a car makes a choice alternative always more attractive. Single family houses are

Table 4.a First step estimation procedure (multinomial logit) for single earner households: interaction parameter estimates

	Amenities	Households characteristics						
		Log (hous. income)	Age	Age sq. /1000	Number of children	Education (medium)	Education (high)	Singles
α's	Employment access with public transport / 1000 * dummy variable indicating no car	-0.005***	-0.001***	0.007***	-0.001***	0.002*	0.001	-0.003***
		(0.001)	(0.0001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Proximity to the nearest metro station (km) * dummy variable indicating no car	-0.062	0.019*	-0.243**	-0.054	-0.069	0.016	-0.109
		(0.062)	(0.009)	(0.096)	(0.042)	(0.057)	(0.059)	(0.068)
	Dummy variable indicating one car	0.501***	-0.033***	0.329**	0.155***	0.211**	0.030	-0.830***
		(0.082)	(0.011)	(0.120)	(0.053)	(0.078)	(0.087)	(0.088)
β's	Dummy variable indicating non-apartment	-0.693***	-0.053***	0.460***	0.404***	0.152	0.001	-1.159***
		(0.084)	(0.012)	(0.125)	(0.054)	(0.086)	(0.096)	(0.092)
	Log (standardized housing price)	2.230***	0.052***	0.109	0.195**	-0.283**	-0.017	1.030***
		(0.111)	(0.016)	(0.168)	(0.070)	(0.116)	(0.126)	(0.122)
	Share of higher educated	2.420***	-0.087***	1.082***	0.178**	2.968***	5.582***	0.732***
		(0.182)	(0.025)	(0.261)	(0.109)	(0.177)	(0.201)	(0.186)
	Number of conserved/protected buildings per sq.m.	-0.262***	0.005	0.071	0.270***	0.021	-0.129*	0.161***
		(0.052)	(0.007)	(0.071)	(0.029)	(0.048)	(0.055)	(0.050)
	Distance to the CBD.	0.013***	0.001**	0.006**	0.011***	-0.008**	-0.024***	0.019***
		(0.002)	(0.0003)	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)
	Social housing (share)	-0.528***	0.018**	-0.072	0.108***	0.096	0.084	-0.189***
		(0.069)	(0.008)	(0.085)	(0.035)	(0.062)	(0.078)	(0.062)
γ's	Dummy variable indicating non-apartment * dummy variable indicating one car	0.285***	0.052***	-0.308***	-0.075***	0.081	0.218***	-0.313***
		(0.052)	(0.007)	(0.072)	(0.031)	(0.048)	(0.055)	(0.053)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.058	-0.007	-0.057	-0.291***	-0.138**	0.004	-0.287***
		(0.054)	(0.008)	(0.083)	(0.039)	(0.061)	(0.057)	(0.059)

Notes: standard errors in parentheses; ***, **, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and 0.10 levels, respectively.

Table 4.b First step estimation procedure (multinomial logit) for dual earners households: interaction parameter estimates

	Amenities	Households characteristics									
		Log (hous. income)	Age, head	Age sq. /1000, head	Number of children	Education (medium), head	Education (high), head	Age, partner	Age sq. /1000, partner	Education (medium), partner	Education (high), partner
α's	Empl. access with public transport / 1000 * dummy indicating no car	-0.006***	0.0004	0.002	0.002***	0.003**	0.003*	-0.0002	0.004	0.004**	0.006***
		(0.002)	(0.001)	(0.008)	(0.001)	(0.001)	(0.002)	(0.001)	(0.009)	(0.001)	(0.002)
	Proximity to the nearest metro station (km) * dummy indicating no car	-0.561***	0.043	-0.439	-0.185***	0.139	0.317***	0.039	-0.668	0.295***	0.273***
		(0.106)	(0.049)	(0.518)	(0.039)	(0.091)	(0.095)	(0.052)	(0.588)	(0.088)	(0.097)
	Dummy variable indicating one car	0.274*	0.036	-0.659	0.485***	0.224*	0.201	0.001	0.436	0.405***	0.162
		(0.144)	(0.071)	(0.728)	(0.054)	(0.132)	(0.142)	(0.074)	(0.820)	(0.128)	(0.148)
	Dummy variable indicating two cars	1.268***	0.005	-0.065	0.301***	0.413*	0.218	0.143	-1.331	0.410**	0.335*
		(0.184)	(0.098)	(0.999)	(0.073)	(0.177)	(0.186)	(0.103)	(1.126)	(0.173)	(0.190)
β's	Dummy variable indicating non-apartment	0.089	0.028	-0.859	0.701***	0.294***	0.337***	0.247***	-1.695***	0.631***	0.097
		(0.120)	(0.063)	(0.645)	(0.050)	(0.114)	(0.118)	(0.067)	(0.729)	(0.110)	(0.123)
	Log (standardized housing price)	3.656***	0.153**	-0.457	0.048	-0.350***	-0.434**	-0.173**	1.373	-0.524***	-0.112
		(0.136)	(0.077)	(0.775)	(0.059)	(0.131)	(0.132)	(0.081)	(0.870)	(0.126)	(0.137)
	Share of higher educated	3.932***	-0.142	1.574	0.350***	2.680***	5.566***	0.298	-2.441*	2.579***	4.165***
		(0.222)	(0.118)	(1.193)	(0.090)	(0.169)	(0.203)	(0.123)	(1.343)	(0.188)	(0.213)
	Number of conserved/protected buildings per sq.m.	-0.772***	0.075**	-0.912***	0.126***	0.029	0.100*	0.055*	-0.456	0.037	0.028
		(0.062)	(0.029)	(0.303)	(0.021)	(0.046)	(0.052)	(0.031)	(0.340)	(0.044)	(0.053)
	Distance to the CBD.	0.010***	0.002	-0.022	0.006***	0.006**	0.002	-0.005***	0.056***	-0.003	-0.003
		(0.003)	(0.002)	(0.017)	(0.001)	(0.003)	(0.003)	(0.002)	(0.019)	(0.002)	(0.003)
	Social housing (share)	-0.923***	0.101***	-1.166***	0.001	0.283***	0.304***	0.007	0.137	0.154***	0.252***
		(0.073)	(0.035)	(0.362)	(0.025)	(0.054)	(0.064)	(0.036)	(0.400)	(0.052)	(0.066)
γ's	Empl. access with public transport / 1000 * dummy indicating one car	-0.0004	-0.0001	0.003	-0.0002	0.002***	0.002**	-0.001	0.008	0.002**	0.004***
		(0.001)	(0.001)	(0.005)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.006)	(0.001)	(0.001)
	Proximity to the nearest metro station (km) * dummy indicating one car	-0.368***	-0.002	-0.019	-0.112***	0.113**	0.134**	0.076	-0.889**	0.105**	0.276***
		(0.068)	(0.032)	(0.332)	(0.024)	(0.052)	(0.056)	(0.033)	(0.370)	(0.051)	(0.057)
	Dummy variable indicating non-apartment * dummy indicating one car	-0.633***	-0.059	0.602	-0.205***	0.004	0.056	0.074*	-0.891*	-0.192***	0.056
		(0.086)	(0.040)	(0.418)	(0.032)	(0.075)	(0.079)	(0.043)	(0.479)	(0.073)	(0.082)
	Dummy variable indicating non-apartment * dummy indicating two cars	-1.052***	-0.057	0.448	-0.098	-0.061	0.116	-0.034	0.378	-0.189	0.131
		(0.147)	(0.082)	(0.830)	(0.060)	(0.143)	(0.147)	(0.087)	(0.938)	(0.141)	(0.149)
	Dummy variable indicating parking charging * dummy indicating one car	-0.298***	-0.038	0.454	-0.143***	-0.176***	-0.057	0.004	-0.227	-0.053	-0.027
		(0.077)	(0.034)	(0.359)	(0.028)	(0.067)	(0.061)	(0.037)	(0.419)	(0.065)	(0.062)
	Dummy variable indicating parking charging * dummy indicating two cars	0.493**	0.077	-1.000	-0.277**	0.061	-0.095	-0.338	3.870	-0.128	-0.397
		(0.225)	(0.226)	(2.251)	(0.136)	(0.333)	(0.338)	(0.222)	(2.301)	(0.320)	(0.343)

Notes: standard errors in parentheses; ***, **, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and 0.10 levels, respectively.

preferred to apartments and a higher housing price makes an alternative less attractive. The presence of higher educated households and monuments make an area more attractive. Distance to the CBD is valued positively, perhaps because of the crowding and congestion effects, while the attractive features of city life are reflected already in the share of higher educated and the monuments. The presence of social housing has a negative impact. The interactions of car and neighbourhood characteristics have no significant impact on the average household.

Dealing with the endogeneity issues through IV makes a substantial difference for the estimation results. The larger (in absolute value) size of the price coefficient is a well-known phenomenon that is caused by attributing the impact of unobserved heterogeneity to limited price sensitivity when this is not properly taken into account. The coefficient of the share of higher educated almost doubles, which may have similar reasons. The coefficient for the accessibility of employment by public transport hardly changes.

The results for the dual earners households, presented in Table 3b, are qualitatively similar. Having one or two cars is better than having none, but one car is clearly the situation that is on average most preferred. This is probably related to the high costs of car ownership and use in Denmark and the diminishing returns to ownership of an additional car. The interaction term for having one car and living in a single family house is now significantly positive, which may be related to better parking possibilities (on one's own plot) that are often present with such housing.

5.2 Deviations from the average

Tables 4.a. and 4.b show the coefficients that relate deviations from average utility to household characteristics. Income is clearly important in this respect. Let us first look at the single earner households. Having a higher income makes one less sensitive to the availability of public transport if no car is owned, but owning a car becomes much more attractive. The sensitivity to the housing price decreases, but the presence of higher educated is appreciated more. And the combination of a single family house and a car gets more important with income. The interactions with other household characteristics show that accessibility to public transport as well as owning a car become less important with age although at a decreasing rate, while households with children have stronger preferences for cars and single family houses. The

26

combination of children and living in an area with parking charges is unattractive. Singles are less sensitive to the availability of public transport if no car is owned. Moreover, owning a car is much less attractive for singles but the presence of higher educated and access to monuments are appreciated more. The combination of car ownership and living in an area with parking charges and the combination of car ownership and single family houses are less attractive for singles.

The results for dual-earner households presented in Table 4b confirm the importance of household income. We have included age and education of both workers, which are in many households similar. The estimation results confirm the picture that arises from Table 4a for the single-earner households.

6. The impact of an improved metro network

The metro system in Copenhagen is relatively new. The first stations opened in 2002, a second set of stations followed in 2003 while the third phase (extending an existing line to the airport) was opened in 2007. The metro represented a significant upgrading of public transport with respect to quality and has been quite popular (almost) from the start. It is used daily by many people and has at present more than 56 mio. passengers yearly (in 2014). The metro has at present 22 stations; see the black dots in Figure 5.

The extension of the metro that is currently under construction, and is expected to open in 2019, implies a significant expansion of the network with a city-circle and 18 new stations, most of them in central Copenhagen (see the red dots in Map 5). This contrasts with many extensions of metro networks in other metropolitan areas that aim to link suburbs with the central city.

We use the estimated model to simulate the impact of the extension of the metro network. The estimated model is affected in two different ways. The primary effect of the extended public transport is a change in neighborhood characteristics: the distance to the nearest metro station reduces for many areas in the city of Copenhagen and job accessibility by public transport (travel times by public transport) improves as well. The changes in these variables are available from the Danish National Traffic Model. These changes will affect the utility attached to the choice alternatives concerned and through this on household location behaviour.

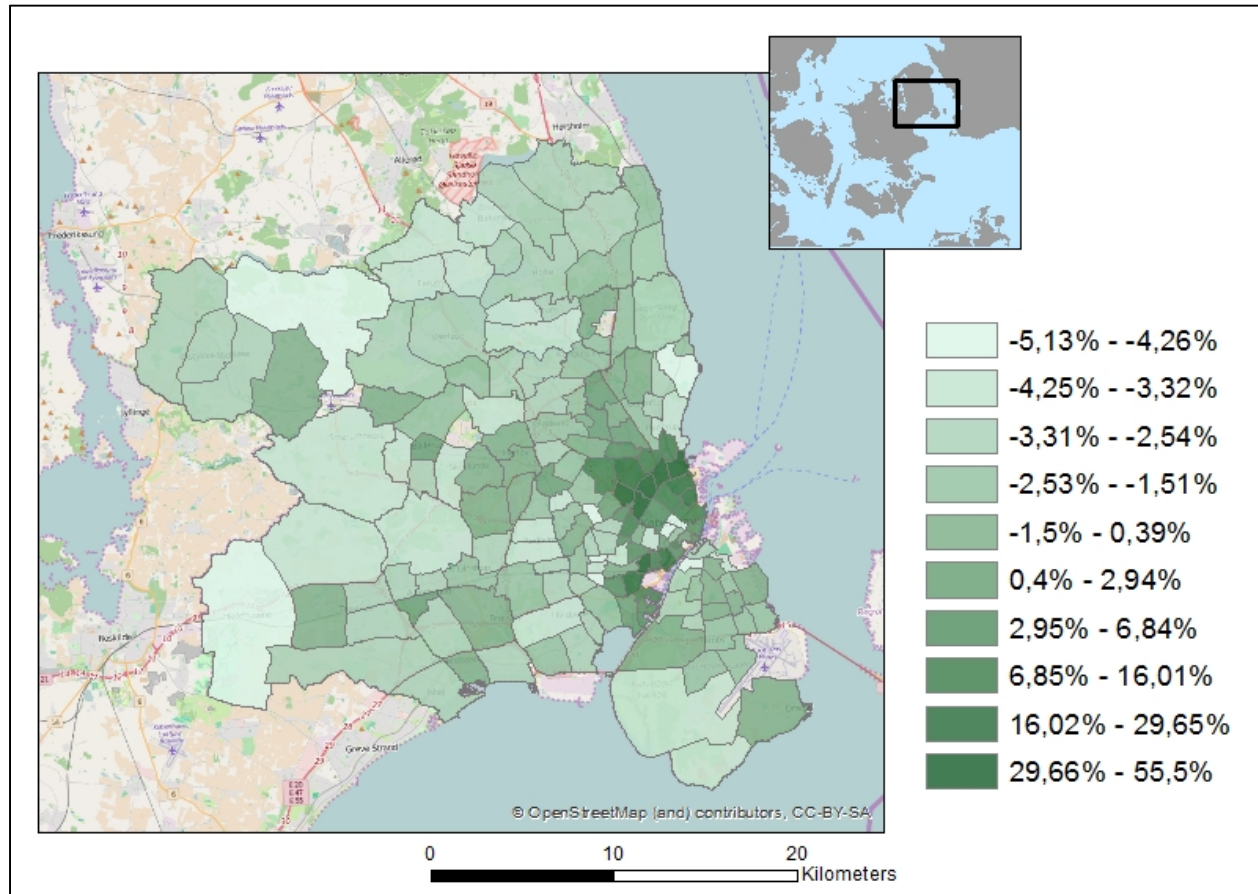**Map 5. Metro system extension**
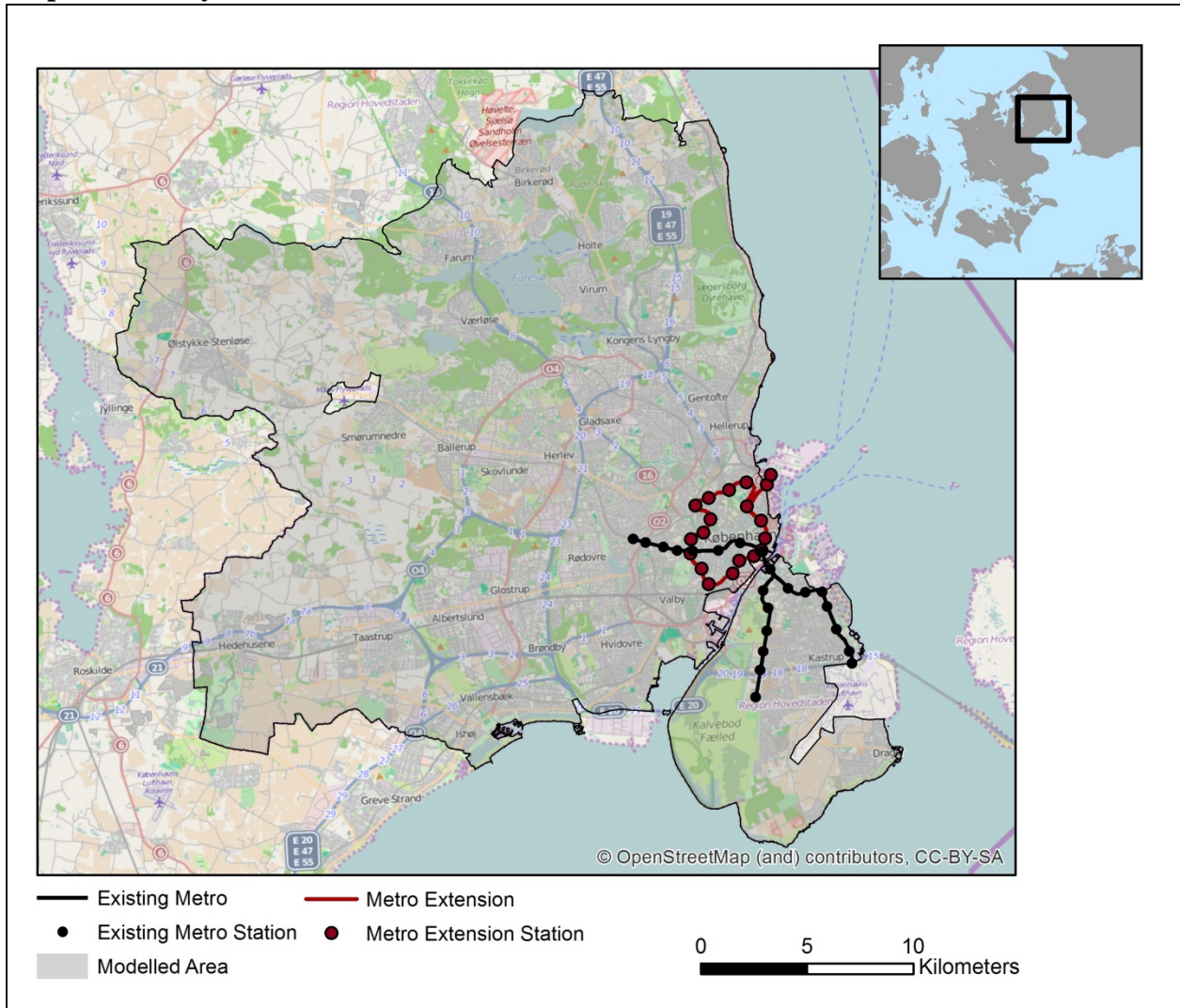
6.1. Excess demand

Our first investigation concerns the changes in housing demand that would occur because of the extension of the metro network if house prices would remain unchanged. These changes in demand can only be realized if housing supply is infinitely elastic, which is obviously not the case in the Copenhagen area, if only because of the fact that so much land is already used for houses and other buildings. The exercise is nevertheless interesting because it shows how people would react to the change in public transport *per se*.

Map 6 shows pct. change in household population per area in the GCA caused by the extension of the metro system. The map suggests that extension of the metro system will have a

Map 6. Pct. change in population of the households in the GCA caused by the metro extension

Map legend:
- -5,13% - -4,26%
- -4,25% - -3,32%
- -3,31% - -2,54%
- -2,53% - -1,51%
- -1,5% - 0,39%
- 0,4% - 2,94%
- 2,95% - 6,84%
- 6,85% - 16,01%
- 16,02% - 29,65%
- 29,66% - 55,5%

© OpenStreetMap (and) contributors, CC-BY-SA

0   10   20 Kilometers

substantial impact on housing demand, especially in the centre of the area along the new metro line. The increase in demand for these areas implies, of course, a decrease elsewhere in the region but since the improvement in accessibility is concentrated in a few areas, the decrease is spread over a much larger area. The extension of the metro will have more or less identical impact on single earner households and dual earners households. Moreover, especially relatively wealthier and higher educated households will be attracted by the extension of the metro system (see Maps A.1.-A.4 in Appendix A.3).

6.2. Housing price adjustments

We have also investigated the housing market equilibrium that would realize after the extension of the metro network has been realized under the assumption that housing supply remains unchanged. This assumption is clearly the opposite of the one used in the previous subsection,

**Map 7. Change in relative house prices caused by the metro extension**
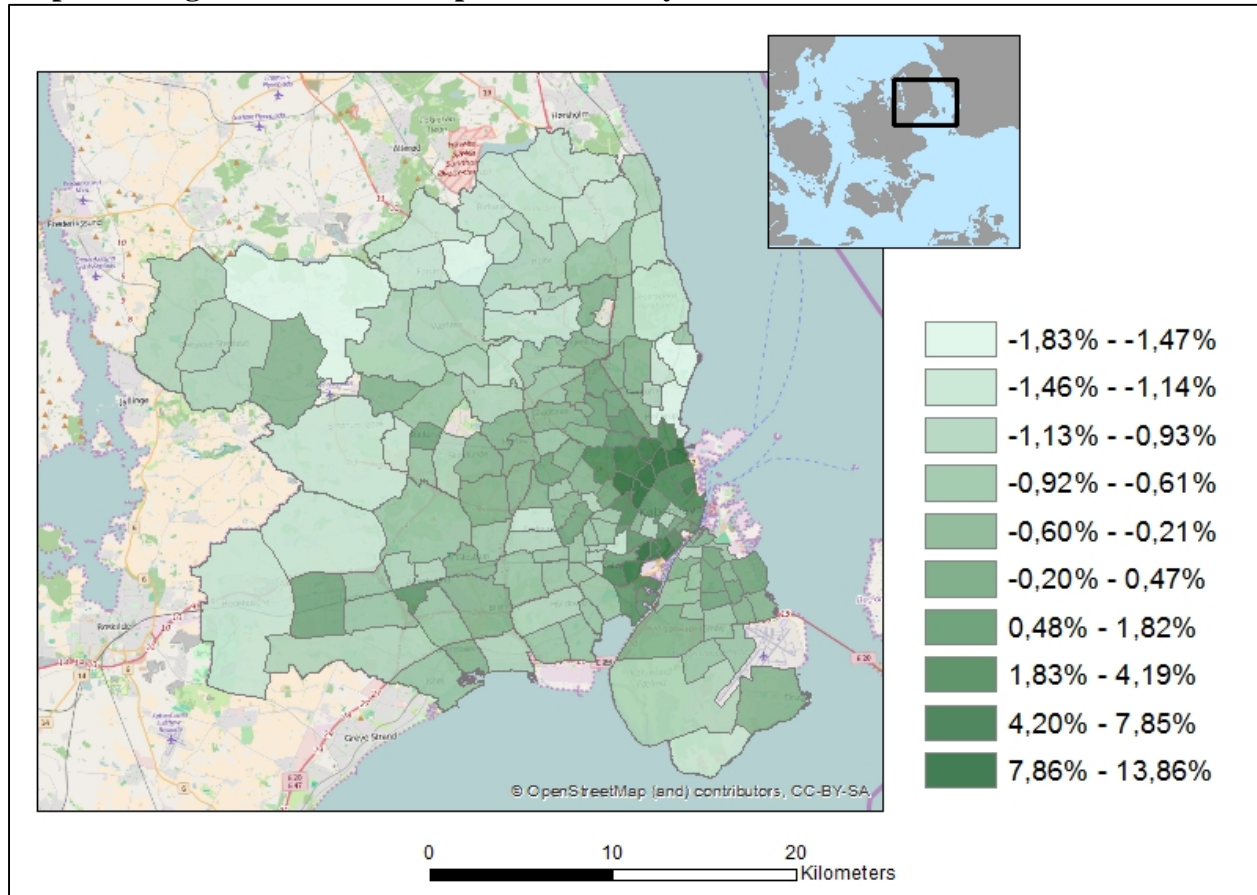
and it is likely that it is also false. However, the possibilities to increase housing supply in the centre of the CGA are clearly limited and a decrease in the housing stock in the suburban regions because of lower house prices also seems unlikely. Hence, the assumption of a zero elasticity of supply should be expected to be much closer to the truth than that of an infinitely inelastic supply. Moreover, a benefit of adopting this assumption, although admittedly somewhat extreme, is that it saves us the efforts of modelling housing supply in the various parts of the CGA under conditions of increasing as well as decreasing house prices.

The reaction of the housing prices to the metro extension is shown on Map 7. Since our model can only deal with relative prices, we assumed that the average price level remains constant in the Greater Copenhagen Area. The housing prices increase in the areas closer to the new metro line and decrease in other areas that become relatively less attractive. The reaction of house prices thus counteracts that of the extension of the metro network since areas that benefit most of this extension will now have higher user costs for housing, while areas that do not

benefit directly from the extension will have lower user costs. The return to housing market equilibrium with fixed housing supply thus acts as a redistribution of the benefits of the metro extension.[18]

Our results suggest also a substantial increase in the interest for living in areas close to the metro network (see Maps A.1-A.4 in Appendix A.3). Moreover, Map 8 shows the impact of the extension of the metro-network on the location choices of high income households, i.e. the relatively wealthier households from the northern part of the GCA are in particular attracted by the improved high quality public transport. Higher educated and households with children are attracted by the extended metro-network as well (see Maps A.5-A.6 in Appendix A.3). Our simulation results suggest that improving high quality public transport significantly affects the

Map 8. Change in household income caused by the metro extension (fixed supply and house prices adjust)

---
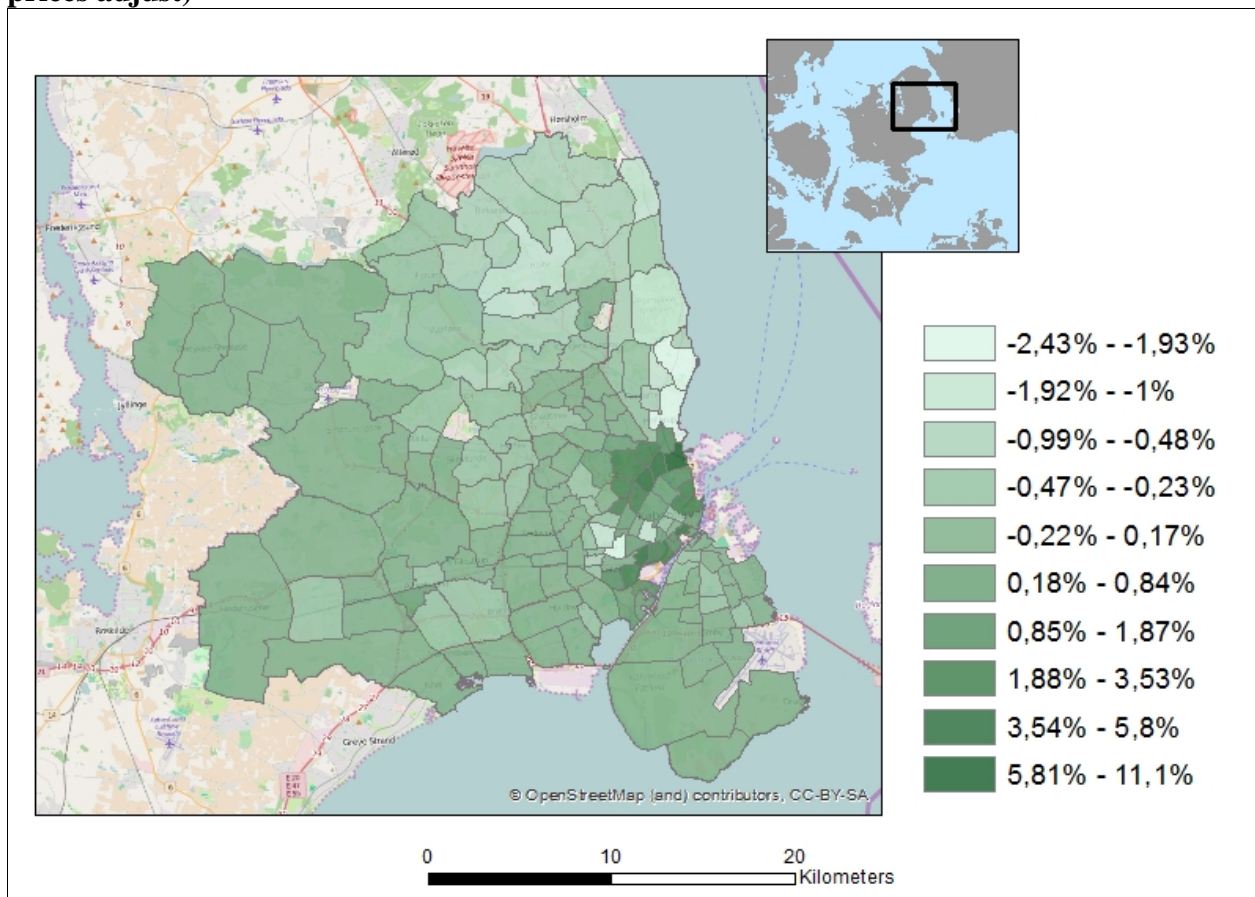
[18] Note, however, that we do not consider the wealth effects of the housing price changes.

demographic composition of neighbourhoods.

6.3 Car ownership

We argued in section 3 that we expect that improving public transport will have a nonpositive impact on car ownership. This is confirmed by the simulation study. The model suggests that the number of car owners will be reduced as a result of the metro extension. Table 5 shows that number of one car owners will decrease by 2.9% if housing supply would be elastic and by 2.3% with inelastic supply (house price adjust). For two car owners the corresponding figures are 4.5% and 3.1%, respectively. Clearly some households that would give up their car (or one of their cars) if they could move to the areas where metro accessibility improved will change this intention when house prices adjust. When interpreting these figures, it should be noted that they refer to the whole GCA. Changes in the shares of car owners are much larger in the neighbourhoods that are directly affected by the extension, see Map 9.

Table 5: Car ownership

	Reference scenario	Scenario 1	Scenario 2
		Fixed prices	Fixed supply
One car households	85,388	82,906	83,389
Two cars households	17,495	16,695	16,949
Total number of cars	120,378	116,295	117,287

6.4 Welfare change

There are various ways to assess the impact of the extension of the Copenhagen metro network. The first possibility is to look at what its impact will be on the welfare of those involved if choice behaviour would remain unchanged. All households therefore stay in their initially chosen alternative. We can compute the compensating variation of these households.

It is, of course, unlikely that choice behaviour remains unchanged. The areas that become more attractive will be chosen more frequently unless prices change. In our second assessment we assume this to be the case. Since we do not yet take into account house price adjustments, this requires housing supply to be infinitely elastic. Alternatively, it can be regarded as indicating the change in pressure on the existing housing stock that is the result of the improved public transport. Note also, that households do not only switch to other areas, but also to other positions with respect to car ownership, and here the assumption of elastic supply is more realistic.

Map 9: Change in car ownership in the GCA caused by the metro extension (percentage point change)
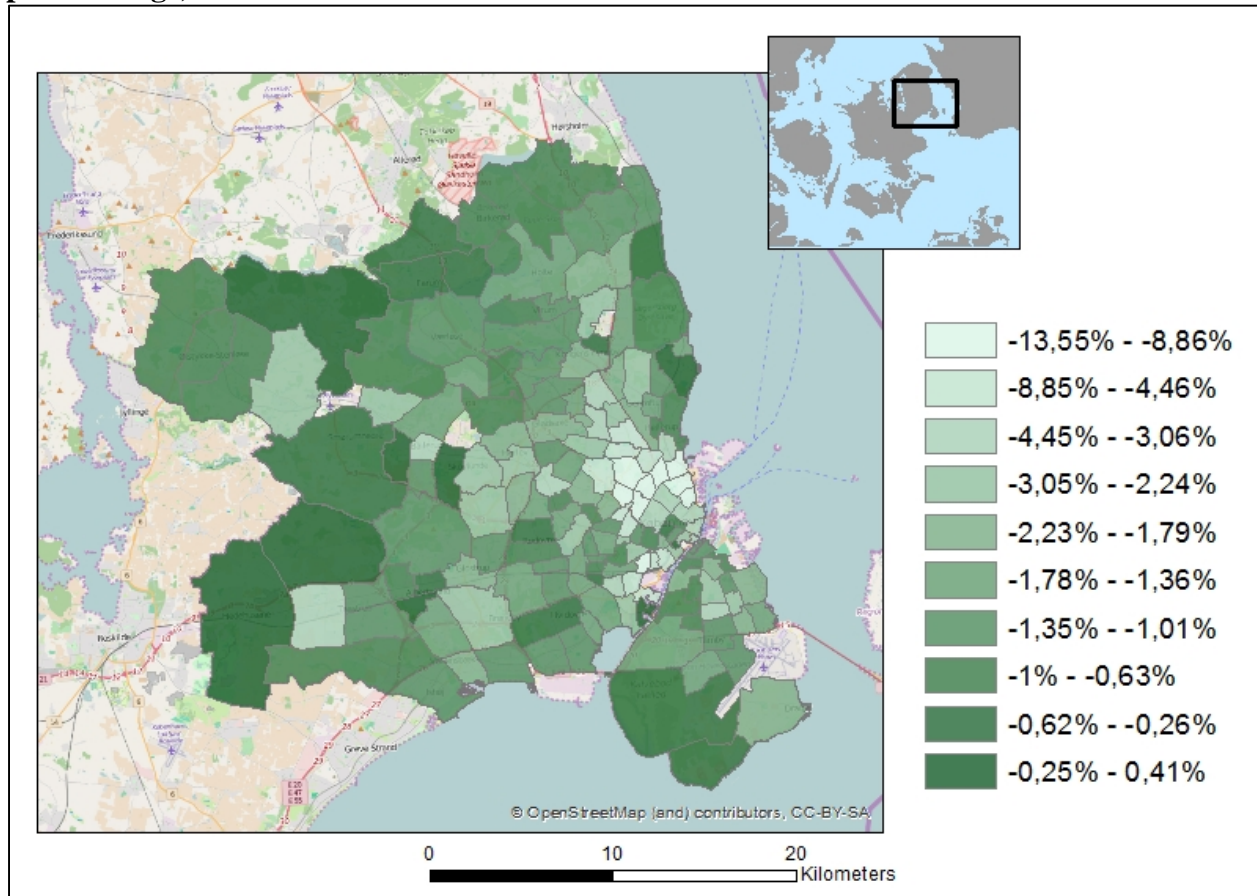
In making this assessment we treat the random parts of the utilities of each choice alternative as household-specific constants. Our computations are based on De Palma and Kilani (2003) who derive the distribution of the compensating variation when income has an impact on choice behaviour, as it clearly has in our estimated model. See the Appendix A.4 for further discussion.

Third we assess the welfare implications of the extended metro network under the contrary assumption that housing supply is completely inelastic. In this case prices will adjust so as to re-establish the equality between supply and demand. We take into account that single- and dual earner households are active on the same housing market and compute the new equilibrium prices. Prices increase in areas that become more attractive because of the extended metro network, which compensates for the initial increase in attractiveness. Similarly, areas that

Table 6: Compensating variations of the extension of the metro network

			[1] No mobility	[2] Elastic supply	[3] House prices adjust
Single earner households	All households	Average CV	11,062	12,026	11,899
		Share of income (%)	*2.8*	*3.1*	*3.0*
	Dir. affected alt. (no car)	Average CV	33,753	34,386	24,324
		Share of income (%)	*8.6*	*8.7*	*6.2*
Dual earners households	All households	Average CV	13,271	13,669	13,012
		Share of income (%)	2,1	2.2	2.1
	Dir. affected alt. (no car)	Average CV	53,156	53,413	38,641
		Share of income (%)	*8.4*	*8.4*	*6.1*
	Dir. affected alt. (one car)	Average CV	12,019	12,412	3,518
		Share of income (%)	*1.9*	*2.0*	*0.6*

[1] Note that the values of the weights do not matter for the case of the MNL.

became initially less attractive because of increased public transport now get lower housing prices. Our computations ignore the wealth effects of the house price changes.

The results for the average single (top panel) and dual earners households (bottom panel) are presented in Table 6. The figures are averages of the compensating variation for households with average characteristics that initially have chosen a particular choice alternative. For both types of households we first present the average compensating variations over all choice alternatives. Column 1 gives the compensating variation if house prices do not change and all household stay where they are. It equals slightly more than 11,000 DKK for single earner households and 2,000 DKK more for two earner households.[19] If we allow households to move, but still keep house prices constant the figures in column 2 result. The possibility to move to a choice alternative that has become more attractive than the one currently chosen (e.g by abandoning the car) causes the moderately larger welfare effect. Column 3 shows the welfare effects if house prices adjust to their new equilibrium values. This implies an additional gain for single earner households but a lower average welfare effect for the dual earners as prices increase most in the areas that are popular among this group.

The second line in the panel referring to the single earner households concerns only those choice alternatives that benefit directly from the metro extension, that is, to those alternatives in which no car is owned and a new metro station is closer than 3 km. There are 89 such alternatives. The welfare gain for households that choose these alternatives is roughly three times as large as the average. However, roughly 50% of this additional gain disappears if house prices

---

[19] 1 DKK is appr. €0.13.

increase so as to equilibrate housing supply and demand after the extension of the metro network.

The second line in the panel referring to the dual earner households also concerns households without a car that gain directly from better access the metro network (81 choice alternatives). Their welfare gain is roughly four times the average. Again, a large part of it disappears when house prices adjust. The third line in this panel refers to dual earner households with one car that live in close proximity to the new metro stations (93 choice alternatives). Their welfare gain is smaller than that of the average dual earner household, which is due to the fact that this average is determined in part by the large welfare gain of those who do not own a car.[20] Little of this gain is left after house prices adjust.[21]

7. Conclusions

In this paper we developed a model for the joint choice of residential location and car ownership, focusing on the interaction with high quality public transport. We estimated the model on register data for the Greater Copenhagen Area (GCA) and used the estimated model to simulate the impact of a planned extension of the metro network in Copenhagen, which is currently under construction. The model predicts a substantial increase in the interest of living in the centre of the area, that is, close to the extended metro network. The predicted change in the overall car ownership rate is -2.3%. The impact of the extension of the metro-network interacts significantly with the location choices of high income households, who often own two cars.

We close the paper with a brief discussion of two issues that could potentially improve the analysis presented above: the use of nesting structures and the construction of the instrument for the housing price. With respect to the former, it may be observed that some of the choice alternatives that we use share important characteristics: owning one or two cars, living in a single family house or in a particular area. The idiosyncratic utilities of these alternatives may reflect these similarities.[22] If so, the multinomial logit model is no longer appropriate. Instead a nested logit model could be formulated, where the nesting could refer to car ownership, housing type or

---

[20] Note that the figures are unweighted averages over the choice alternatives.

[21] Note (again) that the wealth effect of the change in housing prices is not included in these welfare measures.

[22] That is, the $\varepsilon_{a,h,c}^i$'s in (1) may be statistically dependent for alternatives sharing the same *a, h* or *c*.

area. Since it is difficult to choose for a particular nesting structure, the flexibility offered by the cross-nested logit model (see Vovsha, 1997; Bierlaire, 2006) is attractive. In the model the three types of correlation can be present simultaneously[23] and by estimating the model the data show the relative importance of each of them. However, estimation of (cross-)nested logit models is typically more involved than that of estimating a multinomial logit model. For instance, the likelihood function is no longer necessarily concave and the 'logsum parameters' may be difficult to estimate. A grid search over relevant values of these parameters may be a useful strategy, although probably time consuming.

With respect to the instrument for the house price we noted some ambiguity with respect to the use of exogenous characteristics of areas in the vicinity. It is not always clear that such characteristics are excluded from the utility of a particular choice alternative. For instance, amenities in contiguous areas may be easy to visit as those of one's own residential area and therefore contribute as much to the utility of living there. Bayer's procedure for computing the instrument for the price uses exogenous characteristics of all alternatives and therefore does not rely solely on those of the neighbouring areas. But it does not exclude these characteristics from the computation of the instrument which raises questions about its appropriateness. It may therefore be useful to look for other ways to compute an instrument for the price. Reynaert and Verboven (2014) have recently emphasized the importance of using approximations to the optimal instruments. Their methodology appears to be flexible enough to avoid the use of characteristics of neighbouring areas in the construction of the instrument.[24] We are currently working on the incorporation of the nesting structures and the alternative procedures for finding an instrument for the house price in models like that of the present paper.

---

[23] The cross-nested logit may be formulated as a weighted average of the various nested logit models where the weights are determined by some of the parameters to be estimated.

[24] An alternative approach may be based on Belloni et al. (2012) who propose the use of Lasso methods.

References

Anderson, M.L. 2014. Subways, strikes and slowdowns: The impact of public transit on traffic congestion. *American Economic Review*, **104**, 2763-2796.

Bayer, P., F. Ferreira and R. McMillan. 2007. A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, **115**, 588-638.

Bayer, P., R. McMillan and K. Rueben. 2004. An equilibrium sorting model of sorting in an urban housing market. NBERWorking Paper No. 10865.

Bayer, P. and C. Timmins. 2007. Estimating equilibrium models of sorting across locations. *Economic Journal,* **117**(518), 353-374.

Belloni, A., D. Chen, V. Chernozhukov and C. Hansen (2012) Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain. *Econometrica*, **80**, 2369-2429.

Berry, S. 1994. Estimating discrete choice models of product differentiation. *RAND Journal of Economics*, **25**, 242-262.

Berry, S., J. Levinsohn and A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica*, 63, 841-890.

Bierlaire, M. 2006 A Theoretical Analysis of the Cross-Nested Logit Model. *Annals of Operations Research*, **144**, 287-300.

Chamberlain, G. 1987 Asymptotic Efficiency in Estimation with Condtional Moment Restrictions. *Journal of Econometrics*, **34**, 305-334.

Dargay, J.M. 2002. Determinants of car ownership in rural and urban areas: a pseudo-panel analysis. *Ttransportation Research E*, **38**, 351-366.

de Jong, G.C. 1998. An indirect utility model of car ownership and private car use. *European Economic Review*, **34**, 971-985.

De Palma, A. and K. Kilani. 2003. (Un)conditional compensating variation in discrete choice models. Working paper, THEMA, University of Cergy-Pontoise.

Dieleman, F.M., M. Dijst and G. Burghouwt. 2002. Urban form and travel behavior: micro level household attributes and residential context. *Urban Studies*, **39**, 507-527.

Epple, D. and G.J. Platt. 1998. Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes. *Journal of Urban Economics,* **43**, 23-51.

Fernandez, R. and R. Rogerson. 1996. Income distribution, communities, and the quality of public education. *The Quarterly Journal of Economics*, **111**, 135–164.

Goodwin, P.B. 1993. Car ownership and public transport use: revisiting the interaction. *Transportation*, **27**, 21-33.

Judd, K. 1998. *Numerical Methods in Economics*. MIT.

Kahn, M.E. 2006. *Green cities*. Brookings.

Klaiber, H.A. and D.J. Phaneuf. 2010. Valuing open space in a residential sorting model of the twin cities. *Journal of Environmental Economics and Management*, **60**, 57-77.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez. 2011. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, **79**, 651_692.

Kuminoff, N.V., V.K. Smith and C. Timmins. 2013. The new economics of equilibrium sorting and policy evaluation using housing markets. Journal of Economic Literature, **51**(4), pp. 1007-1064.

Mannering, F. and C. Winston. 1985. A dynamic empirical analysis of vehicle ownership and utilization. *RAND journal of Economics*, **16**, 215-236.

Martens, K. 2004. The bicycle as a feedering mode: experiences from three European countries. Transportation Research Part D, **9**, 281–294.

Matas, A., J.-L. Raymond and J.L. Roig. 2009. Car ownership and access to jobs in Spain. *Transportation Research A*, **43**, 67-617.

McFadden, D.L. 1973. Conditional logit analysis of qualitative choice behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York.

Muth, R. 1969. *Cities and housing*. Chicago.

Nechyba, T. J. 2000. Mobility, targeting, and private-school vouchers. *The American Economic Review*, 90, 130–146.

Pyddoke, R. and C. Kreutzer. 2014. Household car ownership in urban and rural areas in Sweden 1999-2008. Working paper, Swedish National Road and Transport Research Institute.

Potoglou, D. and P.S. Kanaroglou. 2006. Modelling car ownership in urban areas: A case study of Hamilton, Canada. Working paper.

Reynaert, M. and F. Verboven, 2014. Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments. *Journal of Econometrics*, **179**, 83-98.

Rouwendal, J. 1998. On housing services. *Journal of Housing Economics*, **7**, 218-242.

Van Duijn, M. and J. Rouwendal. 2013. Cultural heritage and the location choice of Dutch households in a residential sorting model. *Journal of Economic Geography*, 13(3), 473-500.

Van Ommeren, J., D. Wentink and J. Dekkers. 2011. The real price of parking policy. *Journal of Urban Economics,* 70, 25-31.

Vovsha, P. 1997. Cross-nested Logit Model: An Application to Mode Choice in the Tel-Aviv Metropolitan Area. *Transportation Research Record*, **1645**, 133-142.

Appendix

A.1 Hedonic equations for house and apartment prices

Standardized house price has been compiled from the two separated hedonic models with area fixed effect, i.e. one for the houses and one for the apartments. We use data for three years (2006-2008) in order to get a reasonable number of sales per area. The standard (single family) house has the average size and other characteristics (for the whole GCA) that are used in the hedonic price equation for houses. Similarly, the standard apartment has the average characteristics of apartments (for the whole GCA), (which differ from the average characteristics of single family houses). The correlation coefficient for house price indexes for houses and apartments is 0.58.

Table A.1.1. Hedonic price equation for houses with area fixed effect, OLS

	Coef.	Std. Err.
Natural logarithm of square meters	0.408***	0.016
Natural logarithm of number of rooms	0.001	0.015
Dummy indicating conserved property	-0.063**	0.029
Age, years	-0.002***	0.000
Age squared, years	5.52e-06***	1.65e-06
Dummy variable indicating two toilets	0.026***	0.007
Dummy indicating two bathrooms	0.045***	0.007
Dummy variable indicating built-up (flat roof)	-0.095***	0.033
Dummy variable indicating roofing felt with pithed roof	-0.030***	0.032
Dummy variable indicating cement roof tile (incl. asbestos)	-0.052*	0.031
Dummy variable indicating cement stone	-0.013	0.032
Dummy variable indicating tiled roof	0.008	0.031
Dummy variable indicating sheets made of metallic	-0.047	0.038
Dummy variable indicating thashed roof	0.027	0.058
Dummy variable indicating cement roof tile (non asbestos)	-0.032	0.040
Dummy variable indicating PVC roofing	-0.084	0.195
Dummy variable indicating district heating	-0.016	0.047
Dummy variable indicating central heating with oil or nature gas	0.039	0.046
Dummy variable indicating central heating without oil or nature gas	-0.031	0.054
Dummy variable indicating heat pump	-0.034	0.057
Dummy variable indicating electric heating	-0.006	0.047
Dummy variable indicating stoves	0.050	0.089
Dummy variable indicating garage	0.038*	0.020
Dummy variable indicating carport	0.112***	0.033
Dummy indicating year 2007	-0.031***	0.006
Dummy indicating year 2008	-0.121***	0.007
Area fix effect	yes	
Constant	12.844***	0.086
R-squared	0.423	
Number of observations	11,509	

Notes: Dependent variable is natural logarithm of house price. The reference category associated with the dummies representing roofing material is "other materials or unknown roofing". The reference category associated with the dummies representing heating type is "other types of central heating". The reference category associated with the dummies representing years is year 2006. ***,**, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and at the 0.10 level, respectively.

Table A.1.2. Hedonic price equation for apartments with area fixed effect, OLS

	Coef.	Std. Err.
Natural logarithm of square meters	0.735***	0.009
Natural logarithm of number of rooms	0.160***	0.007
Dummy indicating conserved property	-0.019	0.013
Age, years	-0.001***	0.000
Age squared, years	3.74e-06***	5.71e-07
Dummy variable indicating two toilets	0.042***	0.009
Dummy indicating two bathrooms	0.027***	0.009
Dummy variable indicating built-up (flat roof)	0.015	0.045
Dummy variable indicating roofing felt with pithed roof	0.022	0.045
Dummy variable indicating cement roof tile (incl. asbestos)	0.024	0.045
Dummy variable indicating cement stone	0.041	0.046
Dummy variable indicating tiled roof	0.036	0.045
Dummy variable indicating sheets made of metallic	0.092*	0.047
Dummy variable indicating cement roof tile(non asbestos)	0.089*	0.050
Dummy variable indicating roofing with other materials	0.058	0.046
Dummy variable indicating district heating	0.611***	0.208
Dummy variable indicating central heating with oil or natural gas	0.616***	0.209
Dummy variable indicating heat pump	0.599***	0.210
Dummy variable indicating other types of central heating	0.797***	0.230
Dummy variable indicating electric heating	0.605***	0.209
Dummy variable indicating stoves	0.751***	0.218
Dummy variable indicating garage	0.046	0.031
Dummy variable indicating carport	0.143*	0.080
Dummy indicating year 2006	0.252***	0.004
Dummy indicating year 2007	0.133***	0.004
Areae fix effect	yes	
Constant	10.429***	0.219
R-squared	0.802	
Number of observations	18,040	

Notes: Dependent variable is the natural logarithm of apartment price. The reference category associated with the dummies representing roofing material is "glass roofing". The reference category associated with the dummies representing heating type is "no or unknown heating type". The reference category associated with the dummies representing years is year 2008. ***,**, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and at the 0.10 level, respectively.

Table A.2.1 First step IV estimation results for single earner households

	[1] Log (std. house, apartment price)	[2] Share of higher educated	[3] Employment access with public transport for no car owners
Proximity to the nearest metro station (km) * dummy variable indicating no car	0.001 (0.003)	-0.026* (0.014)	15.998*** (2.017)
Number of conserved/protected buildings per sq.km.	0.003 (0.003)	-0.010 (0.012)	-4.101** (1.722)
Distance to the CBD.	0.0003 (0.0002)	0.005*** (0.001)	-0.861*** (0.094)
Dummy variable indicating non-apartment	0.014*** (0.003)	-0.202*** (0.014)	-0.899 (2.022)
Dummy variable indicating non-apartment * dummy variable indicating one car	-0.002 (0.003)	0.004 (0.011)	3.490** (1.546)
Dummy variable indicating parking charging * dummy variable indicating one car	0.003 (0.003)	0.037*** (0.014)	-2.444 (2.002)
Social housing (share)	0.001 (0.003)	-0.079*** (0.014)	5.701*** (2.057)
Dummy variable indicating one car	-0.001 (0.002)	-0.032*** (0.009)	-77.689*** (1.319)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	*0.971**** *(0.004)*	*0.298**** *(0.018)*	*-3.844* *(2.646)*
Distance to the nearest school in 1890 (IV)	*-0.001*** *(0.0002)*	*-0.012**** *(0.001)*	*0.679**** *(0.136)*
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	*-0.0002* *(0.001)*	*-0.022**** *(0.004)*	*-15.167**** *(0.653)*
Constant	0.013*** (0.004)	0.163 (0.018)	85.424*** (2.620)
Partial R-squared	0.3877	0.2578	0.4689
No. of observations	538	538	538

Notes: standard errors in parentheses; standardized house/apartment price, and share of higher educated and employment access with public transport are instrumented; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

Table A.2.2 First step IV estimation results for dual earners households

|	[1] Log (standardized house, apartment price)	[2] Share of higher educated	[3] Employment access with public transport for no car owners	[4] Employment access with public transport for one car owners
Proximity to the nearest metro station (km) * dummy variable indicating no car	-0.0003	-0.024*	14.024***	-4.291**
	(0.003)	(0.013)	(1.803)	(1.953)
Proximity to the nearest metro station (km) * dummy variable indicating one car	-0.0001	-0.029**	-4.350**	10.529***
	(0.003)	(0.013)	(1.846)	(1.999)
Number of conserved/protected buildings per sq.km.	-0.001	-0.005	-3.606***	-2.650*
	(0.002)	(0.010)	(1.377)	(1.491)
Distance to the CBD.	-3.03e-07	0.006***	-0.677***	-0.775***
	(0.001)	(0.001)	(0.080)	(0.087)
Dummy variable indicating non-apartment	0.004	-0.244***	-5.937***	3.311
	(0.003)	(0.014)	(1.943)	(2.104)
Dummy variable indicating non-apartment * dummy variable indicating one car	0.0003	0.019*	6.043***	-2.460
	(0.003)	(0.011)	(1.438)	(1.557)
Dummy variable indicating non-apartment * dummy variable indicating two cars	0.0002	-0.035**	6.033***	-1.524
	(0.004)	(0.014)	(2.060)	(2.230)
Dummy variable indicating parking charging * dummy variable indicating one car	0.00001	0.035***	-0.372	12.941***
	(0.003)	(0.013)	(1.818)	(1.969)
Dummy variable indicating parking charging * dummy variable indicating two cars	0.001	0.078***	-3.083	-2.995
	(0.006)	(0.026)	(3.690)	(3.996)
Social housing (share)	-0.002	-0.054***	2.936*	6.884***
	(0.003)	(0.012)	(1.710)	(1.851)
Dummy variable indicating one car	9.85e-06	-0.016*	-78.845***	76.091***
	(0.002)	(0.009)	(1.313)	(1.421)
Dummy variable indicating two cars	0.00003	-0.009	-78.778***	0.921
	(0.003)	(0.013)	(1.822)	(1.972)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	0.994***	0.349***	0.153	-4.137
	(0.004)	(0.017)	(2.379)	(2.576)
Distance to the nearest school in 1890 (IV)	-0.0002	-0.008***	0.513***	0.556***
	(0.0002)	(0.001)	(0.112)	(0.121)
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	0.0001	-0.032***	-14.112***	-0.422
	(0.001)	(0.004)	(0.628)	(0.680)
*Distance to the nearest train station in 1939 *dummy Indicating one car (IV)*	-0.0001	-0.036***	0.143	-14.971***
	(0.001)	(0.004)	(0.575)	(0.623)
Constant	0.001	0.114***	84.472	7.369***
	(0.004)	(0.016)	(2.250)	(2.436)
Partial R-squared	0.2865	0.1960	0.3915	0.3619
No. of observations	636	636	636	636

Notes: standard errors in parentheses; standardized house/apartment price and, share of higher educated and employment access with public transport are instrumented; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

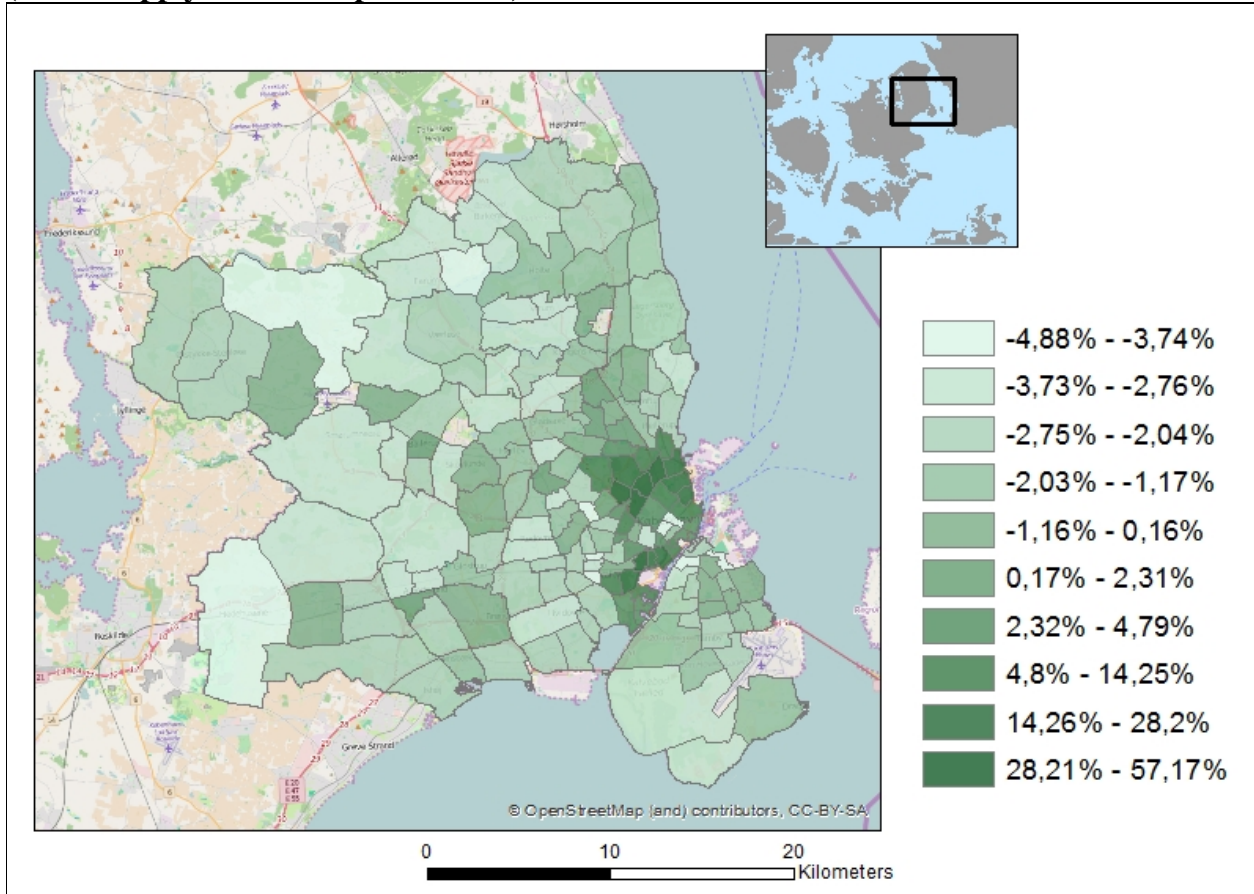**Map A.1. Pct. change in population of the single earner households in the GCA caused by the metro extension (elastic supply and house prices fixed)**
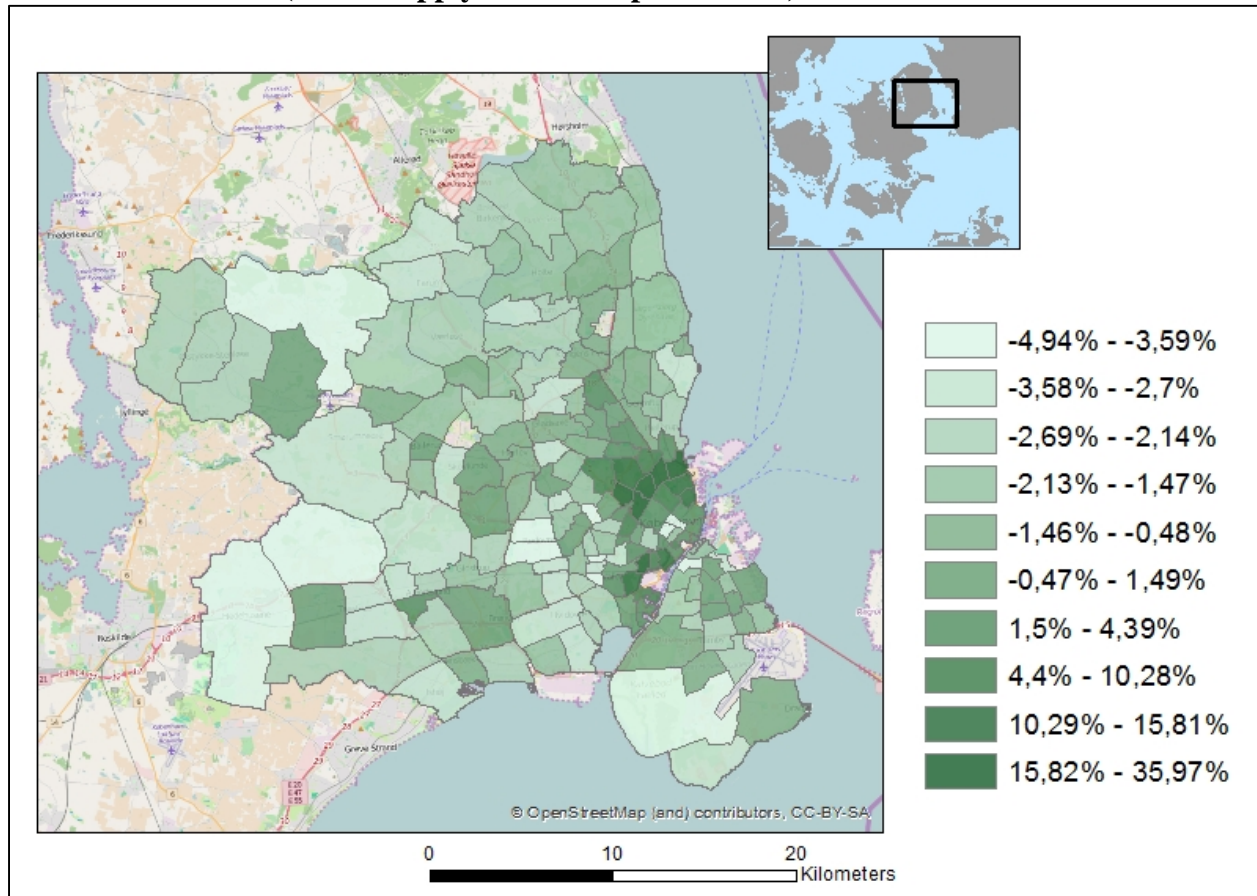
Map A.2. Pct. change in population of the dual earners households in the GCA caused by the metro extension (elastic supply and house prices fixed)
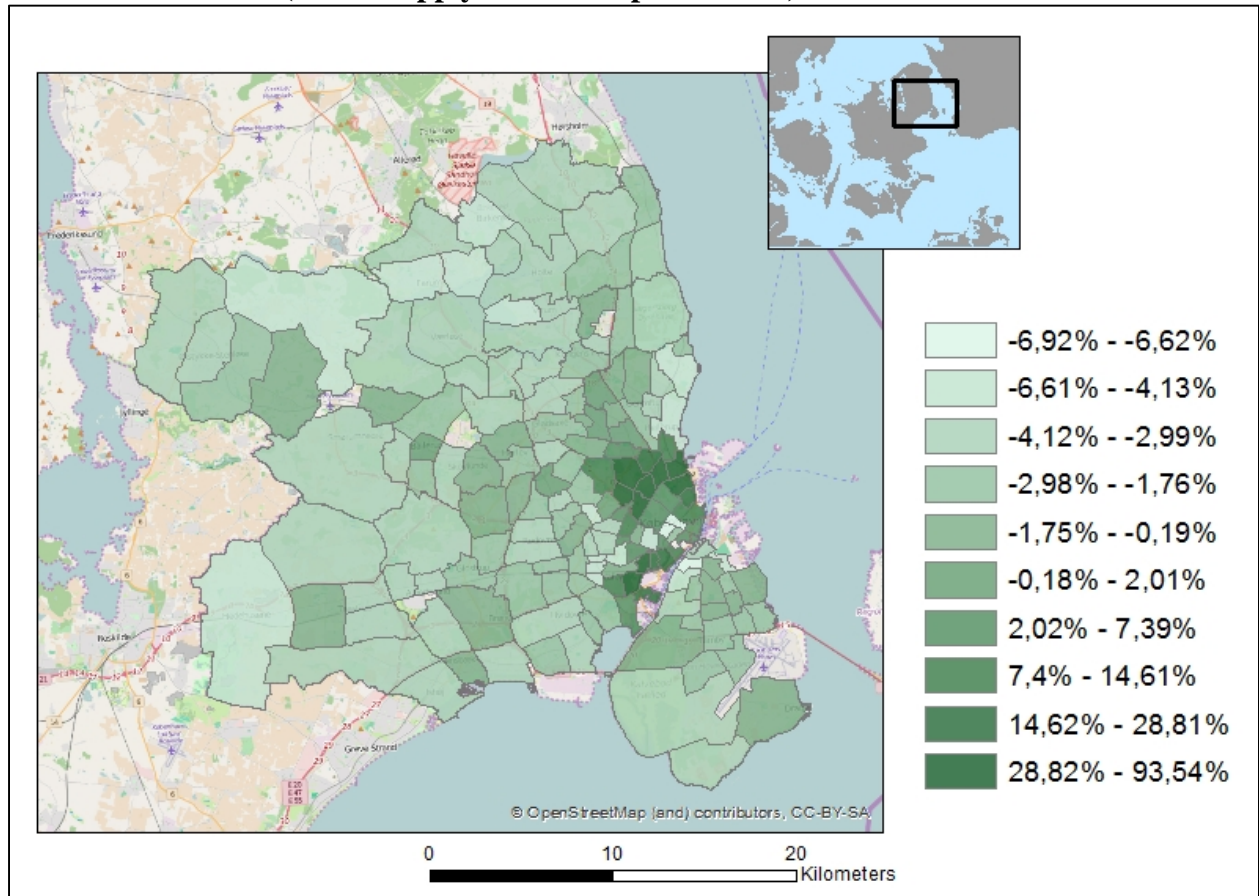
Map A.3. Pct. change in the household income in the GCA caused by the metro extension (elastic supply and house prices fixed)

Map A.4. Pct. change in the share of higher educated in the GCA caused by the metro extension (elastic supply and house prices fixed)
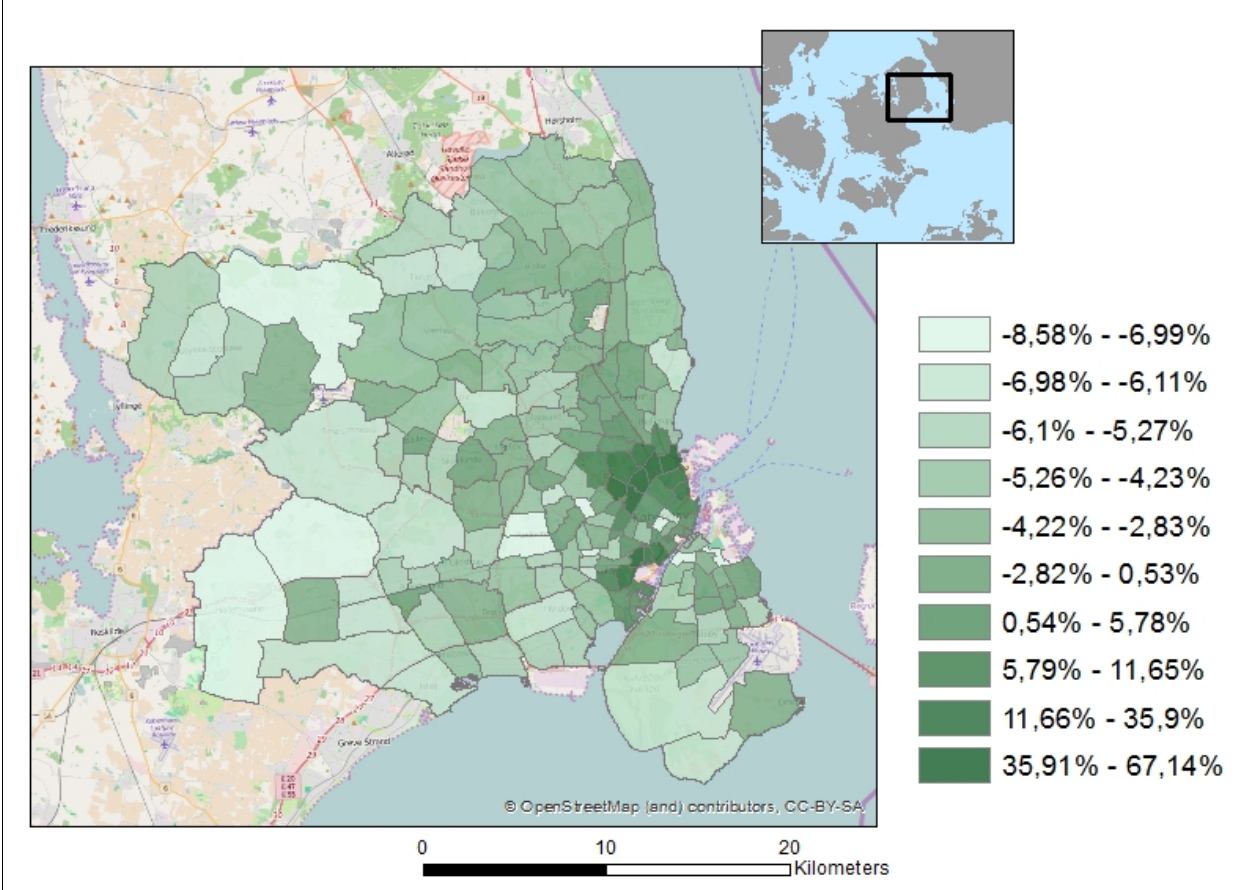
Map A.5. Pct. change in the share of higher educated in the GCA caused by the metro extension (fixed supply and house prices adjust)
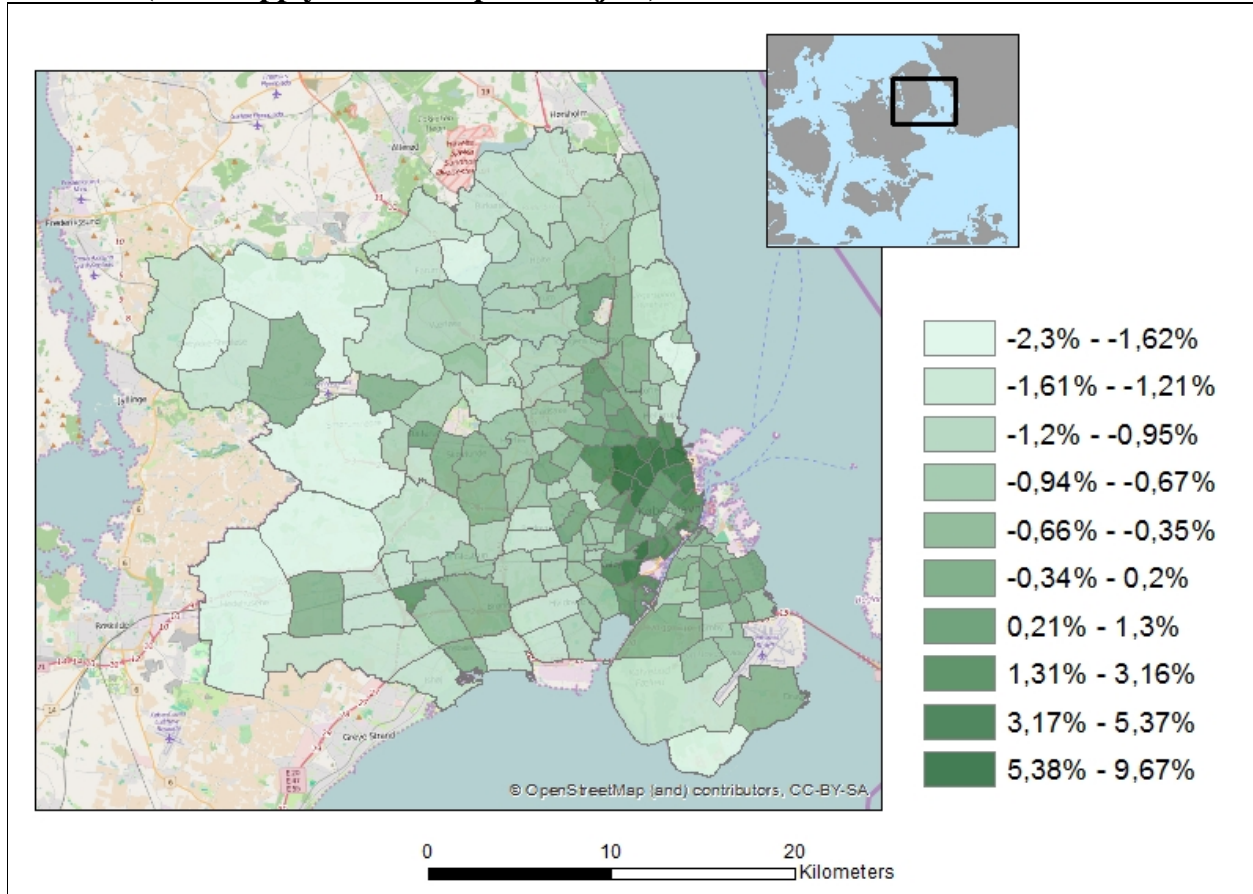
Map A.6. Pct. change in the number of households with children in the GCA caused by the metro extension (fixed supply and house prices adjust)
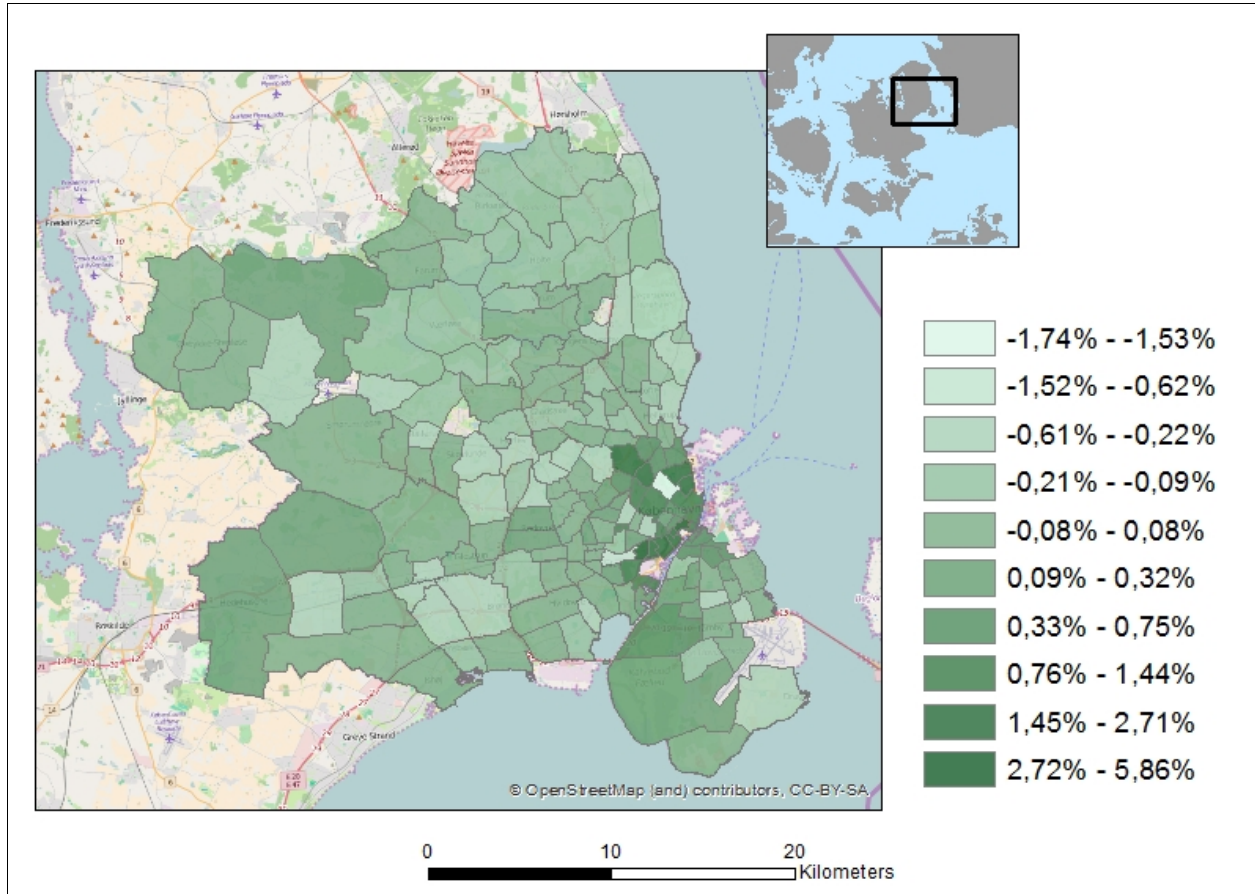
A.4 Computing the value of the extension of the metro network

We are looking for the compensating variation in income. The discussion below refers to the one-earner households, but the treatment for the two-earner households is completely analogous. The compensating variation is negative of the change in income that brings the household back to the utility it experienced before the extension of the metro network. The minus sign is there to get a positive number in case welfare increases due to the extension of the metro network, and a negative one if it decreases.

Our starting point is (9), which we repeat here as:

$$v_{a,h,c}^i\left(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i, Z^i\right) = \alpha_1^i apt_a + \alpha_2^i amt_a + \alpha_3^i d_c +$$

$$\beta_1^i d_h + \beta_2^i P_{h,a} + \beta_3^i X_a + \left(\gamma_1^i apt_a + \gamma_2^i amt_a + \gamma_3^i d_h + \gamma_4^i X_a\right)d_c + \xi_{a,h,c} \qquad \text{(A1)}$$

The extension of the metro network implies that the values of two variables change: apt_a and amt_a. If the house price adjusts to a new equilibrium $P_{h,a}$ will also change. We will denote these changes as Δapt_a, Δamt_a and $\Delta P_{h,a}$, respectively. All other characteristics of the alternatives remain constant.

We will first look at the compensating variation conditional upon the choice of a particular alternative. That is, we want to compute the change in income that is necessary to maintain a given level of utility after the extension of the metro network has been realized. To be able to do this, we recall that all the coefficients in (A1) are linear functions of the logarithm of income as specified for the α's in (10), which we repeat here as:

$$\alpha_j^i = \tilde{\alpha}_j^0 + \tilde{\alpha}_j^1 \ln y^i + \sum_{l=1}^{L} \tilde{\alpha}_j^{l+1} Z_l^i \tag{A2}$$

While similar equations hold for the other coefficients. Substitution of these equations in (A1) and rearranging terms allows us to rewrite (A1) as:

$$v_{a,h,c}^i \left(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i, Z^i \right) =$$

$$\left(\tilde{\alpha}_1^1 + \tilde{\gamma}_1^1 d_c \right) \ln\left(y^i \right) apt_a + \left(\tilde{\alpha}_1^2 + \tilde{\gamma}_1^2 d_c \right) \ln\left(y^i \right) amt_a + \tilde{\beta}_2^1 \ln\left(y^i \right) P_{h,a} + \tag{A3}$$

$$\theta_1 apt_{a_1}^i + \theta_2 amt_{a_1}^i + \theta_3 P_{h,a} + \mu \ln\left(y^i \right) + \rho$$

The second line of (A3) gives the cross terms of log income and the variables that change due to the extension of the metro network; the third line collects the terms that are linear in these variables or in log income and the terms in which neither of these variables occur.

The change in log income, $\Delta \ln\left(y_{a,h,c}^i \right)$ that keeps the utility of a household in (a, h, c) constant after the extension of the metro network can now be determined as:

$$\Delta \ln\left(y_{a,h,c}^i \right) = -\frac{\left(\theta_1 + \tilde{\alpha}_1^1 \ln y^i + \tilde{\gamma}_1^1 d_c \right)\Delta apt_a + \left(\theta_2 + \tilde{\alpha}_1^2 \ln y^i + \tilde{\gamma}_1^2 d_c \right)\Delta amt_a + \left(\theta_3 + \tilde{\beta}_2^1 \ln y^i \right)\Delta P_{h,a}}{\left(\tilde{\alpha}_1^1 + \tilde{\gamma}_1^1 d_c \right)\ln(y^i) apt_a + \left(\tilde{\alpha}_1^2 + \tilde{\gamma}_1^2 d_c \right)\ln(y^i) amt_a + \tilde{\beta}_2^1 \ln(y^i) P_{h,a} + \mu} \tag{A4}$$

In this equation the levels of apt_a, amt_a and $P_{h,a}$ that occur in the numerator refer to the situation before the extension of the metro network.[25] The change in income itself that corresponds with the change $\Delta\ln\left(y_{a,h,c}^i\right)$ in log income can be determined as:

$$\Delta y_{a,h,c}^i = y^i \left(1 - e^{\Delta\ln\left(y_{a,h,c}^i\right)}\right) \tag{A5}$$

If households would not change their location, housing type or car ownership position, this would suffice to compute the welfare impact of the change in the metro network. However, the changes in the utilities of the choice alternatives that occur as a consequence of the extension of the metro network will induce some households to change their residential location, car ownership position or perhaps even their housing type.

To take this into account as well, we use the results of De Palma and Kilani (2003). We start by defining Δy_+^i as the largest of the conditional compensating variations:

$$\Delta y_+^i = \max_{a,h,c}\left\{\Delta y_{a,h,c}^i\right\} \tag{A6}$$

DePalma and Kilani (2003) show that the unconditional compensating variation is always between the minimum and the maximum of the conditional compensating variations.

To be able to compute it exactly we define functions $w_{a,h,c}(r)$ as follows:

$$w_{a,h,c}(r) = v_{a,h,c}^i\left(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i - r, Z^i\right) \tag{A7}$$

where the values of apt_a, amt_a and $P_{h,a}$ refer to the situation with the extended metro network.

Clearly, if $r = \Delta y_{a,h,c}^i$, $w_{a,h,c}(r)$ is equal to the utility household I experienced before the extension of the metro network. Now define: $w_{a,h,c}^*(r) = \max \left\{w_{a,h,c}(r), w_{a,h,c}\left(\Delta y_{a,h,c}^i\right)\right\}$ and

let $\pi_n(w^*)$ denote the logit choice probabilities defined on the basis of these utilities:[26]

---

[25] The income level that occurs in the denominator refers (from the derivation of (A.3)) to the situation after the extension of the metro network, but we assume throughout that household income does not change because of the extension of the metro network.

[26] This equation is analogous to (2) for the generator function of the multinomial logit model.

$$\pi_{a,h,c}\big(w^*(r)\big) = \frac{e^{w^*_{a,h,c}(r)}}{\sum_a \sum_h \sum_c e^{w^*_{a,h,c}(z)}} \tag{A8}$$

This brings us – finally – in the position to define the expected value of the unconditional expected value of the compensating variation (cv) : [27]

$$E[cv] = \Delta y^i_+ - \sum_a \sum_h \sum_c \int_{\Delta y^i_{a,h,c}}^{\Delta y^i_+} \pi_{a,h,c}\big(w^*(r)\big)dr \tag{A9}$$

De Palma and Kilani (2003) suggest to evaluate this expression through simulation. Since the integral has to be computed is one-dimensional, Gaussian quadrature is also feasible and we used this technique (see Judd, 1999, chapter 7).

---

[27] It can be shown that (A8) reduces to the change in the logsum when utility is linear in income.