Does Improving Public Transport Decrease Car Ownership?

Evidence from a Residential Sorting Model for the Copenhagen Metropolitan Area

Mulalic, Ismir ; Rouwendal, Jan

Document Version
Accepted author manuscript

Published in:
Regional Science and Urban Economics

DOI:
10.1016/j.regsciurbeco.2020.103543

Publication date:
2020

Does improving public transport decrease car ownership?

Evidence from a residential sorting model for the Copenhagen metropolitan area

Ismir Mulalic[1] and Jan Rouwendal[2]

Abstract

Car ownership is lower in urban areas, where public transport is of high quality. This suggests that better public transport offers the possibility to relieve the many problems (congestion, pollution, and parking) associated with the presence of cars in urban areas. To investigate this issue, we develop a model for the simultaneous choice of residential location and car ownership by households, and estimate it on Danish data, paying special attention to accessibility of the metro network. We use the estimated model to simulate the impact of an extension of the metro network. We show that for the Greater Copenhagen Area an extension of the metro network decreases car ownership by 2-3%, while the average compensating variation is approximately 3% of household income.

Keywords: car ownership, public transport, residential sorting, cost-benefit analysis.

JEL codes: H4, R4, R1, D1.

Acknowledgement
Earlier versions of the paper have been presented at ITEA conference in Oslo, the workshop on New Developments in Spatial Sorting in Copenhagen, the 62st Annual North American Meetings of the Regional Science (RSAI) and 10th Meeting of the Urban Economics Association, the 56th ERSA Congress, University of Cergy-Pontoise, University of Copenhagen, University of Edinburgh, SFI - The Danish National Centre for Social Research, University of Valencia, TU Dresden, Copenhagen Business School and at Vrije Universiteit Amsterdam. The authors thank Ninette Pilegaard, Thomas Crossley, Gerard de Jong, Dennis Epple, Bo Honoré, Nicolai Kuminoff, Lars Nesheim, Henry Overman, Chris Timmins, two anonymous referees and Gabriel Ahlfeldt (the editor) for useful suggestions on earlier drafts. They are grateful to Statistics Denmark for providing the data. Research support from Kraks Fond – Institute for Urban Economic Research is gratefully acknowledged. The usual disclaimer applies.

[1] Department of Economics, Copenhagen Business School, Porcelænshaven 16A, 2000 Frederiksberg, Denmark.
[2] Department of Spatial Economics, VU University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, and Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands.

1. Introduction

Public transport is, potentially, an important substitute for the car. If parking spaces are difficult to find, or if parking is expensive, as is the case in many city centres, the benefits of owning a car may still be lower (Van Ommeren, Wentink, & Dekkers, 2011; Ostermeijer et al., 2019). Moreover, the availability of many amenities at walking distance in dense 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).[1]

In this paper we investigate this issue by developing and estimating a choice model in which households simultaneously decide on residential location and car ownership, while taking into account the availability of public transport, and use the estimated model to simulate the effects of an extension of the public transport network. The interaction between car ownership and public transport is of considerable interest because road congestion is an important problem in many urban areas, and recent research suggests that the presence of public transport may have an important impact on urban congestion and pollution (Anderson, 2014; Bauernschuster, Hener, & Rainer, 2017). Moreover, pollution by cars associated with health problems and global warming to which CO_2 emissions of cars contribute, 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.

Substitution between cars and public transport presumably depends on the price and availability of both. While prices are a standard element in economic models, availability is often taken for granted. This may be problematic for public transport as the density of railroad stations, metro stations and bus stops, and hence the accessibility of employment, varies with population density and shows substantial differences over space. This suggests that availability of public transport has some of the characteristics of a local public good and is associated with Tiebout sorting (see e.g. Epple and Sieg, 1999).

The close connection between residential location and availability of public transport suggests – quite strongly – that an analysis of the substitution between car and public transport should be integrated with that of location choice. Accordingly, it is the purpose of this paper to

[1] These effects are also found *within* urban areas. We will also see later that the share of car-owners is lower in the central city of Copenhagen compared with the suburbs.

develop a simultaneous model for residential location choice and car ownership while taking the spatial aspect of the availability of public transport properly into account. Doing so means that we fill a gap in the current literatures in economics, planning, geography and transportation.

The interplay between public transport and car ownership, has received little attention in economics.[2] There is an older literature looking at car ownership (see, for instance, Mannering & Winston, 1985) that pays marginal attention to it. For instance De Jong (1990) 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 has been a substantial interest in the relationship between traffic and the built environment among planners (see Ewing & Cervero, 2010; Kenworthy & Laube, 1999 for reviews), but much less for car ownership and its relationship to the provision of public transport. There also exists a small geographic literature on the impact of urban form and urban amenities on car ownership. See for instance Dieleman, Dijst, & Burghouwt (2002); and Potoglou & Kanaroglou (2008). In this literature car ownership is estimated as a binomial choice, conditional on the characteristics of the residential area. For instance, Potoglou & Kanaroglou (2008) find that mixed land use is associated with a lower share of car owners.

It is surprising that even in transportation the impact of accessibility of public transport on car ownership does not seem to be an intensively studied topic. It has been addressed in an older literature (see for instance Goodwin, 1993), but appears to have been almost neglected in recent decades. Matas, Raymond, & Roig (2009) is an exception.

In this paper we develop a structural model in which we treat the choice of car ownership and residential location as a simultaneous decision, that depends on the availability of public transport. That is, we assume that households looking for a residential location contemplate to live in a particular area while owning a car or not. More specifically, our model extends a logit-based 'horizontal' residential equilibrium sorting model (see Kuminoff, Smith, & Timmins 2013) with car ownership. From the viewpoint of transportation economics it may be said that our model extends a standard car ownership model with residential location choice. The choice alternatives in our model are combinations of residential areas, single or multi-family housing and car

[2] See, for instance, Anas et al. (1998), Glaeser and Kahn (2004) and Huang et al. (2018) where the impact of cars on the structure of modern cities is recognized, but the connection with public transport receives little attention.

ownership. Interactions between characteristics of the residential areas and car ownership are the focus of interest. The residential area characteristics include public transport related amenities as well as more traditional urban amenities. We model car ownership in a relatively simple way by focusing on the number of cars per household, while ignoring differences between car brands, makes, et cetera.

The model is estimated on administrative data for Copenhagen, the capital of Denmark. The choice of a European city appears natural, because of the availability of public transport that has sufficient quality to offer a real alternative to the car, at least in some locations. Related to this is the fact that there is a substantial share of households not owning a car, whereas in the US car ownership is the default.[3]

To get an idea of how expansion of public transport would affect car ownership and residential location choice, 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 implied by our model.

The results we report provide important information for cost-benefit analyses of extending subway systems that are often neglected in practice[4]: the potential impacts of changes in demographic composition of neighbourhoods on traffic congestion, pollution of fine particles, CO_2 emissions, and the number of traffic accidents. Although the present paper makes no attempt to evaluate these externalities, it provides important ingredients for the necessary computations.

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 Greater Copenhagen Area (GCA)). In section 3 we present and discuss the theoretical model and the specification we use in our empirical work. Section 4 reports the estimation results. In Section 5 we simulate the response of an extension of the metro network. Section 6 concludes.

[3] More than 90 % of U.S. households in 2017 had at least one car at their disposal (McGuckin and Fucci, 2018). Moreover, there are on average 1.9 vehicles per household in the same year.

[4] See, for instance, Sieg, Smith, Banzhaf, & Walsh (2004) for an analysis that is similar in spirit to that of the present paper by taking into account residential sorting, although it concentrates on air pollution and uses a different analytical framework.

2. Data and descriptives

2.1. The Greater Copenhagen Area

The Greater Copenhagen Area (GCA), the Danish capital, is part of the island Zealand.[5] 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). In the dataset, we divide the GCA into 166 smaller areas that were originally designed for detailed traffic modelling.[6] It is a fair simplification to claim that the GCA constitutes a single spatial labour market as commuting by car between any pair of locations within its boundary is possible.[7] This suggests that workers consider the whole area when looking for a job. Commuting from GCA to other parts of Zealand is negligible, whereas commuting flows inside GCA are relatively large.

Car ownership in Denmark is extremely expensive compared to international standards due to taxation. The purchase-tax of a car is 105% for the value of the car below app. € 10.500 and 180% of the value of the car above.[8] In addition there is an annual ownership tax of app. € 500 (300-900) depending on the characteristics of the car. Consequently, car ownership in Denmark is low relative to other comparable countries (0.81 cars per household in Denmark, 0.71 cars per household in the 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

[5] The geographical area of GCA is rather small (615.7 km^2). See the dark area in map A.1 in the Appendix A.1.

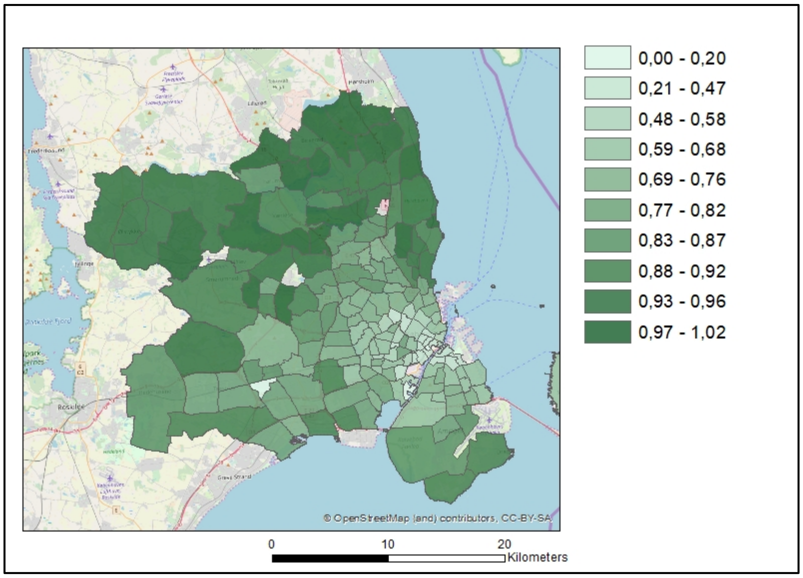
[6] They consist of a few adjacent church parishes. In Denmark a parish is an administrative area consisting of several villages or localities originating from the middle ages. From 1841, the parishes were established as administrative areas and remained in use ever since. Only few alterations to the parishes were made since 1841.

[7] We have also information on travel times between the considered zones for 2008 from the Danish National Travel Model. The travel times include congestion delays and waiting and transition times for public transport. The average 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 travel time with public transport is about 48 min., and the maximum is almost 2 hours (112.7 min.).

[8] For example, the retail price of a luxury car such as e.g. VW Passat is about € 24,000, the purchase tax is about € 29,000, and the VAT is about € 6,000. The list price is then € 59,000. The disposable income for an average household in Denmark in 2010 was about € 50,000. Moreover, car owners have to pay the vehicle excise duty, and annual insurance premium. Mulalic and Rouwendal (2015) show that the mean annual total expenditure associated with ownership and use of a new car purchased in 2004 and used in the period 2004-2008 is about € 11,000.

of course public transport but a bike is also a common mode of transport, especially in Copenhagen and other bigger cities, and among younger people. Map 1 shows that the share of car-owners is, as expected, increasing with the distance from the city centre which is the core of the regional public transport network and where many amenities are available at walking distance. The map also shows a substantial difference between car ownership in the remoter areas in the south that are populated with relatively less wealthier and educated households, where it is relatively low, and in the north on the GCA that is considered as highly attractive, where it is close to 100%. The high cost of car ownership implies that car ownership is often reconsidered when households change residence. Households thus experience an active trade-off between car ownership and housing expenditures.

Map 1: Car ownership in the GCA (number of cars per household)



2.2. Selection of sample

The equilibrium sorting model is estimated on data derived from the Danish administrative registers on the population of both households and vehicles in Denmark for all households with residence in the GCA for the year 2008. We use two main sources. The first is the vehicle license plate register, which contains the vehicle identification number, vehicle attributes, date of registration and owner identification number. The second primary data source is the household register, which contains detailed demographic data at the calendar year-level. These data include the number of members of the household, ages, genders and the highest educational levels obtained of these members, residence address, and income[9] of the household members (including transfers). We combine the data from the various sources to create a final dataset.

We use a 20% sample of the GCA population living in *owner-occupied housing*. 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.[10] The share of owner-occupied residences constitutes just over 50% of the housing stock. The model focuses strictly on the location choices of households that are active both on the labour and the housing market.[11] 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). When estimating model we assume that the supply of owner-occupied residences is fixed, as seems appropriate because most of the area is already built-up and open space that is left is usually protected. However, in the simulations we explore the implications of extending the housing stock in response to market forces.

We distinguish between single earner households (66,012) and dual earner households (87,330) and estimate separate models for these two groups, because these household types are different in many respects.[12] Within these models we distinguish between owning 0 and 1 car for single-earner households and between owning 0, 1 and 2 cars for the dual-earner households.[13]

[9] Information about household income is based on third-party reporting (includes both reporting from firms and banks, mortgage institutions, brokers, etc.) and is considered highly reliable.

[10] The market for rented housing in Denmark is strictly regulated in many ways. 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. Since estimation on a subset of alternatives does not bias the results of a logit model (McFadden, 1978), ignoring the rental part of the housing market is not problematic.

[11] We exclude owner-occupier households when all adult members are either student, unemployed, retired or otherwise inactive on the labour market (23.8% of the population as a whole, 34.0% of the population of owner-occupiers). The majority of these inactive households are pensioners (89.93%). Note that households in which the adult members have reached retirement age or studying are included in our sample when at least one of the two persons in such a couple is still active in the labour market.

[12] For instance, Gutiérrez-i-Puigarnau, Mulalic, & Van Ommeren (2016) show that in Denmark the causal effect of household income on commuting distance is larger for single-earner than for dual-earner households.

[13] The share of two car owners among single-earner households is very low (2.1%).

The total choice set includes 538 and 636 elements for single-earner households and dual-earner households, respectively.[14]

2.3. Household heterogeneity and amenities

To account for the household heterogeneity we include the following socioeconomic variables: i) age (and square of age), ii) three dummy variables indicating the highest education level obtained, 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 the partner.

Table 1. Household characteristics

	Single-earner households		Dual-earner 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
Single family house 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 includes: basic school, general upper secondary school, vocational upper secondary school and vocational education; medium education 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. 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. Car-ownership is also higher for dual earner households and they hold a larger share of higher educated.

[14] 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 expect the different types of households to have different preferences for urban amenities. Table 2 shows the summary statistics for the considered urban amenities.[15]

Table 2. Descriptive statistics for 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	0.001
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.

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. For instance, bus stops are always close by. However, travel times by public transport can be considerably larger than those for cars, in particular between suburban areas. In order to account for these differences we include a measure of employment access (EA) by public transport. For area a $EA_a = \sum_{a'} J_{a'} \, e^{-\delta \, d_{a\,a'}}$, where the summation runs over all locations a', J is employment expressed in full time job equivalents and d denotes distance measured in travel time minutes by public transport. We set δ to 0.05.[16] This implies, for instance, that jobs 'around the corner' have a weight 1, while a job at the distance of 120 min (the max) has weight 0.0025. A job at a distance of 48 min (the mean) has weight 0.09.[17]

---

[15] All households have a universal access to childcare institutions and primary schools. Consequently, the availability of childcare and schools are not central for the residence location decision in Denmark. Moreover, there is no variation in this variable in our sample, so it is not useful in the model estimation.

[16] This decay parameter is within the range of recent estimates in the literature (Ahlfeldt and Wendland, 2016). Moreover, Ahlfeldt (2011) convincingly argues that the employment access measures are successful in empirically establishing a relationship between housing prices and the spatial distribution of employment. Jacob et al. (2019) shows that it is important to consider heterogeneity in measures that include commuting. For a discussion of accessibility measures, see for instance, Spiekermann et al. (2015).

[17] Using this accessibility measure implies that we do not pay specific attention to the location of the current job of the workers. The reason is that household members more frequently change job than house, which makes it likely that they do not only take into account their present commute when choosing residential location. Job mobility is extremely high and even the highest in Europe. This applies to most categories of workers and is not caused by a minor share of (unskilled) workers (Mulalic et al., 2015). Moreover, due to steep housing transaction taxes and rent control, residential mobility rates are moderate and substantially less than for example in the UK and the US. Note also that

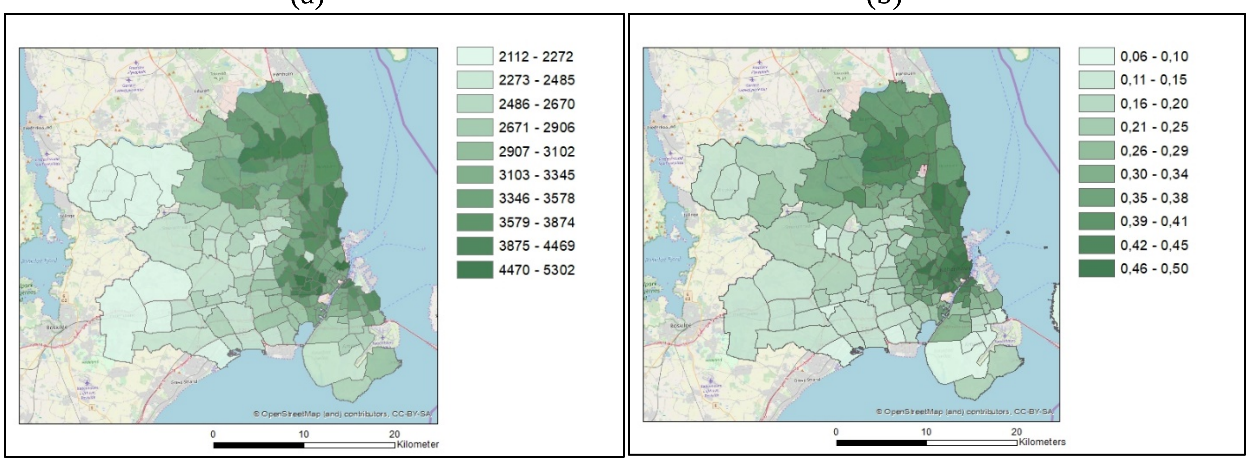
Because of its high frequency, greater reliability and better transportation quality, the significance of access to the metro network is unlikely to be captured completely by its impact on employment accessibility. We capture this by introducing the proximity to the nearest metro station as a separate variable. Specifically, we model proximity to the nearest metro as the average of a linearly decreasing function of the distance from each address in the area to the nearest metro station, assuming a threshold of 3 km:

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

Although the choice of the threshold is somewhat arbitrary, there is evidence that many bikers are willing to travel at most this distance for multimodal commute trips using bike and public transport.[18] Note that biking is very popular in Denmark. Most of the Danes are able to bike and there are facilities for parking bikes at most metro stations.

Map 2. Weighted average standardised housing price in the GCA (1000 DKK) (panel a) and the share of higher educated (panel b)

(a) (b)



Notes: standardized house price has been compiled from the two separated hedonic models with area fixed effect, i.e. one for the single family houses and one for the multi-family houses. Estimation results of these hedonic price functions are reported in the online appendix.

the choice of the work location is likely endogenous -- e.g. Paetzold (2019) shows that commuting subsidies increases the length of commute --, which would imply an endogeneity problem that can perhaps best be solved by developing a simultaneous model for residential and location choice, which is outside the scope of the present paper.

[18] The majority of bike-and-ride users travel up to 3 km to a public transport stop (Martens, 2004). We have also as an additional sensitivity analysis estimated a model with a threshold of 2 km; see section 4.4.

We model the housing market by including an area price index for a standard house (see section 3.1), 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 houses and the other for multifamily houses.[19] Map 2.a 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.

We also include the share of higher educated people 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.[20] Map 2.b shows the share of higher educated in the sample distributed over the considered areas. It is interesting to notice the similarity over maps 2.a and 2.b. Map 2.b 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. It is however 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. We include for similar reasons the share of social housing relative to the total number of houses in the neighbourhood: households may have preferences for (or against) living in the proximity of social housing. It has been shown in the literature that the concentration of historical buildings is important for household location choice (van Duijn & 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). Finally, we include a dichotomous variable indicating whether curb

[19] Estimation results of these hedonic price functions are provided in the online appendix.

[20] 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).

side parking in the area is subject to charges because they may reduce the benefits of owning a car while living in that area (Van Ommeren, Wentink, & Dekkers, 2011).[21]

3. The model

This section presents the theoretical model that underlies the empirical analyses. We introduce the model for single-earner households. The setup for dual-earner households is similar, as we discuss in 3.6 below.

3.1. Choice alternatives

The model we use 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 urban amenities. These characteristics of the area determine the value of having a car while living there. Following this reasoning we develop a discrete choice model in which the choice alternatives are combinations of car ownership and residential areas. A household thus considers living in a residential area with and without having a car and chooses the alternative that offers the highest utility. Car ownership is included as a simple indicator that takes on the values of 0 and 1.[22]

We model the demand for housing using the concept of housing services. Housing services are available at a given price per unit that is specific for the residential area. Households determine their optimal number of units consumed by choosing from the housing stock and, if necessary, adjust to the desired number of housing services (see Muth, 1969; Rouwendal, 1998; Epple & Platt, 1998).[23] However, the neglect of the durable aspects of housing that are difficult to change 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

[21] 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.

[22] This means that we ignore heterogeneity in car types and makes, car sharing and carpooling. We also do not model car use for commuting or other purposes.

[23] Using a neighborhood, instead of single house as the unit of choice 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 (low income) households.

distinguish between these two types of houses, while maintaining the "Muth-framework" for each of the separate stocks of these two types of housing.

In summary, we have for each area in principle four choice alternatives that are determined by the car ownership (yes or no) and housing type (single vs multifamily) decisions. Choice alternatives are therefore defined by three dimensions: area ($a = 1 \ldots n$), house type ($h = s, m$), and car ownership ($c = 0,1$).

3.2. Utility

Indirect 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 natural logarithm of 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 or 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 for household i is:

$$v^i_{a,h,c}\left(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i, Z^i\right) = \alpha^i_1 apt_a + \alpha^i_2 amt_a + \alpha^i_3 d_c +$$

$$\beta^i_1 d_h + \beta^i_2 P_{h,a} + \beta^i_3 X_a + \left(\gamma^i_1 apt_a + \gamma^i_2 amt_a + \gamma^i_3 d_h + \gamma^i_4 X_a\right) d_c + \xi_{a,h,c}\,. \qquad (1)$$

Utility is thus 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 superscript indicates that the coefficients are functions of household characteristics, as will be discussed below. Equation (1) 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. 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 negative. 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.[24] The final term was originally proposed in Berry, Levinsohn, & Pakes (1995) in the context of discrete choice models for car type choice. Bayer, Ferreira, & McMillan (2007) used it in the context of residential sorting behavior and we follow them here. Incorporating this term is helpful in fitting the model[25] 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:

$$\alpha_j^i = \tilde{\alpha}_j^0 + \tilde{\alpha}_j^1 \ln y^i + \sum_{l=1}^{L} \tilde{\alpha}_j^{l+1} Z_l^i \quad (2)$$

and analogous expressions for the βs and γs. Note that for the coefficients with a tilde, the superscript 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.

The total utility attached to a choice alternative is the sum of the deterministic part, discussed thus far, and a random part ε_j^i. The assumption that these random parts are independent and identically Extreme Value Type I distributed leads to the multinomial logit model (McFadden, 1973).

3.3. Estimation

To estimate the model we use a two-step procedure adapted from Berry et al. (1995) and Bayer et al. (2007). We substitute (2) and the analogous expressions for the βs and γs into (1) and write the

[24] 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).
[25] It implies the use of an alternative specific constant and this guarantees that the share of households choosing a particular alternative as predicted by the estimated model equals the observed share of households doing so in the data. See Berry et al. (1995).

result as the sum of the average utility $\delta^0_{a,h,c}$ of the alternative (that only includes the coefficients $\tilde{\alpha}^0_j$ $\tilde{\beta}^0_j$ $\tilde{\gamma}^0_j$ 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 multinomial logit (MNL) model.

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

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

$$\tilde{\beta}^0_1 d_h + \tilde{\beta}^0_2 P_{h,a} + \tilde{\beta}^0_3 X_a + \quad (\tilde{\gamma}^0_1 apt_a + \tilde{\gamma}^0_2 amt_a + \tilde{\gamma}^0_3 d_h + \tilde{\gamma}^0_4 X_a) d_c + \xi_{a,h,c}. \qquad (3)$$

Eq. (3) can be estimated using methods for linear equations. In the context of the present paper OLS is not appropriate, since the housing price, which equilibrates supply and demand, should be expected to reflect the impact of the unobserved neighborhood characteristics $\xi_{a,h,c}$. We therefore use an instrumental variables approach.

3.4. The implied model for car ownership

In the model developed above 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^i_1 $\left(= max\left\{v^i_{a,h,c} + \varepsilon^i_{a,h,c} | c = 1\right\}\right)$ with the εs denoting the random parts of the utilities, is:

$$U^i_1 = \ln\left(\sum_a \sum_h e^{v^i_{a,h,1}}\right) + \varepsilon^i_{c=1}. \qquad (4)$$

For the utility U^i_0 of not having a car we have similarly:

$$U^i_0 = \ln\left(\sum_a \sum_h e^{v^i_{a,h,0}}\right) + \varepsilon^i_{c=0} . \qquad (5)$$

The first terms on the right-hand side of (4) and (5) are the expected values of the maximum utility a household will be able to reach when owning a car, or not owning one, respectively. The random terms $\varepsilon^i_{c=1}$ and $\varepsilon^i_{c=0}$ are also 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 expected values in (4) and (5) are the deterministic parts of the utilities. Denoting the probability of car ownership as $\pi_{c=1}^{i}$ we thus have:[26]

$$\pi_{c=1}^{i} = \frac{e^{\ln\left(\sum_a \sum_h e^{v_{a,h,1}^{i}}\right)}}{e^{\ln\left(\sum_a \sum_h e^{v_{a,h,1}^{i}}\right)} + e^{\ln\left(\sum_a \sum_h e^{v_{a,h,0}^{i}}\right)}} \qquad (6)$$

The expression reflects that households can choose any location and housing type when owning or not owning a car. This model differs from the one typically used in the literatures discussed above that concentrates on binomial models that take location as given. When a logit model is used, its formulation would be:

$$\pi_{c=1|a,h}^{i} = \frac{e^{v_{a,h,1}^{i}}}{e^{v_{a,h,1}^{i}} + e^{v_{a,h,0}^{i}}} \qquad (7)$$

This equation results from (6) if the choice set for location and housing types is restricted to a single alternative. While model (6) allows the consumer to choose a different neighbourhood and housing type depending on whether or not a car will be owned, (7) only compares the utility a household would be able to reach with and without owning a car in a given neighbourhood and housing type. This implies, among other things, that the model of (6) is able to explain household's relocation decisions in response to improvements in public transport such as the introduction of a metro network, whereas (7) is not. Summarizing, the model of the present paper does not only generalize existing Tiebout-sorting models to incorporate public transport and car ownership, but it also generalizes existing car ownership models to include the choice of the residential location.[27]

3.5. The effects of better public transport

In our model the availability of public transport is an amenity that increases the utility of living in a particular zone.[28] If estimation results confirm our conjecture that the impact of the availability

---

[26] This equations can also be rewritten to $\pi_{c=1}^{i} = \frac{\sum_a \sum_h e^{v_{a,h,1}^{i}}}{\sum_a \sum_h e^{v_{a,h,1}^{i}} + \sum_a \sum_h e^{v_{a,h,0}^{i}}}$.

[27] We also compared our model with the conventional car ownership models. Notice that the traditional car ownership models are silent about the choice of neighborhood or housing type. Estimation results for a number of such traditional car ownership morels are provided in the online appendix.

[28] 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.

of public transport is larger for the choice alternatives in which no car is owned, improving public transport will have a nonpositive impact on car ownership.[29]

To see how this works in the model discussed above, observe that the immediate effect of the opening of a metro network will be an increase in the attractiveness of areas close to its stops. This effect will be different for choice alternatives with and without car-ownership. Not owning a car will become more attractive in a relative sense. This will induce some households to abandon their car or to change their residential location to benefit from improved public transport, or both.[30] As a result, demand for cars will decrease while demand for housing will increase in areas close to metro stations and decrease elsewhere. Housing supply or housing prices will react in response to this to restore equilibrium. These effects, as implied by our estimated model will be discussed in the simulation exercises reported below.

Since public transport is an amenity, its improvement increases utility and the associated compensating variation is expected to be positive. The welfare analysis of public transport improvement follows De Palma & Kilani (2003).[31]

3.6. Extension to dual-earner households

The setup of our model for dual earner households is entirely similar to that of single earner households discussed above, so we can be brief. The main differences are the extension of the choice set by now have to consider the choice of one residential location and housing type, but two job locations and by including ownership of 2 cars for each of these combinations. We used the same structure of the household utility function (see eq. (3)), but now make the parameters functions of the characteristics of both partners. We distinguish between head and partner based on the definition of Statistics Denmark and allow the parameters for both spouses to be different. This gives reasonably flexible specification of the model that allows for comparison with single earners. Unfortunately, we have no information about the use of the car for commuting purposes in dual earner households owning a single car.

[29] Under ceteris paribus conditions. It is possible that adjustments on the housing market following the improvement of public transport have an impact on car ownership that partly counteracts the initial effect.

[30] Recall here that we treat car ownership similarly to house ownership. That is, we treat car owners as if they lease their vehicles and pay an annualized price of car ownership to an absentee car dealer.

[31] See also McFadden (1999) and Dagsvik and Karlström (2005).

Just like in the case of single earner households, the model allows for substitution between all choice alternatives in response to, for instance, the opening of the metro network. So households may – for instance - switch from owning two cars to owning one car while simultaneously moving to a neighbourhood with a metro station in response to the realisation of this network.

4. Estimation results

We estimate two models: one for single-earner households and another for dual-earner households. For both samples we first estimate a multinomial logit model with alternative-specific constants, while in the second stage we use methods for linear equations to link these constants to the characteristics of the choice alternatives.

4.1. Endogeneity and the selection of instruments

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 (3). 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 alternative 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.[32] In the context of residential sorting some researchers have used characteristics of alternatives that are geographically close (see Klaiber & 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 amenities.[33] Bayer et al. (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

[32] 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).
[33] van Duijn & Rouwendal (2018) develop a model in which this is explicitly taken into account.

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.[34]

Van Duijn & Rouwendal (2018) have shown that the equilibrium housing price depends on an entropy-based measure of inequality of the choice probabilities. The variation in choice probabilities is caused by the differences in actor characteristics which is exogenous information that is not employed in the second stage of the estimation procedure and thus can be exploited to construct an instrument. Bayer et al. (2007)'s instrument are the equilibrium prices that are predicted by the model if the unobserved heterogeneity is removed from it. These counterfactual prices depend on counterfactual choice probabilities that transform the distribution of actor characteristics to alternative-specific variables by interacting it in a nonlinear way with the characteristics of the choice alternatives.[35]

Van Duijn & Rouwendal (2018) employ the methodology of Belloni, Chen, Chernozhukov, & Hansen (2012) to choose the preferred instrument among a large number of nonlinear transformations of the exogenous characteristics of the choice alternatives. They find that Bayer et al.'s instrument for the price is selected whenever it is included in the set of candidate instruments. We have here used the same procedure and obtained the same result.[36] Our interpretation is that the superior performance of Bayer's instrument is due to the use of the additional exogenous information about the heterogeneity of the population that is not employed in the other candidate instruments.

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. 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

[34] This instrument is thus a function of all exogenous area characteristics (urban amenities). It may be observed that this requires area characteristics to be excluded from the equation for the average utility (3).

[35] See Bernasco, de Graaff, Rouwendal, & Steenbeek (2017) for the use of a similar instrument in a different setting.

[36] A detailed description of the construction of the instrument variable candidates is provided in the online appendix.

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 to 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 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.

4.2. 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 household to the various choice alternatives. Table 3a refers to the single-earner households and 3b to the dual-earner households. Tables 4.a and 4.b report the coefficients that show deviations from the average utilities related to household characteristics, for the same groups of households.

Tables 3.a and 3.b show the results of the second step of the estimation procedure based on (3). 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 single-earner households or dual-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 always makes a choice alternative more attractive. Single family houses are preferred to multi family houses 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 already reflected 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.

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[1]	0.008***	0.007*
		(0.003)	(0.004)
	Proximity to the nearest metro station (km) * dummy variable indicating no car[1]	0.454**	0.547**
		(0.207)	(0.230)
	Dummy variable indicating one car	0.960***	0.889***
		(0.227)	(0.304)
β's	Dummy variable indicating non-apartment	1.432***	1.980***
		(0.235)	(0.353)
	Log (standardized house price)	-2.178***	-3.032***
		(0.324)	(0.517)
	Share of higher educated	1.874***	3.130***
		(0.532)	(1.043)
	Number of conserved and protected buildings per sq.m.	0.937***	0.903***
		(0.167)	(0.167)
	Distance to the CBD.	0.020**	0.016*
		(0.008)	(0.009)
	Social housing (share)	-0.418**	-0.410*
		(0.206)	(0.219)
γ's	Dummy variable indicating non-apartment * dummy variable indicating one car	0.128	0.126
		(0.151)	(0.152)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.168	-0.179
		(0.194)	(0.196)
	Constant	-1.189***	-0.937**
		(0.324)	(0.392)
	Anderson canon. corr. LM statistic		137.887
	Cragg-Donald Wald F statistic		60.423
	R-squared	0.214	
	No. of observations	538	538

Notes: standard errors in parentheses; standardized house 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, 0.05 and 0.10 levels, respectively.

[1] Note that the multiplication with the no-car dummy $(1 - d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

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.

Table 3.b. Second step estimation results for dual-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[1]	0.012***	0.010*
		(0.003)	(0.005)
	Proximity to the nearest metro station (km) * dummy variable indicating no car[1]	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 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)
	Anderson canon. corr. LM statistic		131.118
	Cragg-Donald Wald F statistic		40.188
	R-squared	0.570	
	No. of observations	636	636

Notes: standard errors in parentheses; standardized house 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, 0.05 and 0.10 levels, respectively.

[1] Note that the multiplication with the no-car dummy $(1 - d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

The results for the dual-earner households, presented in Table 3b, are qualitatively similar. Having one or two cars is better than having none, but given that cars are costly, ownership of one car is clearly the situation that is on average most preferred. This reflects 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.

4.3. 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 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.

Table 4.a. First step estimation procedure (MNL) 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[1]	-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[1]	-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)	(0007)	(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.

[1] Note that the multiplication with the no-car dummy $(1-d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

Table 4.b. First step estimation procedure (MNL) for dual-earner 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[1]	-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[1]	-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 and 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.

[1] Note that the multiplication with the no-car dummy $(1 - d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

4.4. Sensitivity Analyses

We *have* performed a number of sensitivity analyses. Most of the robustness checks focus on the model specification. 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.[37] We have therefore also estimated a mixed logit model. The results of the mixed logit model are very similar to the ones with the MNL. We conclude that, despite the apparently plausible *a priori* arguments for correlation between the random parts of the utilities that are similar in number of cars, housing type or geographical area, the empirical importance of this phenomenon appears to be limited. We have also estimated a model in which we have replaced the variable representing "proximity to the nearest metro station" with "proximity to the nearest bus stop". This can be considered as a placebo test. We find that, conditional on the measure of employment access (EA) by public transport that also includes bus connectivity, proximity to the nearest bus stop does not have significant effect on the household utility. Then we have estimated a model when the threshold for the distance to the nearest metro station is not 3 km, but 2 km as in e.g. Gibbons and Machin (2005). The results of this model are very similar to the basic model. Moreover, when we use an alternative instrument for the housing price, the share of house sales following a divorce in the year before the sale in the total number of house sales in the area, it appears that the estimates are qualitatively similar to those in the basic model.[38] Finally, we have also tested an alternative instrument for the share of higher educated based on the methodology proposed by Bayer et al. (2007). It is the counterfactual equilibrium share of higher educated predicted by the model when the term that reflects the unobserved characteristics is absent. This instrument is by construction independent of the unobserved heterogeneity terms and most likely strongly correlated with the observed share of higher educated. Estimation results with this alternative instrument are also qualitatively similar to those in the basic model. Results of the sensitivity analyses are available in the online appendix.

---

[37] That is, the $\varepsilon^{i}_{a,h,c}$'s may be statistically dependent for alternatives sharing the same *a, h* or *c*.

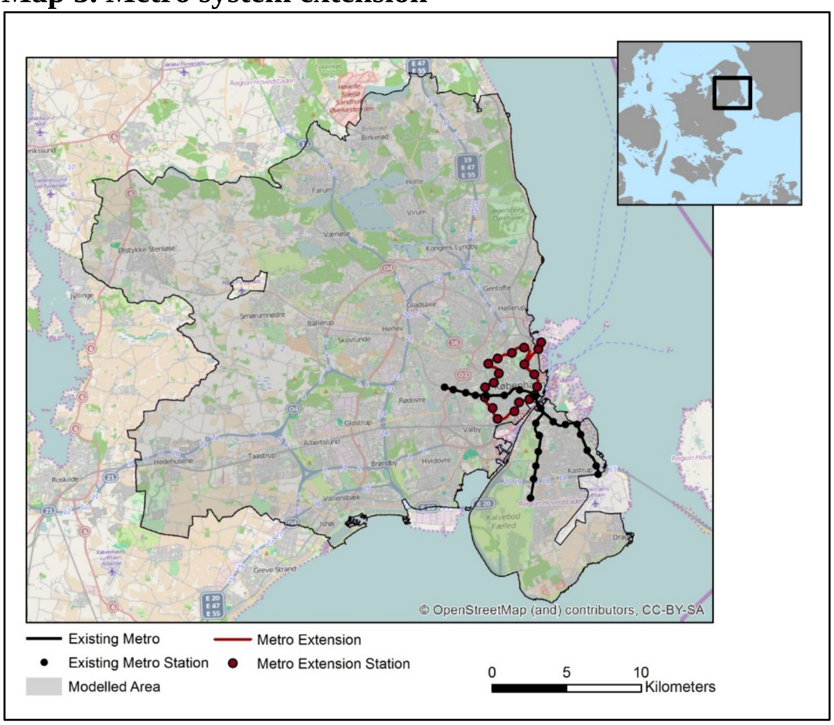
[38] The idea behind this instrument is that a larger share of divorces will lead to an increase in the number of houses for sale and hence, along with the limited time-window acceptable for a house sale, will have a negative effect on the price. A more detailed discussion of this instrument is provided in the online appendix.

5. 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 more than 63 mio. passengers yearly (in 2017). The metro has at present 22 stations; see the black dots in map 3.

The extension of the metro opened 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 3). This contrasts with many extensions of metro networks in other metropolitan areas that aim to link suburbs with the central city.

Map 3. Metro system extension



We use the estimated model to simulate the impact of the extension of the metro network. 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.

We compute the counterfactual equilibria implied by our model after the extension of the metro network under two different assumptions with respect to housing supply and compare them with the original situation.[39] This comparison is facilitated by the fact that estimated logit models with alternative-specific constants predict exactly the observed number of households choosing a particular alternative in the sample. It is well-known that without 'social interaction' effects the price equilibrium of a logit model of the kind discussed here is unique (see, for instance, Rouwendal, 1990). However, the fact that the attractiveness of neighbourhoods is determined partly by the demographic composition of the households living there causes complications. Bayer & Timmins (2005) show that if the presence of one group of households – the higher educated in our model – makes a neighbourhood more attractive, multiple equilibria may occur. Although we cannot completely exclude this possibility for our particular application, in our simulations the model always converged to the same equilibrium.[40]

5.1. 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

[39] Since the allocation system on the rental part of the Copenhagen housing market differs substantially from the price system, our simulations refer only to the owner-occupied part of the market. That is, we assume that the population of owner-occupying households does not change because of the metro and compute counterfactual equilibria for that part of the market.

[40] To compute the equilibrium we started by computing the demand for each choice alternative in the new situation (i.e. after the extension of the metro network). Then we adjusted the price (in the direction of reduced excess demand, only if housing supply is inelastic) and share of higher educated (using their share in the predicted demand for each alternative). We recomputed demand with the new values of the local prices and amenities, and so on. This procedure seems reasonably close to what one could expect of the actual adjustment process. See the online appendix for a detailed description of the different transition path algorithms that we used to test for the presence of multiple equilibria.

buildings. The exercise is nevertheless interesting because it shows how people would react to the change in public transport *per se*.

The simulation results suggests that extension of the metro system will have a 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-earner households. Moreover, especially households with higher incomes and more education will be attracted by the extension of the metro system (see the maps A.2. and A.3. in Appendix A.1).

5.2. Housing price adjustments

Next, we investigate 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 also unlikely to be completely true, but 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 very unlikely. Hence, the assumption of a zero elasticity of supply should be expected to be much closer to reality than that of an infinitely inelastic supply.

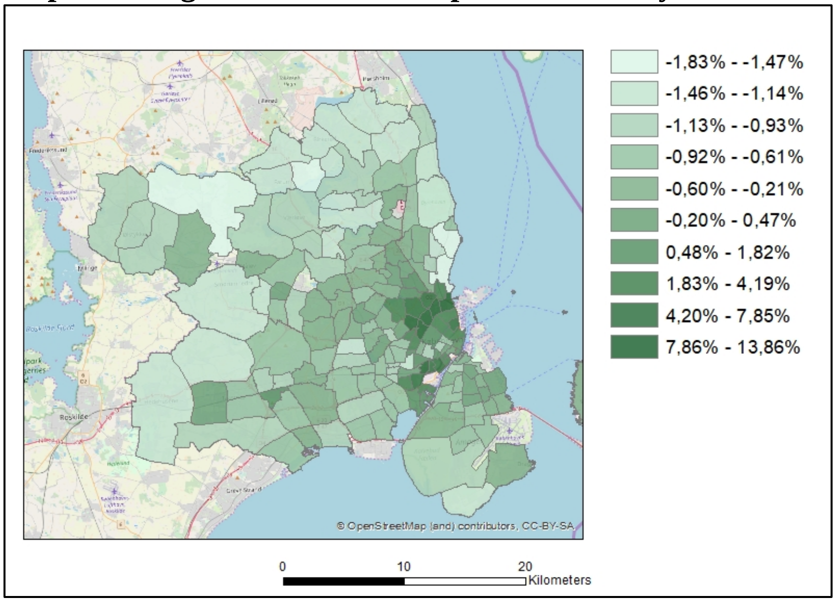
The reaction of the housing prices to the metro extension is shown on Map 4. Since our model can only deal with relative prices, we assumed that the average price level remains constant in the Greater Copenhagen Area.[41] The housing prices increase in the areas closer to the new metro line and decrease in other areas that become relatively less attractive. The reaction of house prices thus counteracts that of the extension of the metro network. The return to housing market equilibrium with fixed housing supply thus acts partly as a redistribution of the benefits of the metro extension.[42]

---

[41] Since our model does only refer to the Copenhagen area, it cannot predict the possible effect of the metro on migration from other parts of Denmark that could change the price level.

[42] Note, however, that we do not consider the wealth effects of the housing price changes. Although these are redistribution effects, they may have an impact on the behavior of the actors involved that we do not take into account.

Map 4. Change in relative house prices caused by the metro extension



Our results suggest also a substantial increase in the interest for living in areas close to the metro network. Moreover, the simulation results show the impact of the extension of the metro-network on the location choices of high income households, i.e. households from the northern part of the GCA who are on average more well-to-do are in particular attracted by the improved access to high quality public transport (see map A.4 in Appendix A.2). Higher educated workers and households with children are attracted by the extended metro-network as well. Our simulation results suggest that improving high quality public transport significantly affects the demographic composition of neighbourhoods.[43]

5.3. 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 prices adjust). For two car owners the corresponding figures are 4.5% and

[43] See online appendix for additional maps of the impact of the extension of the metro network.

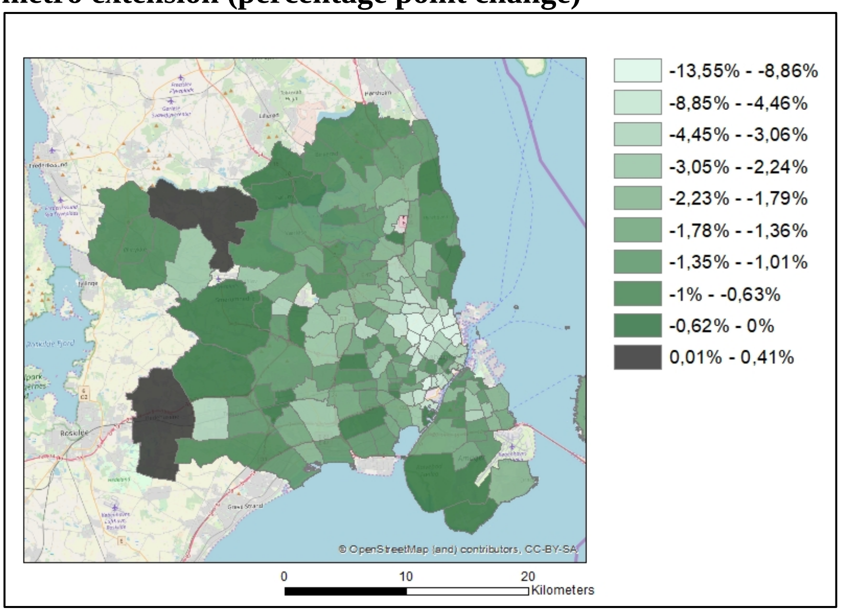
3.1%, respectively. Clearly some households that would give up their car (or one of their cars) if they could move to the areas where metro accessibility improved will change this intention when house prices adjust.[44] When interpreting these figures, it should be noted that they refer to the whole GCA. Changes in the shares of car owners are much larger in the neighbourhoods that are directly affected by the extension, see Map 5. In some areas further away from the city centre the car ownership rates even increase slightly. The reason is that the composition of the population shifts towards households that attach a relatively small value access to the metro network and are more inclined to own a car.

Table 5. Car ownership

	Reference scenario	Scenario 1	Scenario 2
		Fixed prices	Fixed supply
One car households	85,388	82,906 (-2.9%)	83,389 (-2.3%)
Two cars households	17,495	16,695 (-4.6%)	16,949 (-3.1%)
Total number of cars	120,378	116,295 (-3.4%)	117,287 (-2.6%)

Notes: percent changes of number of car owners are in parentheses.

Map 5: Change in car ownership in the GCA caused by the metro extension (percentage point change)



[44] We have also estimated the model with Bayer et al. (2007)'s instrument. The simulation results based on this version of the model are similar to those in Table 5, see the online appendix.

5.4. Welfare change

To compute the welfare implication of the change in the metro network, we assume that the random parts of the utilities - the $\varepsilon^{i}_{a,h,c}$'s – are individual-specific constants. Our utility function is nonlinear income, which complicates the computation of the compensating variation (McFadden, 1999). We use the approach of De Palma & Kilani (2003) who show that the compensating variation can be expressed as a single-dimensional integral (see online appendix for a detailed description of our application).

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, house prices do not change. We can thus compute the compensating variation of the households that results directly from the change in the quality of public transport.

It is, of course, likely that households will react. In our second assessment we assume this to happen, while housing supply is infinitely elastic. Note that households do not only switch to other areas, but also to other positions with respect to car ownership, and with respect to this change, the assumption of elastic supply is clearly more realistic.

Third we assess the welfare implications of the extended metro network under the contrary assumption that housing supply is completely inelastic and prices adjust so as to re-establish the equality between supply and demand. We take into account that single-earner households and dual-earner households are active on the same housing market when computing 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 became initially less attractive because of increased public transport now get lower housing prices.

Note that our computations ignore the wealth effects of the house price changes. In our model higher house prices translate in higher user costs that make a neighbourhood less attractive. This seems realistic for explaining (re)location choices of households. For incumbents the increase in house prices implies an additional welfare benefit that should ideally be taken into account in a full cost-benefit analysis. Since wealth is not included in our model, we do not present such calculations in this paper.[45]

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

The results for the average single (top panel) and dual earner 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 households 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.[46] 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-earner households 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, those alternatives in which no car is owned and a new metro station is closer than 3 km. There are 89 such alternatives.

[45] It may be argued that the wealth effect of the metro has potentially an impact on neighborhood choice and the demand for housing types, housing services and cars. Assessing this would require a separate study.
[46] 1 DKK is appr. € 0.13.

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 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 to 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.[47] Little of this gain is left after house prices adjust.[48] Moreover, when we use the Bayer et al. (2007) type instrument, it seems the implied welfare changes are of the same order of magnitude as those resulting from simulating the basic model, see the online appendix.

Since our model refers only to owner-occupiers, it is useful to ask if the results have anything to say about the welfare effects for renters. It seems reasonable to assume that renting households experience similar benefits from the opening of the metro network as owners. Since the price mechanism is of limited importance on the regulated rental market of Copenhagen, incumbents will probably not have to pay directly for the increased attractiveness of the neighbourhoods in which they live, which makes them somewhat comparable to incumbent owners. Relocating households who want to rent will probably be faced with longer waiting times, which makes them less accessible. Again, this effect is to some extent comparable to that experienced by non-incumbent owners. To the extent that waiting times play a similar role on the rental market as price adjustments on the owner-occupied market, the welfare effects may be similar. In the longer run, this situation may result in upgrading (followed by substantially higher rent) or sale of rental housing, which decreases the stock of rental housing in the areas that benefit most of the metro network. Summarizing: our welfare calculations probably have some relevance for the rental sector, but is unable to address issues specific for that segment of the housing market.

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

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

7. Conclusions

In this paper we develop a model for the joint choice of residential location and car ownership, focusing on the interaction with high quality public transport. While existing models of car ownership take the residential location as given, our approach shows that the presence of high quality public transport, which acts like a local public good and induces Tiebout-type sorting, can offer a good substitute for car ownership for some households. We estimated the model on register data for the Greater Copenhagen Area (GCA) and used the estimated model to simulate the impact of an extension of the metro network in Copenhagen. 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, especially among the higher income households, while reducing the overall car ownership rate by 2.3%. The results of the model are robust to alternative specifications.

Our results are relevant when considering policies concerned with urban area development. Place-based policies, such as the improvement of the public transport, which aim to improve some areas within a city, are frequently criticized in the economic literature because they improve places rather than the households' welfare. Our results suggest that a place-based policy which focuses on areas close to attractive city centres will attract relatively wealthier households and most likely cause more segregation. However, our model also predicts a significant increase in the relative housing prices in the areas in proximity to the new metro line and a decrease in other areas. The reaction of house prices thus acts partly as a redistribution of the benefits within an urban area after the improvement of the public transport.

The connection between public transport and gentrification suggested by our findings may come as a surprise. It has been argued in the literature that public transport only acts as a good substitute for cars among the poor (see e.g. Glaeser et al., 2008). The probable explanation is that European cities, often with historical cores that are major consumer amenities, differ substantially from most of their American counterparts as was noted by Brueckner, Thisse and Zenou (1999) and confirmed empirically by van Duijn and Rouwendal (2013). The connection between high quality public transport and the current strong revival of interest in inner city living is a topic that deserves more interest.

Future work may extend the results of the present paper in several directions. For instance, our study hints at implications of high quality public transport on vehicle kilometres travelled and their impact on congestion and pollution but does not quantify them. To do so, more attention should be paid to aspects that had to be treated in a relatively crude way here like the costs associated with car ownership and use and those associated with public transport use. Other suggestions already mentioned earlier in the paper are the extension of the sorting framework to explain the choice of combinations of residential and job locations, and to undertake a similar analysis for the rental market.

However, note that – these lose threads notwithstanding – the present paper suggests important additions to conventional cost benefit analyses of improvements in urban public transportation networks by providing an analytical tool for quantifying its effects – including that on welfare – on household location choices and on the number of cars owned in the context of an urban equilibrium model with heterogeneous households. Since the relationships between the number of cars on the one hand and congestion, pollution and traffic accidents on the other are usually well known in specific metropolitan areas, this will allow for meaningful calculations of the social benefits due to decreased car ownership caused by the metro.

References

Ahlfeldt, G.M. 2011. If Alonso was right: modelling accessibility and explaining the residential land gradient. *Journal of Regional Science*, 51 (2), pp. 318-338.

Ahlfeldt, G.M. and N. Wendland. 2016. The spatial decay in commuting probabilities: Employment potential vs. commuting gravity. *Economics Letters*, 143, pp. 125-129.

Anas, A., R. Arnott and K.A. Small. 1998. Urban spatial structure. *Journal of Economic Literature*, **36**, pp. 1426-1464.

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

Bauernschuster, S., T. Hener and H. Rainer. 2017. When Labor Disputes Bring Cities to a Standstill: The Impact of Public Transit Strikes on Traffic, Accidents, Air Pollution, and Health. *American Economic Journal: Economic Policy*, vol **9**, issue 1, pp. 1-37.

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. and C. Timmins. 2005. On the equilibrium properties of residential sorting models. *Journal of Urban Economics*, **57**, 462-477.

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.

Bernasco, W., T. de Graaff, J. Rowendal and W. Steenbeek. 2017. Social interactions and crime revisited: an investigation using individual offender data in Dutch neighborhoods. *The Review of Economics and Statistics*, **99**(4), 622-636.

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

Brueckner, J., J. Thisse and Y. Zenou (1999) Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review*, **43**, 91-107.

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

Dagsvik, J.K. and Karlström, A., 2005. Compensating variation and Hicksian choice probabilities in 30 random utility models that are non-linear in income. *Review of Economic Studies,* **72**(1), 57-76.

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

de Jong, G.C. 1990. 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 H. Sieg. 1999. Estimating equilibrium models of local jurisdictions. *Journal of Political Economy*, **107**(4), 645-681.

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.

Ewing, R. and R. Cervero (2010) Travel and the built environment. *Journal of the American Planning Association*, **76**, 265-294.

Gibbons, S. and S. Machin. 2005. Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1), pp. 148-169.

Glaeser, E.L. and M.E. Kahn (2004) Sprawl and urban growth. In: J.V. Henderson and J.-F. Thisse *Handbook of Regional and Urban Economics*, **IV**, ch. 56, 2481-2527.

Glaeser, E.L, M.E. Kahn and J. Rappoport (2008) Why do the poor live in cities? The role of public transport. *Journal of Urban Economics*, **63**, 1-24.

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

Gutiérrez-i-Puigarnau, E., I. Mulalic and J.N. van Ommeren. 2016. Do rich households live farther away from their workplaces? *The Journal of Economic Geography*, 16, pp. 177-201.

Huang, N., J. Li and A. Ross. 2019. The impact of the cost of car ownership on the house price gradient in Singapore. *Regional Science and Urban Economics*, 68, pp. 160-171.

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

Jacob, N., L. Munford, N. Rice and J. Roberts. 2019. The disutility of commuting? The effect of gender and local labor markets. *Regional Science and Urban Economics*, 77, pp. 264-275.

Kenworthy, J.R. and F.B. Laube. 1999. Patterns of automobile dependence in cities: An overview of key physical and economic dimensions with some implications for urban policy. *Transportation Research A*, **33**, 691-723.

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.

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.

McFadden, D.L. 1978. The choice of the residential location. In: A. Karlqvist, F. Snickars and J. Weibull (eds) *Spatial Interaction Theory and Planning Models*, 75-96, North Holland, Amsterdam.

McFadden, D.L. 1999. Computing willingness-to-pay in random utility models. In: J. Moore, R. Riezman and J. Mevlin (eds) *Trade, Theory and Econometrics: Essays in honor of John Chipman*, Routledge.

McGuckin N. and A. Fucci. 2018. Summary of Travel Trends: 2017 National Household Travel Survey. Report No. FHWA-PL-18-019, U.S. Department of Transportation, Federal Highway Administration.

Mulalic I. and J. Rouwendal. 2015. The impacts of fixed and variable costs on automobile demand: Evidence from Denmark. *Economics of Transportation*, **4**, 227-240.

Mulalic, I., J.N. van Ommeren and N. Pilegaard. 2014. Wages and commuting: quasi-natural experiments' evidence from firms that relocate. *The Economic Journal*, 124, pp. 1086-1105.

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

Ostermeijer, F., H.R.A. Koster and J. van Ommeren. 2019. Residential parking costs and car ownership: Implications for parking policy and automated vehicles. *Regional Science and Urban Economics*, 77, pp. 276-288.

Paetzold, J. 2019. Do commuting subsidies increase commuting distances? Evidence from a Regression Kink Design. *Regional Science and Urban Economics*, 75, pp. 136-147.

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

Rouwendal J. 1990. Existence and Uniqueness of Stochastic Price Equilibria in Heterogeneous Markets. *Regional Science and Urban Economics*, **21**, 23-42.

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

Sieg, H., V.K. Smith, H.S. Banzhaf and R. Walsh. 2004. Estimating the general equilibrium benefits of large changes in socially delineated public goods. *International Economic Review*, **45**, 1047-1077.

Spiekermann, K., Wegener, M., Kveton, V., Marada, M., Schürmann, C., Biosca, O., Ulied Segui, A., Antikainen, H., Kotavaara, O., Rusanen, J., Bielanska, D., Fiorello, D., Komornicki, T., Rosik, P., Stepniak, M. 2015. Transport Accessibility at Regional/Local Scale in Europe TRACC Scientific Report. Luxembourg: ESPON.

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.

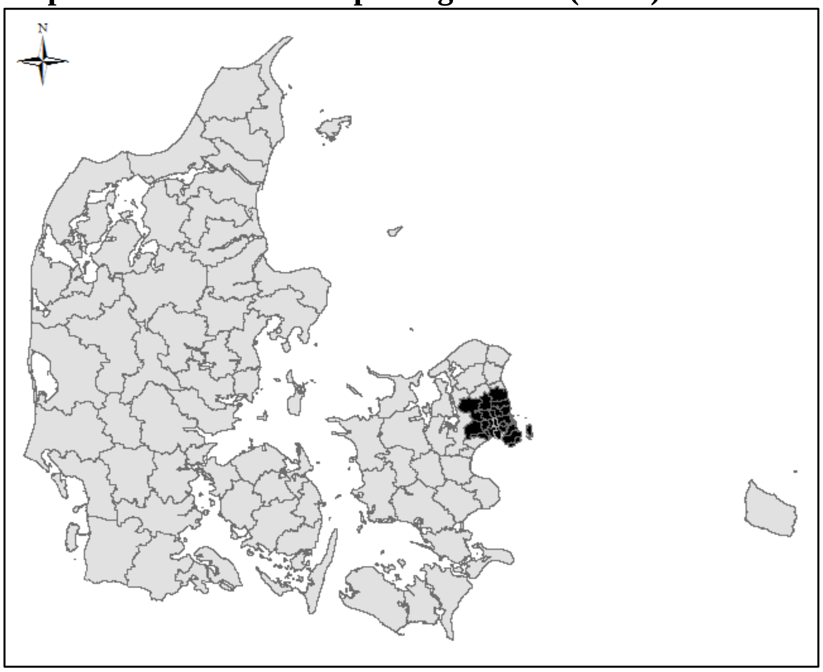
Van Duijn, M. and J. Rouwendal. 2018. Sorting based on urban heritage and income: Evidence from the Amsterdam metropolitan area. TI Discussion paper 15-030/VII (revised)

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

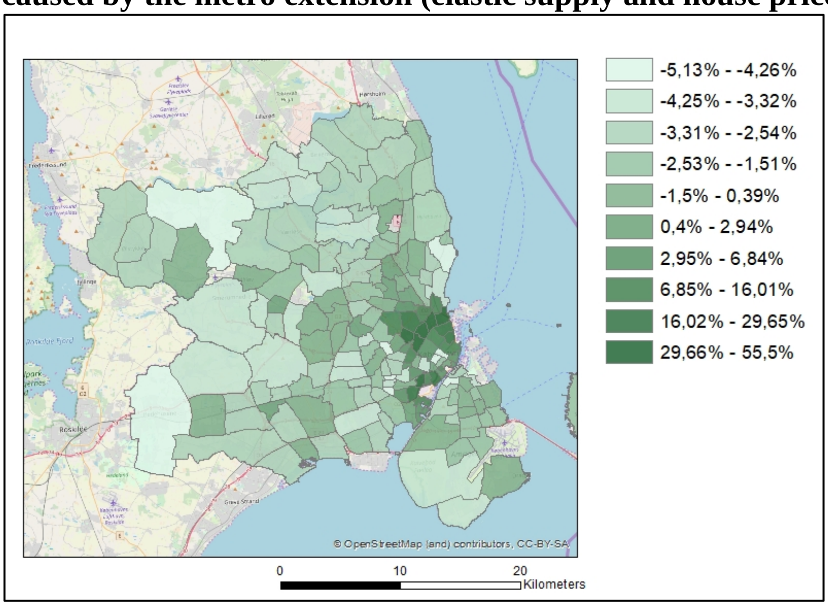
Appendix

A.1 Maps

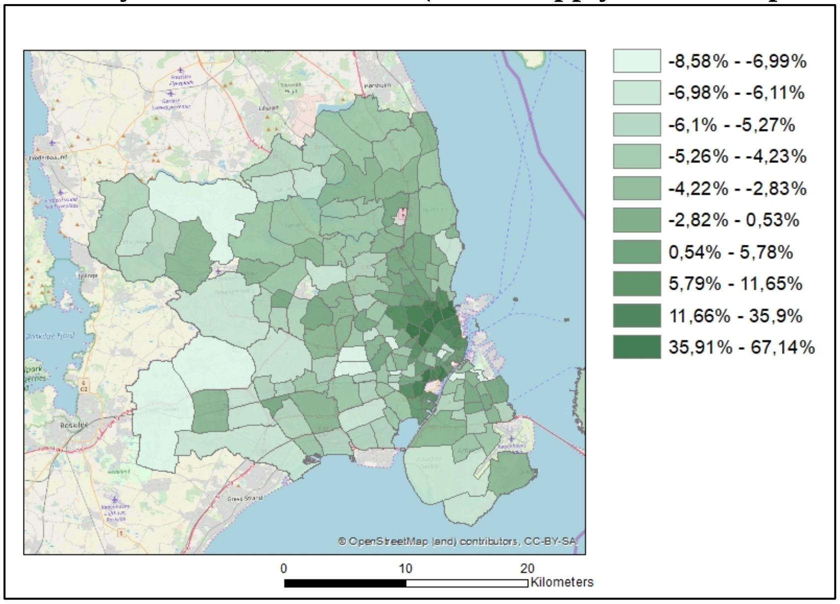
Map A.1. The Greater Copenhagen Area (GCA)



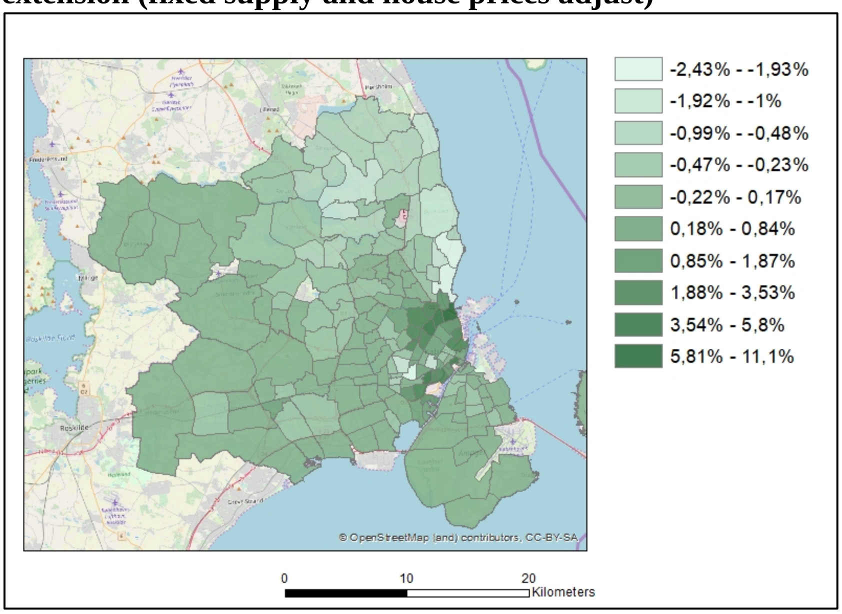
Map A.2. Pct. change in population of the households in the GCA caused by the metro extension (elastic supply and house prices fixed)



Map A.3. Pct. change in the share of higher educated in the GCA caused by the metro extension (elastic supply and house prices fixed)



Map A.4. Change in household income caused by the metro extension (fixed supply and house prices adjust)



A.2 First step results for the IV regressions

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

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

Notes: standard errors in parentheses; standardized house 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. Durbin (score) $\chi^2(3) = 263.321$ and Wu-Hausman $F(3{,}532) = 167$.

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

	[1] Log (std. house 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	0.3034	0.2150	0.4059	0.3770
No. of observations	636	636	636	636

Notes: standard errors in parentheses; standardized house 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. Durbin (score) $\chi^2(4) = 634.483$ and Wu-Hausman $F(4{,}615) = 285$.

Does improving public transport decrease car ownership?

Evidence from a residential sorting model for the Copenhagen metropolitan area

Online Appendix

Ismir Mulalic[1] and Jan Rouwendal[2]

Contents

[1] Department of Economics, Copenhagen Business School, Porcelænshaven 16A, 2000 Frederiksberg, Denmark.
[2] Department of Spatial Economics, VU University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, and Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands.

1. Data

1.1. Hedonic equations for single- and multi-family houses

Standardized house prices per area have been compiled from the two separate hedonic models with area fixed effects, i.e. one for the single-family houses and one for the multifamily housing. 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 Greater Copenhagen Area (GCA)) of the sample used in the hedonic price equation for houses. Similarly, the standard multifamily house 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 single and multi-family houses is 0.58.

Table 1.1. Hedonic price equation for single-family houses with area fixed effect, OLS

	Coeff.	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 1.2. Hedonic price equation for multifamily houses with area fixed effect, OLS

	Coeff.	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
Area 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.

2. Estimation results

2.1. Conventional car ownership models

Table 2.1. Logit model for car ownership conditional on the choice of residential area and housing type for single-earner households

	[1]	[2]	[3]	[4]	[5]
Log (household income)	0.929***	1.066***	0.937***	1.116***	1.116***
	(0.027)	(0.039)	(0.029)	(0.044)	(0.044)
Age	0.040***	0.032***	0.040***	0.030***	0.030***
	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)
Age sq. / 1000	-0.217***	-0.198***	-0.213***	-0.184***	-0.184***
	(0.035)	(0.054)	(0.037)	(0.058)	(0.058)
Number of children	0.150***	0.163***	0.158***	0.180***	0.180***
	(0.014)	(0.023)	(0.019)	(0.025)	(0.025)
Education (medium)	0.147***	0.148***	0.147***	0.144***	0.144***
	(0.023)	(0.031)	(0.024)	(0.034)	(0.034)
Education (high)	0.075***	0.026	0.072***	0.008	0.008
	(0.026)	(0.033)	(0.028)	(0.037)	(0.037)
Singles	-0.849***	-0.619***	-0.859***	-0.599***	-0.599***
	(0.024)	(0.037)	(0.026)	(0.041)	(0.041)
Dummy variable indicating non-apartment	0.128***	3.307***	0.122***	3.832***	3.832***
	(0.025)	(0.664)	(0.025)	(0.702)	(0.702)
Dummy indicating parking charging			-0.568	1.230	-2.134
			(1.024)	(1.083)	(2.299)
Log (household income) * Dummy variable indicating non-apartment		-0.300***		-0.345***	-0.345***
		(0.053)		(0.056)	(0.056)
Age * Dummy variable indicating non-apartment		0.024***		0.026***	0.026***
		(0.007)		(0.007)	(0.007)
Age sq. / 1000 * Dummy variable indicating non-apartment		-0.101		-0.113	-0.113
		(0.072)		(0.075)	(0.075)
Number of children * Dummy variable indicating non-apartment		0.018		0.003	0.003
		(0.030)		(0.031)	(0.031)
Education (medium) * Dummy variable indicating non-apartment		-0.007		-0.004	-0.004
		(0.045)		(0.047)	(0.047)
Education (high) * Dummy variable indicating non-apartment		0.102**		0.118**	0.118**
		(0.052)		(0.054)	(0.054)
Singles * Dummy variable indicating non-apartment		-0.379***		-0.398***	-0.398***
		(0.049)		(0.052)	(0.052)
Log (household income) * Dummy indicating parking charging			-0.067	-0.220***	-0.220***
			(0.079)	(0.084)	(0.084)
Age * Dummy indicating parking charging			0.003	0.011	0.011
			(0.012)	(0.012)	(0.012)
Age sq. / 1000 * Dummy indicating parking charging			-0.064	-0.084	-0.084
			(0.127)	(0.132)	(0.132)
Number of children * Dummy indicating parking charging			-0.124*	-0.137**	-0.137**
			(0.057)	(0.059)	(0.059)
Education (medium) * Dummy indicating parking charging			-0.001	0.006	0.006
			(0.079)	(0.081)	(0.081)
Education (high) * Dummy indicating parking charging			0.0198	0.075	0.075
			(0.075)	(0.078)	(0.078)
Singles * Dummy indicating parking charging			0.109	-0.106	-0.106
			(0.083)	(0.087)	(0.087)
Employment access with public transport / 1000					0.059*
					(0.033)
Proximity to the nearest metro station (km)					-2.978***
					(1.473)
Area fixed effect	yes	yes	yes	yes	yes
Constant	-12.989***	-14.560***	-11.749***	-13.836***	-14.823***
	(0.371)	(0.517)	(0.370)	(0.560)	(0.844)
Log likelihood	-36,149	-36,080	-36,143	-36,073	-36,073
No. of observations	66,012	66,012	66,012	66,012	66,012

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

2.2. Nesting structures

In this section we discuss an alternative specification of the model that we have tested to check the robustness of the results: the use of nesting structures. 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.[1] If so, the multinomial logit (MNL) model is no longer appropriate. Instead, a nested logit model could be formulated, where the nesting could refer to car ownership, housing type, or area. Various nesting structures are possible and it is difficult to choose between them on *a priori* grounds. Indeed, it seems reasonable that all choice alternatives with the same number cars have correlated random parts of the utilities, but also that alternatives referring to the same type of housing (single-family or multi-family) or the same geographical neighbourhood do so. The three level nested logit model force one to impose a particular hierarchical structure (often pictured as a tree with branches) on these correlations that are hard to motivate.

Since it is difficult to choose for a particular nesting structure, the flexibility offered by the cross-nested logit model (Bierlaire, 2006; Vovsha, 1997) is attractive. It is possible to formulate a cross-nested logit model in which the three types of correlation can be present simultaneously[2] and by estimating the model the data show the relative importance of each of them. However, in our experience estimation of (cross-)nested logit models is complex and time consuming. We have therefore adopted a third approach in which we re-specify the random part of utility as the sum of four terms. That is, we re-specify the modes as:

$$u_{a,h,c}^{i} = v_{a,h,c}^{i} + \left(\theta_{a}^{i} + \mu_{h}^{i} + \sigma_{c}^{i} + \varepsilon_{a,h,c}^{i}\right). \tag{2.1}$$

The random part of the utility is placed in parentheses. The random variable σ_c^i takes on the same value for all alternatives with the same number of cars c and thus introduces correlation between the random parts of the utility function referring to a given number of cars. Similar remarks hold for the other two random variables that are specific for housing type (μ_h^i) and area (θ_a^i). We maintain the assumption that the $\varepsilon_{a,h,c}^{i}$'s are iid Extreme Value Type I distributed, which implies that (2.1) defines a mixed logit model with three random parameters ($\theta_a^i + \mu_h^i + \sigma_c^i$). The structure

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

[2] 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.

of this model is close to the nested logit model in which the random part of the utilities in single-level nested logit models can be written as the sum of two components of which one is iid Extreme Value Type I, while the other is common to all alternatives in a given nest.

We specify each of the random variables θ_a^i, μ_h^i and σ_c^i as the product of a parameter to be estimated and a standard normal random variable and estimate the model by simulation using Hess et al.'s (2006) sampling method. We evaluate the likelihood at N sets of numbers. That is, for each random coefficient we use N different numbers. We first construct numbers based on the uniform distribution and then transform them to numbers based on the standard normal distribution. For a given coefficient we start by setting the first number equal to ρ/N, where ρ is a random draw from the uniform distribution. The second number is equal to $(\rho+1)/N$, the third to $(\rho+2)/N$,…, the n-th to $(\rho+(n-1))/N$. We do this for every random coefficient, each time starting with another independent draw ρ. To avoid correlation between the draws, we shuffle them using Durstenfeld's modification of the Fisher-Yates algorithm. That is, we randomly choose one of the integers 1 ... N, say m. Then we remove the m-th number and move the $m+1$-th to the N-th numbers one place towards the front of the sequence. The N-th position is then filled by, what was originally, the m-th number. Then we randomly choose one of the integers 1 ... $N-1$ say k. We place the $k+1$-th until the $N-1$-th numbers one position towards the front and the k-th draw to position $N-1$. And so on, $N-1$ times. Finally, the reshuffled draws from the uniform distribution are then transformed into draws from the normal distribution.

Table 2.2. MNL and mixed logit: the coefficients for the random terms

	Single-earner households		Dual-earner households	
	MNL	Mix logit	MNL	Mix logit
One car		0.076		-0.010
		(0.218)		(0.213)
Two cars				0.132
				(0.109)
House		0.024		0.025
		(0.231)		(0.254)
Area		0.092		-0.024
		(0.105)		(0.094)
Log likelihood	-372,013.28	-372,012.73	-481,038.17	-481,037.71

Notes: standard errors are in parentheses; both models include also control cross terms as models presented in tables 4.a and 4.b in the main text.

The results of the alternative model formulation are very similar to the ones with the MNL. The coefficients for the new random terms, which we report in Table 2.2, are small and none of them is significant at the 10% level, both in the single-earner and dual-earner models. We conclude that, despite the apparently plausible *a priori* arguments for correlation between the random parts of the utilities that are similar in number of cars, housing type or geographical area, the empirical importance of this phenomenon appears to be limited.[3]

2.3. Estimation results with the alternative instrument for house prices

We noted that there may exist some reservations with respect to the constructed instrument for the housing price proposed by Bayer et al. (2007). We have therefore tested an alternative instrument. It is the share of house sales following a divorce in the year before the sale in the total number of house sales in the area. The idea behind this instrument is that a larger share of divorces will lead to an increase in the number of houses for sale and hence, along with the limited time-window acceptable for a house sale, will have a negative effect on the price. For the instrument to be valid, the share of divorce-related sales should not be correlated with the unobserved characteristics of a neighbourhood. According to Wehner et al. (2001) the divorce rate in Denmark has stabilised around a level of 1.2-1.3%.[4] The two most common reasons for dissolution of marriage reported in the literature are “growing apart” and “not able to talk together” (see e.g. Hawkins et al., 2012, and Amato and Previti, 2003). Unobserved neighbourhood characteristics may have an impact on the divorce rate if they attract or abhor households which are more or less prone to divorce than the average household. One possible channel would be that such characteristics attract households with children or households that are highly educated. However, the negative relationship between children and divorce rates, that is often presumed, is not always confirmed in empirical work and Svarer and Verner (2008) even argue, using Danish register data, that children have in fact a small negative effect on relationship duration. There is also no consensus about the relationship between education and divorce rates. Moreover, note that even if the relationship between sorting of household types based on unobserved characteristics and divorce rates would be clear, this does not necessarily imply a relationship between the *share of divorce-related* sales. For instance, it could be that in areas with a relatively high divorce rate, turnover of houses is also high for other

[3] Notice also that the model presented in eq. (1) and (2) in the main text is similar to the error component formulation of the mixed logit model as the cross terms can be interpreted as resulting from a random draw of a household from the population.

[4] For every 1000 married couples, 12-13 marriages got dissolved.

reasons. For instance, in neighbourhoods where many households with children live, housing market mobility could be relatively high because increasing household size is an important driver of housing mobility. In summary, although we admit that one may tell stories in which there is a relationship between unobserved neighbourhood characteristics and the share of divorce related sales, it is not obvious that they are realistic.[5] In our view this is sufficient reason to investigate the performance of this alternative instrument and compare it with the constructed instrument that was proposed by Bayer et al. (2007) which has been criticised.

Table 2.3. Second step estimation results for single-earner households: decomposition of the household's mean indirect utilities

		[1] OLS	[2] IV (2SLS)	[3] IV (2SLS) new IV
α's	Employment access with public transport / 1000 * dummy variable indicating no car	0.008***	0.007*	0.007
		(0.003)	(0.004)	(0.005)
	Proximity to the nearest metro station (km) * dummy variable indicating no car	0.454**	0.547**	0.704**
		(0.207)	(0.230)	(0.278)
	Dummy variable indicating one car	0.960***	0.889***	0.931***
		(0.227)	(0.304)	(0.331)
β's	Dummy variable indicating non-apartment	1.432***	1.980***	3.573***
		(0.235)	(0.353)	(1.271)
	Log (standardized house price)	-2.178***	-3.032***	-5.489***
		(0.324)	(0.517)	(1.194)
	Share of higher educated	1.874***	3.130***	6.306**
		(0.532)	(1.043)	(2.629)
	Number of conserved/protected buildings per sq.km.	0.937***	0.903***	0.807***
		(0.167)	(0.167)	(0.197)
	Distance to the CBD.	0.020**	0.016*	0.006
		(0.008)	(0.009)	(0.012)
	Social housing (share)	-0.418**	-0.410*	-0.494**
		(0.206)	(0.219)	(0.250)
γ's	Dummy variable indicating non-apartment * dummy variable indicating one car	0.128	0.126	0.092
		(0.151)	(0.152)	(0.168)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.168	-0.179	-0.160
		(0.194)	(0.196)	(0.214)
	Constant	-1.189***	-0.937**	-0.347
		(0.324)	(0.392)	(0.640)
	Anderson canon. corr. LM statistic		137.887	17.120
	Cragg-Donald Wald F statistic		60.423	5.763
	R-squared	0.214		
	No. of observations	538	538	538

Notes: standard errors are in parentheses; standardized house price, share of higher educated and employment access with public transport are instrumented; see Table 2.5 for first-stage regression estimates of the 2SLS related to [3]; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

Estimation results with this alternative instrument are qualitatively similar to those in the basic model. Table 2.3 reports the second step results of the single-earners model and is similar to

[5] Map 2.1 shows the share of house sales following a divorce in the year before the sale for the GCA.

Table 3a in the main text. Table 2.4 reports the second step results for the dual-earner model and is similar to Table 3.b in the main text. The last two tables report the first stage regressions of the IV procedure for both models. The signs of the estimated coefficients never change, standard errors are often larger, but so are the absolute values. The level of significance of the coefficients hardly changes. We conclude that the model presented before is robust against two relevant changes in formulation.

Table 2.4. Second step estimation results for dual-earner households: decomposition of the household's mean indirect utilities

		[1] OLS	[2] IV (2SLS)	[3] IV (2SLS) new IV
α's	Employment access with public transport / 1000 * dummy variable indicating no car	0.012***	0.010*	0.008
		(0.003)	(0.005)	(0.006)
	Proximity to the nearest metro station (km) * dummy variable indicating no car	0.712***	0.800***	1.010***
		(0.215)	(0.236)	(0.330)
	Dummy variable indicating one car	1.728***	1.770***	2.008***
		(0.298)	(0.392)	(0.489)
	Dummy variable indicating two cars	1.033***	0.912**	0.861*
		(0.327)	(0.444)	(0.477)
β's	Dummy variable indicating non-apartment	2.743***	3.428***	5.294***
		(0.277)	(0.463)	(1.992)
	Log (standardized house price)	-2.321***	-3.357***	-6.092**
		(0.361)	(0.651)	(2.907)
	Share of higher educated	2.644***	3.880***	7.674**
		(0.586)	(1.255)	(4.084)
	Number of conserved/protected buildings per sq.km.	0.897***	0.848***	0.765***
		(0.159)	(0.161)	(0.192)
	Distance to the CBD.	0.039***	0.027**	0.007
		(0.009)	(0.012)	(0.024)
	Social housing (share)	-0.370*	-0.443**	-0.483**
		(0.199)	(0.215)	(0.238)
γ's	Employment access with public transport / 1000 * dummy variable indicating one car	0.004	0.002	-0.002
		(0.003)	(0.005)	(0.007)
	Proximity to the nearest Metro station (km) * dummy variable indicating one car	0.243	0.300	0.488
		(0.217)	(0.235)	(0.314)
	Dummy variable indicating non-apartment * dummy variable indicating one car	0.495***	0.471***	0.363*
		(0.168)	(0.174)	(0.219)
	Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.147	-0.142	-0.173
		(0.236)	(0.245)	(0.268)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.130	-0.122	-0.149
		(0.212)	(0.214)	(0.233)
	Dummy variable indicating parking charging * dummy variable indicating two cars	-0.072	-0.143	-0.320
		(0.424)	(0.431)	(0.495)
	Constant	-2.854***	-2.370***	-1.676*
		(0.368)	(0.498)	(0.871)
	Anderson canon. corr. LM statistic		131.118	11.141
	Cragg-Donald Wald F statistic		40.188	2.759
	R-squared	0.570		
	No. of observations	636	636	636

Notes: standard errors in parentheses; standardized house price, share of higher educated and employment access with public transport are instrumented; see Table 2.6 for first-stage regression estimates of the 2SLS related to [3]; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

Table 2.5. First step IV estimation results for single earner households

	[1] Log (std. house 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	0.046	-0.011	15.859***
	(0.032)	(0.017)	(2.012)
Number of conserved/protected buildings per sq.km.	-0.089***	-0.038***	-3.725**
	(0.027)	(0.014)	(1.703)
Distance to the CBD.	0.003**	0.006***	-0.854***
	(0.001)	(0.001)	(0.094)
Dummy variable indicating non-apartment	0.669***	-0.007	-4.257***
	(0.020)	(0.010)	(1.236)
Dummy variable indicating non-apartment * dummy variable indicating one car	-0.013	0.001	3.549**
	(0.024)	(0.013)	(1.544)
Dummy variable indicating parking charging * dummy variable indicating one car	0.089***	0.064***	-2.762
	(0.031)	(0.017)	(1.986)
Social housing (share)	-0.233***	-0.150***	6.762***
	(0.031)	(0.017)	(1.962)
Dummy variable indicating one car	-0.030	-0.040***	-77.551***
	(0.021)	(0.011)	(1.316)
Share of divorced couples (IV)	*-0.408****	*-0.055*	*10.301**
	(0.098)	*(0.052)*	*(6.193)*
Distance to the nearest school in 1890 (IV)	*-0.023****	*-0.019****	*0.740****
	(0.002)	*(0.001)*	*(0.120)*
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	*-0.045****	*-0.036****	*-14.988****
	(0.010)	*(0.005)*	*(0.641)*
Constant	0.765***	0.387	81.668***
	(0.029)	(0.016)	(1.836)
Partial R-squared	0.032	0.048	0.468
No. of observations	538	538	538

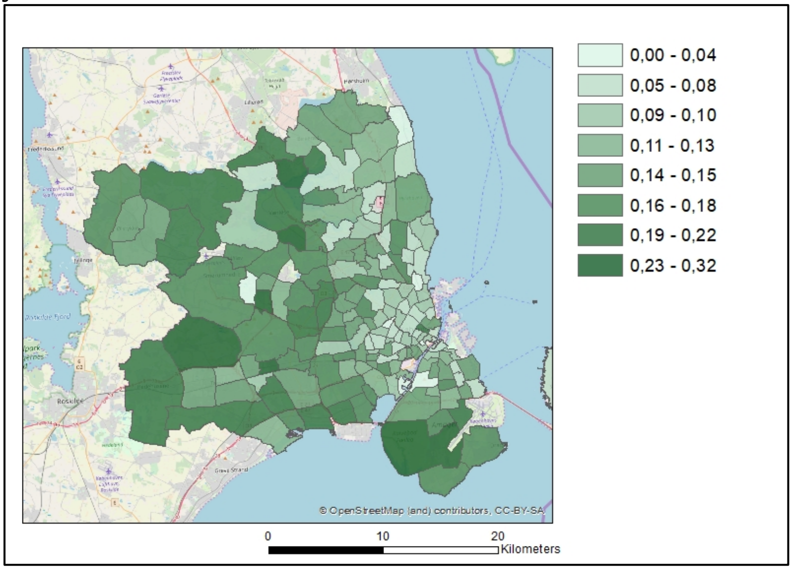
Notes: standard errors are in parentheses; standardized house price, and share of higher educated and employment access with public transport are instrumented; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively. Durbin (score) $\chi^2(3) = 3.621$ and Wu-Hausman $F(3{,}532) = 1.181$.

Table 2.6. First step IV estimation results for dual-earner households

	[1] Log (standardized house 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	0.075**	0.003	13.997***	-4.652**
	(0.030)	(0.016)	(1.793)	(1.946)
Proximity to the nearest metro station (km) * dummy variable indicating one car	0.028	-0.017	-4.446**	10.288***
	(0.031)	(0.017)	(1.847)	(2.004)
Number of conserved/protected buildings per sq.km.	-0.054**	-0.025**	-3.503**	-2.294
	(0.023)	(0.013)	(1.374)	(1.491)
Distance to the CBD.	0.004***	0.007***	-0.682***	-0.799***
	(0.001)	(0.001)	(0.080)	(0.086)
Dummy variable indicating non-apartment	0.706***	-0.001	-5.351***	0.967
	(0.020)	(0.011)	(1.211)	(1.314)
Dummy variable indicating non-apartment * dummy variable indicating one car	-0.031	0.008	6.001***	-2.374
	(0.024)	(0.013)	(1.436)	(1.558)
Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.120***	-0.078***	6.063***	-0.963
	(0.034)	(0.019)	(2.038)	(2.212)
Dummy variable indicating parking charging * dummy variable indicating one car	0.038	0.048***	-0.388	12.754***
	(0.030)	(0.017)	(1.815)	(1.969)
Dummy variable indicating parking charging * dummy variable indicating two cars	0.094	0.111***	-3.117	-3.443
	(0.062)	(0.034)	(3.681)	(3.994)
Social housing (share)	-0.205***	-0.125***	2.799*	7.600***
	(0.028)	(0.015)	(1.647)	(1.787)
Dummy variable indicating one car	0.030	-0.006	-78.827***	75.982***
	(0.022)	(0.012)	(1.310)	(1.422)
Dummy variable indicating two cars	0.051*	0.010	-78.815***	0.655
	(0.030)	(0.017)	(1.816)	(1.971)
Share of divorced couples (IV)	-0.273***	-0.063	-4.225	-3.944
	(0.077)	(0.042)	(4.573)	(4.963)
Distance to the nearest school in 1890 (IV)	-0.022***	-0.015 ***	0.518***	0.659***
	(0.002)	(0.001)	(0.098)	(0.106)
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	-0.066***	-0.055***	-14.121***	-0.145
	(0.010)	(0.006)	(0.607)	(0.659)
*Distance to the nearest train station in 1939 *dummy Indicating one car (IV)*	-0.055***	-0.055***	0.140	-14.734***
	(0.009)	(0.005)	(0.560)	(0.607)
Constant	0.697***	0.355***	85.067***	5.060**
	(0.028)	(0.015)	(1.676)	(1.819)
Partial R-squared	0.018	0.024	0.381	0.274
No. of observations	636	636	636	636

Notes: standard errors are in parentheses; standardized house 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. Durbin (score) $\chi^2(4) = 3.609$ and Wu-Hausman $F(4,615) = 0.877$.

Map 2.1. The share of house sales following a divorce in the year before the sale for the GCA



2.4. Estimation results with the Bayer et al. (2007) type instrument for the share of higher educated

We have also tested an alternative instrument for the share of higher educated based on the methodology proposed by Bayer et al. (2007).[6] It is the counterfactual equilibrium share of higher educated predicted by the model when the term that reflects the unobserved characteristics is absent. This instrument is by construction independent of the unobserved heterogeneity terms and most likely strongly correlated with the observed share of higher educated.

Estimation results with this alternative instrument are qualitatively similar to those in the basic model. Table 2.7 reports the second step results of the single-earners model and is similar to Table 3a in the main text. Table 2.8 reports the second step results for the dual-earner model and is similar to Table 3.b in the main text. The last two tables report the first stage regressions of the IV procedure for both models. The signs of the estimated coefficients remain the same and the absolute values, but also the standard errors, are often larger. We conclude that the model presented before is robust.

[6] This methodology has been also discussed in Bayer and Timmins (2005).

Table 2.7. Second step estimation results for single-earner households: decomposition of the household's mean indirect utilities

		[1] OLS	[2] IV (2SLS)	[3] IV (2SLS) Bayer type instrument for the share of higher educated
α's	Employment access with public transport / 1000 * dummy variable indicating no car	0.008***	0.007*	0.004*
		(0.003)	(0.004)	(0.001)
	Proximity to the nearest metro station (km) * dummy variable indicating no car	0.454**	0.547**	0.039***
		(0.207)	(0.230)	(0.055)
	Dummy variable indicating one car	0.960***	0.889***	0.577*
		(0.227)	(0.304)	(1.728)
β's	Dummy variable indicating non-apartment	1.432***	1.980***	7.289***
		(0.235)	(0.353)	(1.120)
	Log (standardized house price)	-2.178***	-3.032***	-3.789***
		(0.324)	(0.517)	(0.573)
	Share of higher educated	1.874***	3.130***	7.345***
		(0.532)	(1.043)	(1.245)
	Number of conserved/protected buildings per sq.km.	0.937***	0.903***	0.678**
		(0.167)	(0.167)	(0.327)
	Distance to the CBD.	0.020**	0.016*	-0.018
		(0.008)	(0.009)	(0.018)
	Social housing (share)	-0.418**	-0.410*	1.316
		(0.206)	(0.219)	(1.512)
γ's	Dummy variable indicating non-apartment * dummy variable indicating one car	0.128	0.126	0.246
		(0.151)	(0.152)	(0.298)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.168	-0.179	-0.715*
		(0.194)	(0.196)	(0.392)
	Constant	-1.189***	-0.937**	1.156
		(0.324)	(0.392)	(0.849)
	Anderson canon. corr. LM statistic		137.887	145.854
	Cragg-Donald Wald F statistic		60.423	13.920
	R-squared	0.214		
	No. of observations	538	538	538

Notes: standard errors in parentheses; standardized house price, share of higher educated and employment access with public transport are instrumented; see Table 2.9 for first-stage regression estimates of the 2SLS related to [3]; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

Table 2.8. Second step estimation results for dual-earner households: decomposition of the household's mean indirect utilities

		[1] OLS	[2] IV (2SLS)	[3] IV (2SLS) Bayer type instrument for the share of higher educated
α's	Employment access with public transport / 1000 * dummy variable indicating no car	0.012*** (0.003)	0.010* (0.005)	0.007*** (0.003)
	Proximity to the nearest Metro station (km) * dummy variable indicating no car	0.712*** (0.215)	0.800*** (0.236)	1.769 *** (0.642)
	Dummy variable indicating one car	1.728*** (0.298)	1.770*** (0.392)	1.615 *** (0.932)
	Dummy variable indicating two cars	1.033*** (0.327)	0.912** (0.444)	1.199* (1.454)
β's	Dummy variable indicating non-apartment	2.743*** (0.277)	3.428*** (0.463)	5.769*** (1.957)
	Log (standardized house price)	-2.321*** (0.361)	-3.357*** (0.651)	-3.033*** (1.303)
	Share of higher educated	2.644*** (0.586)	3.880*** (1.255)	6.524 *** (1.962)
	Number of conserved/protected buildings per sq.km.	0.897*** (0.159)	0.848*** (0.161)	0.880 (0.489)
	Distance to the CBD.	0.039*** (0.009)	0.027** (0.012)	0.016*** (0.002)
	Social housing (share)	-0.370* (0.199)	-0.443** (0.215)	-1.631** (0.635)
γ's	Employment access with public transport / 1000 * dummy variable indicating one car	0.004 (0.003)	0.002 (0.005)	0.002 (0.022)
	Proximity to the nearest Metro station (km) * dummy variable indicating one car	0.243 (0.217)	0.300 (0.235)	0.156*** (0.666)
	Dummy variable indicating non-apartment * dummy variable indicating one car	0.495*** (0.168)	0.471*** (0.174)	0.146 (0.401)
	Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.147 (0.236)	-0.142 (0.245)	-0.102 (0.614)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.130 (0.212)	-0.122 (0.214)	-0.155 (0.491)
	Dummy variable indicating parking charging * dummy variable indicating two cars	-0.072 (0.424)	-0.143 (0.431)	-0.408 (0.478)
	Constant	-2.854*** (0.368)	-2.370*** (0.498)	-3.587** (1.641)
	Anderson canon. corr. LM statistic		131.118	111.551
	Cragg-Donald Wald F statistic		40.188	2.759
	R-squared	0.570		
	No. of observations	636	636	636

Notes: standard errors are in parentheses; standardized house price, share of higher educated and employment access with public transport are instrumented; see Table 2.10 for first-stage regression estimates of the 2SLS related to [3]; ***, ** indicate that estimates are significantly different from zero at the 0.01 and 0.05 levels, respectively.

Table 2.9. First step IV estimation results for single-earner households

	[1] Log (std. house 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	0.002	-0.054***	17.741***
	(0.003)	(0.015)	(2.022)
Number of conserved/protected buildings per sq.km.	0.004	-0.010	-4.950**
	(0.003)	(0.012)	(1.750)
Distance to the CBD.	-0.0001	-0.0001	-0.573***
	(0.0001)	(0.0001)	(0.080)
Dummy variable indicating non-apartment	0.017***	-0.245***	2.123
	(0.003)	(0.014)	(1.954)
Dummy variable indicating non-apartment * dummy variable indicating one car	-0.004*	-0.006	3.935**
	(0.002)	(0.012)	(1.581)
Dummy variable indicating parking charging * dummy variable indicating one car	0.003	0.028*	-1.855
	(0.003)	(0.015)	(2.037)
Social housing (share)	0.002	-0.072*	5.415**
	(0.003)	(0.015)	(2.101)
Dummy variable indicating one car	0.018***	0.030**	-79.811***
	(0.003)	(0.015)	(1.982)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	*0.969****	*0.379****	*-9.172****
	(0.004)	*(0.018)*	*(2.422)*
Bayer type instrument for the share of higher educated (IV)	*0.064****	*0.252****	*-9.867***
	(0.007)	*(0.036)*	*(4.952)*
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	*0.003****	*0.017****	*-15.142****
	(0.001)	*(0.005)*	*(0.725)*
Constant	-0.005*	0.040*	91.539***
	(0.004)	(0.020)	(2.787)
Partial R-squared	0.013	0.076	0.270
No. of observations	538	538	538

Notes: standard errors are in parentheses; standardized house 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. Durbin (score) $\chi^2(3) = 3.621$ and Wu-Hausman $F(3,532) = 1.181$.

Table 2.10. First step IV estimation results for dual-earner households

	[1] Log (standardized house 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	0.001	-0.035***	14.707***	-2.713
	(0.003)	(0.013)	(1.793)	(1.958)
Proximity to the nearest metro station (km) * dummy variable indicating one car	-0.001	-0.045***	-3.203**	11.453***
	(0.003)	(0.013)	(1.842)	(2.007)
Number of conserved/protected buildings per sq.km.	0.0001	0.004	-4.267***	-3.135**
	(0.003)	(0.010)	(1.378)	(1.501)
Distance to the CBD.	-0.0002*	0.003***	-0.482***	-0.621***
	(0.0001)	(0.0001)	(0.072)	(0.079)
Dummy variable indicating non-apartment	0.012***	-0.269***	-4.515**	8.805***
	(0.003)	(0.013)	(1.869)	(2.036)
Dummy variable indicating non-apartment * dummy variable indicating one car	-0.004	0.010	6.741***	-3.261**
	(0.002)	(0.010)	(1.459)	(1.589)
Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.005	-0.051***	7.284***	-2.242
	(0.003)	(0.015)	(2.087)	(2.274)
Dummy variable indicating parking charging * dummy variable indicating one car	0.001	0.033**	-0.302	13.323***
	(0.030)	(0.013)	(1.826)	(1.989)
Dummy variable indicating parking charging * dummy variable indicating two cars	0.001	0.071***	-2.628	-2.402
	(0.006)	(0.027)	(3.704)	(4.035)
Social housing (share)	-0.001	-0.048***	2.456	6.780***
	(0.003)	(0.012)	(1.717)	(1.870)
Dummy variable indicating one car	0.028***	0.059***	-84.725***	80.836***
	(0.003)	(0.014)	(1.994)	(2.173)
Dummy variable indicating two cars	0.028***	0.061***	-84.264***	6.260**
	(0.004)	(0.017)	(2.375)	(2.587)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	0.986***	0.402***	-3.176	-11.445***
	(0.003)	(0.015)	(2.146)	(4.964)
Bayer type instrument for the share of higher educated (IV)	0.085***	0.233***	-18.100***	14.230 ***
	(0.007)	(0.033)	(4.557)	(4.964)
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	0.005***	-0.025***	-14.751***	0.928
	(0.001)	(0.005)	(0.681)	(0.741)
*Distance to the nearest train station in 1939 *dummy Indicating one car (IV)*	-0.001	-0.046***	0.838	-14.483***
	(0.001)	(0.004)	(0.566)	(0.616)
Constant	-0.022 ***	0.017***	92.516***	7.535***
	(0.004)	(0.004)	(2.420)	(2.636)
Partial R-squared	0.078	0.049	0.165	0.108
No. of observations	636	636	636	636

Notes: standard errors are in parentheses; standardized house 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. Durbin (score) $\chi^2(4) = 3.609$ and Wu-Hausman $F(4{,}615) = 0.877$.

2.5. The choice of preferred instruments and instrument variable candidates

Van Duijn & Rouwendal (2018) employ the methodology of Belloni et al. (2012) to choose the preferred instrument among a large number of nonlinear transformations of the exogenous characteristics (amenities) of the choice alternatives. The basis set of exogenous amenities used for the computation of candidate-instruments are: (i) number of conserved/protected buildings per sq.m., (ii) distance to the CBD, (iii) social housing (share), and (iv) proximity to the nearest metro station (km). We first follow Dubé et al. (2012) and construct candidate-instruments by taking the squares, cubes and second and third order cross products of the four characteristics. Then we follow the BLP methodology and use the sums of amenities of zones that are close and therefore likely substitutes. We use three threshold distances: 1 km, 5 km and 10 km. We also take the sums of amenities of zones that are not close, using the same thresholds, in order to avoid the spatial interaction. We then follow Bayer et al. (2007) and construct *donut* instruments, i.e. we use again the same distance thresholds and take the sums of amenities of zones located between 1 and 5 km, 5 and 10 km and 1 and 10 km from the chosen one. We also take the sum of amenities of zones excluded those of the chosen one. We also exploit the idea developed by e.g. Bayer et al. (2007) that the equilibrium-housing price in a neighbourhood reflects its relative position in the housing market. Following this argument we construct three additional types of candidate-instruments: (i) we use the ranking of zones on specific amenities, (ii) we define an ideal point by combining the best scores of zones on all amenities and computing a measure of the distance of the actual set of amenities of a zone from this ideal point, and (iii) we use an indicator of the number of zones that are similar by requiring such a distance measure to be smaller than a particular threshold. In total this gives us 94 candidate instruments. Finally, we also use the original instrument variables computed from the historical information (distance to the nearest school in 1890 and distance to the nearest train station in 1939) and the share of divorced couples.

2.6. Estimation results with "proximity to the nearest bus stop" - a placebo test

In order to account for the accessibility to public transport -- including buss connectivity -- we include in our models a measure of employment access (EA) by public transport. Moreover, because of the high frequency, greater reliability and better transportation quality, we also include the proximity to the nearest metro station as a separate variable, because the significance of access to the metro network is unlikely to be captured completely by its impact on employment accessibility. One may also argue that the bus network is as important as the metro network. We have therefore executed a placebo test by replacing the variable representing "proximity to the nearest metro station" with "proximity to the nearest bus stop". Table 2.11 reports the second step results of the single-earners model and is similar to Table 3.a in the main text. Table 2.12 reports the second step results for the dual-earners model and is similar to Table 3.b in the main text. They show that, conditional on the measure of employment access (EA) by public transport, proximity to the nearest bus stop does not have significant effect on the household utility. We conclude that a measure of employment access (EA) by public transport sufficiently account for the bus connectivity and that it is not necessary to include the proximity to the nearest bus stop as a separate variable. Note that the accessibility to public transport is highly developed in the GCA and that bus stops are always close by.

Table 2.11. 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[1]	0.010***	0.008
		(0.004)	(0.006)
	Proximity to the nearest bus stop (km) * dummy variable indicating no car[1]	**-0.179**	**-0.354**
		(0.500)	**(0.652)**
	Dummy variable indicating one car	0.838**	0.674
		(0.387)	(0.615)
β's	Dummy variable indicating non-apartment	1.283***	1.752***
		(0.236)	(0.342)
	Log (standardized house price)	-2.046***	-2.781***
		(0.326)	(0.499)
	Share of higher educated	1.630***	2.588***
		(0.529)	(0.987)
	Number of conserved and protected buildings per sq.m.	0.895***	0.849***
		(0.170)	(0.173)
	Distance to the CBD.	0.018**	0.013
		(0.008)	(0.009)
	Social housing (share)	-0.546**	-0.593***
		(0.204)	(0.216)
γ''s	Dummy variable indicating non-apartment * dummy variable indicating one car	0.204	0.205
		(0.152)	(0.154)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.207	-0.219
		(0.195)	(0.197)
	Constant	-1.024**	-0.644
		(0.471)	(0.722)
	R-squared	0.198	
	No. of observations	538	538

Notes: standard errors are in parentheses; standardized house/apartment price, share of higher educated and employment access with public transport are instrumented; see Table 2.13 for first-stage regression estimates of the 2SLS; ***, **, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and 0.10 levels, respectively.

[1] Note that the multiplication with the no-car dummy $(1-d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

Table 2.12. Second step estimation results for dual-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[1]	0.017***	0.006
		(0.004)	(0.011)
	Proximity to the nearest bus stop (km) * dummy variable indicating no car[1]	**0.605**	**-0.475**
		(0.585)	**(0.698)**
	Dummy variable indicating one car	2.032***	2.063*
		(0.564)	(1.093)
	Dummy variable indicating two cars	1.273***	0.287
		(0.476)	(1.051)
β's	Dummy variable indicating non-apartment	2.592***	5.586***
		(0.279)	(0.791)
	Log (standardized house price)	-2.118***	-3.595***
		(0.364)	(0.775)
	Share of higher educated	2.255***	4.102***
		(0.582)	(1.359)
	Number of conserved/protected buildings per sq.m.	0.843***	0.573***
		(0.163)	(0.206)
	Distance to the CBD.	0.031***	-0.016
		(0.009)	(0.021)
	Social housing (share)	-0.544***	-0.860***
		(0.197)	(0.246)
γ's	Employment access with public transport / 1000 * dummy variable indicating one car	0.004	-0.007
		(0.004)	(0.010)
	Proximity to the nearest bus stop (km) * dummy variable indicating one car	**0.222**	**-0.475**
		(0.483)	**(0.698)**
	Dummy variable indicating non-apartment * dummy variable indicating one car	0.520***	0.393**
		(0.170)	(0.198)
	Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.088	-0.047
		(0.240)	(0.302)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.077	0.008
		(0.203)	(0.248)
	Dummy variable indicating parking charging * dummy variable indicating two cars	-0.121	-0.463
		(0.432)	(0.519)
	Constant	-2.956***	-0.577
		(0.515)	(1.165)
	R-squared	0.559	
	No. of observations	636	636

Notes: standard errors are in parentheses; standardized house/apartment price, share of higher educated and employment access with public transport are instrumented; see Table 2.14. for the first-stage regression estimates of the 2SLS; ***, **, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and 0.10 levels, respectively.

[1] Note that the multiplication with the no-car dummy $(1 - d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

Table 2.13. First step IV estimation results for single earner households

	[1] Log (std. house price)	[2] Share of higher educated	[3] Employment access with public transport for no car owners
Proximity to the nearest bus stop (km) * dummy variable indicating no car	0.005	-0.056*	-45.639***
	(0.008)	(0.033)	(4.717)
Number of conserved/protected buildings per sq.km.	0.003	-0.007	-6.313**
	(0.003)	(0.012)	(1.673)
Distance to the CBD.	0.0002	0.005***	-0.933***
	(0.0001)	(0.001)	(0.088)
Dummy variable indicating non-apartment	0.014***	-0.203***	-1.563
	(0.003)	(0.014)	(1.964)
Dummy variable indicating non-apartment * dummy variable indicating one car	-0.002	0.004	3.508**
	(0.002)	(0.011)	(1.505)
Dummy variable indicating parking charging * dummy variable indicating one car	0.003	0.040***	-3.912**
	(0.003)	(0.014)	(1.931)
Social housing (share)	0.001	-0.069***	-1.589
	(0.003)	(0.014)	(2.008)
Dummy variable indicating one car	-0.0001	-0.010	-93.787***
	(0.003)	(0.012)	(1.674)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	*0.971****	*0.295****	*-2.509*
	(0.004)	*(0.018)*	*(2.575)*
Distance to the nearest school in 1890 (IV)	*-0.001***	*-0.012****	*0.851****
	(0.0002)	*(0.001)*	*(0.130)*
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	*-0.001*	*-0.026****	*-12.063****
	(0.001)	*(0.005)*	*(0.751)*
Constant	0.012***	0.139***	103.198***
	(0.004)	(0.019)	(2.644)
Partial R-squared	0.4209	0.2827	0.3442
No. of observations	538	538	538

Notes: standard errors are in parentheses; standardized house 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. Durbin (score) $\chi^2(3) = 262.638$ and Wu-Hausman $F(3{,}532) = 166.278$.

Table 2.14. First step IV estimation results for dual-earner households

	[1] Log (standardized house 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 bus stop (km) * dummy variable indicating no car	-0.016	0.017	-47.043***	15.585***
	(0.029)	(0.038)	(4.628)	(4.900)
Proximity to the nearest bus stop (km) * dummy variable indicating one car	0.002	0.017	7.931**	-38.276***
	(0.024)	(0.033)	(3.934)	(4.165)
Number of conserved/protected buildings per sq.km.	-0.0002	0.017	-7.084***	-6.130***
	(0.009)	(0.011)	(1.377)	(1.459)
Distance to the CBD.	-0.0002	0.004***	-0.562***	-0.624***
	(0.0004)	(0.001)	(0.068)	(0.072)
Dummy variable indicating non-apartment	0.065***	-0.238***	-5.410***	1.049
	(0.012)	(0.016)	(1.950)	(2.065)
Dummy variable indicating non-apartment * dummy variable indicating one car	0.012	0.048***	1.666	-7.739***
	(0.011)	(0.014)	(1.722)	(1.823)
Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.011	-0.050***	6.033***	-0.749
	(0.012)	(0.017)	(2.002)	(2.121)
Dummy variable indicating parking charging * dummy variable indicating one car	0.003	0.023**	-2.425	14.927***
	(0.010)	(0.014)	(1.654)	(1.751)
Dummy variable indicating parking charging * dummy variable indicating two cars	0.011	0.086***	-3.431	-3.430
	(0.022)	(0.030)	(3.607)	(3.819)
Social housing (share)	-0.031**	-0.063***	1.089	6.940***
	(0.011)	(0.014)	(1.753)	(1.856)
Dummy variable indicating one car	-0.003	-0.015	-97.770***	93.709***
	(0.011)	(0.014)	(1.716)	(1.817)
Dummy variable indicating two cars	-0.004	-0.002	-94.338***	6.632***
	(0.013)	(0.017)	(2.033)	(2.152)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	0.901***	0.345***	-0.101	-1.588
	(0.015)	(0.020)	(2.371)	(2.510)
Distance to the nearest school in 1890 (IV)	-0.011**	-0.026***	2.151***	4.941***
	(0.005)	(0.007)	(0.862)	(0.912)
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	-0.008*	-0.048***	-10.239***	-0.513
	(0.004)	(0.006)	(0.693)	(0.794)
*Distance to the nearest train station in 1939 *dummy Indicating one car (IV)*	-0.006	-0.044***	-0.460	-12.520***
	(0.004)	(0.005)	(0.662)	(0.701)
Constant	0.059***	0.085***	102.364***	1.790***
	(0.014)	(0.019)	(2.320)	(2.456)
Partial R-squared	0.1166	0.0741	0.1839	0.1737
No. of observations	636	636	636	636

Notes: standard errors are in parentheses; standardized house 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. Durbin (score) $\chi^2(4) = 634.137$ and Wu-Hausman $F(4{,}615) = 644.2$.

2.7. Estimation results when the threshold for the distance to the nearest metro station is 2 km

We model proximity to the nearest metro as the average of a linearly decreasing function of the distance from each address in the area to the nearest metro station, assuming a threshold of 3 km. The choice of the threshold is somewhat arbitrary. We have therefore estimated, as an additional sensitivity analysis, a model when we assume that the threshold is 2 km, as suggested by Gibbons and Machin (2005):

$$\text{Proximity} = \begin{matrix} \frac{2-distance}{2} & if & distance\ to\ metro\ st. \leq 2\ km \\ 0 & if & distance\ to\ metro\ st. > 2\ km \end{matrix}$$

Tables 2.15 and 2.16 reports the second step results of the single-earners model and of the dual-earners model, respectively. These tables are similar to Tables 3.a and 3.b in the main text. The results of this model are very similar to the basic model. Note that biking is very popular in Denmark and that most of the Danes are able to bike and there are facilities for parking bikes at most metro stations.

Table 2.15. Second step estimation results for single earner households: decomposition of the household's mean indirect utilities - the threshold for the distance to the nearest metro station is 2 km

		[1] OLS	[2] IV (2SLS)
α's	Employment access with public transport / 1000 * dummy variable indicating no car[1]	0.009***	0.009**
		(0.003)	(0.004)
	Proximity to the nearest metro station (km) * dummy variable indicating no car[1]	0.485**	0.550**
		(0.229)	(0.246)
	Dummy variable indicating one car	0.892***	0.917***
		(0.228)	(0.299)
β's	Dummy variable indicating non-apartment	1.338***	1.899***
		(0.236)	(0.353)
	Log (standardized house price)	-2.121***	-2.981***
		(0.325)	(0.517)
	Share of higher educated	1.806***	3.042***
		(0.532)	(1.043)
	Number of conserved and protected buildings per sq.m.	0.949***	0.923***
		(0.169)	(0.168)
	Distance to the CBD.	0.020**	0.018**
		(0.008)	(0.009)
	Social housing (share)	-0.433**	-0.432*
		(0.204)	(0.217)
γ''s	Dummy variable indicating non-apartment * dummy variable indicating one car	0.193	0.179
		(0.152)	(0.153)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.172	-0.168
		(0.194)	(0.197)
	Constant	-1.158***	-1.013**
		(0.325)	(0.391)
	R-squared	0.205	
	No. of observations	538	538

Notes: standard errors are in parentheses; standardized house/apartment price, share of higher educated and employment access with public transport are instrumented; see Table 2.17 for the first-stage regression estimates of the 2SLS; ***, **, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and 0.10 levels, respectively.

[1] Note that the multiplication with the no-car dummy $(1 - d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

Table 2.16. Second step estimation results for dual-earner households: decomposition of the household's mean indirect utilities - the threshold for the distance to the nearest metro station is 2 km

		[1] OLS	[2] IV (2SLS)
α's	Employment access with public transport / 1000 * dummy variable indicating no car[1]	0.013***	0.009
		(0.003)	(0.009)
	Proximity to the nearest metro station (km) * dummy variable indicating no car[1]	0.666***	1.067***
		(0.273)	(0.332)
	Dummy variable indicating one car	1.749***	2.125***
		(0.298)	(0.437)
	Dummy variable indicating two cars	1.040***	0.860
		(0.327)	(0.674)
β's	Dummy variable indicating non-apartment	2.719***	3.856***
		(0.277)	(0.919)
	Log (standardized house price)	-2.330***	-3.931***
		(0.362)	(0.724)
	Share of higher educated	2.555***	4.838***
		(0.584)	(1.789)
	Number of conserved/protected buildings per sq.m.	0.916***	0.784***
		(0.160)	(0.187)
	Distance to the CBD.	0.037***	-0.001
		(0.009)	(0.021)
	Social housing (share)	-0.423**	-0.526**
		(0.197)	(0.297)
γ's	Employment access with public transport / 1000 * dummy variable indicating one car	0.004	-0.003
		(0.003)	(0.009)
	Proximity to the nearest Metro station (km) * dummy variable indicating one car	0.218	0.576*
		(0.240)	(0.320)
	Dummy variable indicating non-apartment * dummy variable indicating one car	0.505***	0.335*
		(0.168)	(0.197)
	Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.131	-0.162
		(0.236)	(0.300)
	Dummy variable indicating parking charging * dummy variable indicating one car	-0.120	-0.154
		(0.213)	(0.242)
	Dummy variable indicating parking charging * dummy variable indicating two cars	-0.078	-0.388
		(0.426)	(0.523)
	Constant	-2.829***	-1.478**
		(0.368)	(0.747)
	R-squared	0.556	
	No. of observations	636	636

Notes: standard errors are in parentheses; standardized house/apartment price, share of higher educated and employment access with public transport are instrumented; see Table 2.18 for first-stage regression estimates of the 2SLS; ***, **, * indicate that estimates are significantly different from zero at the 0.01, 0.05 and 0.10 levels, respectively.

[1] Note that the multiplication with the no-car dummy $(1 - d_c)$ implies that we have in fact specified the term as $\alpha X + \gamma X d_c$, where X is employment access or proximity to metro, while imposing the restriction $\gamma = -\alpha$.

Table 2.17. First step IV estimation results for single earner households

	[1] Log (std. house 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	0.002	-0.023	17.691***
	(0.004)	(0.015)	(2.249)
Number of conserved/protected buildings per sq.km.	0.003	-0.010	-3.591**
	(0.003)	(0.012)	(1.731)
Distance to the CBD.	0.0003	0.005***	-0.895***
	(0.0002)	(0.001)	(0.093)
Dummy variable indicating non-apartment	0.014***	-0.202***	-0.591
	(0.003)	(0.014)	(2.028)
Dummy variable indicating non-apartment * dummy variable indicating one car	-0.002	0.004	3.491**
	(0.003)	(0.011)	(1.547)
Dummy variable indicating parking charging * dummy variable indicating one car	0.003	0.038***	-2.518
	(0.003)	(0.014)	(2.002)
Social housing (share)	0.001	-0.078***	5.425***
	(0.003)	(0.014)	(2.053)
Dummy variable indicating one car	-0.001	-0.029***	-78.772***
	(0.002)	(0.009)	(1.269)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	*0.971****	*0.298****	*-4.323*
	(0.004)	*(0.018)*	*(2.653)*
Distance to the nearest school in 1890 (IV)	*-0.001***	*-0.012****	*0.719****
	(0.0002)	*(0.001)*	*(0.135)*
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	*-0.0002*	*-0.021****	*-15.628****
	(0.001)	*(0.004)*	*(0.646)*
Constant	0.013***	0.159***	86.760***
	(0.004)	(0.018)	(2.572)
Partial R-squared	0.3910	0.2613	0.4956
No. of observations	538	538	538

Notes: standard errors are in parentheses; standardized house 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. Durbin (score) $\chi^2(3) = 263.353$ and Wu-Hausman $F(3{,}532) = 167.164$.

Table 2.18. First step IV estimation results for dual-earner households

	[1] Log (standardized house 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	0.007	-0.025	15.981***	-4.202*
	(0.012)	(0.016)	(2.067)	(2.175)
Proximity to the nearest metro station (km) * dummy variable indicating one car	0.002	-0.034*	-3.216	12.853***
	(0.012)	(0.016)	(2.067)	(2.210)
Number of conserved/protected buildings per sq.km.	0.0004	0.011	-4.941***	-5.040***
	(0.008)	(0.012)	(1.455)	(1.556)
Distance to the CBD.	-0.0001	0.004***	-0.551***	-0.667***
	(0.0004)	(0.001)	(0.072)	(0.077)
Dummy variable indicating non-apartment	0.064***	-0.244***	-5.469***	1.909
	(0.012)	(0.016)	(2.007)	(2.146)
Dummy variable indicating non-apartment * dummy variable indicating one car	0.010	0.045***	3.038*	-8.127***
	(0.010)	(0.014)	(1.748)	(1.869)
Dummy variable indicating non-apartment * dummy variable indicating two cars	-0.012	-0.046**	6.335***	-1.366
	(0.012)	(0.017)	(2.083)	(2.227)
Dummy variable indicating parking charging * dummy variable indicating one car	0.003	0.032**	-0.797	12.455***
	(0.011)	(0.015)	(1.845)	(1.973)
Dummy variable indicating parking charging * dummy variable indicating two cars	0.011	0.082***	-3.188	-3.567
	(0.022)	(0.030)	(3.738)	(3.996)
Social housing (share)	-0.027**	-0.070***	4.482**	9.614***
	(0.011)	(0.014)	(1.797)	(1.921)
Dummy variable indicating one car	0.002	-0.015	-80.268***	77.119***
	(0.007)	(0.010)	(1.249)	(1.335)
Dummy variable indicating two cars	0.0004	-0.012	-79.096***	2.230
	(0.011)	(0.014)	(1.806)	(1.931)
Prices that would clear the market if there were no unobserved heterogeneity (IV)	0.902***	0.352***	-1.000	-2.329
	(0.014)	(0.019)	(2.429)	(2.596)
Distance to the nearest school in 1890 (IV)	-0.009*	-0.023***	2.675***	4.749***
	(0.0005)	(0.006)	(0.861)	(0.920)
*Distance to the nearest train station in 1939 *dummy Indicating no car (IV)*	0.010***	-0.045***	-13.788***	0.666
	(0.004)	(0.005)	(0.628)	(0.671)
*Distance to the nearest train station in 1939 *dummy Indicating one car (IV)*	-0.007**	-0.043***	0.177	-15.770***
	(0.004)	(0.005)	(0.604)	(0.645)
Constant	0.051***	0.097***	86.435***	6.302***
	(0.013)	(0.017)	(2.196)	(2.347)
Partial R-squared	0.0871	0.0541	0.1957	0.1565
No. of observations	636	636	636	636

Notes: standard errors are in parentheses; standardized house 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. Durbin (score) $\chi^2(4) = 634.433$ and Wu-Hausman $F(4{,}615) = 663.3$.

3. Simulation

3.1. The uniqueness of the sorting equilibrium

Bayer and Timmins (2005) study the equilibrium properties of the residential location sorting models and provide a test for the uniqueness in empirical analyses of the sorting equilibrium.[7] They develop an algorithm that can be used to test for the uniqueness in a particular application using the estimated parameters of the model. To describe the Bayer and Timmins' algorithm consider the iterative application of the mapping of the share of higher educated agents, denoted here as X, that choose each alternative $j = 1, \dots, J$.[8,9] Start from the initial vector such that $X_j^0 = 1$, $X_k^0 = 0 \ \forall k \neq j$ for each $j = 1, \dots, J$, respectively. Then compute the model equilibrium. Finally, repeat this for all elements in the choice set J. Bayer and Timmins (2005) show that given any set of the estimated model coefficients, finite vector of exogenous amenities and any agglomeration effect (in our case the estimated coefficient associated with the share of higher educated is strictly positive), the iterative sequence described above converges to at least two distinct fixed points whenever multiple fixed points of the mapping exist. The uniqueness can only be guaranteed analytically in the presence of a congestion effect (negative spillovers). This implies that for any set of parameters and any specific draw of the exogenous data, it is possible to determine whether a unique equilibrium obtains. We apply this methodology and our model converges to the same equilibrium in all cases. It comes as no surprise. Bayer and Timmins (2005) show that the probability for uniqueness is increasing in i) the number of alternatives, ii) the variation in the contribution to utility made by the exogenous amenities, and iii) the heterogeneity in individual preferences. In our case, the total choice set includes 538 and 636 elements for single-earner households and dual-earner households, respectively, and the models include a large number of exogenous amenities and substantial amount of household heterogeneity.

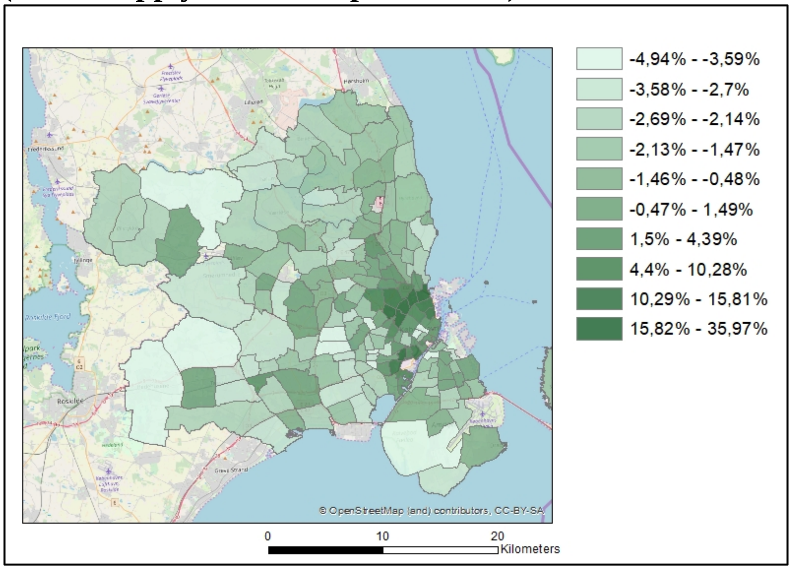
[7] The estimation strategy applied in this paper does not require a unique equilibrium for estimation purposes. However, uniqueness is important for the simulation studies, e.g. the equilibrium responses to a large policy change such as the analyzed extension of the metro network.

[8] Choice alternatives in our model are defined by three dimensions: area ($a = 1 \dots n$), house type ($h = s, m$), and car ownership ($c = 0,1$). We ignore here the heterogeneity in the choice set in order to simplify the notation.

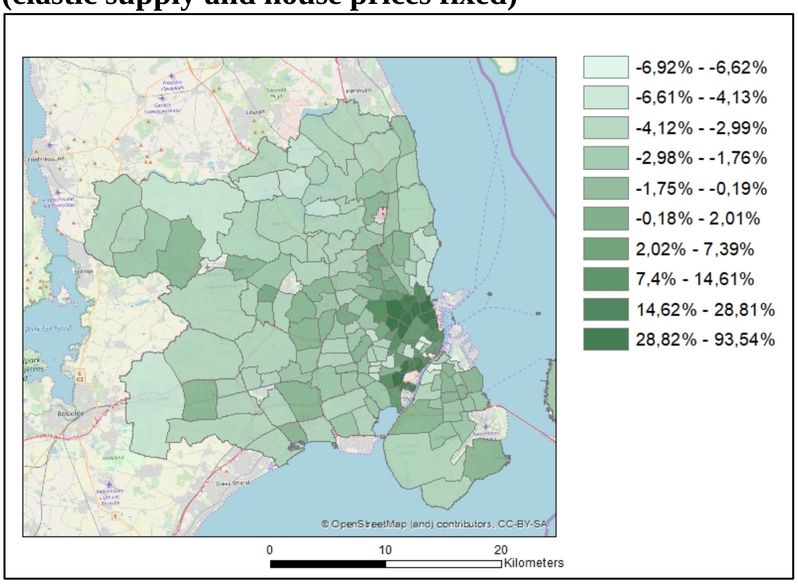
[9] We consider here the iterative mapping ψ defined as the function that implicitly defines the share of higher educated that chose each alternative $\psi_j = \sum_i \Pi_{i,j}$, where $\Pi_{i,j}$ is the probability of choice of alternative j for household i. Notice here that $X_j^t = \psi_j(X^{t-1})$. Starting from an initial vector, $\bar{X}^0$, the iterative application of this mapping gives rise to the sequence $\Omega(\bar{X}^0) = \{\bar{X}^0, \psi(\bar{X}^0), \psi(\psi(\bar{X}^0)), \dots, \psi^t(\bar{X}^0), \dots\}$.

3.2. Maps of the impact of the extension of the metro network

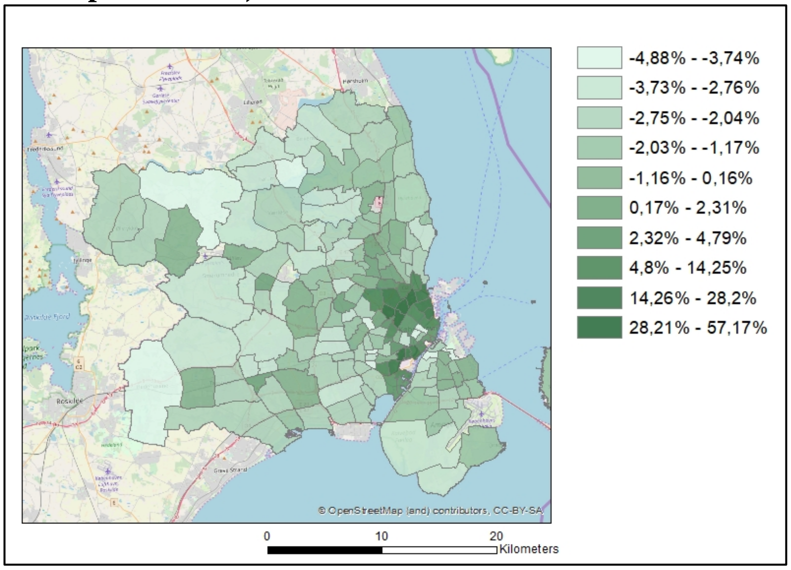
Map 3.1. Pct. change in population of the single-earner households in the GCA caused by the metro extension (elastic supply and house prices fixed)



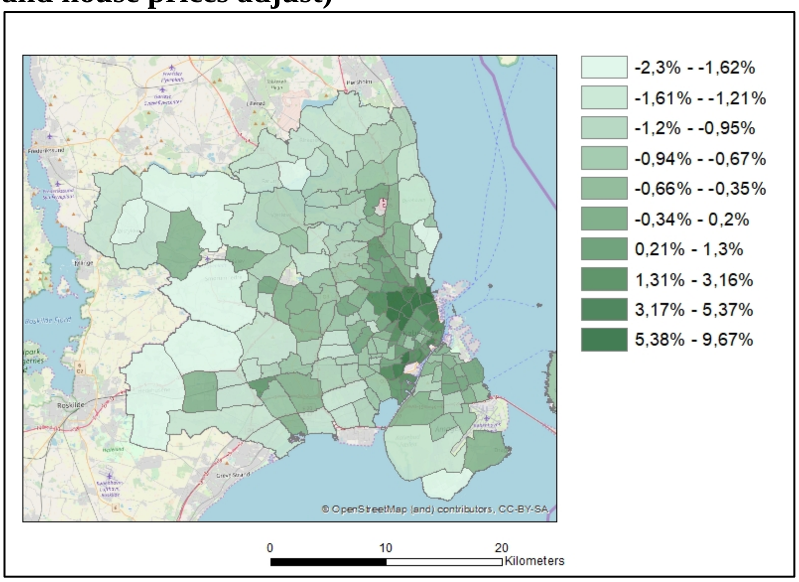
Map 3.2. Pct. change in population of the dual-earner households in the GCA caused by the metro extension (elastic supply and house prices fixed)



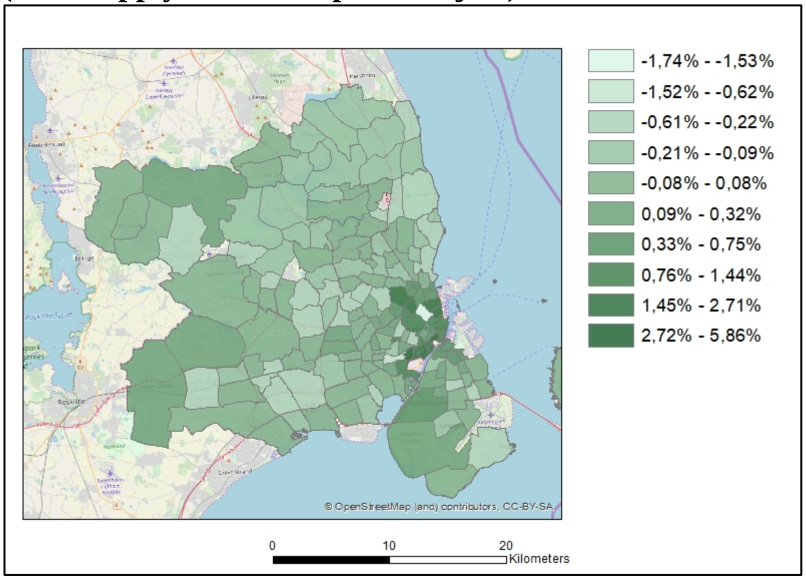
Map 3.3. Pct. change in the household income in the GCA caused by the metro extension (elastic supply and house prices fixed)



Map 3.4. Pct. change in the share of higher educated in the GCA caused by the metro extension (fixed supply and house prices adjust)



Map 3.5. Pct. change in the number of households with children in the GCA caused by the metro extension (fixed supply and house prices adjust)



3.3 Computing the welfare change of the extension of the metro network

We are looking for the compensating variation in income. The discussion below refers to the single-earner households, but the treatment for the dual-earner households is completely analogous. The compensating variation is the negative of the change in income that brings the household back to the utility it experienced before the extension of the metro network. Hence, the compensating variation will be positive in case welfare increases due to the extension of the metro network, and negative if it decreases.

Our starting point is (1), which we repeat here as:

$$v^i_{a,h,c}(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i, Z^i) = \alpha^i_1 apt_a + \alpha^i_2 amt_a + \alpha^i_3 d_c + \beta^i_1 d_h + \beta^i_2 P_{h,a} + \beta^i_3 X_a + (\gamma^i_1 apt_a + \gamma^i_2 amt_a + \gamma^i_3 d_h + \gamma^i_4 X_a) d_c + \xi_{a,h,c} \quad \text{(A1)}$$

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, while disregarding the possibility of relocation. 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 (2), 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:

$$\begin{aligned} v_{a,h,c}^i\left(apt_a, amt_a, d_c, d_h, P_{h,a}, X_a; y^i, Z^i\right) = \\ (\tilde{\alpha}_1^1 + \tilde{\gamma}_1^1 d_c) \ln\left(y^i\right) apt_a + (\tilde{\alpha}_2^1 + \tilde{\gamma}_2^1 d_c) \ln\left(y^i\right) amt_a + \tilde{\beta}_2^1 \ln\left(y^i\right) P_{h,a} + \\ \theta_1 apt_{a_1}^{\ i} + \theta_2 amt_{a_1}^{\ i} + \theta_3 P_{h,a} + \mu \ln\left(y^i\right) + \rho \end{aligned} \tag{A3}$$

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 coefficients in this line are determined by collecting the appropriate terms:

$$\theta_1 = \left(\tilde{\alpha}_1^0 + \sum_{l=1}^{L} \tilde{\alpha}_1^{l+1} Z_l^i + \left(\tilde{\gamma}_1^0 + \sum_{l=1}^{L} \tilde{\gamma}_1^{l+1} Z_l^i\right)\right)$$

$$\theta_2 = \left(\tilde{\alpha}_2^0 + \sum_{l=1}^{L} \tilde{\alpha}_2^{l+1} Z_l^i + \left(\tilde{\gamma}_2^0 + \sum_{l=1}^{L} \tilde{\gamma}_2^{l+1} Z_l^i\right)\right)$$

$$\theta_3 = \tilde{\beta}_2^0 + \sum_{l=1}^{L} \tilde{\beta}_2^{l+1} Z_l^i$$

$$\mu = \tilde{\alpha}_3^1 + \tilde{\beta}_1^1 + \tilde{\beta}_3^1 + (\tilde{\gamma}_3^1 d_h d_c + \tilde{\gamma}_4^1 X_a d_c)$$

$$\rho = \left(\tilde{\alpha}_3^0 + \sum_{l=1}^{L} \tilde{\alpha}_3^{l+1} Z_l^i\right) d_c + \left(\tilde{\beta}_1^0 + \sum_{l=1}^{L} \tilde{\beta}_1^{l+1} Z_l^i\right) d_h + \left(\tilde{\beta}_3^0 + \sum_{l=1}^{L} \tilde{\beta}_3^{l+1} Z_l^i\right) X_a + \left(\tilde{\gamma}_3^0 + \sum_{l=1}^{L} \tilde{\gamma}_3^{l+1} Z_l^i\right) d_h d_c + \left(\tilde{\gamma}_4^0 + \sum_{l=1}^{L} \tilde{\gamma}_4^{l+1} Z_l^i\right) X_a d_c$$

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:

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

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.[10] The change in income itself that corresponds with the change $\Delta\ln\left(y^i_{a,h,c}\right)$ in log income can be determined as:

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

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:

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

De Palma 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:

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

where the values of apt_a, amt_a and $P_{h,a}$ refer to the situation with the extended metro network. Clearly, if $r = \Delta y^i_{a,h,c}$, $w_{a,h,c}(r)$ is equal to the utility household I experienced before the extension of the metro network. Now define: $w^*_{a,h,c}(r) = \max\left\{w_{a,h,c}(r), w_{a,h,c}\left(\Delta y^i_{a,h,c}\right)\right\}$ and let $\pi_n(w^*)$ denote the logit choice probabilities defined on the basis of these utilities:

[10] 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.

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

This brings us – finally – in the position to define the expected value of the unconditional compensating variation (cv):[11]

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

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).

3.4 Car ownership and welfare changes with the Bayer et al. (2007) type instrument

In this section, we use the estimated model when using Bayer et al. (2007) type instrument to simulate the impact of the extension of the metro network. We focus on car ownership and welfare changes.

Table 3.1 shows the simulation results for car ownership. Most importantly, the simulation results are similar to results in Table 5 in the main text. We find again that households would give up their car(s) if they could move to the areas where metro accessibility improved.

Table 3.1. Car ownership change with the Bayer et al. (2007) type instrument

	Reference scenario	Scenario 1	Scenario 2
		Fixed prices	Fixed supply
One car households	89,235	87,240 (-2.2%)	86,622 (-2.0%)
Two cars households	19,301	18,408 (-4.6%)	18,362 (-3.5%)
Total number of cars	127,837	124,055 (-3.0%)	123,347 (-2.5%)

Notes: percent changes of number of car owners are in parentheses.

[11] It can be shown that (A8) reduces to the change in the expected value of the maximum utility when utility is linear in income.

Table 3.2 shows the results for welfare changes. Here the results differ more from those found for the basic model. For instance, we now find that the negative impact of price adjustments for single-earner households without a car is larger than the initial positive effect of better accessibility. However, this is the only occasion where the sign of a welfare effect changes and all effects are of the same order of magnitude as found for the basic model.

Table 3.2. Compensating variations of the extension of the metro network with the Bayer et al. (2007) type instrument

			[1] No mobility	[2] Elastic supply	[3] House prices adjust
Single earner households	All households	Average CV (DKK x 1000)	4.75	4.78	5.69
		Share of income (%)	*1.2*	*1.2*	*1.4*
	Dir. affected alt. (no car)	Average CV (DKK x 1000)	9.40	9.42	-2.37
		Share of income (%)	*2.4*	*2.4*	*0.6*
Dual earners households	All households	Average CV (DKK x 1000)	14.88	16.65	15.53
		Share of income (%)	2.4	2.6	2.5
	Dir. affected alt. (no car)	Average CV (DKK x 1000)	80.23	81.38	67.91
		Share of income (%)	*12.7*	*12.9*	*10.8*
	Dir. affected alt. (one car)	Average CV (DKK x 1000)	7.66	9.48	0.97
		Share of income (%)	*1.2*	*1.5*	*0.6*

Notes: 1 DKK I appr. € 0.13.

References

Amato, P. R. and D. Preveti. 2003. People's reasons for divorcing: gender, social class, the life course, and adjustment. *Journal of Family Issues*, **24(5)**, 602-626.

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. and C. Timmins. 2005. On the equilibrium properties of locational sorting models. *Journal of Urban Economics*, **57**, 462-477.

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.

Bierlaire, M. 2006. A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*, **144(1)**, 287-300.

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

Dubé, J.-P., Fox, J. T., & Su, C.-L. 2012. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficient demand estimation. *Econometrica*, **80(5)**, 2231–2267. http://doi.org/10.3982/ECTA8585.

Gibbons, S. and S. Machin. 2005. Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1), pp. 148-169.

Hawkins , A.J., B. J. Willoughby and W. J. Doherty. 2012. Reasons for Divorce and Openness to Marital Reconciliation, Journal of Divorce and Remarriage, 53,6, 453-463

Hess, S., K.E. Train and J.W. Polak. 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice. *Transportation Research B*, **40**, 147-163.

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

Svarer M. and M. Verner. 2008. Do children stabilize relationships in Denmark? *Journal of Population Economics* 21, pp. 395–417.

Van Duijn, M. and J. Rouwendal. 2018. Sorting based on urban heritage and income: Evidence from the Amsterdam metropolitan area. TI Discussion paper 15-030/VII (revised).

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

Wehner, C., P. Abrahamson and M. Kambskard, M. 2002. Demography of the Family: The Case of Denmark. København: FAOS, Sociologisk Institut, Københavns Universitet.