Drivers of habit: Evidence of persistence from car inspections and neighborhood moves*

PRELIMINARY, PLEASE DO NOT CIRCULATE

Click here for latest version

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

*Department of Economics, Aalto University / Helsinki GSE

December 12, 2024

Abstract

How do our past residential neighborhoods shape our current habits? We study this in the context of driving in Helsinki. We exploit granular data on the universe of cars and individual residential locations in Finland to show that variation in car ownership and vehicle kilometers driven are explained to some extent by where drivers currently reside and to a much greater extent by where they used to reside at the age of 18. Growing up in a neighborhood with high car ownership makes drivers likely to drive more today. These results are suggestive of the importance of neighborhood exposure during early formative years in helping form habits that persist across residential locations and across decades.

*Corresponding author: pablo.warnes@aalto.fi

We thank seminar participants at Universitat Autonoma Barcelona, VATT Institute of Economic Research, University of Turku, and the Finnish Economic Association meeting for helpful comments and feedback. All errors are our own.

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? Do residents drive less in areas with worse driving infrastructure or better travel mode alternatives?

Differences across locations in how much people drive reflect both features of the neighborhoods and tastes of the households that reside there. In this paper, we decompose the role of neighborhoods versus individual attributes in explaining variation in individual driving behavior over time and across locations. This decomposition is key to understanding the effectiveness of place-based investments in shaping individual driving habits. For instance, if people who reside in car-friendly neighborhoods (e.g. with wider streets and poor public transit options) would drive the same amount in a less car-friendly neighborhood, then targeted investments in these neighborhoods are unlikely to change driving behavior. But the opposite is true if the amount people drive largely reflects the constraints they face in their neighborhood, such as poor public transit access to amenities of interest or relative ease of parking.

More specifically, we study driving behavior in Helsinki between 2013 and 2021, a period of widespread public transit service expansions. We use our novel granular data to document substantial persistence in individuals' driving decisions over time and across residential neighborhoods. To rationalize these observations, we propose a model of travel mode choices and kilometers driven as a function of features of the residential neighborhood as well as of individual driving habits that persist across locations. Taking the model to data, we find that most of the variation in an individual's kilometers driven (and car ownership) is in fact attributable to individual habits that persist across neighborhoods rather than to individuals changing residential neighborhoods. When we decompose these persistent individual habits further, we find that past residential locations explain a large share of the differences in driving habits. Individuals who grew up in neighborhoods with high car ownership rates tend to drive more today long after moving from the original neighborhoods. We find the power of past neighborhoods in explaining current habits to be strongest when we focus on an individual's residential neighborhood during their teenage years. Our most striking finding is that a driver's location at age 18 explains six times more variation in kilometers driven than where they reside currently. These results are suggestive of the importance of neighborhood exposure during early formative years in helping form habits that persist across residential locations and across decades.

To derive these insights, we leverage detailed administrative data for the universe of residents of Helsinki together with odometer readings from mandatory vehicle inspections for all car owners in the city. This allows us to follow the kilometers driven, car ownership and residential location choices of individual residents every year over a long period of time and associate childhood exposure to later-life outcomes in a clean way. The data also allows us to derive several novel motivating empirical regularities. First, we show that movers who move to neighborhoods with lower (higher) average driving than their origin neighborhood tend to decrease (increase) their driving from the year before the move to the year after the move, suggesting that neighborhood characteristics do play a role in shaping driving behavior. However, we also find that those moving from high-driving to low-driving neighborhoods continue to drive substantially more on average than the destination neighborhood incumbents several years after the move, while those that move from low-driving to high-driving neighborhoods converge to the destination neighborhood habits almost immediately after the move. Exposure to high-driving neighborhoods thus seems to shape driving habits even after an individual has moved to a location with lower driving. Indeed, we document a positive relationship between car ownership in the neighborhood one lived in at age 18 and current kilometers driven, even conditional on residing in a different neighborhood today.

Our theoretical framework and its empirical application helps us tease apart the role of current residential locations from persistent individual habits in explaining how much car owners drive. In particular, we follow a two-stage approach, where in the first stage we estimate a high-dimensional fixed effects model (in the spirit of Abowd et al., 1999), and 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 1% of the variation. We then regress the driver fixed effects on fixed effects for the neighborhood the driver lived in when they were

17 (restricting our attention to drivers who no longer live in the same neighborhood, and come from origins shared by at least 10 drivers in our sample), finding them to explain more than 12% of the variation in the driver fixed effects, which translates to more than 6% of the total variation in driving – many times more than the variation explained by the current neighborhood.

Our contributions to the academic literature are three-fold. First, we contribute to the large and growing literature that studies how travel behavior (such as the modes and amount of travel) is affected by location features of where travelers currently reside (and/or work). This literature has studied the effects of changes in transport infrastructure and pricing (e.g., Allen and Arkolakis, 2022; Severen, 2023; Tsivanidis, 2022) as well as changes in nearby economic activity such as jobs and shopping centers due to other shocks (e.g., Almagro et al., 2024; Gorback, 2020). We shed light on an important under-explored channel through which residential locations shape travel behavior: they build driving habits that persist long after the travellers have relocated to locations with very different features and travel activity.

Second, our result that the variation in driving behavior is explained more by features of drivers' former childhood neighborhoods than of their current one is consistent with the literature on long-term neighborhood effects. Like much of this literature, our results imply that while place-based policies may appear to have minimal impact on adults' contemporary choices and outcomes, there are large long-term gains from targeting these investments at children in their most formative years (e.g., Chetty et al., 2014, 2016; Chetty and Hendren, 2018).

Third, we contribute to the literature on the role of urban form and the age-old debate on targeting investments at people versus places (e.g., Kline and Moretti, 2014; Gaubert, 2018; Duranton and Venables, 2021; Lyubich, 2022). Given people's ability to sort into locations that match their habits (and continue relocating) in the long run, how effective are place-based investments in shaping their habits and outcomes in the long term? Our findings suggest that they could be very effective - significantly more than investments targeting current residential locations and current outcomes.

Our findings also have important implications for the effectiveness of various urban planning policies at moderating driving behavior. Most traditional policy interventions, 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? Of course, the Helsinki region is relatively uncongested even in the densest areas and well-served by public transit even in more remote areas, meaning that the cost-differential between driving and alternative modes may be smaller in most neighborhoods there than in many other places around the world. Nevertheless, policies that encourage driving can have long-standing effects through habit formation, and even inheritance of habits between generations.

The remainder of the paper is structured as follows. The next section discusses our data on driving behavior and individual characteristics 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 decomposes the variation in driving to the contributions of neighborhoods versus individual habits, and 6 discusses the role of origin neighborhoods in shaping the habits. Finally, section 7 concludes.

2 Data and Geographic Scope

We combine annual observations on individuals' residential locations at the level of a 250 meters by 250 meters grid with data on car ownership and kilometers driven from the vehicle registry of the Finnish Transport and Communications Agency, *Traficom*. The residential locations are obtained from the administrative database of Statistics Finland alongside individual demographic characteristics, such as disposable income, age, and household size. The Statistics Finland data covers the years 1990-2021, and includes a pseudonymized personal identifier, which allows us to link it to the vehicle registry data.

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.[1] 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 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.

[1]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.

Note that the ownership and driving spells do not correspond to calendar years, while the rest of our data is at that level. We classify individuals as car owners in a given year if the total length of the intersection between their ownership spells and that year is more than half of the year's length. Meanwhile, to map the driving spells to kilometers driven by an individual within a calendar year, we compute the length of the year-ownership intersection for a given car, multiply it by the mean daily kilometers driven by that car within driving spells that overlap the intersection, weighted by the length of the overlap, and sum over all cars linked to an individual.[2]

In addition to the detailed information included in the vehicle registry, the Statistics Finland data contains a dummy for individual car ownership for the years 1990-2018. We use this variable to construct car ownership rates in origin neighborhoods, as we observe a much longer history for it than for the ownership information in the vehicle registry.

2.2 Geographic scope

Our analysis focuses on the Helsinki capital region – Finland's 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 region constitutes a single commuting area around central Helsinki, with several smaller centers connected to the central business district by radial roads and railways.

We observe the residential location of each (adult) inhabitant in the region at the level of a 250 meters by 250 meters grid cell. In our data, the region contains 6 505 populated grid cells, which vary substantially in population density from a few to a few thousand people per cell. The interquartile range extends from 16 to 178 adults, with a median of 77, and mean around 147.

This variation in population density naturally comes with significant heterogene-

[2]In some cases, the intersections of ownership spells and years are not fully, or even at all, covered by the driving spells – either because the car has not yet been inspected for the first time, or because we have dropped a driving spell due to clearly erroneous kilometers driven (namely negative or more than 1000 per day), or because we were not able to allocate it to a single individual. If an intersection is partially covered, we simply extrapolate the daily kilometers from the covered part to the rest of the intersection. We assume zero kilometers for intersections that do not overlap with any driving spells, and drop individual-year-level observations for whom we observe no kilometers for any of their cars. The use-spell-year intersections are at least partially covered for all cars for 85% of the observations, with only 7% totally missing, while 69% of the intersections are fully covered by the driving spells, with the average coverage being 86% of the intersection.

ity in the level of public transit service, car infrastructure, and other neighborhood amenities that one would expect to have an impact on driving. Indeed, figure 1 shows that kilometers driven are on average clearly lower in the municipality of Helsinki – which contains the densest urban areas in the region – than in the more suburban municipalities of Espoo, Vantaa, and Kauniainen. Figure 2 shows a similar story for car ownership, with the highest rate in the small municipality of Kauniainen, which is relatively high-income and located in the middle of suburban Espoo. Kaunianen having substantially higher car ownership than Espoo and Vantaa but smaller average kilometers driven suggests that driving decisions at the intensive and extensive margins are to some extent made for different reasons. We find that the former in particular is explained more by past neighborhoods than current ones.[3]

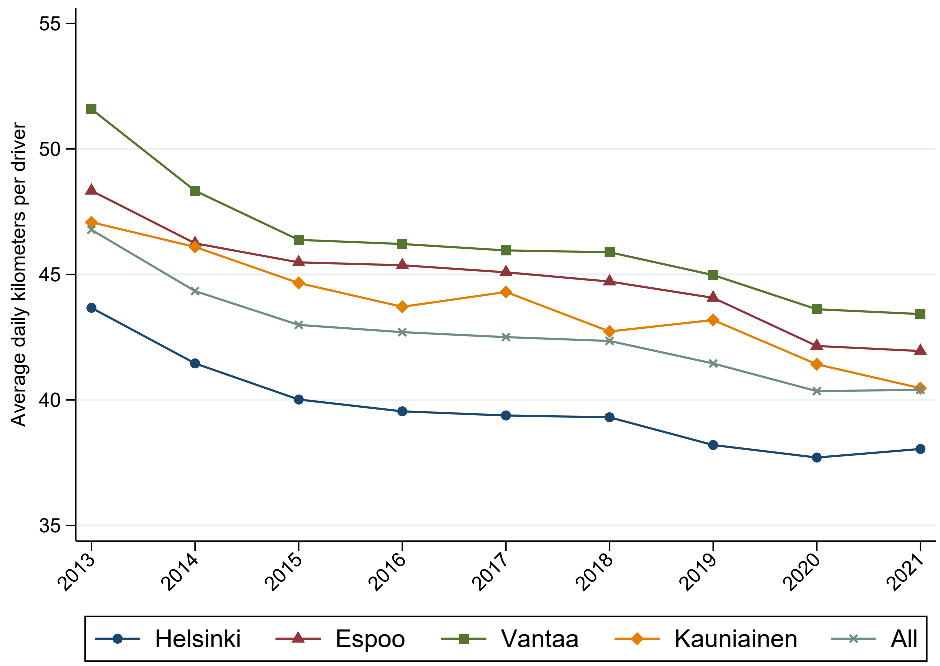

Figure 1: **Average daily kilometers driven (y-axis) per year (x-axis) by drivers in the different municipalities of the Helsinki metropolitan area, and the region as a whole ("All").**

[3]Figures 1 and 2 suggest that car ownership has remained quite stable throughout the study period, while average kilometers driven have somewhat decreased. However, the decrease in kilometers driven especially after 2018 is partially due to us not observing the first inspection for new cars, which tend to be driven more than older cars.

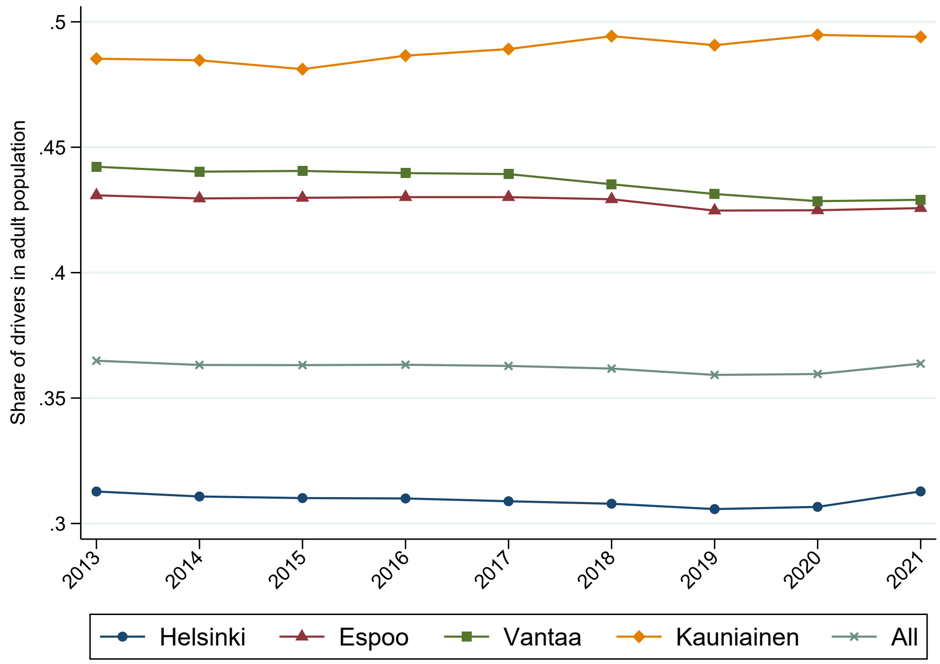

Figure 2: **Car ownership rate of adults (y-axis) per year (x-axis) in the different municipalities of the Helsinki metropolitan area, and the region as a whole ("All").**

3 Empirical Regularities

We exploit our granular data on individual driving behavior and residential location changes to document three key empirical regularities about driving habit persistence.

3.1 Driving behavior changes with residential location

Features of residential locations may be important determinants of how much people drive and own cars. We might drive more in some neighborhoods than others because these places are further away from destinations of interest, because they have more driving-friendly infrastructure (such as easy access to parking, highways, etc.) or worse alternative travel modes, or simply because we see our neighbors driving more. In this section, we leverage our granular data on residential location changes to document how driving behavior changes across neighborhoods.

Figure 3A follows car owners who moved residential neighborhoods between 2014 and 2020 to compare how much they drove in the years right before and right after the move. The vertical axis depicts changes in average (log) daily kilometers driven. The horizontal axis distinguishes the difference in the movers' origin and destination neighborhood driving norms, where norms refer to the average daily kilometers driven by car owners in the neighborhood in 2013. More specifically, points on the left (of the red vertical line) of the graph correspond to moves to neighborhoods where residents drive less (on average) than in the neighborhood of origin, and points on the right correspond to moves to neighborhoods where residents drive more.

The figure highlights a clear pattern in movers' driving behavior. First, note that when movers move to a neighborhood with the same driving norm as their previous neighborhood (the point on the vertical red line at 0), they reduce their kilometers driven slightly, which is consistent with the nationwide downward trend over time in how much people drive. But net of the nationwide downward trend, drivers who move to neighborhoods where people drive more tend to increase their own driving in the year right after the move. For instance, after moving to a neighborhood where residents drive 10km more (than residents in the origin neighborhood), the movers themselves start driving roughly 8% more. In contrast, drivers who move to neighborhoods where people drive less tend to decrease their driving after the move. For instance, after moving to a neighborhood where residents drive 10km less (than residents in the origin neighborhood), the movers themselves start driving roughly 14% less. The larger the difference in driving norms between the mover's origin and destination (i.e. the farther away the move is from the red vertical line), the larger is the change in the mover's own driving behavior.

We see a similar pattern of changes in the mover's likelihood of owning a car. In Figure 3B, the vertical axis measures the change in the rate of car ownership among movers between the years right before and right after their move. The horizontal axis continues to measure the difference in average driving norms between the origin and destination of the move. Note that car ownership is increasing among all movers. But net of this upward trend in car ownership, movers to neighborhoods with higher driving norms are more likely to become car owners than movers to neighborhoods with lower driving norms.[4] For instance, after moving to a neighborhood where residents drive 10km more (than residents in the origin neighborhood) movers are 3 percentage

[4]Note, however, that this pattern flattens out left of the red line: movers to neighborhoods where residents drive 10km less (than residents in the origin neighborhood) are no less likely to become car owners than movers to neighborhoods where residents drive 5km less. We suspect this is because people are much slower at giving up their car than buying a new car.

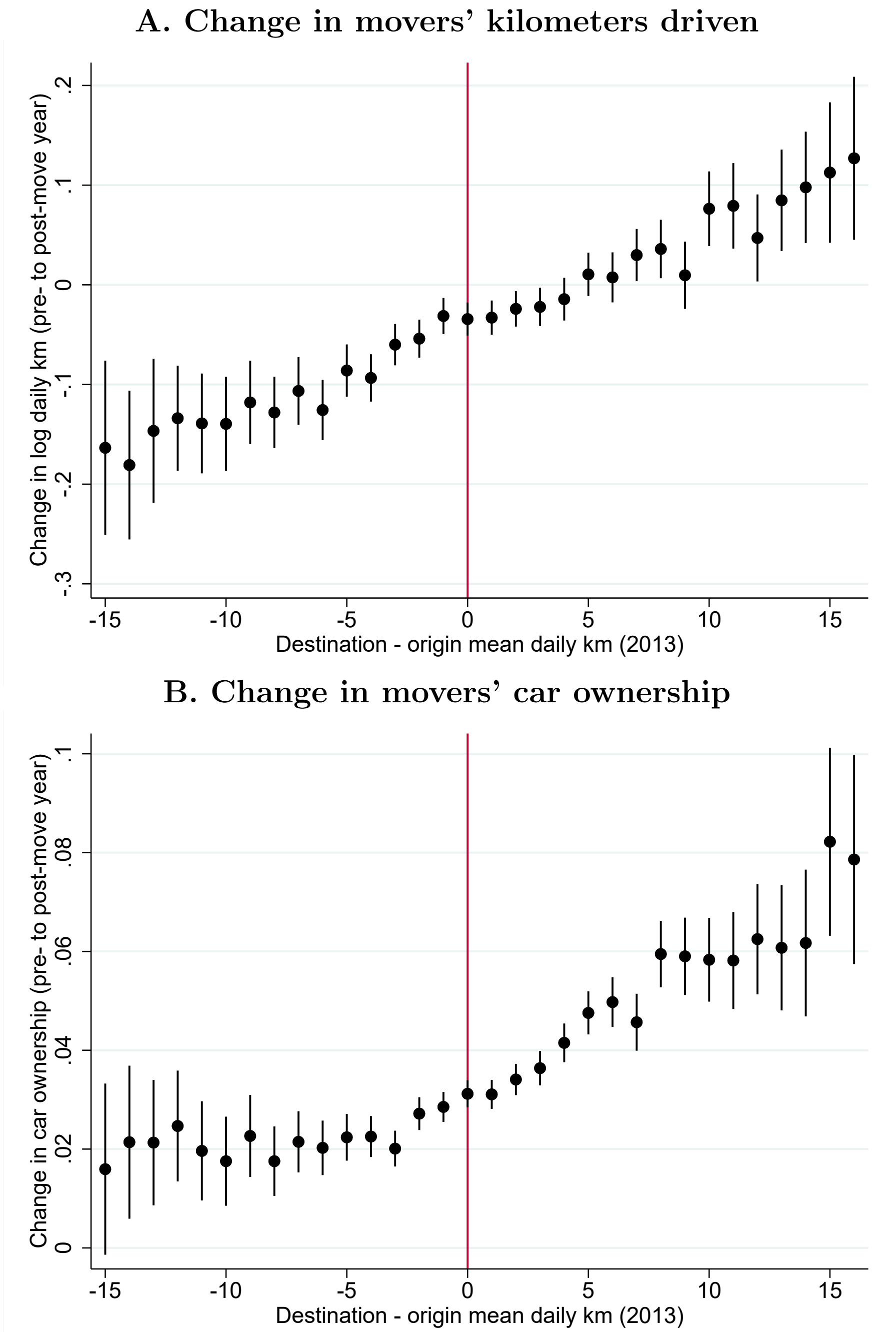

Figure 3: **Change in kilometers driven and car ownership after moving residential neighborhoods (y-axis) by difference in mean kilometers driven among residents of the origin and destination neighborhoods (x-axis).** The y-axis measures the change in either average (log) kilometers driven (A) or car ownership (B) of movers between the year before and the year after their move. Sample in (B) includes all neighborhood moves between 2014 and 2020. Sample in (A) restrict to car owners. The x-axis measures differences (between the movers' origins and destinations) in the average daily kilometers driven by car owners in the neighborhoods in 2013. Each point at $x = X$ corresponds to a bin $[X - 0.5, X + 0.5)$ of width 1. Points to the left of the red vertical line correspond to moves to neighborhoods where residents drive less. Points to the right of the red line correspond to moves to neighborhoods where residents drive more.

points more likely to become car owners than movers between neighborhoods with similar driving norms (i.e. movers at the red line).

In short, we document that driving behavior changes with residential locations, and movers quickly update their driving habits in the direction of their new neighborhood's driving norms. Next, we show that despite updating how much they drive, movers may also hold on to some old habits and continue to differ in their driving behavior from incumbent residents of their new neighborhood.

3.2 Driving behavior persists across residential locations

The graphs in Figure 4 follow the average kilometers driven by neighborhood movers in the years leading up to, of, and following their move. Our sample includes all moves between 2014 and 2020, with the year of the move depicted by the vertical red line at 0. For comparison, we also include lines showing the average kilometers driven by incumbent residents (i.e. those who do not move between 2013 and 2021) of the mover's origin and destination neighborhoods. For these incumbent residents, the averages are weighted by the fraction of moves from/to the neighborhood. We distinguish two types of neighborhoods and two types of movers. We define a "high-driving neighborhood" as one whose average resident drives more than the average resident of the median neighborhood, and a "low-driving neighborhood" as one whose average resident drives less. The graph on the left of Figure 4 follows movers who moved from a high-driving neighborhood to a low-driving one, while the graph on the right follows movers who moved from a low-driving neighborhood to a high-driving one. In both cases, we plot the average kilometers driven by the movers, the origin residents, and the destination residents.

First, consider movers from low- to high-driving neighborhoods on the right of Figure 4. Already several years before the move, movers' driving habits start to differ notably from that of the average non-mover resident in the origin neighborhood. The closer they are to the year of move, the more their daily kilometers driven differs from the origin neighborhood's. After the move, their daily kilometers driven continue to increase but also flatten out to roughly resemble that of the average non-mover resident in the destination neighborhood.

Now, consider movers from high- to low-driving neighborhoods on the left of Figure 4. We see a similar pattern in the years before the move: those who are about to

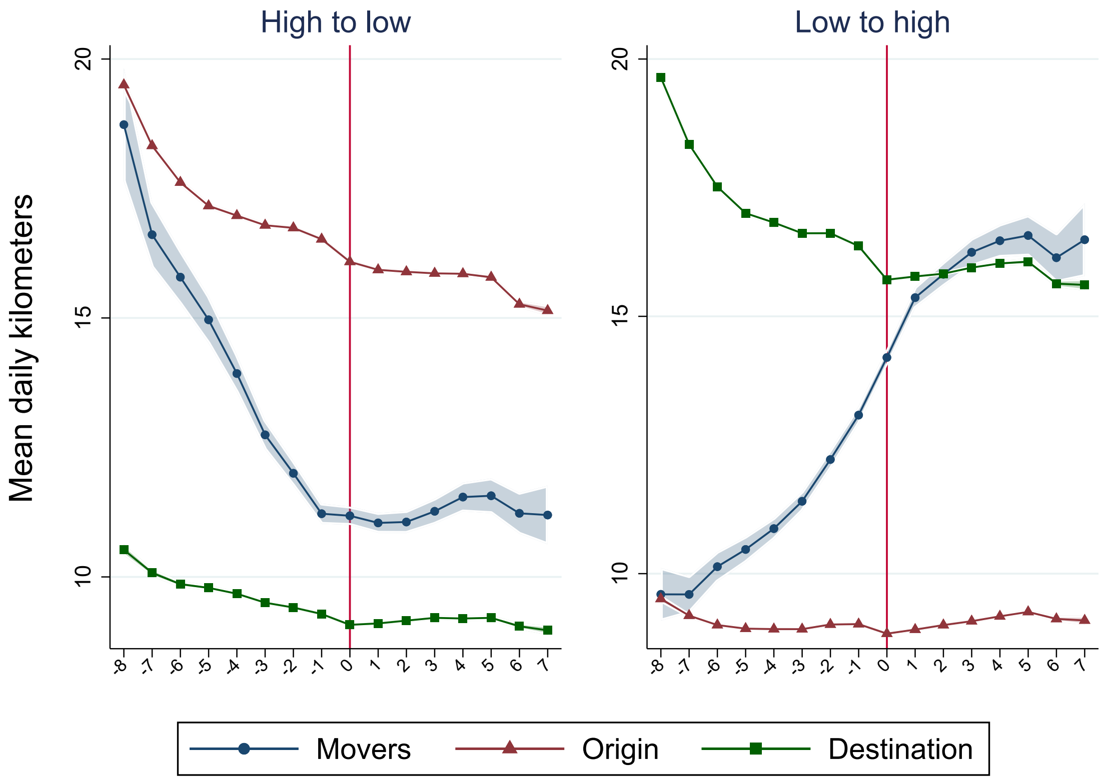

Figure 4: **Average daily kilometers driven by residential neighborhood movers and incumbent residents, for two types of moves: from a neighborhood with high driving mean to one with a low driving mean (left) or from a neighborhood with low driving mean to one with a high driving mean (right).** Lines on each graph depict three groups of car owners: (i) "Movers" between residential neighborhoods, (ii) incumbent residents of "Origin" neighborhoods where the movers move from, and (iii) incumbent residents of "Destination" neighborhoods where the movers move to. We restrict movers to those who move between years 2014 and 2020, with the horizontal axis depicting the years before, of, and after the move. We restrict origin and destination residents to those who did not move between 2013 and 2021. Means of kilometers driven are weighted by the fraction of moves from/to the neighborhood.

move drive less than the average non-mover in their origin neighborhood. The closer they are to the year of the move, their daily kilometers driven looks more like the residents of the destination neighborhoods and less like the residents of the origins. However, the kilometers driven by these movers never end up converging to those of the destination residents after the move, unlike movers from low- to high-driving neighborhoods. Movers from high- to low-driving neighborhoods continue to drive more than incumbent residents of their new neighborhood even seven years after moving. This persistence of high kilometers driven in low-driving neighborhoods suggests that past exposure to high-driving neighborhoods might continue to shape current and long-term driving habits.

Consistent with this idea, we show next that current individual driving habits correlate strongly with car usage in the neighborhood where the individual resided at an early age but have since moved elsewhere.

3.3 Driving habits today correlate with past exposure to car ownership

Figure 5 leverages our data on individuals' residential locations and car ownership since 1990 to show how current average kilometers driven (2013-2021) relates to car ownership in the driver's former residential neighborhood when they were 17 years old. In particular, the horizontal axis measures the car ownership rate in the neighborhood but presented as a percentile of neighborhood car ownership rates in that year.[5] The vertical axis depicts average (log) daily kilometers driven by the individual between 2013 and 2021. We focus on individuals who no longer reside in the neighborhood where they resided at age 17 so that we are not picking up the effect of current neighborhood features on driving behavior.

We see right away that individuals who resided in neighborhoods with high car ownership at age 17 drive more today than those who grew up in neighborhoods with low car ownership. This is consistent with what we showed earlier: even after changing residential locations, some old driving behavior continue to persist as long-term habits.

In summary, we showed in this section that (i) changes in driving behavior are strongly correlated with changes in residential neighborhoods, (ii) old driving behavior can persist across neighborhood changes, and (iii) exposure to car ownership in former residential neighborhoods continue to correlate with driving habits today. We also noted that (future) movers differ systematically from non-movers in their driving habits, suggesting that residential relocation decisions are not random but highly correlated with characteristics of individual drivers. In the following sections, we present a theoretical framework to formulate driving habits as a function of features of both residential neighborhoods and of driver characteristics, and map it to an

[5] By letting car ownership percentiles be across neighborhoods in the same year (rather than across years), we essentially control for differences across age cohorts in their past exposure to car ownership.

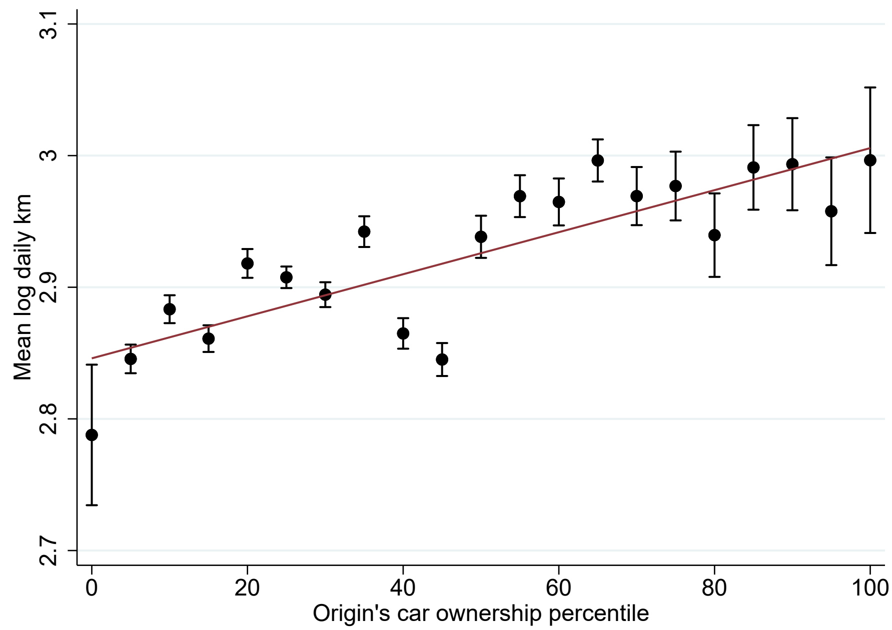

Figure 5: **Average (log) km driven in 2013-2021 (y-axis) by car prominence in driver's residence of origin at age 17 (x-axis).** An origin neighborhood's car prominence is defined as its car ownership rate as a percentile across all neighborhoods in the given year. The percentiles are rounded to the nearest 5 percentile point. The red line depicts the linear best fit. We restrict to car owners who reside in a different neighborhood in 2013-2021 than the one they resided in at the age of 17.

empirical strategy to decompose the extent to which these habits are shaped by the neighborhoods individuals have resided in.

4 A Model of Driving Choice

Consider a city with a fixed population of travelers. Each traveler faces a measure 1 of different trips (such as shopping, commuting, etc.), indexed by q. Conditional on residing in neighborhood n in year t, traveler 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 $m \in \{\textsc{car}, \textsc{alt}\}$ (either driving or mass transit).

4.1 Travel mode choices

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

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

where $\bar{c}^{m}_{njt}$ incorporates costs of using mode m on trips from n to j that are invariant across all travelers (such as trip distances and gas prices) and $\bar{\nu}_{it}$ denotes subjective preferences of individual travelers 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 traveler i chooses to drive is:

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

Note that the first multiplicative component is a traveler-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 travelers. Next, we elaborate further on both the objective and the subjective components of the probability of driving.

We decompose the objective costs of choosing travel mode m as follows:

$$\bar{c}^{m}_{njt} \equiv \tau^{m}_{njt} + \gamma^{m}_{jt} + \kappa^{m}_{nt} \tag{3}$$

where τ^{m}_{njt} captures travel time costs (as well as other correlated costs such as fuel consumption) between n and j 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. 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).

Next, we decompose individual preferences for driving $\bar{\nu}_{it}$. While these preferences are independent of features of the neighborhoods where the driver resides or travels to, we allow them to be shaped by features of past residential neighborhoods. To do so, we formulate $\bar{\nu}_{it}$ as follows:

$$\bar{\nu}_{it} \equiv \nu_{it} + \sum_{t'<t} \zeta_{o(i,t')} \tag{4}$$

where $\zeta_{o(i,t')}$ captures any effect on preferences for (or against) driving that individual i has today due to having previously resided in neighborhood $o(i, t')$ in year t'. We assume this expression is zero for years t' before the driver was born. The expression ν_{it} captures other time-varying subjective preferences for driving.

We will return to this decomposition of subjective preferences for driving shortly. But, we can plug in the decomposition of the objective preferences (Equation 3) into Equation 2 to rewrite the probability of choosing to drive as:

$$\bar{\Phi}_{injt} = \exp\left(\bar{\nu}_{it}\right) \cdot \exp\left(\Delta\kappa_{nt}\right) \cdot \exp\left(\Delta\gamma_{jt}\right) \cdot \exp\left(\Delta\tau_{njt}\right) \tag{5}$$

where $\Delta\kappa_{nt} \equiv \kappa_{nt}^{\text{ALT}} - \kappa_{nt}^{\text{CAR}}$, $\Delta\gamma_{jt} \equiv \gamma_{jt}^{\text{ALT}} - \gamma_{jt}^{\text{CAR}}$, and $\Delta\tau_{njt} \equiv \tau_{njt}^{\text{ALT}} - \tau_{njt}^{\text{CAR}}$. Next, we derive an expression for how much individuals drive.

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 traveled by residents of neighborhood n is a weighted sum of the driving distances l_{nj} to all destinations:

$$L_n = \sum_j \psi_j \cdot l_{nj} \tag{6}$$

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

$$d^{\star}_{int} \equiv d_{int}(m^{\star}) = \sum_{j} \bar{\Phi}_{injt} \cdot \psi_j \cdot l_{nj} \tag{7}$$

In each year t, traveler 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:

$$d^{\star}_{int} = \exp\left(\bar{\nu}_{it}\right) \cdot \exp\left(\Delta\kappa_{nt}\right) \cdot \delta_{nt} \cdot W_n \tag{8}$$

where

$$\delta_{nt} \equiv \sum_{j} \left(\frac{\omega_{nj}}{W_n} \cdot \exp\left(l_{nj} \cdot \Delta\tau_{njt}\right) \right) \tag{9}$$

$$\omega_{nj} \equiv \exp\left(\Delta\gamma_j\right) \cdot \psi_{jt} \cdot l_{nj} \text{ and } W_n \equiv \sum_{j} \omega_{nj}$$

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:

$$\ln(d^{\star}_{int}) = \ln(\delta_{nt}) + \Delta\kappa_{nt} + \ln(W_n) + \bar{\nu}_{it} \tag{10}$$

4.3 Empirical interpretation

We want to use our formulation of kilometers driven to empirically isolate the roles of current and past neighborhoods in explaining driving behavior using observational data on kilometers driven by the same individuals between residential neighborhood changes. To ease empirical interpretation, we can decompose the neighborhood-specific parameters as the sum of a non-time-varying mean ρ_n, a common temporal shock $\xi^{\Delta\kappa}_t$ and a neighborhood-specific temporal shock $\epsilon^{\Delta\kappa}_{nt}$ with a mean of zero:

$$\Delta\kappa_{nt} + \ln(\delta_{nt}) + \ln(W_n) \equiv \rho_n + \xi^{\Delta\kappa}_t + \epsilon^{\Delta\kappa}_{nt}$$

We can similarly decompose the parameter ν_{it} governing individual preferences:

$$\nu_{it} \equiv \nu_i^0 + \xi_t^\nu + \alpha X_{it} + \epsilon_{it}^\nu$$

where ν_i^0 is a time-invariant individual-specific driving mean, ξ_t^ν measure evolving driving norms that are common across all individuals, X_{it} denote observable time-varying individual characteristics, and ϵ_{it}^ν are unobservable idiosyncratic temporal shocks to preferences with a mean of zero.

These decompositions, together with Equation 4, allow us to re-write the distance driven by individual i in year t (Equation 10) as:

$$\ln(d_{int}^\star) = \rho_n + \sum_{t' \in [1990,2013)} \zeta_{o(i,t')} + \nu_i^0 + \alpha X_{it} + \rho_t + \varepsilon_{int} \tag{11}$$

where $\rho_t \equiv \xi_t^\nu + \xi_t^{\Delta\kappa}$ and $\varepsilon_{int} \equiv \epsilon_{nt}^\kappa + \epsilon_{it}^\nu$. Note that ρ_t denote year fixed effects that capture both any common temporal shocks across neighborhoods ($\xi_t^{\Delta\kappa}$), as well as any evolving trends in driving behavior that are common across individuals (ξ_t^ν). Note also that we restricted the effect of past residential neighborhoods $\zeta_{o(i,t')}$ to residences in years $[1990, 2013)$. This restriction reflects two important empirical constraints.

First, in our data, we do not observe kilometers driven by individuals ($d_{int}^\star$) before 2013. But because we observe their residential neighborhoods since 1990, we can still consider $\zeta_{o(i,t')}$ on the right-hand side for $t' \geq 1990$ to identify the role of residential neighborhoods from before 2013 in explaining driving behavior after 2013. Because we cannot identify the effect of residential neighborhoods from before 1990, these effects are likely to be reflected in ν_i^0 for individuals born before 1990.

Second, if past residential neighborhoods did not matter ($\zeta_{o(i,t')} = 0$), we could identify parameters ρ_n and ν_i^0 in our data as neighborhood and individual fixed effects (conditional on year effects ρ_t and time-varying characteristics X_{it}) of kilometers driven between 2013 and 2021. However, assuming a neighborhood n can affect both current and future driving behavior, we cannot distinguish its current effect on driving, ρ_n, from its past effect on driving, ζ_n, when both are informed by mean kilometers driven over the same set of years. So, we do not estimate past neighborhood effects $\zeta_{o(i,t')}$ for $t' \geq 2013$ to avoid temporal overlap with current neighborhood effects ρ_n when $n = o(i, t')$.

Since $\zeta_{o(i,t')}$ do not vary over time $t \in [2013 - 2021]$, they will be subsumed by any individual fixed effects that we estimate. Naturally, we interpret these fixed effects as the individuals' driving "habits" between years 2013 and 2021. More specifically, we let

$$\rho_i = \sum_{t' \in [1990,2013)} \zeta_{o(i,t')} + \nu_i^0 \tag{12}$$

reflect both the driving habits explained by past residential locations as well as any time- and location-invariant habits that persist across residential moves. To back out the effect of past residential locations on individual driving habits, we will need to further decompose our estimated individual fixed effects.

5 The role of current residential neighborhoods

How much of the variation in driving that we observe is explained by the neighborhood individuals currently reside in versus persistent individual behavior across neighborhoods? To answer this question, we estimate a high-dimensional fixed effects model and perform an variance decomposition exercise in the stile of (Abowd et al., 1999). 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 average daily kilometers driven for each driver.

We start by observing that equation 11 can be re-written as

$$\ln(d^{\star}_{int}) = \rho_n + \rho_i + \rho_t + \alpha X_{nit} + \varepsilon_{int}, \tag{13}$$

where ρ_n is a fixed effect for a driver's current neighborhood of residence; ρ_i is an individual-level fixed effect that does not vary through time or neighborhood of residence, following the theoretical framework we have developed, we will interpret this term as capturing the component of individual preferences for driving that are not time varying (within the period analyzed), which will include any neighborhood-level effects for formative years; ρ_t are year fixed effects that capture temporal shocks across all neighborhoods and individuals (e.g., a common declining trend in average kilometers driven); finally, X_{it} is a matrix of observed variables at the level of the individual or the neighborhood that are time-varying, and ε_{int} is the residual variation in driving behavior across time, neighborhoods and individuals.

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,[6] 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 estimate equation 13 on the leave-one-out connected set of neighborhoods[7] for the years 2013 to 2021. We then estimate the unbiased estimator of the covariance matrix proposed by Kline et al. (2020).[8]

Table 1: Decomposing the variation in daily kilometers driven

	Controls		No Controls	
	Value	Share	Value	Share
Total variance (log daily km driven)	0.89	1.000	0.92	1.000
Var. of driver effects	0.45	0.51	0.47	0.51
Var. of neighborhood effects	0.007	.008	0.006	0.007
Covariance neighborhood-driver effects	0.005	0.006	0.005	0.005
Num. of observations (leave out sample)	2,559,818		2,935,930	
Num. of drivers	439,559		464,652	
Num. of movers	171,093		179,887	
Num. of neighborhoods	5,546		5,577	

Notes: The specification with controls partials out the contribution of income and household size - total variance is the residual variance that remains.

Table 1 summarizes the results of decomposing the variation in average log daily kilometers driven into a neighborhood effect and a driver fixed effect. We can see that 51 per cent of the variation can be attributed to idiosyncratic differences in driving, which we will interpret as a drivers current driving habit. However, only between 0.7 and 0.8 per cent of the variation in kilometers driven can be attributed to the current neighborhood where each driver is residing. These results indicate that most of the variation in driving is, in fact, accounted for by the individual fixed

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

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

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

effects. At the same time, the low covariance between the neighborhood fixed effects and the individual fixed effects (between 0.5 and 0.6 per cent) indicates a low degree of positive sorting. Drivers with a high propensity for driving don't seem to sort disproportionately more into neighborhoods with a high degree of driving on average.

6 The Role of Origin on Individual Driving Behavior

In section 5 we established that a large fraction of the variance in kilometers driven can be accounted for by the variation in individual fixed effects. These driver-specific fixed effects capture the average driving habits for each driver within this time period. In this section, we will analyze the association between these measures of driving habits and observable characteristics of the individual drivers.

Table 2: Understanding Individual Fixed Effects for kilometers driven

	Indiv. FE - log daily km. driven			
Neighborhood car ownership at 17	0.13***	0.12***	0.18***	0.17***
	(0.012)	(0.012)	(0.012)	(0.012)
Current neighborhood car ownership		0.32***	0.27***	0.13***
		(0.15)	(.015)	(.016)
Age			0.004***	-0.001***
			(0.0003)	(0.0003)
Income				0.004***
				(0.0001)
Household size				0.03***
				(0.001)
Constant	2.10***	1.97***	1.82***	1.86***
	(0.006)	(0.008)	(0.012)	(0.012)
N	195,187	191,697	191,697	191,636

In Table 2, we can see the resulting coefficients of regressing the individual-level fixed effects from equation 13 on a set of covariates[9]. We can see that there is a positive and significant relationship between our measure of driving habits and the

[9]Since we only have one individual-level fixed effect for each driver in our sample, these regressions will be cross-sectional in nature. As such, we will take the average value of each covariate for the period where we observe kilometers driven (2013 to 2023).

car ownership rate (the share of households that own a car) in the neighborhood where the individual was living at the age of seventeen[10]. This relationship persists even when controlling for the car ownership rate of the current neighborhood of residence. We interpret this as evidence that exposure to an environment with a higher degree of car usage (as measured by the share of households that own a car in the neighborhood) in an individual's formative years may play an important role in determining their future driving habits.

From Table 2, we can also see that the individual fixed effects obtained from estimating equation 13 are positively associated with income and household size. Indicating that drivers that have higher income or live in larger households are more likely to have higher levels of idiosyncratic driving.

To understand if this relationship between exposure to driving at a younger age and current driving habits is specific to a certain age range, we perform the exact same analysis but we vary the age at which we measure the neighborhood-level car ownership rate. In other words, we regress the individual-level fixed effects associated with kilometers driven on a set of covariates (age, income, household size, and car ownership rate in the current neighborhood of residence) as well as the car ownership rate in the neighborhood where the driver lived when they were a certain age a. We repeat this analysis for all ages from $a = 1$ year to $a = 25$. Figure 6 shows the resulting coefficient associated with the car ownership rate at origin for each of these regressions. We can see that, after age two, the relationship between car ownership at the neighborhood of origin and current driving habits is relatively stable up to the age of 19. After 19 years of age, the relationship starts to decrease. This pattern reinforces the idea that there might be a formative period up to the age of 19 where exposure to high levels of driving may shape an individual's driving habits in future.

The car ownership rate at the neighborhood-level is, of course, only a proxy for all the neighborhood characteristics that are associated with exposure to driving a private vehicle. To quantify how much of the variation in driving habits can be explained by the neighborhoods of origin[11], we regress them on a set of origin neighborhood fixed effects. To avoid attributing explanatory power of the current neighborhood to

[10]Our data on car ownership at the neighborhood-level starts in 1990, which means that we can only observe the car ownership rate in the neighborhood where the individual was living at the age of seventeen for individuals who were born after 1973.

[11]For this exercise, we define the neighborhood of origin as the neighborhood where the individual was living at age a for some a between 10 years and 25 years.

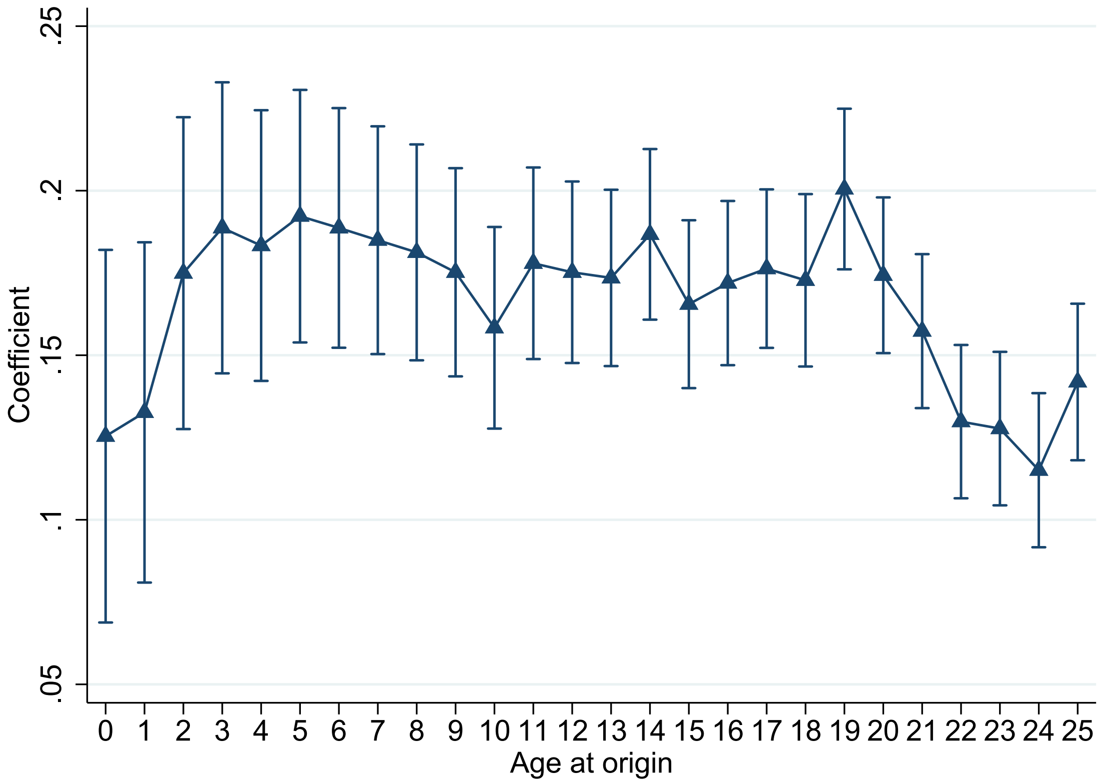

Figure 6: Effect of early-stage exposure to cars on current individual-level driving

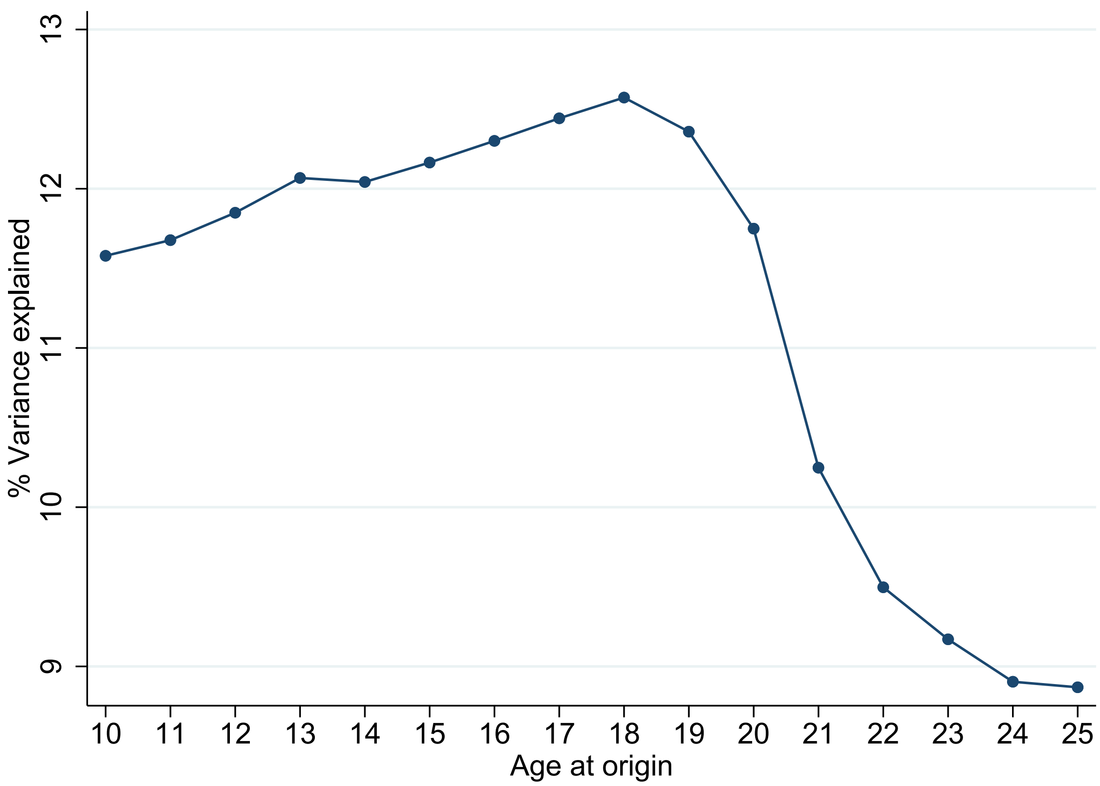

Figure 7: Variation in individual driving explained by "origin" FE. Fixed sample (same sample for all origin ages), movers only

the neighborhood of origin, we consider only drivers who no longer live in the same neighborhood as when they were seventeen. We also drop neighborhoods of origin that are shared by less than 10 drivers in our sample. Figure 7 shows that the variance explained by the neighborhood of origin peaks at around 12.5% when the origins are defined at age 18, and starts diminishing sharply afterwards, again suggesting a larger role for formative years. Since the individual habits capture 51% of the variation in total driving, this translates to around 6% of the total variance in log daily kilometers – more than sixfold the variance explained by current neighborhoods.

7 Conclusions

Individual habits (instead of current residential locations) account for most of the variation in driving behavior. While people on average adjust their driving towards their destination's norms when moving from one neighborhood to another, substantial heterogeneity within neighborhoods remains. This is the case despite us observing neighborhoods at a very granular level, and with or without controlling for income and household size.

But past residential locations matter a lot for driving habit formation, especially at early ages. The car ownership rate in origin neighborhoods predicts current driving habits even conditional on car ownership in current neighborhood, in fact more so than the current neighborhood car ownership when the origin neighborhood is defined at ages 0 to 20, and one controls for the driver's age, income, and household size. Given that individuals in Finland cannot obtain a driver's license before the age of 18, this suggests that habits are also passed down from generation to generation.

These habits formed in early ages persist and explain much of the variation in current driving. For the group of drivers that no longer live in the same neighborhood as when they were 18 (and for whom we can observe both neighborhoods), the origin neighborhoods account for a substantial share of the variation in current habits, representing more than six times the variation in total driving than what is captured by current neighborhoods.

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 20.

—, —, **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 21.

Allen, Treb and Costas Arkolakis, "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 2022, *89* (6), 2911–2957. Cited on page 4.

Almagro, Milena, Felipe Barbieri, Juan Camilo Castillo, Nathaniel G Hickok, and Tobias Salz, "Optimal Urban Transportation Policy: Evidence from Chicago," Technical Report, National Bureau of Economic Research 2024. Cited on page 4.

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 21.

—, **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 21.

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 21.

Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162. Cited on page 4.

—, —, **and Lawrence F Katz**, "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment," *American Economic Review*, 2016, *106* (4), 855–902. Cited on page 4.

— , — , **Patrick Kline, and Emmanuel Saez**, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The quarterly journal of economics*, 2014, *129* (4), 1553–1623. Cited on page 4.

Duranton, Gilles and Anthony J Venables, "Place-based policies: principles and developing country applications," in "Handbook of regional science," Springer, 2021, pp. 1009–1030. Cited on page 4.

Gaubert, Cecile, "Firm sorting and agglomeration," *American Economic Review*, 2018, *108* (11), 3117–3153. Cited on page 4.

Gorback, Caitlin, "Your uber has arrived: Ridesharing and the redistribution of economic activity," *Job Market Paper*, 2020. Cited on page 4.

Kline, Patrick and Enrico Moretti, "People, places, and public policy: Some simple welfare economics of local economic development programs," *Annu. Rev. Econ.*, 2014, *6* (1), 629–662. Cited on page 4.

— , **Raffaele Saggio, and Mikkel Sølvsten**, "Leave-out estimation of variance components," *Econometrica*, 2020, *88* (5), 1859–1898. Cited on page 21.

Lyubich, Eva, "The role of people vs. places in individual carbon emissions," Technical Report, Technical Report, HAAS Working Paper 324 2022 2022. Cited on page 4.

Severen, Christopher, "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *Review of Economics and Statistics*, 2023, *105* (5), 1073–1091. Cited on page 4.

Tsivanidis, Nick, "Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio," *Unpublished manuscript*, 2022, *18.* Cited on page 4.

Appendices

A Figures

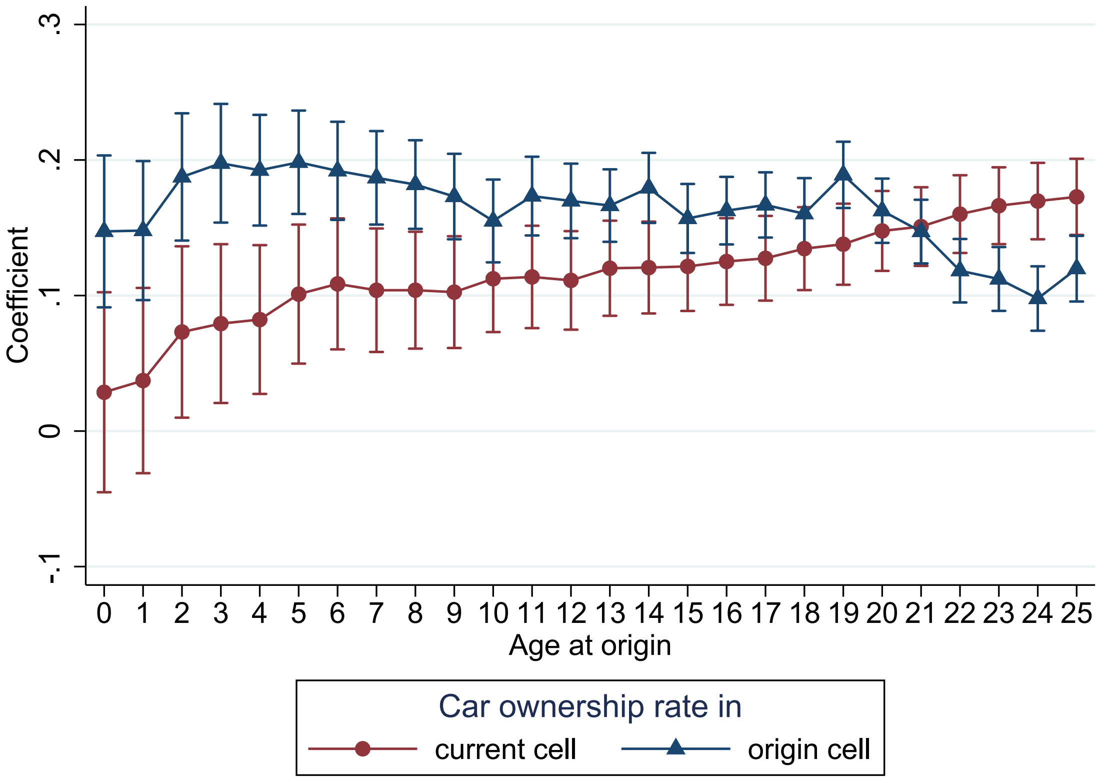

Figure A1: Current Cell vs Origin Cell

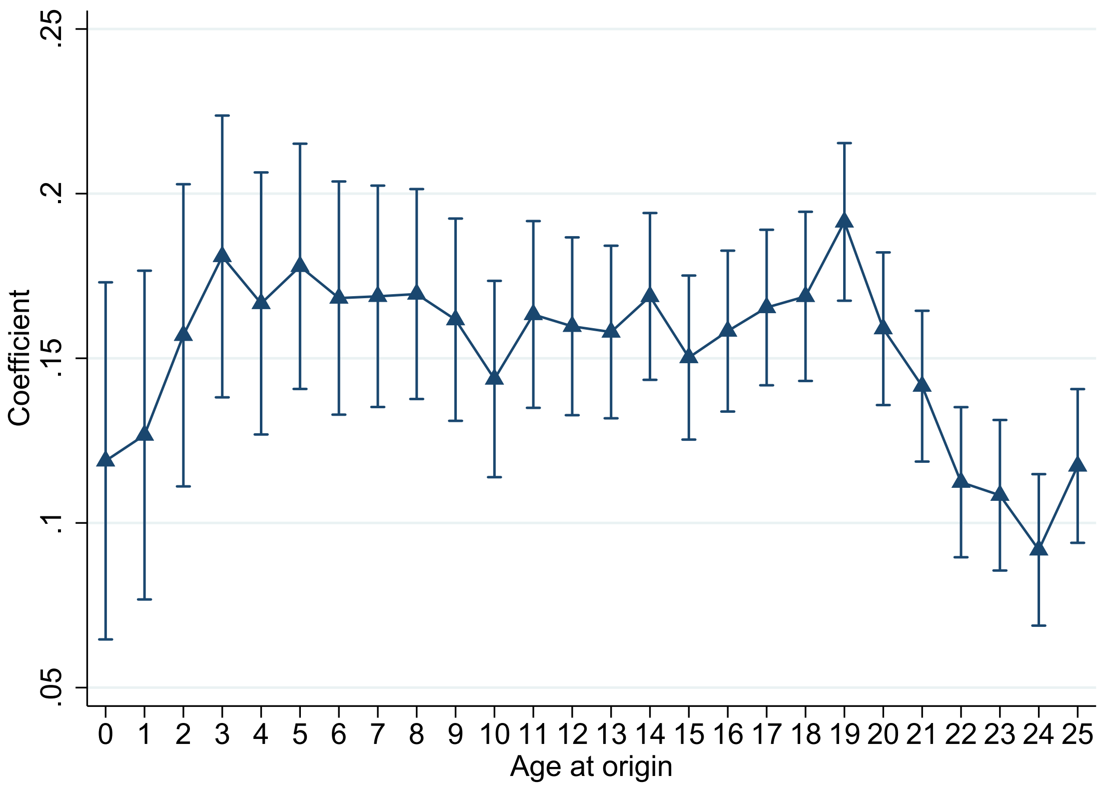

Figure A2: No controls

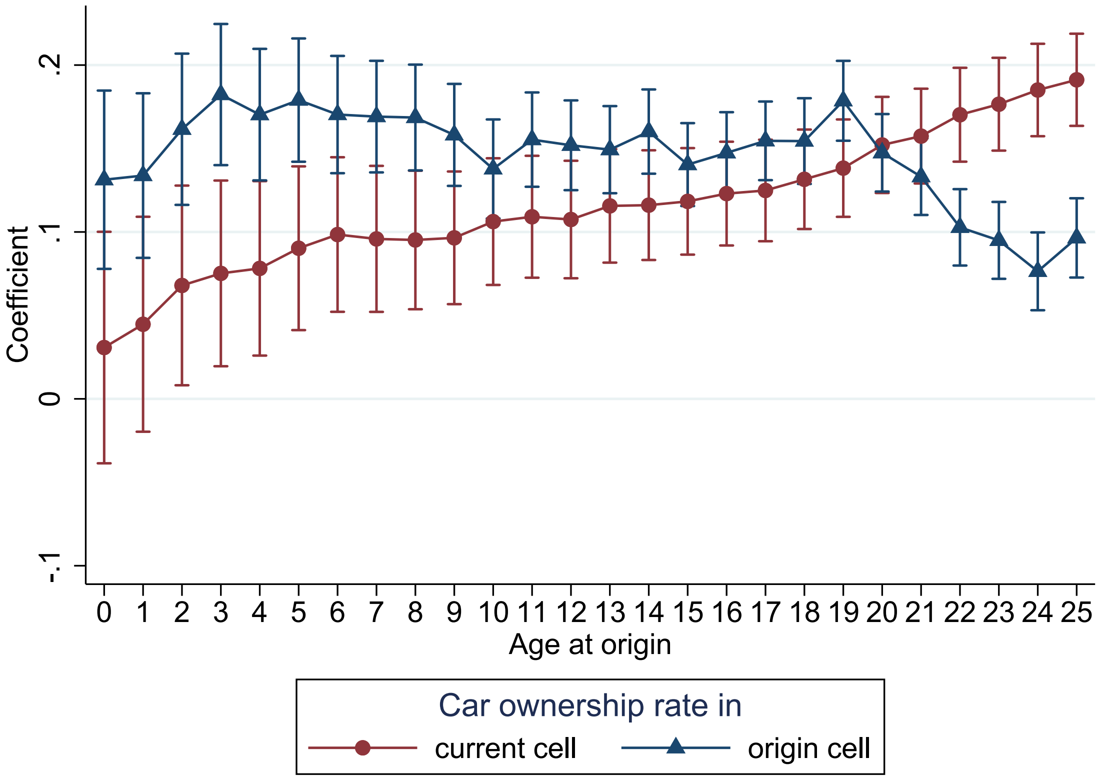

Figure A3: No controls

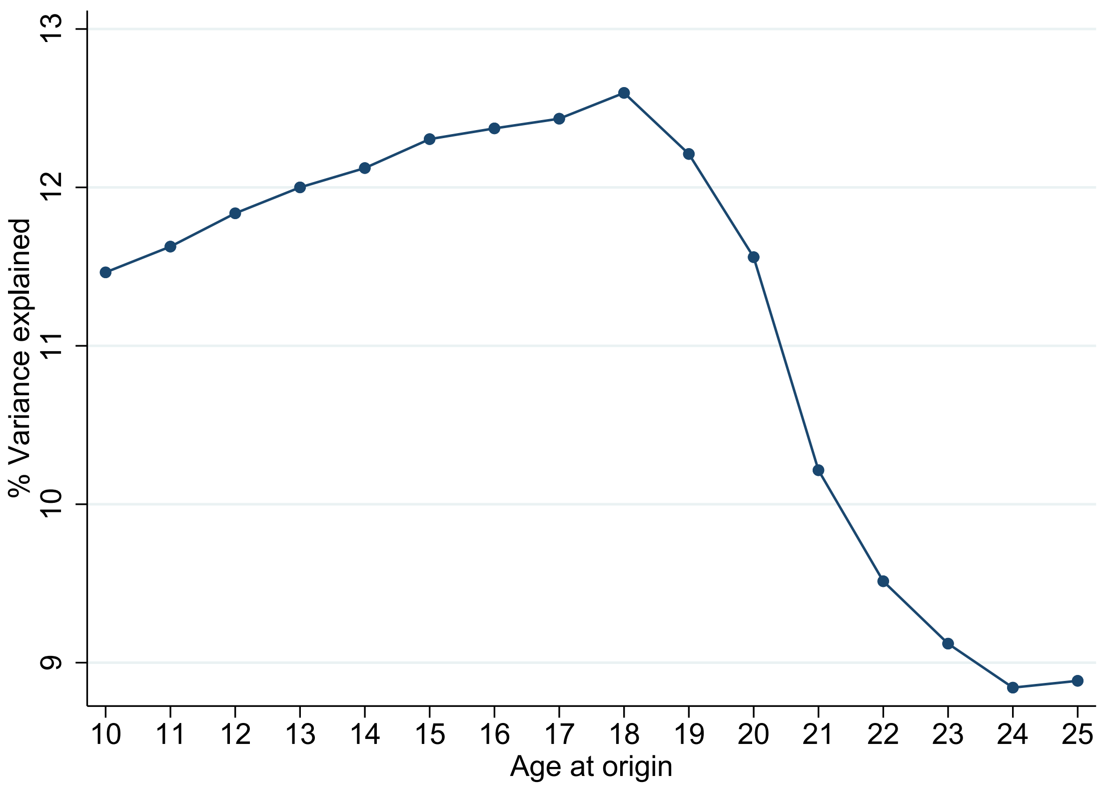

Figure A4: No controls

B Tables

Table A1: Decomposing the variation in car ownership

	Controls		No Controls	
	Value	Share	Value	Share
Total variance (log daily km driven)	0.23	1.000	0.23	1.000
Var. of driver effects	0.17	0.74	0.17	0.74
Var. of neighborhood effects	0.0005	.002	0.0005	0.002
Covariance neighborhood-driver effects	0.001	0.004	0.001	0.004
Num. of observations (leave out sample)	6,963,046		8,089,451	
Num. of drivers	1,063,493		1,120,407	
Num. of movers	535,779		558,108	
Num. of neighborhoods	6,026		6,057	

Notes: The specification with controls partials out the contribution of income and household size - total variance is the residual variance that remains.