Do I drive further or move closer? Evidence from car inspections and neighborhood sorting in Helsinki*

Arttu Ahonen*, Prottoy A. Akbar* and Pablo Warnes*

*Department of Economics, Aalto University / Helsinki GSE

March 15, 2024

Abstract

We study how residents of Helsinki respond to changes in the costs of urban mobility in terms of their decisions on how much to drive and on car ownership. Cities around the world are increasingly investing in policies and infrastructure aimed at promoting alternative modes of transportationwhile also implementing measures to discourage driving.As urban areas become less conducive to driving, do drivers reduce how much they drive? Do fewer residents opt to own cars? We document two conflicting empirical regularities. First, individuals facing lower travel times by transit (relative to driving) tend to reduce their car usage, both in the extensive margin (kilometers driven) and the intensive margin (car ownership). Second, individual driving habits tend to persist in the long term even as their residential neighborhood becomes relatively less driving friendly.

*Corresponding author: pablo.warnes@aalto.fi

1 Introduction

Driving of private vehicles generates large negative externalities in the form of road congestion, air pollution and carbon emissions. Cities around the world are increasingly investing in policies and infrastructure aimed at promoting alternative modes of transportation, like expanding public transit and building bike lanes, while also implementing measures to discourage driving, such as via traffic calming initiatives and congestion pricing schemes. But how do city dwellers respond to these investments? Specifically, as urban areas become less conducive to driving, do fewer residents opt to own cars? Do car owners drive less?

We study driving behavior in Helsinki between 2013 and 2018, a period of widespread public transit service expansions, and document two conflicting empirical regularities. First, individuals facing lower travel times by transit (relative to driving) from their residential neighborhood tend to reduce both their kilometers driven and their likelihood of car ownership. Second, individual driving habits tend to persist in the long term even as their residential neighborhood becomes relatively less driving friendly. As a result, we do not see large changes in the average kilometers driven or car ownership in the city despite a large reduction in average travel times by transit relative to driving. We present a theoretical framework of travel mode choices and kilometers driven as a function of features of the residential neighborhood (such as how far it is from destinations of interest and how connected it is by public transit versus driving) as well as of individual driving habits. Taking the model to data, we find that most of the variation in kilometers driven and car ownership is in fact attributable to individual habits rather than neighborhood attributes.

We combine detailed administrative data for the universe of residents of Helsinki with vehicle inspection data for all car owners in the city, which allows us to measure the driving habits of each individual across time. We complement this individual-level data with a large set of location specific characteristics, with which we can calculate travel times by mode of travel, as well as an index of "driving friendliness" that measures how attractiveness a location is as a destination to drive to. The fact that we can observe kilometers driven and car ownership decisions at the individual level provides us with a unique opportunity to study how driving behavior changes when the cost of driving (relative to public transit) increases.

We highlight several motivating patterns in the data. We find that between 2013 and 2018, travel times by transit have gone down notably relative to driving times throughout the city. At the same time, total kilometers driven and car ownership in the city have not changed much relative to trends in the rest of Finland, even though we observe that individuals and neighborhoods facing more favorable transit travel times do exhibit lower levels of driving and car ownership. To help reconcile this discrepancy, we look at changes in the driving behavior of those who either moved into or out of Helsinki during our period of analysis. We find that those who move out to locations with higher average car usage tend to increase their own car usage. On the other hand, those who move in to Helsinki from locations higher car usage tend to continue to drive more than incumbent Helsinki residents and at comparable levels to the average at their original location even several years after the move to Helsinki. This asymmetric behavior is suggestive of the persistence of old driving habits but also of the role of locational features in the adoption of new driving habits.

From our theoretical framework of travel mode choices and kilometers driven, we derive an equation that relates the amount of kilometers driven by a resident to a function of the differences in travel times by car and by transit[1]. This function can be interpreted as a weighted measure of how much longer it takes to ride mass transit (rather than drive) to destinations of interest from a given neighborhood. We estimate the coefficients implied by this equation using our data for Helsinki. From this estimation, we find that, when the average difference in travel times in their residential neighborhood is reduced (transit becomes relatively faster), drivers respond by reducing their kilometers driven, and in general, residents respond by reducing their car ownership.

How much of the variation in driving that we observe is explained by the neighborhood individuals reside in versus persistent individual behavior across neighborhoods? To answer this question, we estimate a high-dimensional fixed effects model (in the spirit of Abowd et al., 1999), and we decompose the variation in kilometers driven into a driver fixed effect and a neighborhood effect. Surprisingly, we find that the driver fixed effects account for more than half of the variation in kilometers driven, while the neighborhood effects only account for 0.5% of the variation.

Our findings have important implications for the effectiveness of various urban planning policies at moderating driving behavior. Most traditional policy interven-

[1] We technically use the fastest alternative mode to driving, which can be public transit or walking, but for any distance that is not very short, this tends to be public transit.

tions, such as congestion pricing and public transit provision, target locations rather than travellers. But how responsive is the amount of driving to expanding public transit services or increased walkability across locations? How much do people substitute to these alternative modes? Or does the amount of driving depend primarily on individual characteristics so that the more effective policy interventions are ones that can target marginal drivers based on individual attributes?

The remainder of the paper is structured as follows. The next section discusses our data on driving behavior and characteristics of individuals and locations. Section 3 documents key patterns in the observed driving behavior that motivate our analysis. Section 4 develops a model of travel mode choice and car usage to disentangle the roles of individual habits and locational determinants of driving. Section 5 describes our approach to empirically estimate parameters of the model and interprets our results. Section 6 discusses their implications for the aggregate patterns observed earlier and Section 7 concludes.

2 Data

We combine annual observations of individual level demographic characteristics and driving behavior for the full population of Finland with grid level data on travel times and other location characteristics for the country's capital region – its largest metropolitan area consisting of the municipalities of Helsinki, Espoo, Vantaa, and Kauniainen, with a population of 1.2 million amounting to around a fifth of the entire country.

The individual demographic characteristics, obtained from the administrative database of Statistics Finland, cover the years 1995-2021 and include disposable income, age, family and household size,[2] and the number of children under 18 years old in the household. The data also contains the individual's residential location at the level of a 250 meters by 250 meters grid cell, which we use to compute location characteristics from spatial data, and a pseudonymized personal identifier, which links it to data on car ownership and kilometers driven from the vehicle registry of the Finnish Transport and Communications Agency, *Traficom*.

[2] A family consists of parents and children living in the same house, while a household includes any people living in the same house. Thus, a household can include multiple families, but a family does not include multiple households.

2.1 Car ownership and kilometers driven

The Traficom data covers all vehicles in use within mainland Finland between 2013 and 2023, each linked to all of its owners during that time, with dates recording when the ownership spell started and ended.[3] We limit our attention to passenger cars, vans, and trucks (hereafter referred to collectively as "cars") owned by private individuals.

The vehicle registry also includes odometer readings recorded at vehicle inspections. The inspections are mandatory for all cars registered for road use in Finland, and take place at regular intervals, the length of which depends on the car's age. Before 2018, new cars faced their first inspection after three years, second after five years, and an annual inspection after that (the "3-2-1-1 model"), while after 2018 the intervals became longer, with the first one taking place after four years and then every other year until ten years, after which the inspections are again annual (the "4-2-2-2-1 model"). In addition to the readings recorded at the inspections, we impute a zero-reading for all cars on the date that they are first registered in use in Finland or abroad.

Since a car might change hands in the middle of an inspection interval, it is not always possible to attribute the kilometers driven between two inspections to a single individual. To account for this, we also use odometer readings reported by sellers at *nettiauto.com*, Finland's largest online car marketplace, which covers most of the cars on sale in the country by both private individuals and dealerships. The sales listings contain the car's registration number, which is used to link them to the vehicle registry data. The listings data, provided by the company that runs the site, extends from 2006 to 2023, though we only use observations after 2013 to keep the time period consistent with the vehicle registry data.

The period between two odometer readings – regardless of the source – constitutes a driving spell, which we compare to the ownership spells to determine which individual to attribute the kilometers driven to. In particular, we link the driving spell to the individual who is responsible for the majority of the days within it, disregarding

[3] We use the term "owner" here to refer to the person or firm registered as the possessor of the vehicle, distinct from the registered owner (though usually these two are the same), since the possessor is the one owning the right to use the vehicle. The start date of the possessor relationship is directly recorded in the data, while we assign the end date to be either when the next possessor takes control of the vehicle, when the vehicle is decommissioned for the last time, or the end of the last quarter during which the car is observed in the data if no decommission date is observed.

the 3% of spells where no individual is responsible for more than half of the days. 74% of the driving spells are fully covered by one owner, while the average overlap is 93% of the driving spell.

Note that the ownership and driving spells do not correspond to calendar years, while the rest of our data is at that level. To map the driving spells to kilometers driven by an individual within a calendar year, we first compute the length (in days) of the intersection between a given year and an ownership spell, and then multiply that by the mean daily kilometers driven by the car in question within driving spells that overlap the intersection, weighted by the length of the overlap, finally summing over all cars linked to an individual. This means that the same driving spell can contribute to the kilometers driven -estimate for multiple calendar years, which may bias our estimates of the impacts of travel times towards zero, as changes in driving are not fully captured by the measure in the short term.

2.2 Travel times and driving friendliness

From the Statistics Finland data, we observe each individual's residential location at the level of a 250 meters by 250 meters grid. We link this residential location data to the Helsinki Region Travel Time Matrix Tenkanen and Toivonen (2019), an open data set recording the travel times by different modes between the centroids of any two such grid cells in the Helsinki metropolitan area for 2013, 2015, and 2018. We use mid-day travel times by car, public transit, and walking (as these are available for all three years), to construct measures of average travel time differences between car and the fastest alternative from an individual's residential location to potential destinations (see section 5.2 for a detailed description).

In constructing the measures, we weight the potential destinations by proxies for their popularity and driving friendliness. For the former, we use population density, calculated simply as the number of (adult) individuals residing in a given cell, while for the latter, we construct an index based on several characteristics of the location that we find to have the most predictive power in a machine learning model of mode choice using travel survey data from the Helsinki region transport authority (HSL) for the year 2018.[4] The HSL travel survey includes information on 17,137 journeys that originated and culminated within the Helsinki metropolitan area. For these

[4]See section 5.1 for a description of how we construct this driving friendliness measure.

journeys, the survey record the primary mode of travel, a large set of demographic characteristics of the respondents, the exact coordinates of the origin and destination of each journey, and the travel time and distance of the journey among other things.

Most of the location characteristics entering the machine learning model are computed using the Finnish Transport Infrastructure Agency's national road and street database, Digiroad. These include measures of transit stop and intersection density[5], and the total lengths of roads intersecting a given square kilometer grid cell, as well as the lengths of roads with a speed limit over 40 kilometers per hour, and roads that exclude motorized traffic. In addition to road lengths, we compute the area of the square kilometer cell covered by roads for which width-information is available, and include the traffic count on the major road nearest to the 250m-cell centroid, as well as the distance to that road.[6]

Meanwhile, some location characteristics are based on geo-located data from the Finnish Environment Institute. First, we use their definition and map of central and trade areas, computing the distance from each square kilometer cell centroid to the nearest such area. We also use their Corine Land Cover 2018 raster data to divide each square kilometer grid cell to 2500 smaller cells, and compute the share of the square kilometer covered by different types of land. We aggregate some of the several detailed natural land cover categories included in the CLC 2018 data, while keeping the built environment categories at a disaggregate level (see table TTT in section C.1 of the Appendix for the full classification).

3 Empirical regularities

We start by noting several notable patterns in our data on travel times, vehicle kilometers driven and car ownership in Helsinki and in comparison to the rest of Finland.

[5]Specifically, the mean distances to the 10 and 50 nearest transit stops or intersections from the center of the 250m-cell centroid, and the number of transit stops or intersections at the level of a square kilometer grid cell. Only intersections of roads available for non-motorized traffic are counted, and intersections within less than ten meters of each other are counted as one.

[6]The Finnish Transport Infrastructure Agency only performs traffic counts for relatively large roads. For example, in the Helsinki region the data is only available for the seven freeways coming into the city and the three ring roads between them.

3.1 Travel times by transit relative to driving

Like most cities around the world, Helsinki has been increasingly investing in alternative travel modes to driving. Over the last several years, the city has seen several major expansions of its mass transit network, including expanding the metro network's coverage by over 60% in 2017. Consistent with these transport infrastructure expansions, travel times on Helsinki's mass public transit have improved greatly in recent years relative to driving times. To see this, we compute the (population-weighted) average travel time between each pair of neighborhoods in the Helsinki region by driving and by transit.[7] The average difference between transit and driving travel times moved from 30 minutes in 2013 to 31 minutes in 2015 before dropping sharply to 28 minutes in 2018 (possibly in response to Helsinki's West Metro expansion).

The trend of improved relative transit travel times between 2015 and 2018 holds not just between the average neighborhoods but across neighborhoods throughout the city. To illustrate, from each neighborhood, we compute the (population-weighted) average of the difference between transit and driving travel times to every other neighborhood. Figure 1 shows the change in this average travel time from 2015 to 2018, with positive values and bluer shades indicating transit times decreasing relative to driving times and with negative values and redder shades indicating the opposite.[8] In most parts of the city, relative transit travel times have decreased and more so in central and densely populated areas.

Because the population weights on the trips corresponding to these travel times are exogeneously fixed across years, the travel time changes reflect either improvements in the transit network or slowdowns in the driving network. That said, the changes in average travel times look similar even when we allow population weights to vary across years. We may worry that these trips are not representative of trips actually taken. So we also compute average travel times on observed residence-work location pairs in Helsinki and find that relative transit travel times have improved significantly even as commuting patterns have evolved over the years.

[7] We fix the population weights to the neighborhood populations in 2013. For neighborhood pairs where the travel time by transit appears larger than the travel time by walking, we assume the transit time equals the walking time.

[8] Appendix Figure A.1 reproduces this figure using rush-hour travel times instead of the mid-day travel time estimates from the Helsinki Region Travel Time Matrix. Relative transit travel times appear to have improved even more during the rush hours. The rest of the paper will focus on the mid-day travel times.

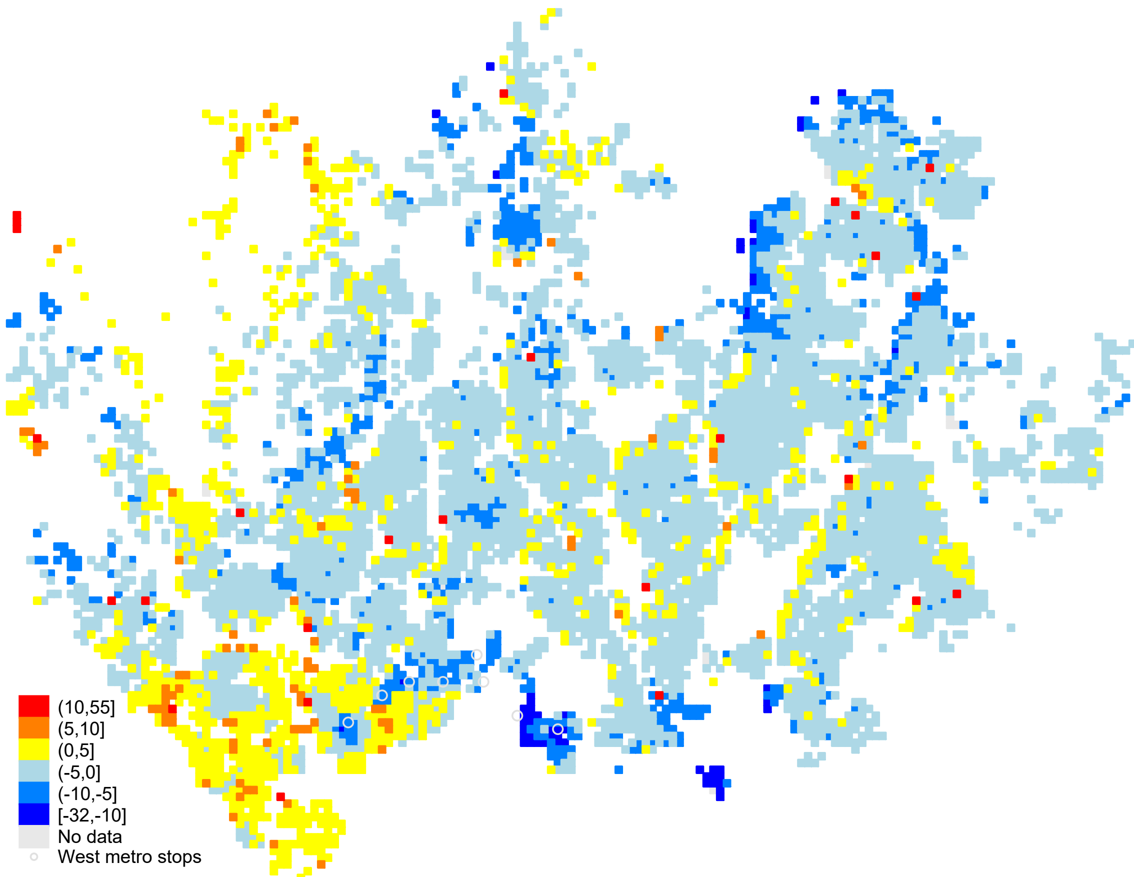
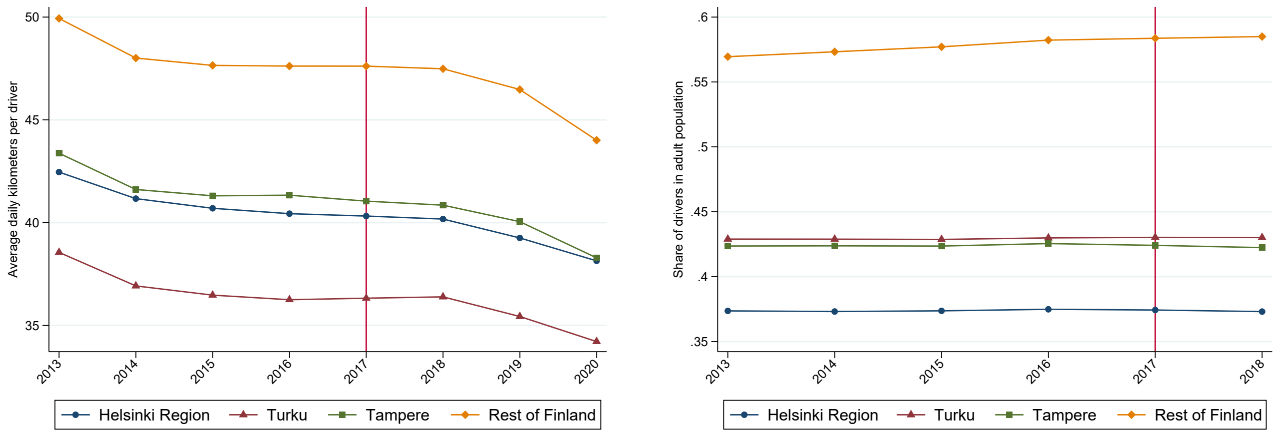


Figure 1: Changes in transit travel times relative to driving between 2015 and 2018.



A. Vehicle kilometers driven **B.** Car ownership

Figure 2: **Trends in aggregate driving behavior**: in Helsinki, in the next two largest metropolitan areas, and in the rest of Finland.

3.2 Driving and car ownership

If travellers care about travel times, we would expect them to respond to these widespread improvements in transit travel times relative to driving times by transitioning away from driving. Figure 2 panel A plots the average distance driven per day by car owners between 2013 and 2020 in Helsinki as well as, for comparison, in the next two largest metropolitan areas (Turku and Tampere) and the rest of Finland. The red vertical line indicates the year of Helsinki's grand West Metro expansion. While car owners throughout the country are driving less, the trends in Helsinki are not notably different from trends elsewhere in Finland. In particular, average vehicle kilometers driven between 2015 and 2018 seem to have hardly changed despite the large travel time changes around the same time.

We wonder if this is because fewer people choose to own cars in later years, and especially those who would be most responsive to travel time changes. Figure 2 panel B plots the rate of car ownership between 2013 and 2018. Car ownership rates have remained mostly unchanged in Helsinki and the other large cities while increasing slightly for the rest of Finland. So, the average urban resident's choice of car ownership does not appear to be responsive to the widespread travel time changes between 2015 and 2018.

However, at the neighborhood level, we do find that improvements in relative transit travel times coincide with decreases in kilometers driven and car ownership.

3.3 Place effects and individual habits

One possible explanation for the discrepancy between the neighborood- and city-level changes in driving and car ownership may lie in a third margin of adjustment in response to travel time changes: residential relocation. When a neighborhood becomes more accessible by transit relative to driving, avid resident drivers may avoid switching travel modes by relocating to a different neighborhood. On the other hand, drivers who move into more transit-accessible neighborhoods may be ones who already drove little. Such re-sorting of low- and high-frequency drivers across neighborhoods would yield lower kilometers driven in the more-transit accessible neighborhoods even in the absence of aggregate changes in kilometers driven across neighborhoods. We will show that such systematic sorting of drivers across neighborhoods in Helsinki is not common.

A second explanation is that locational features, such as public transit accessibility, are not sufficient on their own to alter entrenched individual driving habits. To understand how individuals' location choices affect their driving behavior, consider Figure 3 where we follow the average kilometers driven by drivers who moved in to Helsinki between 2014 and 2017 before and after their move. Each graph depicts with solid red line the average kilometers driven for a different cohort of movers. The year of move is shown as a vertical red line. In addition, for comparison, blue lines depict the average kilometers driven by those who resided in Helsinki the entire time (2013-2020) and green lines depict the average of those who continue to reside in the municipalities where the movers move from. For the latter, we compute the average across all municipalities (outside of the Helsinki region) weighted by the share of moves that originate there.

Across all cohorts of car owners who move, three empirical regularities stand out. First, movers to Helsinki on average tend to come from regions with higher levels of driving, but their average kilometers driven drop sharply right after the move to Helsinki. We do not observe such sharp changes for non-movers. This suggests that location-specific features and norms may be important determinants of current driving behavior. When we plot an equivalent set of graphs for car owners who *move out* of Helsinki (Appendix Figure A.3), we find that, consistent with the reasoning above, they typically move out to regions with higher average driving and their vehicle kilometers rise sharply right after the move. We observe similarly consistent sharp

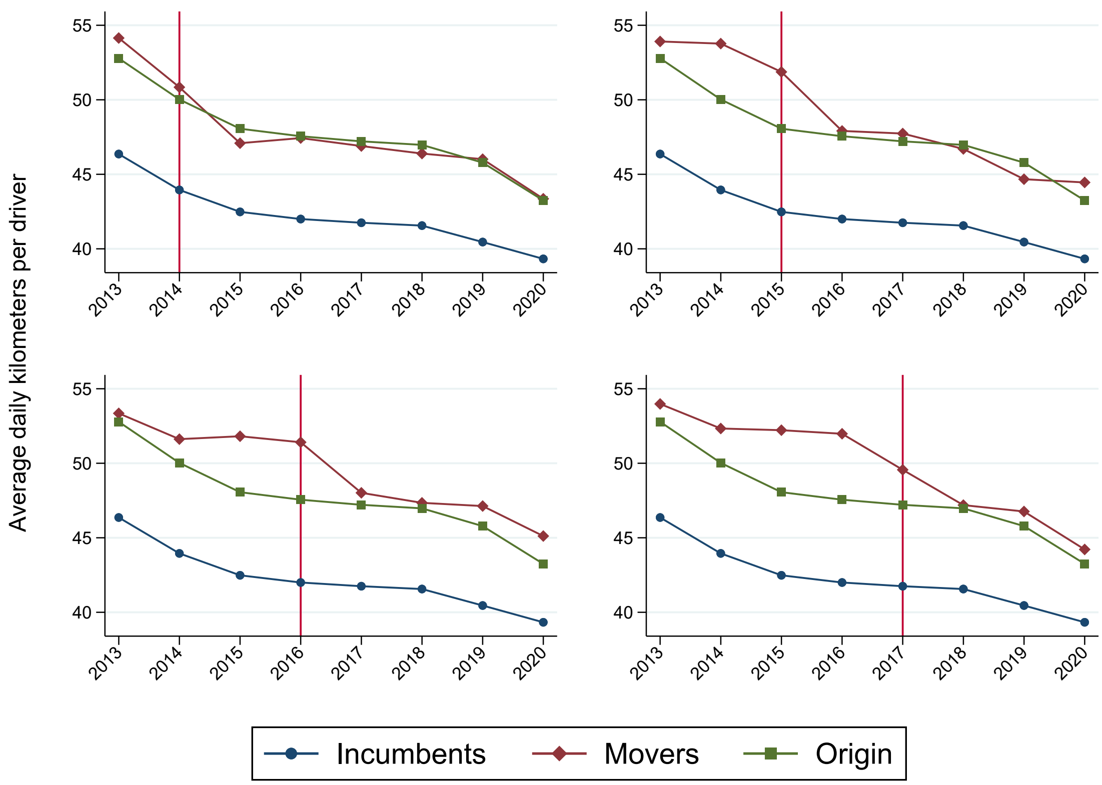


Figure 3: **Average vehicle kilometers driven per day** among three groups of car owners: (i) "Incumbents": those who who reside in Helsinki the entire time (2013-2020), (ii) "Movers" those who move in to Helsinki in either 2014 (top-left), 2015 (top-right), 2016 (bottom-left) or 2017 (bottom-right), and (iii) "Origin": those residing in the municipalities where movers are moving from (average across municipalities weighted by fraction of movers). Red vertical line indicates the year of the move.

changes in car ownership when people move into Helsinki (Appendix Figure A.4) or out of Helsinki (Appendix Figure A.5).

Second, movers differ notably from incumbent residents of Helsinki (i.e. those who have resided in Helsinki 2013 through 2020). They tend to drive significantly more both before and after the move to Helsinki. Even 4-5 years after the move, they continue to drive more and do not appear to be converging to the driving behavior of long-term Helsinki residents. Such behavior is suggestive of the persistence of past driving habits. Moving and settling in a region where the norm is to drive much less is not sufficient to offset individual habits.

Third, kilometers driven by movers tend to resemble average kilometers driven in the municipalities they moved from. With the exception of the years of and just leading up to the move (which could be due to the circumstances of the move), movers' driving behavior tend to mimic the levels observed among car owners in the original municipalities.[9] This seems to suggest that original location plays an important role in the formation of long-term driving habits.

While we have so far presented evidence from inter-city moves, we expect these insights to also generalize to moves across residential neighborhoods within a city. Within Helsinki, we observe widespread changes across neighborhoods in travel times (as in Figure 1), vehicle kilometers driven and car ownership. We also observe 38% of all adult residents and 31% of all car owners in Helsinki changing residential neighborhoods at least once between 2013 and 2018. Such variation in individuals' exposure to neighborhoods allows us to distinguish the extent to which their current driving behavior can be explained by norms and characteristics of the individuals as opposed to the neighborhoods they reside in. Section 4 proposes a theoretical framework to help us quantify the role of each in rationalizing both local and citywide changes in driving behavior.

Before moving on, we note that there may be reason to suspect that relocation decisions are not random but highly correlated with pre-move driving behavior and car ownership. For instance, in the case of inter-city moves, we observe that those who move out of Helsinki were already driving a lot more than the average non-mover in Helsinki (Figure A.3). They also had a higher car ownership rate (Figure A.5).

[9] This does not seem to be the case for car owners who move out of Helsinki (as in Appendix Figure A.3). But those who move out of Helsinki are a very selected sample in that they already drove a lot more in Helsinki prior to moving out than those who continue to reside in Helsinki.

Whereas, those who move into Helsinki were much less likely to own cars relative to the average resident (Figure A.4). Yet, when comparing movers to incumbents within residential neighborhoods in Helsinki, we find very little systematic residential sorting based on pre-move driving behavior. So, we currently abstract away from this potential endogeneity of residential location choices in the theoretical framework presented below.

4 Theoretical Framework

Consider a city with a fixed population of travellers. Each traveller faces a measure 1 of different trips (such as shopping, commuting, etc.), indexed by q. Conditional on residing in neighborhood n in year t, traveller i faces a travel cost $c_{injtq}(m)$ to each destination j of interest within the city that depends on their choice of travel mode $m \in \{\text{CAR}, \text{ALT}\}$ (either driving or mass transit).

4.1 Travel mode choices

For each trip q to destination j, travellers choose mode m to minimize the following cost function:

$$c_{injtq}(m) = \bar{c}^m_{njt} + \left[\nu_{it} + \phi_q\right] \cdot \mathbf{I}_{m=\text{ALT}} \tag{1}$$

where $\bar{c}^m_{njt}$ incorporates costs of using mode m on trips from n to j that are invariant across all travellers (such as trip distances and gas prices) and ν_{it} denotes idiosyncratic preferences of individual travellers for choosing CAR over ALT. Trips via mass transit may be subject to different waiting times before the trip can be started. The associated waiting and scheduling costs are denoted by ϕ_q. We assume mass transit departures happen at a constant rate following a Poisson process such that ϕ_q are drawn from an exponential distribution with c.d.f. $\Phi(x) = 1 - e^{-x}$.

Let $m^\star$ denote the optimal mode choices on each trip. Then, between any origin n and destination j, the share (and the number) of trips where traveller i chooses to drive is:

$$\bar{\Phi}_{injt} \equiv \Pr[m^\star_{injtq} = \text{CAR}] = 1 - \Pr\left[c_{injtq}(\text{ALT}) < c_{injtq}(\text{CAR})\right]$$

$$
\begin{aligned}
&= 1 - \Pr\left[\bar{c}^{\text{ALT}}_{njt} + \nu_{it} + \phi_q < \bar{c}^{\text{CAR}}_{njt}\right] \\
&= 1 - \Phi\big(\bar{c}^{\text{CAR}}_{njt} - \bar{c}^{\text{ALT}}_{njt} - \nu_{it}\big) \\
&= \exp\left(\nu_{it} + \bar{c}^{\text{ALT}}_{njt} - \bar{c}^{\text{CAR}}_{njt}\right) \\
&= \exp\left(\nu_{it}\right) \cdot \exp\left(\bar{c}^{\text{ALT}}_{njt} - \bar{c}^{\text{CAR}}_{njt}\right) \qquad (2)
\end{aligned}
$$

Note that the first multiplicative component is a traveller-specific determinant of driving irrespective of the trip origin and destination. Whereas, the second component is 'objective' in that it depends only on trip locations and is invariant across travellers.

We can further decompose the objective component of the probability of driving as follows:

$$\bar{c}^{m}_{njt} \equiv \left(l_{nj}/s^{m}_{njt}\right) + \gamma^{m}_{jt} + \kappa^{m}_{nt} \qquad (3)$$

where l_{nj} is the driving distance between n and j, s^{m}_{njt} is the travel speed on mode m, and κ^{m}_{nt} and γ^{m}_{jt} incorporate other non-time costs of using travel mode m from trip origin n and destination j.[10] Note that the last two cost parameters incorporate the fixed costs of travel mode m (such as of vehicle ownership or a seasonal bus pass) as well as costs that vary across space (such as ease of parking or proximity to mass transit stops) and across years (such as when gas prices change or new transit routes are introduced).

Then

$$\bar{\Phi}_{injt} = \exp\left(\nu_{it}\right) \cdot \exp\left(\Delta\kappa_{nt}\right) \cdot \exp\left(\Delta\gamma_{jt}\right) \cdot \exp\left(l_{nj} \cdot \Delta s^{-1}_{njt}\right) \qquad (4)$$

where $\Delta\kappa_{nt} \equiv \kappa^{\text{ALT}}_{nt} - \kappa^{\text{CAR}}_{nt}$, $\Delta\gamma_{jt} \equiv \gamma^{\text{ALT}}_{jt} - \gamma^{\text{CAR}}_{jt}$, and $\Delta s^{-1}_{njt} \equiv \left(1/s^{\text{ALT}}_{njt}\right) - \left(1/s^{\text{CAR}}_{njt}\right)$.

4.2 Distance driven

Destinations differ in their amenity value ψ_{jt}, which is proportional to the probability of a trip being taken to them. So, the (expected) aggregate distance travelled by residents of neighborhood n is a weighted sum of the distances to all destinations:

[10] Because l_{nj} is the driving distance and is invariant across modes, we interpret s^{ALT}_{njt} as incorporating both how fast travellers move while on mass transit but also how efficiently the mass transit network connects locations n and j relative to the driving road network.

$$L_n = \sum_j \psi_j \cdot l_{nj} \tag{5}$$

Given residential location n and mode choices $m^\star$ across trips, the (expected) total distance driven by traveller i is

$$d^\star_{int} \equiv d_{int}(m^\star) = \sum_j \bar{\Phi}_{injt} \cdot \psi_j \cdot l_{nj} \tag{6}$$

In each year t, traveller i chooses to drive a total distance of d_{int}, while the rest of the aggregate travel distance, $L_{in} - d_{int}$, is covered by alternative travel modes. We can plug in the probability of driving into the formulation of total distance driven to get:

$$d^\star_{int} = \exp\big(\nu_{it}\big) \cdot \exp\big(\Delta\kappa_{nt}\big) \cdot \delta_{nt} \cdot W_n \tag{7}$$

where

$$\delta_{nt} \equiv \sum_j \left(\frac{\omega_{nj}}{W_n} \cdot \exp\big(l_{nj} \cdot \Delta s^{-1}_{njt}\big) \right) \tag{8}$$

$$\omega_{nj} \equiv \exp\big(\Delta\gamma_j\big) \cdot \psi_{jt} \cdot l_{nj} \text{ and } W_n \equiv \sum_j \omega_{nj}$$

The term δ_{nt} is a weighted measure of how much longer it takes to ride mass transit (rather than drive) to destinations of interest from neighborhood n, and where destination weights ω_{nj} incorporate their amenity value (ψ_j) and driving friendliness ($\Delta\gamma_j$) as well as their driving proximity (l_{nj}). The neighborhood-specific multiplier W is a re-scaling of the weights so that they add up to 1.

Then the log of distance driven can be written as:

$$\ln(d^\star_{int}) = \ln(\delta_{nt}) + \Delta\kappa_{nt} + \ln(W_n) + \nu_{it} \tag{9}$$

4.3 Empirical interpretation

Empirically, we only observe $d^\star_{int}$, l_{nj} and s^m_{njt}. To ease interpretation, we can rewrite the unobservable parameters ν_{it} and $\Delta\kappa_{nt} + \ln(W_n)$ as the sum of a non-time-varying mean, a common temporal shock and an idiosyncratic temporal shock:

$$\nu_{it} \equiv \nu^0_i + \xi^\nu_t + \epsilon^\nu_{it}$$

$$\Delta\kappa_{nt} + \ln(W_n) \equiv \Delta\kappa_n^0 + \xi_t^{\Delta\kappa} + \epsilon_{nt}^{\Delta\kappa}$$

This decomposition allows us to re-write Equation 9 as:

$$\ln(d_{int}^{\star}) = \ln(\delta_{nt}) + \Delta\kappa_n^0 + \nu_i^0 + (\xi_t^{\nu} + \xi_t^{\Delta\kappa}) + (\epsilon_{nt}^{\kappa} + \epsilon_{it}^{\nu}) \tag{10}$$

where log distance is a sum of weighted access to destinations, a neighborhood-specific effect, an individual-specific effect, a year-specific effect, and some idiosyncratic temporal shocks. Equation 10 forms the basis for our empirical specification.

5 Empirical Estimation

5.1 Creating a measure of driving friendliness

In order to estimate the impact of changes in travel times on driving behavior, we start by operationalizing our theoretical measure δ_{nt}, which is a weighted average of the exponential difference in travel times between driving and riding mass transit (see equation 8). One key component of this δ_{nt} measure is the difference in the non-time costs to travelling to a destination j that are associated with each mode: $\Delta\gamma_{jt} = \gamma_{jt}^{CAR} - \gamma_{jt}^{ALT}$. We will interpret this $\Delta\gamma_{jt}$ as a measure of how driving friendly a destination is, relative to travelling by the fastest alternative mode (walking or mass transit).

From equation 4, we can derive an expression for the probability of choosing driving as the mode of travel for a given trip from an origin o to a destination d for an individual i:

$$d_{i,o,d} = \exp\left(\nu_i + \beta\left(t_{od}^{ALT} - t_{od}^{CAR}\right) + \left(\gamma_d^{ALT} - \gamma_d^{CAR}\right) + \left(\kappa_o^{ALT} - \kappa_o^{CAR}\right)\right). \tag{11}$$

Where $t_{id}^{m} = \left(l_{nj}/s_{njt}^{m}\right)$ is the travel time from o to d by mode m. We then parameterize this expression by allowing ν_i to depend on a set of individual level characteristics,[11] X_i, so that $\nu_i = \beta_1 X_i$. In the same way, we will model the difference in non-travel related costs at the origin level as depending on a set of origin-specific covariates,[12] Ω_o,

[11] We will use ages, gender, a person's household size, and their level of education as covariates.
[12] For now, we use the total population at the origin.

so that $\left(\kappa_o^{ALT} - \kappa_o^{CAR}\right) = \alpha_o \Omega_o$. Finally, the driving friendliness of each destination is modeled as a function of a series of covariates that vary at the level of the destination, Ω_d. Under these assumption, equation 11 becomes:

$$d_{i,o,d} = \exp\left(\beta_1 X_i + \beta_2 \left(t_{od}^{ALT} - t_{od}^{CAR}\right) + \alpha_o \Omega_o + \alpha_d \Omega_d\right). \tag{12}$$

Within this equation, we are particularly concerned with estimating properly $\alpha_d \Omega_d$. To do so, we assemble a large set of eighty-two different indicator variables that describe each location[13], including information on the land cover, population, the road traffic near each cell, the road infrastructure, among other things (see section C.1 of the Appendix for a detailed description of the variables included).

We combine these data with information on 17,137 journeys recorded within the Helsinki metro area in 2018 obtained from the Helsinki Regional Transport Authority's travel survey. Since we have the precise coordinates for the origin and destination of each of these trips, we can match them to a destination and an origin cell. We also use the demographic information from this survey to construct our vector of individual level characteristics for each individual in our sample, X_i. Finally, we use classify each journey according to the mode of transport used, and build our outcome variable $d_{i,o,d}$ as an indicator that takes value one if the journey was made by car, and value zero otherwise.

We estimate equation 12 by implementing the penalized Poisson pseudo-maximum likelihood method (PPML) proposed by Breinlich et al. (2022). This method combines a lasso method for variable selection with a classical PPML estimator.[14] Through this process, we are left with a parsimonious set of thirty-six of the original eighty-two variables in Ω_d, we will call this selected subset $\tilde{\Omega}_d$.[15] By combining the penalized PPML estimated coefficients for the destination-specific variables, $\hat{\alpha}_d$, and the set of selected variables $\tilde{\Omega}_d$, we can construct our driving friendly index:

$$\mathcal{F}_d = \exp(\hat{\alpha}_d \tilde{\Omega}_d). \tag{13}$$

This index of driving friendliness ($\mathcal{F}_d$) will function as our proxy for $\exp\left(\Delta\gamma_d^m\right)$.

[13] A location is defined by a 250m-by-250m grid cell within the Helsinki Metro Area.

[14] PPML estimators are often used in situations where log- linearized models estimated by OLS lead to biased estimates (Silva and Tenreyro, 2006).

[15] See section C.1 of the Appendix for a list of all the variables included in $\tilde{\Omega}_d$.

5.2 Measuring average differences in travel times by mode

The term δ_{nt} in equation 8 is a function of how much longer it would take to ride public transit (or walk if this is faster) rather than drive from an origin n to all destination j within the city. If we operationalize $\exp\left(\gamma_d^{ALT} - \gamma_d^{CAR}\right)$ with $\mathcal{F}_d$, as described in subsection 2.2, and ψ_{jt} with the population share in destination j: $\frac{L_{jt}}{L_t}$, we can re-write this equation as:

$$\Delta\mathrm{WTT}_{nt} = \sum_{j=1}^{J} \left(\mathcal{F}_j \frac{L_{jt}}{L_t}\right) \times l_{nj} \times \exp(t_{nj}^{ALT} - t_{nj}^{CAR}), \tag{14}$$

where $\Delta\mathrm{WTT}_{nt}$ is our empirical counterpart to δ_{nt}, l_{nj} is the driving distance from n to j, and t_{nj}^{m} is the travel time from n to j by mode m (either driving or the fastest alternative).

5.3 Effect of changes in travel times on driving behavior

In this section, we will estimate different version of equation 10. We will rely on the $\Delta\mathrm{WTT}_{nt}$ measure described in subsection 5.2 as our way of capturing the changes in accessibility by transit relative to driving. Because this measure is monotonically increasing in transit travel time (and decreasing in driving time), if a neighborhood becomes better connected by public transit, relative to driving, $\Delta\mathrm{WTT}_{nt}$ will decrease. If this decrease leads to a reduction in driving by residents of this neighborhood, then we would expect a positive relationship between the changes in the average daily kilometers driven by individuals in a neighborhood and changes in $\Delta\mathrm{WTT}_{nt}$. We start by testing this relationship at the level of the neighborhood.

5.3.1 Effects at the neighborhood level

In order to test the relationship between the average daily kilometers driven and $\Delta\mathrm{WTT}_{nt}$, we start by estimating the following equation, which is based on equation 10, on data at the level of the neighborhood:[16]

$$Y_{n,t} = \alpha + \beta \log(\Delta WTT_{n,t}) + \rho_n + \rho_t + X_{n,t} + \varepsilon_{n,t}. \tag{15}$$

[16]We define a neighborhood in these regressions as a 250m-by-250m grid cell.

Where $Y_{n,t}$ is the log of the average daily kilometers driven by all drivers living in neighborhood n at time t and $X_{n,t}$ is a vector of control variables: the average (across drivers) of the number of children that are less than 18 years old in their household, the average income and the average household size. ρ_n and ρ_t are fixed effects at the neighborhood level and year level respectively. We estimate this regression using data for the Helsinki metro area for the years 2013, 2015 and 2018.

From table 5.3.1, we can see that a 1% increase in $\log(\Delta WTT_{n,t})$ leads to a increase in average kilometers driven of 16.6%.[17] This relationship is significant at a 99% level. A positive coefficient for $\log(\Delta WTT_{n,t})$ means that, when driving becomes more appealing (driving times decrease relative to transit), people respond on average by increasing their driving. Conversely, when transit times decrease relative to driving, people respond, on average, by decreasing their daily kilometers driven.

Table 1: Effects of travel time differences - Neighborhood level

	log of daily km driven	
log travel time diff. ($\log(\Delta WTT_{n,t})$)	0.159^{***}	0.166^{***}
	(0.029)	(0.029)
N	18,014	18,001
Year FE	Yes	Yes
Cell FE	Yes	Yes
Dem. Controls	No	Yes

Notes:

5.3.2 Individual level regressions

One concern that might arise from analyzing the effects of changes in travel times on the aggregate driving at the neighborhood level is that of sorting. If high-driving individuals sort into neighborhoods where transit becomes relatively slower, and low-driving individuals sort into neighborhoods that see an improvement in transit times (relative to driving), these sorting patterns could also explain the positive relationship between reductions in travel times by transit and reductions in kilometers driven.

[17] The point estimate is 15.9% if we take the estimates from the regression without demographic controls, although the difference is not statistically significant.

To address this concern, we leverage the full granularity of our data and estimate a version of equation 10 at the individual level:

$$Y_{i,n,t} = \alpha + \beta \log(\Delta WTT_{n,t}) + \rho_n + \rho_t + \rho_i + X_{i,t} + \varepsilon_{i,t}. \tag{16}$$

Where $Y_{i,t}$ is the log of the average kilometers driven for individual i, living in neighborhood n at time t, $X_{i,t}$ is the same vector of time-varying controls as in the previous section, but now at taken at the individual level (i.e., the number of children that are less than 18 years old in driver i's household, driver i's income and driver i's household size). ρ_n, ρ_t and ρ_i are fixed effects at the neighborhood, year and individual level respectively. As with the previous estimation, we take all drivers in the Helsinki metro area for the years 2013, 2015 and 2018.

From table 5.3.2, we can see that the results of estimating equation 16 are very similar to what we got when estimating the effects of $\log(\Delta WTT_{n,t})$ on driving behavior at the neighborhood level. A 1% increase in $\log(\Delta WTT_{n,t})$ leads to a increase in the average daily kilometers driven for an individual living in n at time t of 16.1%.[18] This relationship is significant at a 99% level.

Table 2: Effects of travel time differences - Individual level

	log of daily km driven	
log travel time diff. ($\log(\Delta WTT_{n,t})$)	0.167***	0.161***
	(0.026)	(0.026)
N	854,662	849,503
Year FE	Yes	Yes
Cell FE	Yes	Yes
Indiv. FE	Yes	Yes
Dem. Controls	No	Yes

Notes:

5.4 Decomposing the role of neighborhoods

How much of the variation in driving that we observe is explained by the neighborhood individuals reside in versus persistent individual behavior across neighborhoods? To

[18] The point estimate is 16.7% if we take the regression without the demographic controls, but the difference is between these estimates is not statistically significant.

answer this question, we perform an AKM decomposition (Abowd et al., 1999) of the variation in driving. In the literature of labor economics, this type of decomposition is often performed to quantify the contributions of workers and firms to the variation in wage earnings. In this case, we will quantify the contribution of the place of residence (neighborhood) and the individual driver's contribution to the variation in daily kilometers driven.

Because most drivers do not move between residential locations that often, this is clearly a setting that might suffer from the incidental parameters model that often arises from these models,[19] which is often described as a "limited mobility" bias (Abowd et al., 2004; Andrews et al., 2008, 2012). To address this potential "limited mobility" bias, we implement the leave-out estimator developed by Kline et al. (2020).

We start by estimating equation 16 on the leave-one-out connected set of neighborhoods[20] for the years 2013, 2015 and 2017. We then estimate the unbiased estimator of the covariance matrix proposed by Kline et al. (2020).[21]

From Table 3, we see that the driver fixed effects account for 50.5% of the variation in kilometers driven, while the neighborhood effects only account for 0.5% of the variation. This result suggests that, controlling for the difference in travel times (through $\Delta WTT_{n,t}$), most of the variation in driving in our sample does not come from variation across residential locations, but instead is driven by a large degree of dispersion in terms of driving behavior between drivers.

When we look at the covariance between the driver effects and the neighborhood fixed effects (Table 3), we see that, although it is positive, it is extremely low[22] (0.001). This suggests that, although there might be some degree of sorting, where drivers who drive more sort into neighborhoods where the average driving is higher, this sorting appears to be very limited.

[19]See Bonhomme et al. (2023) for a survey of employer-employee matched data sets where this bias is present in the labor economics literature.

[20]This is the set of neighborhoods that remains connected when any one individual-year (i, t) combination is removed.

[21]We use the executable version of the *VarianceComponentsHDFE* package developed by Paul Corcuera, which can be accessed here (as of March 14, 2024).

[22]The correlation between the two fixed effects is 2.5%

Table 3: Decomposing the variation in daily kilometers driven

	Value	Share
Total variance (log daily km driven)	0.804	1.000
Var. of driver effects	0.406	0.505
Var. of neighborhood effects	0.004	.005
Covariance neighborhood-driver effects	0.001	0.001
Num. of observations (leave out sample)	796,766	
Num. of drivers	297,924	
Num. of movers	83,096	
Num. of neighborhoods	4,972	

Notes:

5.5 Car Ownership Regression

Although we do not model the decision to own a car in our theoretical framework, in this section we will test if there is a relationship between the likelihood of owning a car and the accessibility of an individual's residential location by public transit, relative to driving. To do so, we will estimate a version of equation 16, but where the outcome of interest ($Y_{i,n,t}$) is now an indicator variable that takes value one if an individual i living in location n at time t owns a car, and takes value zero otherwise. As with our previous regressions at the individual level, we include fixed effects at the level of the individual, at the level of the neighborhood, and for each year. We estimate this regression equation on the sample of all residents of the Helsinki metro area of age 18 or older for the years 2013, 2015 and 2018.

From table 4, we can see that a 1% increase in $\log(\Delta WTT_{n,t})$ is associated with an increase of 3 percentage points in the probability of owning a car.[23] If we consider that the car ownership rate in the city of Helsinki is 37%, a average change of 3 percentage points would imply an 8.1% change relative to the average rate of car ownership.

[23]The point estimate is slightly lower, at 2.5 percentage points for the regression without individual level demographic controls.

Table 4: Effects of travel time differences on car ownership - Individual level

	Car ownership	
log travel time diff. ($\log(\Delta WTT_{n,t})$)	0.025***	0.030***
	(0.006)	(0.006)
N	2,473,112	2,397,797
Year FE	Yes	Yes
Cell FE	Yes	Yes
Indiv. FE	Yes	Yes
Dem. Controls	No	Yes

Notes:

6 Understanding Aggregate Trends

Explaining Aggregate Facts from Estimation Results

7 Conclusions

References

Abowd, John M, Francis Kramarz, and David N Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333. Cited on pages 3 and 22.

_ , _ , Paul Lengermann, and Sébastien Pérez-Duarte, "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States," *Unpublished manuscript*, 2004, *5.* Cited on page 22.

Andrews, Martyn J, Len Gill, Thorsten Schank, and Richard Upward, "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 2008, *171* (3), 673–697. Cited on page 22.

—, Leonard Gill, Thorsten Schank, and Richard Upward, "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias," *Economics Letters*, 2012, *117* (3), 824–827. Cited on page 22.

Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler, "How much should we trust estimates of firm effects and worker sorting?," *Journal of Labor Economics*, 2023, *41* (2), 291–322. Cited on page 22.

Breinlich, Holger, Valentina Corradi, Nadia Rocha, Michele Ruta, JMC Santos Silva, and Thomas Zylkin, "Machine learning in international trade research-evaluating the impact of trade agreements," 2022. Cited on page 18.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten, "Leave-out estimation of variance components," *Econometrica*, 2020, *88* (5), 1859–1898. Cited on page 22.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *The Review of Economics and statistics*, 2006, *88* (4), 641–658. Cited on page 18.

Tenkanen, Henrikki and Tuuli Toivonen, "Helsinki Region Travel Time Matrix," 2019. Cited on page 6.

A Supplementary figures

Figure A.1: Mid-day travel times

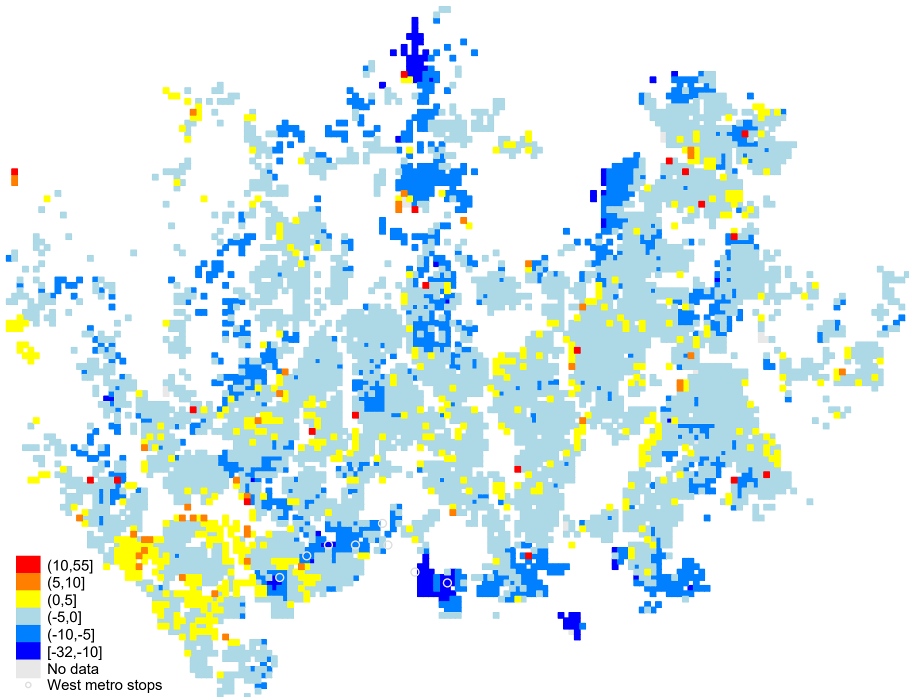


Figure A.2: Changes in transit travel times relative to driving (rush hours) between 2015 and 2018.

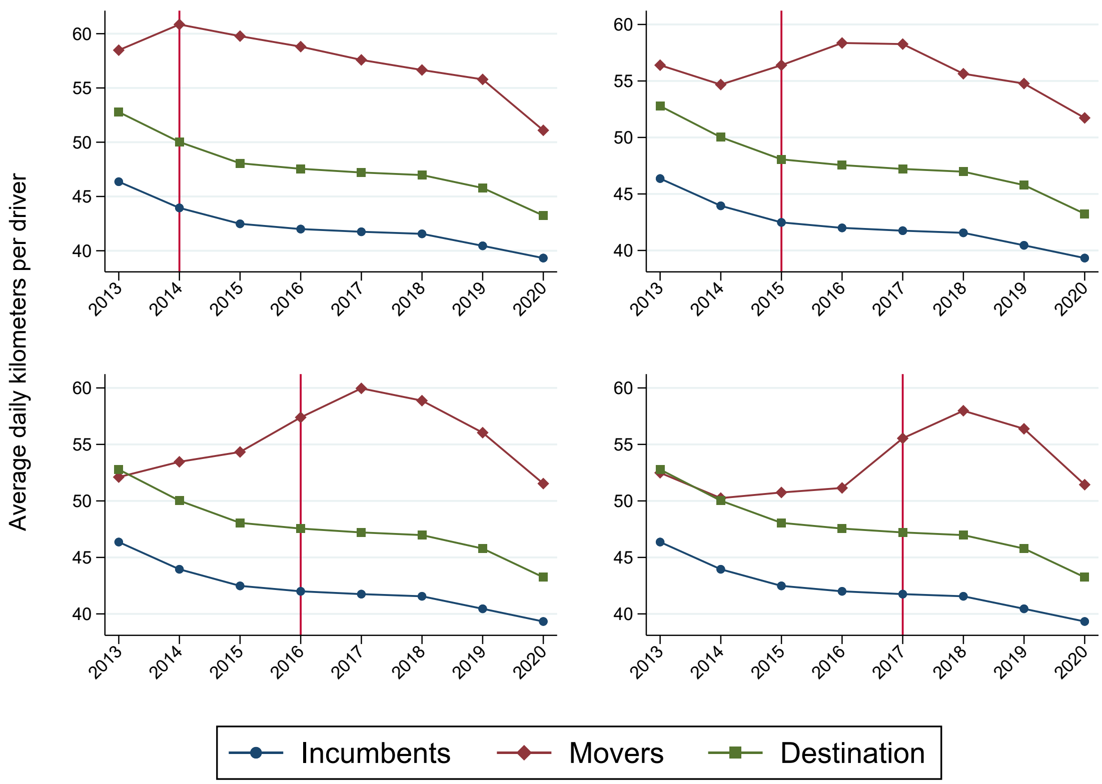


Figure A.3: **Average vehicle kilometers driven per day** among three groups of car owners: (i) "Incumbents": those who who reside in Helsinki the entire time (2013-2020), (ii) "Movers" those who move out from Helsinki in either 2014 (top-left), 2015 (top-right), 2016 (bottom-left) or 2017 (bottom-right), and (iii) "Destination": those residing in the municipalities where movers are moving to (average across municipalities weighted by fraction of movers). Red vertical line indicates the year of the move.

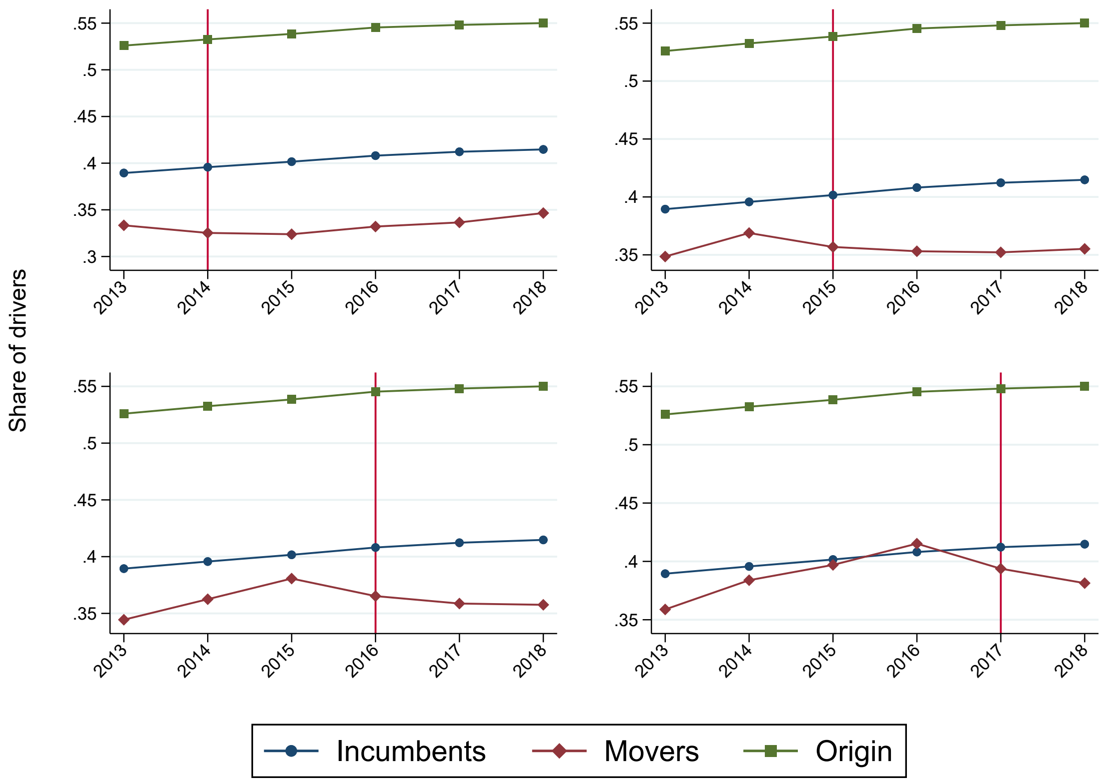


Figure A.4: **Car ownership rates** among three groups: (i) "Incumbents": those who who reside in Helsinki the entire time (2013-2020), (ii) "Movers" those who move in to Helsinki in either 2014 (top-left), 2015 (top-right), 2016 (bottom-left) or 2017 (bottom-right), and (iii) "Origin": those residing in the municipalities where movers are moving from (average across municipalities weighted by fraction of movers). Red vertical line indicates the year of the move.

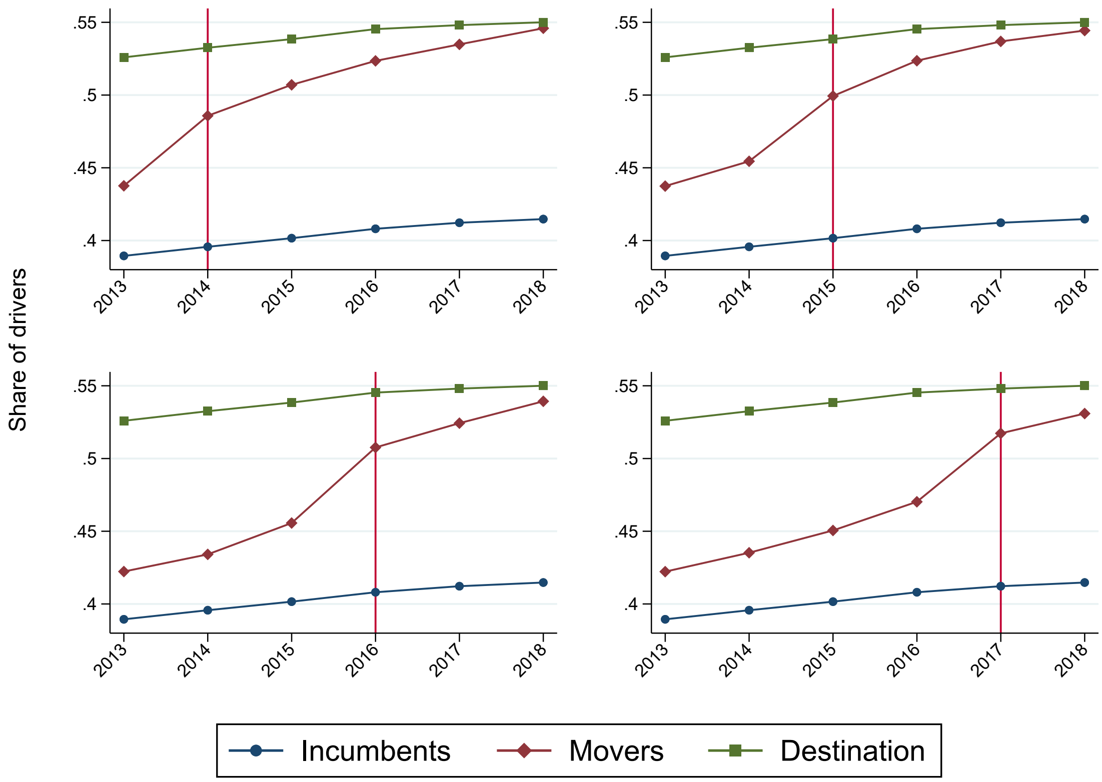


Figure A.5: **Car ownership rates** among three groups: (i) "Incumbents": those who who reside in Helsinki the entire time (2013-2020), (ii) "Movers" those who move out from Helsinki in either 2014 (top-left), 2015 (top-right), 2016 (bottom-left) or 2017 (bottom-right), and (iii) "Destination": those residing in the municipalities where movers are moving to (average across municipalities weighted by fraction of movers). Red vertical line indicates the year of the move.

B Supplementary Tables

C Data

C.1 Data used for estimating driving friendly index

Table A.1: Coefficients for Driving Friendly Index ($\mathcal{F}_n$)

Variable	Coefficient	Included in index
Land cover type 1, q =1	-0.0969811332	1
Land cover type 2, q =1	0.2369453908	1
Land cover type 2, q =2	0.2922627517	1
Land cover type 2, q =3	0.2381429595	1
Land cover type 2, q =4	0.2971155068	1
Land cover type 3, q =1	-0.0973670183	1
Land cover type 3, q =2	0	0
Land cover type 3, q =3	0.0233422567	1
Land cover type 3, q =4	0	0
Land cover type 4, q =1	-0.038047608	1
Land cover type 4, q =2	0.0699950828	1
Land cover type 5, q =1	0	0
Land cover type 5, q =2	0.0588867841	1
Land cover type 5, q =3	0	0
Land cover type 5, q =4	0	0
Land cover type 6, q =1	0	0
Land cover type 7, q =1	-0.1382652215	1
Land cover type 8, q =1	0	0
Land cover type 10, q =1	0	0
Land cover type 11, q =1	0	0
Land cover type 12, q =1	-0.0146583647	1
Land cover type 13, q =1	0	0
Land cover type 14, q =1	0	0
Land cover type 15, q =1	0	0

Land cover type 16, q =1	0	0
Land cover type 17, q =1	0.0393745787	1
Land cover type 18, q =1	0	0
Land cover type 18, q =2	0	0
Land cover type 18, q =3	0	0
Land cover type 18, q =4	0	0
Land cover type 19, q =1	0	0
Land cover type 20, q =1	0	0
Number of transit stops, q =1	0	0
Number of transit stops, q =2	0	0
Number of transit stops, q =3	0.0592847283	1
Number of transit stops, q =4	0.0121690125	1
Total length of roads for populated cells, q =1	0	0
Total length of roads for populated cells, q =2	0	0
Total length of roads for populated cells, q =3	0	0
Total length of roads for populated cells, q =4	0	0
Distance to nearest center area, q =1	-0.0819930176	1
Distance to nearest center area, q =2	-0.100481555	1
Distance to nearest center area, q =3	0	0
Distance to nearest center area, q =4	0	0
Distance to nearest trade area, q =1	0	0
Distance to nearest trade area, q =2	-0.0475672727	1
Distance to nearest trade area, q =3	-0.0440646752	1
Distance to nearest trade area, q =4	-0.3288916483	1
Number of pedestrian intersections, q =1	0	0
Number of pedestrian intersections, q =2	0	0
Number of pedestrian intersections, q =3	0	0
Number of pedestrian intersections, q =4	-0.0601422157	1
Population, q =1	-0.2886014388	1
Population, q =2	0	0
Population, q =3	0	0
Population, q =4	-0.0858845409	1
Length of walking and cycling roads, q =1	0	0
Length of walking and cycling roads, q =2	-0.0377694209	1

Length of walking and cycling roads, q =3	0.0293524061	1
Length of roads with speed limit over 40 km/h, q =1	0	0
Length of roads with speed limit over 40 km/h, q =2	0.0132460274	1
Traffic on major roads, q =1	-0.1077223248	1
Traffic on major roads, q =2	0.017322719	1
Traffic on major roads, q =3	0	0
Traffic on major roads, q =4	-0.0855777684	1
Distance to major roads, q =1	0	0
Distance to major roads, q =2	0	0
Distance to major roads, q =3	-0.0785392531	1
Distance to major roads, q =4	-0.0611667527	1
Distance to nearest 50 intersections, q =1	0.0308420577	1
Distance to nearest 50 intersections, q =2	0	0
Distance to nearest 50 intersections, q =3	0	0
Distance to nearest 50 intersections, q =4	0	0
Distance to nearest 10 intersections, q =1	0	0
Distance to nearest 10 intersections, q =2	0	0
Distance to nearest 10 intersections, q =3	0	0
Distance to nearest 10 intersections, q =4	1.1096848892	1
Distance to nearest 50 traffic stops, q =1	0	0
Distance to nearest 50 traffic stops, q =2	0	0
Distance to nearest 50 traffic stops, q =3	-0.2018945128	1
Distance to nearest 50 traffic stops, q =4	0	0
Distance to nearest 10 traffic stops, q =1	0	0
Distance to nearest 10 traffic stops, q =2	0	0
Distance to nearest 10 traffic stops, q =3	0	0
Distance to nearest 10 traffic stops, q =4	0	0