Population Density and Public Transit Funding: Political Equilibrium and Social Optimum

Still being polished, please do not circulate yet

Nils-Petter Lagerlöf[†]

March 2026

Abstract

In most U.S. counties, the share of the population who drive for their daily commute is close to 100%. When population density reaches a certain threshold that share starts to decline with population density, while traffic congestion continues to worsen. This paper applies a median voter model with heterogeneous preferences over driving, and endogenous congestion and public transit funding, to match these patterns. The equilibrium outcome tends to have more driving than what is socially optimal, while the tax rate for funding public transit can be higher or lower in equilibrium compared to the socially optimal tax rate.

[†]Department of Economics, York University. E-mail: lagerlof@yorku.ca.

1 Introduction

The vast majority of U.S. commuters drive to work, even those who live in relatively densely populated urban areas. As a consequence, several U.S. cities suffer from severe traffic congestion. The most commonly proposed remedy for this is probably congestion charges, which tend to be politically controversial. For example, congestion pricing in New York City met staunch opposition before it was finally introduced in 2025 (The Economist, 2025), and has been challenged in court by the U.S. Department of Transportation (Shepardson, 2026).

At the same time, many urban areas subsidize public transit to mitigate congestion, which can be costly to taxpayers. Parry and Small (2009, Table 1) find that such subsidies cover well over half of most transit systems' operating costs.[1] Transit fares can also be subject to political meddling (Hughes, 2026). Even those who want to abolish transit subsidies in the U.S. often seem resigned to their continued survival (Winston and Maheshri, 2007, p. 381).

The observation that transit subsidies seem so popular, at least compared to congestion charges, calls for a political-economy theory of what determines their size, and why they exist at all. This is particularly relevant in the U.S. context, where transit users are usually a political minority.

This paper proposes a model where agents face a binary choice between driving or using public transit, and where public transit is funded exclusively by income taxes. These taxes are levied uniformly across the population (i.e., on both drivers and public transit users), at a rate set by a median voter, who in equilibrium can be a driver or a public transit user. A key feature of the model is that agents are heterogeneous in their preferences for driving: some would almost never give up driving, while others are more willing to consider alternatives when they become available.

The political equilibrium in this model depends on population density. When population density is very low, and congestion is not a problem, all agents drive in equilibrium, and the median voter does not support taxes to fund public transit. At the other extreme,

[1] Including subsidies to capital investments would presumably make those numbers even larger.

when density is very high, some still drive, but congestion is bad enough that the median voter uses public transit, and thus also supports high taxes to fund it.

For intermediate levels of population density, the median voter still drives, but because congestion is relatively high she supports some tax-funded subsidies for public transit to induce others to get off the road, although lower than when the median voter herself uses public transit. The model can thus explain the political viability of (modest) transit subsidies in an environment where the majority drives.

The model's results are also consistent with observed variation in population density and driving across U.S. counties. For the vast majority of counties, the share of the population driving for their daily commute is close to 100%. However, once population density reaches about 1800 people per square kilometer, that share declines steeply with population density. The model can match this pattern well quantitatively.

The model can also explain why traffic congestion continues to increase with population density also after driving starts to decline. That is, the share who drives does not decline fast enough to fully countervail the increase in density. To capture this, the model relies on heterogeneity in driving preferences. This makes it increasingly difficult to induce agents to give up driving, since those left driving are the ones who value it the most.

The equilibrium outcome can also be compared to the social optimum, which we define as the amount of transit use and the subsidy rate that maximize an agent's utility under a "veil-of-ignorance" assumption, i.e., before she knows what preference type she is. The socially optimal amount of driving tends to be lower than the equilibrium outcome, and the threshold density level at which driving starts to decline is also lower. However, the optimal tax rate can be either lower or higher than the equilibrium outcome, and it varies continuously with population density, rather than making discrete jumps at certain density levels.

The rest of this paper continues with a literature overview in Section 2. Section 3 then presents some of the facts that our model can explain. Section 4 sets up the model, and explains why it cannot match the data without heterogeneity. Section 5 presents the quantitative analysis, first for a setting where per-capita incomes are kept constant across locations (Section 5.2), and then letting them vary as in the data (Section 5.3). Section 6

concludes.

2 Existing Literature

The observation that higher population density is associated with less driving is well known from, e.g., Newman and Kenworthy (1989), Kenworthy and Laube (1999), and McIntosh et al. (2014). This body of research typically compares different large cities worldwide, rather than locations in the same country. They do identify policies than can reduce car dependence, including public transit investment, but not why some jurisdictions seem to choose different policies than others in that regard.

There are many papers on the costs of road congestion. Some focus on quantifying the costs without assessing possible remedies (e.g., Seidel and Wickerath, 2020), while others document that some often suggested solutions do *not* seem to work, like building more roads (Duranton et al., 2011; Duranton and Turner, 2018).

The perhaps most obvious solution to congestion is corrective taxes, i.e., congestion charges (see, e.g., Vickrey, 1969; Hall, 2018; Yang, 2024; Almagro et al., 2024; Ostrovsky and Schwarz, 2025). Closest to this paper may be Almagro et al. (2024), who study the socially optimal combination of congestion charges and public transit subsidies. Considering Chicago, they find that the optimal policy may be fully subsidized transit (i.e., zero transit fares), funded with congestion charges. Here we only consider funding through income taxes, which (as discussed) tends to be the most common source of transit funding, and focus on both the political equilibrium and the social optimum.[2]

Papers that specifically study public transit subsidies as a way to reduce driving and/or congestion include Proost and Van Dender (2008), Parry and Small (2009), Basso and Silva (2014), and Basso et al. (2021). These start with the so-called Mohring effect (after Mohring, 1972), which says that public transit exhibits positive scale effects, as increased ridership allows for better and more frequent service, thus benefiting all riders, arguably the opposite of congestion externalities. However, none of these papers explores what political-

[2] Another difference compared to some of these (in particular Almagro et al., 2024; Yang, 2024) is that this paper does not study any specific city, but is more interested in broad patterns across U.S. locations.

economy or voting mechanisms may explain variation in public transit use and driving behaviors across the U.S.

The mitigating effect of public transit on congestion is otherwise evidenced by research on how congestion and related externalities increase during public transit strikes (Anderson, 2014; Adler and van Ommeren, 2016; Bauernschuster et al., 2017; Spyropoulou, 2020).

Finally, there is a large theoretical literature on public transit subsidies and how these are determined politically and/or optimally (see, e.g., Vickrey, 1980; Parry and Small, 2005, 2009; De Borger and Proost, 2012, 2015). The model presented here differs from all these in its direct focus on population density in a U.S. context.

3 Patterns in U.S. Data

This section sums up some of the patterns that our model is able to capture. Unless otherwise stated, the observations refer to U.S. counties in 2010. Population density is measured as people per square kilometer. More information about data sources is provided in Section 5.1 below and in the Online Appendix.

3.1 Driving

Figure 1 plots the share of the population who report driving for their daily commute against logged population density. In the vast majority of counties, well over 90% drive. The main exceptions are a few very densely populated counties, such as those in and around New York City. Specifically, there seems to be a break around a log population density level of approximately 7.5, corresponding to about 1800 people per square kilometer, close to the county of Nassau, New York. The share driving is relatively constant up until that point, and then starts to fall with further increases in population density.

We also note from Figure 1 that the share who drives exceeds 50% in most counties. That is, drivers seem to mostly form a political majority.

Figure 2 plots vehicle miles per capita against population density, both in logs. Vehicle miles does not correspond closely to any specific model variable, but the plot confirms that

counties with higher population densities see less driving overall, and that the negative relationship is steeper when log population density exceeds 7.5.

3.2 Congestion

Figure 3 offers a different way of looking at the same data as in Figure 1, by constructing a measure of congestion, or what we can call "driver density," i.e., the share of the population who drives multiplied by population density. Log driver density thus equals log population density plus the log of the share who drive, the latter being close to zero for most counties, since the (non-logged) share who drives is close to one. When plotting driver density against population density, both logged, most counties are clustered around the 45-degree line. The exception is the same cluster of counties that we noted in Figure 1, with log population density above 7.5. Interestingly, driver density—which may serve as a crude proxy for congestion—is still increasing with population density among these counties. In other words, despite the decline in driving, congestion is still becoming worse with rising population density.

This last point can be corroborated using the so-called Roadway Congestion Index made available by the Bureau of Transportation Statistics. This measures vehicle travel density on major roadways in select urban areas, many of which overlap with the more densely population counties in Figures 1 and 3, although often spanning several counties and/or states. Figure 4 plots this congestion index against log population density in these urban areas, showing a clear positive association. This confirms that congestion gets worse with rising population density, despite the decline in driving.

3.3 Public Transit Funding

Even though congestion tends to be worse in more densely populated locations, there is also some evidence that such locations tend to spend more on providing alternatives to driving, in particular public transit. This may seem obvious, since public transit is largely an urban phenomenon, but it seems to hold also when comparing different urban areas that have transit systems. To see this, we can use data on expenses for different public tran-

sit systems across urban areas (UZAs), allowing us to compute transit expenses as a share of total income. Figure 5 correlates population density with this public transit spending share across urban areas, finding a positive relationship: more densely populated places spend more on public transit.

4 The Model

Let an agent j living in location i care about consumption (C_{ij}) and a good that I call just transportation (T_{ij}). Preferences are described by

$$V_{ij} = C_{ij}^{\alpha} T_{ij}^{1-\alpha}, \tag{1}$$

where $\alpha \in (0, 1)$.

4.1 Transportation

The agent's choice of transportation mode is binary: she can either drive, or use public transit. For an agent who drives, transportation is given by

$$T_{ij} = \frac{\eta_j X_i}{z_i P_i}, \tag{2}$$

where X_i is the amount of land (or space) at location i, P_i is population size, z_i is the share of the population who drives, and $\eta_j \geq 1$ is agent j's preference parameter for driving, explained below. That is, drivers are worse off with less space per driver, or lower $X_i/(z_i P_i)$, capturing a congestion effect from driving.

Public transit is a tax-funded non-rivalrous public good, the amount of which is assumed to depend on aggregate tax revenue per unit of land; dividing by land area captures the idea that public transit users benefit from factors such as network connectivity and train/bus frequency, which are easier to supply when the transit service area is smaller for a given population (cf. Proost and Van Dender, 2008; Mohring, 1972).

Let y_i denote pre-tax income per agent in location i, and τ_i the public-transit tax, levied uniformly on drivers and public transit users. With each of the P_i agents paying $\tau_i y_i$ in taxes, total tax revenues equals $\tau_i y_i P_i$, and public transit spending per unit of land becomes

$\tau_i y_i P_i / X_i$. We then assume that an agent using public transit obtains transportation equal to

$$T_{ij} = \left(\frac{\Omega \tau_i y_i P_i}{X_i} \right)^{\lambda}, \tag{3}$$

where Ω is a productivity parameter, and where $\lambda \in (0, 1)$ captures a decreasing marginal effect of tax revenue, the role of which is discussed further below.[3]

By construction, land and population in (2) and (3) only matter through the ratio P_i / X_i, which we will refer to as population density. This exerts opposite effects on drivers' and public transit users' payoffs. That is, density benefits public transit users (due to transit being a non-rivalrous public good, and because of the scale effect in funding per land area), and harms drivers (due to the congestion externality). This set-up is meant to simplify the analysis, but is not strictly necessary for the main results to hold. For example, allowing for some congestion in public transit would not alter the results qualitatively if that congestion effect is weak enough relative to the scale effect in funding.[4] Similarly, we could introduce tax funding for roadways for drivers, thus allowing tax revenues and population density to benefit drivers as well, but that would not change the mechanics of the model as long as the congestion effect dominates. In short, the payoffs defined in (2) and (3) are meant to keep the analysis simple and tractable.

Letting $L_i = P_i / X_i$ denote population density, we can write transportation consumption more compactly as

$$T_{ij} = \begin{cases} \frac{\eta_j}{z_i L_i} & \text{if agent } j \text{ drives,} \\ (\Omega \tau_i y_i L_i)^{\lambda} & \text{if agent } j \text{ uses public transit.} \end{cases} \tag{4}$$

[3]Equivalently, letting $T_{ij} = \widetilde{\Omega} \left(\tau_i y_i P_i / X_i \right)^{\lambda}$ for some productivity parameter $\widetilde{\Omega} > 0$, we can define $\Omega = \widetilde{\Omega}^{1/\lambda}$ to get (3).

[4]Specifically, instead of (3), we could let public transit users' transportation consumption equal

$$T_{ij} = \left(\frac{\tau_i y_i P_i}{X_i} \right)^{\lambda} \left(\frac{[1 - z_i] P_i}{X_i} \right)^{-\kappa}$$

for some $\kappa \in [0, 1)$. This would make T_{ij} decreasing in transit user density, $[1 - z_i] P_i / X_i$, but also rule out closed form solutions for equilibrium z_i. Here we consider the case where $\kappa = 0$.

4.2 Consumption

Recall that y_i denotes pre-tax income per agent in location i, and τ_i the public-transit tax. (We abstract from other taxes.) Letting D denote the exogenous private cost of driving (insurance, repairs, the car itself, etc.), non-transport consumption can be written

$$C_{ij} = \begin{cases} (1-\tau_i)y_i - D & \text{if agent } j \text{ drives,} \\ (1-\tau_i)y_i & \text{if agent } j \text{ does not drive.} \end{cases} \tag{5}$$

It is assumed throughout that $y_i \geq D/\alpha$, which reduces the number of cases we need to consider, by ensuring that some agents will always drive in an equilibrium with public transit taxes.

Incomes do not differ between agents in a location, so all agent heterogeneity is captured by the preference parameter η_j. One way to introduce an agent-specific income component without changing any results would be to let both the driving cost and incomes vary proportionally between agents, leaving the driver/non-driver consumption ratio the same across agents.[5]

4.3 Equilibrium

The equilibrium in location i is such that (i) agents choose transport mode to maximize individual utilities; and (ii) the (uniform) tax rate, τ_i, maximizes utility of the median voter. Part (ii) captures the model's political-economy mechanism. Part (i) means we can substitute (4) and (5) into (1), to write the utility of an agent in location i as

$$V_{ij} = \max\left\{V_{ij}^D, V_{ij}^{ND}\right\}, \tag{6}$$

where V_{ij}^D and V_{ij}^{ND} are utilities associated with driving and not driving, respectively, defined as follows:

$$V_{ij}^D = \left[(1-\tau_i)y_i - D\right]^\alpha \left(\frac{\eta_j}{z_i L_i}\right)^{1-\alpha}, \tag{7}$$

and

$$V_{ij}^{ND} = \left[(1-\tau_i)y_i\right]^\alpha \left(\Omega\tau_i y_i L_i\right)^{\lambda(1-\alpha)}. \tag{8}$$

[5]That is, suppose that an agent j in location i earns income $y_{ij} = A_i h_j$ and faces a driving cost of $h_j D$. Then the ratio $[(1-\tau_i)A_i h_j - h_j D]/[(1-\tau_i)A_i h_j]$ boils down to the same as if using (5), but with A_i replacing y_i.

4.3.1 Without Heterogeneity

It is helpful to first analyze a setting without heterogeneity in driving prefences, by setting $\eta_j = 1$ for all j. We here focus on the most interesting equilibrium, where some agents use public transit, meaning $z_i < 1$. (It is easy to show that this must exist for large enough L_i.) In this equilibrium, since all agents are identical and some drive and others use public transit, they must be indifferent between the two modes of transport, and thus obtain the same utility, $V_{ij}^D = V_{ij}^{ND}$. This in turn means they must prefer the same tax rate. Maximizing V_{ij}^{ND} in (8) with respect to τ_i gives the optimal tax rate as

$$\tau_i^{**} = \frac{\lambda(1-\alpha)}{\alpha + \lambda(1-\alpha)}, \tag{9}$$

which depends only on α and λ, and is independent of population density, L_i. Substituing the tax rate in (9) into (8), we can now write the utility of all agents, both drivers and public transit users, as

$$V_{ij}^D = V_{ij}^{ND} = (1-\tau_i^{**})^{\alpha} \, (\tau_i^{**})^{\lambda(1-\alpha)} y_i^{\alpha+\lambda(1-\alpha)} (\Omega L_i)^{\lambda(1-\alpha)}, \tag{10}$$

which is increasing in L_i. That is, both drivers and public transit users are better off in more densely populated locations. Intuitively, the fraction of the population who drive (z_i) decreases when transit options become more available, which alleviates congestion, in turn benefiting those who still drive.

To see this more clearly, we can use the expression for V_{ij}^D in (7) with $\eta_j = 1$ to back out the total number of drivers per area:

$$z_i L_i = \left[\frac{(1-\tau_i^{**})y_i - D}{(1-\tau_i^{**})y_i} \right]^{\frac{\alpha}{1-\alpha}} \left(\frac{1}{\Omega \tau_i^{**} y_i L_i} \right)^{\lambda}. \tag{11}$$

We now see that $z_i L_i$ is decreasing in L_i, since $\lambda > 0$. In other words, an increase in L_i is offset by a greater than proportional reduction in z_i. Since $z_i L_i$ is a measure of congestion, this setting (i.e., without heterogeneity in driving preferences) thus predicts a negative association between population density and congestion when public transit is operational ($z_i < 1$), contradicting what we see in the data. As we shall see in the next section, allowing for heterogeneity in driving preferences can render the model qualitatively consistent with the data in this regard.

4.3.2 With Heterogeneity

The Driving Preference Distribution We let the preference parameter η_j be Pareto distributed, with cumulative distribution function

$$F(\eta_j) = 1 - \eta_j^{-\frac{1}{\delta}}, \tag{12}$$

where $\delta > 0$. It is easy to see (by setting $1 - \eta_j^{-\frac{1}{\delta}} = 1/2$) that the median η_j equals 2^{δ}.[6] As $\delta \to 0$ heterogeneity goes away, and all agents become identical with $\eta_j = 1$ for all j.

It is also useful to note a link between the share who drive and the preference of the marginal driver. Let $\widetilde{\eta}$ be the preference parameter for the marginal driver, such that all agents drive who have $\eta_j > \widetilde{\eta}$. Then the population share with $\eta_j < \widetilde{\eta}$ must equal $1 - z_i$, i.e., the share who uses public transit. From (12), it follows that

$$F(\widetilde{\eta}) = 1 - \widetilde{\eta}^{-\frac{1}{\delta}} = 1 - z_i, \tag{13}$$

implying

$$\widetilde{\eta} = z_i^{-\delta}. \tag{14}$$

Intuitively, the lower is the share who drives, z_i, the greater is the preference for driving by the marginal driver, and thus all agents who still drive. The implication is that, as population density increases, and agents start to substitute driving for public transit, it becomes increasingly difficult to induce the remaining drivers to change behavior.

Tax rates Since this is our benchmark setting, it is helpful to give a full characterization of the types of equilibria that can arise. Section A.1 in the Online Appendix derives expressions for τ_i and z_i in three possible types of equilibrium, each of which previals for different levels of population density, L_i. The equilibrium tax rates can be described as follows.

First, it is straightforward to see that an equilibrium where all agents drive ($z_i = 1$) must be such that taxes are zero ($\tau_i = 0$), since there is no point funding transit if no one uses it.

[6] A mean for η_j exists only if $\delta < 1$, but we do not need to use that here, only the median.

Next, let τ_i^* be the tax rate that drivers prefer if some other agents use public transit. This will be the tax rate in a voting equilibrium where drivers make up a majority, but not 100% of the population, i.e., $z_i \in (1/2, 1)$. Intuitively, improved public transit eases congestion, so drivers may prefer a positive tax rate. As shown in Section A.1.2 of the Online Appendix, τ_i^* is defined from $G(\tau_i^*) \equiv 0$, where

$$G(\tau) = -\frac{\alpha}{1-\tau} - \frac{\delta\alpha y_i}{(1-\tau)\, y_i - D} + \frac{\lambda(1-\alpha)}{\tau}. \tag{15}$$

This implicitly defines $\tau_i^* > 0$ in terms of exogenous variables other than L_i.

Finally, the tax rate preferred by those who choose to use public transit is the same as in (9). This is the equilibrium tax rate if non-drivers are a majority, $z_i < 1/2$, and maximizes V_{ij}^{ND} in (8).

Population Thresholds The type of equilibrium depends on population, L_i. Let $\widehat{L}_i$ denote the level of L_i below which the equilibrium is such that all agents drive. I show in the appendix that we can write this as

$$\widehat{L}_i = \left(\frac{B_i^*}{z_i^*}\right)^{\frac{1+\delta}{1+\lambda}}, \tag{16}$$

where

$$B_i^* = \left[\frac{(1-\tau_i^*)y_i - D}{(1-\tau_i^*)y_i}\right]^{\frac{\alpha}{(1-\alpha)(1+\delta)}} \left(\frac{1}{\Omega\tau_i^* y_i}\right)^{\frac{\lambda}{1+\delta}}, \tag{17}$$

and

$$z_i^* = \left[\frac{(1-\tau_i^*)y_i - D}{y_i - D}\right]^{\frac{\alpha}{1-\alpha}} < 1, \tag{18}$$

where the inequality comes from $\tau_i^* > 0$. We can interpret z_i^* as the level to which z_i must fall to induce those agents who still drive to support a tax rate of τ_i^*, which (recall) is leveled uniformly on both drivers and non-drivers. The factor B_i^* contains variables, other than L_i, that determine z_i; see (23) below.

Next, let $\widehat{\widehat{L}}_i$ denote the level of L_i above which the equilibrium is one with non-drivers in majority, $z_i < 1/2$. Section A.1.5 of the Online Appendix shows that

$$\widehat{\widehat{L}}_i = (2B_i^{**})^{\frac{1+\delta}{1+\lambda}}, \tag{19}$$

where B_i^{**} is defined similarly as B_i^* in (17), except that τ_i^* is replaced by τ_i^{**}:

$$B_i^{**} = \left[\frac{(1-\tau_i^{**})y_i - D}{(1-\tau_i^{**})y_i}\right]^{\frac{\alpha}{(1-\alpha)(1+\delta)}} \left(\frac{1}{\Omega\tau_i^{**}y_i}\right)^{\frac{\lambda}{1+\delta}}. \tag{20}$$

The intuition behind how $\widehat{L}_i$ and $\widehat{\widehat{L}}_i$ are derived, and what B_i^{**} and B_i^* mean, will become more clear soon.

Having defined the population thresholds $\widehat{L}_i$ and $\widehat{\widehat{L}}_i$, we can now write the equilibrium tax rate as

$$\tau_i = \begin{cases} 0 & \text{if } L_i < \widehat{L}_i, \\ \tau_i^* & \text{if } L_i \in [\widehat{L}_i, \widehat{\widehat{L}}_i), \\ \tau_i^{**} & \text{if } L_i \geq \widehat{\widehat{L}}_i, \end{cases} \tag{21}$$

and the equilibrium fraction driving as

$$z_i = \begin{cases} 1 & \text{if } L_i < \widehat{L}_i, \\ Z^{\text{eq}}(y_i, \tau_i^*, L_i) & \text{if } L_i \in [\widehat{L}_i, \widehat{\widehat{L}}_i), \\ Z^{\text{eq}}(y_i, \tau_i^{**}, L_i) & \text{if } L_i \geq \widehat{\widehat{L}}_i, \end{cases} \tag{22}$$

where

$$Z^{\text{eq}}(y_i, \tau_i, L_i) = \left[\frac{(1-\tau_i)y_i - D}{(1-\tau_i)y_i}\right]^{\frac{\alpha}{(1-\alpha)(1+\delta)}} \left(\frac{1}{\Omega\tau_i y_i}\right)^{\frac{\lambda}{1+\delta}} L_i^{-\frac{1+\lambda}{1+\delta}}. \tag{23}$$

Note from (17) and (23) that $B_i^* = Z^{\text{eq}}(y_i, \tau_i^*, 1)$, i.e., the first two factors on the right-hand side in (23) evaluated at $\tau_i = \tau_i^*$. It can now be seen from (16), (17), (22), and (23) that a population level of $\widehat{L}_i$, and a tax rate of τ_i^*, generate a fraction driving of $Z^{\text{eq}}(y_i, \tau_i^*, \widehat{L}_i) = z_i^*$, where z_i^* is given by (18).[7]

Analogously, to understand where $\widehat{\widehat{L}}_i$ comes from, note from (19), (20), (22), and (23) that $Z^{\text{eq}}(y_i, \tau_i^{**}, \widehat{\widehat{L}}_i) = 1/2$. That is, $\widehat{\widehat{L}}_i$ is the population level where non-drivers can form a majority and implement their preferred tax rate, τ_i^{**}.

[7]To see it another way, use (22) and (23) to note that $z_i^* = Z(y_i, \tau_i^*, \widehat{L}_i) = Z(y_i, \tau_i^*, 1)\widehat{L}_i^{-\frac{1+\lambda}{1+\delta}} = B_i^*\widehat{L}_i^{-\frac{1+\lambda}{1+\delta}}$. Solving for $\widehat{L}_i$ gives (16).

Figure 6 plots several variables against logged population density, based on the same numerical example as in Section 5 below. The top two panels show the tax rate and the share drivers, τ_i and z_i. Note the discrete jumps when L_i reaches the two thresholds, $\widehat{L}_i$ and $\widehat{\widehat{L}}_i$, as described by (21) and (22).

The bottom left panel shows our measure of congestion, defined as the ratio of total drivers to land, $z_i L_i$. (Recall that $L_i = P_i / X_i$ measures population per unit of land.) For $L_i > \widehat{L}_i$ (i.e., for levels of population density where public transit is operative), we see from (22) and (23) that $z_i L_i$ is proportional to $L_i^{\frac{\delta-\lambda}{1+\delta}}$, which is increasing in L_i as long as $\lambda < \delta$. This contrasts with the results in Section 4.3.1, where we abstracted from heterogeneity. In other words, heterogeneity allows the model to replicate the empirical pattern that congestion tends to worsen when population density increases, even when public transit is operational.

4.4 Social Optimum

Next we compare the equilibrium outcomes to the social optimum. We let the social planner assign all agents' utilities equal weight. This corresponds to maximizing the expected utility of an agent "under the veil of ignorance," i.e., before her preference for driving (η_j) is drawn from the Pareto distribution in (12).

Recall that we let $\widetilde{\eta}$ denote the preference parameter for the marginal driver, such that all agents drive who have $\eta_j > \widetilde{\eta}$. We can thus write the total utility of those who drive by integrating (7) over η_j, which gives

$$V_i^{D,\text{tot}} = \int_{\widetilde{\eta}}^{\infty} V_{ij}^D f(\eta_j) d\eta_j, \tag{24}$$

where

$$f(\eta_j) = F'(\eta_j) = \frac{\eta_j^{-\frac{1+\delta}{\delta}}}{\delta}, \tag{25}$$

is the probability density function associated with the cumulative density function in (12). Section A.2 of the Online Appendix solves the integral in (24), and uses (14), to show that

$$V_i^{D,\text{tot}} = \frac{z_i^{\alpha-\delta(1-\alpha)}}{1-\delta(1-\alpha)} \left[(1-\tau_i)y_i - D\right]^{\alpha} \left(\frac{1}{L_i}\right)^{1-\alpha}. \tag{26}$$

Note that, if $\alpha - \delta(1-\alpha) < 0$, then $V_i^{D,\text{tot}}$ is decreasing in z_i. That is, if δ is sufficiently large, and α sufficiently small, then driving preferences increase so steeply as z_i falls that the total utility of all drivers increases. Intuitively, reducing the number of drivers leaves more space for the remaining drivers, who value that space more. To ensure that $V_i^{D,\text{tot}}$ is increasing in z_i, it is sufficient to assume that $\delta < \alpha/(1-\alpha)$.

The corresponding total utility of all non-drivers is obtained by simply multiplying their population share, $1 - z_i$, by each agent's utility in (8), which is independent of type, η_j. This gives

$$V_i^{ND,\text{tot}} = (1 - z_i) \left[(1-\tau_i)y_i\right]^{\alpha} \left(\Omega \tau_i y_i L_i\right)^{\lambda(1-\alpha)}. \tag{27}$$

The social planner's objective function can now be written

$$\begin{aligned} W_i &= V_i^{D,\text{tot}} + V_i^{ND,\text{tot}} \\ &= \tfrac{z_i^{\alpha-\delta(1-\alpha)}}{1-\delta(1-\alpha)} \left[(1-\tau_i)y_i - D\right]^{\alpha} \left(\tfrac{1}{L_i}\right)^{1-\alpha} \\ &\quad + (1 - z_i) \left[(1-\tau_i)y_i\right]^{\alpha} \left(\Omega \tau_i y_i L_i\right)^{\lambda(1-\alpha)}. \end{aligned} \tag{28}$$

The social planner sets z_i and τ_i to maximize W_i, subject to $z_i \in [0,1]$ and $\tau_i \in [0,1]$. While it is not feasible to find analytical solutions for optimal τ_i, we can find an explicit expression for the interior candidate for optimal z_i in terms of τ_i. Interestingly, the result resembles the corresponding equilibrium outcome in (23). Specifically, if optimal $z_i < 1$ (and optimal $\tau_i > 0$), then $z_i = Z^{\text{opt}}(y_i, \tau_i, L_i)$, where

$$Z^{\text{opt}}(y_i, \tau_i, L_i) = \Psi \left[\frac{(1-\tau_i)y_i - D}{(1-\tau_i)y_i}\right]^{\frac{\alpha}{(1-\alpha)(1+\delta)}} \left(\frac{1}{\Omega \tau_i y_i}\right)^{\frac{\lambda}{1+\delta}} L_i^{-\frac{1+\lambda}{1+\delta}} = \Psi Z^{\text{eq}}(y_i, \tau_i, L_i) \tag{29}$$

where we have defined

$$\Psi = \left[\frac{\alpha - \delta(1-\alpha)}{1-\delta(1-\alpha)}\right]^{\frac{1}{(1-\alpha)(1+\delta)}} < 1, \tag{30}$$

and where the inequality follows from $\alpha < 1$ (and $\alpha - \delta[1-\alpha] > 0$).

In other words, for any given tax rate τ_i, the share of the population who drives is higher in a political equilibrium than what is socially optimal. Of course, the tax rate need not be identical between the socially optimal and equilibrium outcomes. However, when matching the model to data in Section 5 below, the differences in tax rates tend to quantitatively small, so (29) and (30) explains most of the gap in fraction drivers.

5 Quantitative Analysis

Next we examine how well the model can match some of the facts presented in Section 3. In the data, counties vary both in terms of population density and per-capita income levels, but the model explains variation in driving mostly through differences in population density. To make the results easier to interpret, we start in Section 5.2 below by treating per-capita incomes as constant across locations, so that all model variation is driven by differences in population density across locations. Section 5.3 then lets both population density and per-capita incomes vary as in the data.

5.1 Data

5.1.1 Driving, Population Densities, and Incomes by County

We use the census to measure population density, and the Bureau of Economic Analysis (BEA) for per-capita incomes, both referring to 2010. Data on the share who drive to work, corresponding to z_i in the model, are from the American Community Survey, also for the year 2010.

For ease of exposition, we consider counties with data on both incomes and population density, and with densities above median in the data. More sparsely populated counties would not provide public transit anyhow, in either the equilibrium or optimal outcomes in the model. This gives us $1,527$ counties.

We normalize population density in the data to equal one (zero in logs) when actual population density is 7.5 in logs, or about 1800 people per square kilometer. Recall that this is the approximate level of density above which the share drivers starts to decline with further increases in density; cf. Figure 1.

We also normalize the data on per-capita incomes, y_i, to average 100 across the counties considered.

5.1.2 Transit Expenditures by UZA

We interpret transit taxes, τ_i in the model, as total transit spending divided by total income. Our starting point is data from the National Transit Database (NTD) on operating expenses for different public transit systems across the U.S. in 2010. These are reported by urbanized areas (UZAs), which are different from counties.

Helpfully, the NTD also reports population density and total population at the UZA level.[8] All UZAs in the NTD data have population densities above the level at which the share drivers starts to fall below 100% in Figure 1, suggesting some level of consistency between the two datasets.

NTD only reports operating expenditures. We lack data on capital expenditures (the other major expenditure component), but across UZAs these sum up to approximately 41% of operating expenditure, according to Zhao et al. (2021, Table 1). Based on this, we impute total transit expenditures as 1.41 times operating expenditures. However, the main quantitative results of interests are very similar if using only operating expenditures.

To estimate total income by UZA, we use the product of the total UZA population reported by the NTD, and an estimate of per-capita income from the BEA for Metropolitan Statistical Areas (MSAs) that we can match to UZAs by name. Transit taxes can then be estimated as total transit spending divided by total income.

All in all, we end up with 224 UZAs with data on per-capita incomes, population density, and transit taxes.

5.2 Benchmark

5.2.1 Benchmark Calibration

In our benchmark analysis, we set $y_i = 130$ for all locations, which is 30% above the (normalized) sample mean. This matches the level observed in counties with population densities close to $\widehat{L}_i$ in the model, i.e., around the "kink" in Figure 1. Those counties are more densely populated than the sample mean, and because per-capita incomes and density are positively correlated in the data they also have higher per-capita incomes.

[8] All measures are converted to people per square kilometers to conform with the county-level data.

We set the cost of driving, D to 10, which is 10% of the (normalized) per-capita income mean. This is in line with BEA data on consumer expenditures related to driving.

We set the productivity parameter, Ω, so that $\widehat{L}_i = 1$ (i.e., $\ln \widehat{L}_i = 0$) when $y_i = 130$. Recall that this corresponds to a non-normalized population density of 7.5, the level above which the share drivers starts to decline in the data.

We then set the remaining parameters—α, δ, and λ—to minimize the unweighted sum of squared deviations of z_i and τ_i in the model from the corresponding measures in the data. More precisely, for a given set of values for α, δ, and λ, we loop across counties to calculate z_i, compute square deviations of these from the share driving in the data, and sum up across counties. Then we run the same loop across UZA's to calculate τ_i in the model, compute the square deviations of τ_i from data, and sum up across UZAs. We divide each sum by the square of the number of counties and UZAs, respectively, and add the two sums up. We do this for multiple combinations of α, δ, and λ, and choose the set of values that produces the smallest sum of squared deviations.

The resulting benchmark values are summed up in Table 1, together with some of the endogenous outcomes.

5.2.2 Benchmark Results

Figure 7 compares the share drivers in the data to the equilibrium z_i in the model, and the corresponding outcome when z_i is determined optimally, i.e., when maximizing the social utility function W_i in (28).

The share who drives makes a discrete drop from 1 to $z_i^* = .92$ when density reaches $\widehat{L}_i$ (which we normalized to one, or zero in logs), and then declines, matching the data nicely. The correlation coefficient between z_i produced by the model and the corresponding measure in data is about 0.91.

Comparing the equilibrium path to the optimal, we see that more people tend to drive than what is socially optimal. Driving starts to decline at lower density levels, about 40% lower than $\widehat{L}_i$. When population density reaches $\widehat{L}_i$, only approximately 60% of commuters would drive if decisions were made optimally.

A different way of interpreting the same patterns as in Figure 7 is to plot the share driv-

ing in the data against the corresponding equilibrium and optimal outcomes produced by the model. This is presented in Figure A.1 in the Online Appendix.

Table 3 presents details about the share driving in the 25 most densely populated counties according to the data, as well as the model's equilibrium and optimal z_i.

Figure 8 compares driver density, $z_i L_i$, to the corresponding data in Figure 3. Not surprisingly, the patterns here mirror those in Figure 7: both the data and the model follow the 45-degree line when z_i is close to one. More interestingly, the model can match a continued rise in congestion also when z_i starts to decline. Recall that this hinges on agents being heterogenous ($\delta > 0$).

Figure 9 shows the paths for the tax rate. This is harder to compare to data, especially the model's stepwise pattern in the equilibrium tax rate, but it matches the broad observation that denser population is associate with higher per-capita spending on public transit. Interestingly, the optimal tax rate can be higher or lower than the equilibrium rate, depending on the level of population density. Intuitively, the social planner controls z_i directly and does not need to induce drivers to switch mode of transport.

5.3 Letting Per-Capita Incomes Vary

5.3.1 Recalibration

When letting per-capita incomes vary, most parameters are recalibrated following the same procedure as in the benchmark setting. Specifically, we loop across counties to calculate the squared deviations of z_i from the corresponding values in the data, and across UZAs to calculate squared deviations of τ_i from data. The difference is that z_i and τ_i are calculated based on both densities and per-capita incomes in each county/UZA. We then set α, δ, and λ to minimize the sum of the squared deviations.

Given these parameter values, Ω is set to make $\widehat{L}_i = 1$.

The cost of driving, D, is set to 10, which is the same in the benchmark setting. The news is that the cost of driving relative to income now varies across counties and UZAs.

The resulting parameter values are summed up in Table 2, together with endogenous variable outcomes of interest.

5.3.2 Results

Figure 10 and Figure 11 show the outcomes for z_i and τ_i when simulating the model using the parameter values in Table 2, and letting per-capita incomes vary across locations as in the data. For comparison, we include the graphs for z_i and τ_i based on $y_i = 130$ for all counties or UZAs, corresponding to those in Figures 7 and Figure 9, but here based on different parameter values.

The discrete drop in z_i when population density reaches $\widehat{L}_i$ is smaller in Figure 10 compared to Figure 7, but the paths are qualitatively similar. Optimal z_i tends to fall below the equilibrium level, and the share drivers in the data, at least for higher population densities. When density reaches $\widehat{L}_i$, the optimal share drivers is a little over half.

Figure 10 also shows that allowing per-capita incomes to vary across locations does not improve the model fit much. Figure A.2 in the Online Appendix illustrates this by plotting the share driving in the data against the corresponding equilibrium and optimal outcomes, similar to Figure A.1 for the benchmark case. Even though the parameters have been recalibrated to maximize model fit, the fit in Figure A.1 is marginally better than that in Figure A.2. When incomes are allowed to vary, the correlation coefficient between z_i produced by the model and the corresponding measure in data is about 0.89, slightly below 0.91 in the benchmark case when we held incomes fixed.

Table 4 presents information on data and model outcomes for the 25 most densely populated counties in the setting where we let incomes vary, similar to the benchmark case in Table 3. The major news here is that the model-generated z_i are not monotonically decreasing in population density.

6 Conclusion

TBW

Acknowledgement: This paper has received no financial support to date, and the author has no financial relationships to declare, or other potential conflicts of interest.

References

Adler, M. W. and van Ommeren, J. N. (2016). Does public transit reduce car travel externalities? Quasi-natural experiments' evidence from transit strikes. *Journal of Urban Economics*, 92:106–119.

Almagro, M., Barbieri, F., Castillo, J. C., Hickok, N. G., and Salz, T. (2024). Optimal urban transportation policy: Evidence from Chicago. *NBER Working Papers*, 32185.

Anderson, M. L. (2014). Subways, strikes, and slowdowns: the impacts of public transit on traffic congestion. *American Economic Review*, 104(9):2763–96.

Basso, L. J., Navarro, M., and Silva, H. E. (2021). Public transport and urban structure. *Economics of Transportation*, 28:100232.

Basso, L. J. and Silva, H. E. (2014). Efficiency and substitutability of transit subsidies and other urban transport policies. *American Economic Journal: Economic Policy*, 6(4):1–33.

Bauernschuster, S., Hener, T., and Rainer, H. (2017). When labor disputes bring cities to a standstill: The impact of public transit strikes on traffic, accidents, air pollution, and health. *American Economic Journal: Economic Policy*, 9(1):1–37.

De Borger, B. and Proost, S. (2012). A political economy model of road pricing. *Journal of Urban Economics*, 71(1):79–92.

De Borger, B. and Proost, S. (2015). The political economy of public transport pricing and supply decisions. *Economics of Transportation*, 4(1-2):95–109.

Duranton, G. and Turner, M. A. (2018). Urban form and driving: Evidence from US cities. *Journal of Urban Economics*, 108:170–191.

Duranton, G., Turner, M. A., Arnott, R., Borenstein, S., Desmet, K., Brueckner, J., Couture, V., Glaeser, E., Kohlhagen, S., Kopp, A., Levinson, D., Michaels, G., and Mun, K. (2011). The Fundamental Law of Road Congestion: Evidence from US Cities. *American Economic Review*, 101(6):2616–52.

Hall, J. D. (2018). Pareto improvements from Lexus Lanes: The effects of pricing a portion of the lanes on congested highways. *Journal of Public Economics*, 158:113–125.

Hughes, S. (2026). Urban expansion in the age of liberalism. *Works in Progress*.

Kenworthy, J. R. and Laube, F. B. (1999). Patterns of automobile dependence in cities: an international overview of key physical and economic dimensions with some implications for urban policy. *Transportation Research Part A: Policy and Practice*, 33(7-8):691–723.

McIntosh, J., Trubka, R., Kenworthy, J., and Newman, P. (2014). The role of urban form and transit in city car dependence: Analysis of 26 global cities from 1960 to 2000. *Transportation Research Part D: Transport and Environment*, 33:95–110.

Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *American Economic Review*, 62(4):591–604.

Newman, P. W. and Kenworthy, J. R. (1989). Gasoline consumption and cities. *Journal of the American Planning Association*, 55(1):24–37.

Ostrovsky, M. and Schwarz, M. (2025). Congestion pricing, carpooling, and commuter welfare. *NBER Working Papers*, 34261.

Parry, I. W. and Small, K. A. (2005). Does Britain or the United States Have the Right Gasoline Tax? *American Economic Review*, 95(4):1276–1289.

Parry, I. W. and Small, K. A. (2009). Should Urban Transit Subsidies Be Reduced? *American Economic Review*, 99(3):700–724.

Proost, S. and Van Dender, K. (2008). Optimal urban transport pricing in the presence of congestion, economies of density and costly public funds. *Transportation Research Part A: Policy and Practice*, 42(9):1220–1230.

Seidel, T. and Wickerath, J. (2020). Rush hours and urbanization. *Regional Science and Urban Economics*, 85:103580.

Shepardson, D. (2026). US judge says Trump effort to end New York City congestion pricing unlawful | Reuters.

Spyropoulou, I. (2020). Impact of public transport strikes on the road network: The case of Athens. *Transportation Research Part A: Policy and Practice*, 132:651–665.

The Economist (2025). Congestion pricing in Manhattan is a predictable success. *The Economist*.

Vickrey, W. (1969). Congestion theory and transport investment. *American Economic Review*, 59(2):260.

Vickrey, W. (1980). Optimal transit subsidy policy. *Transportation*, 9(4):389–409.

Winston, C. and Maheshri, V. (2007). On the social desirability of urban rail transit systems. *Journal of Urban Economics*, 62(2):362–382.

Yang, N. (2024). The political economy of congestion pricing: Evidence from London voters. *Mimeo, Columbia University*.

Zhao, J. Z., Lou, S., Fonseca, C., Feiock, R., and Shen, R. (2021). Explaining transit expenses in US urbanised areas: Urban scale, spatial form and fiscal capacity. *Urban Studies*, 58(2):280–296.

Tables and Figures

Exogenous variables		
Variable	**Value**	**Comment**
y_i	130 for all i	Per-capita incomes in counties with $L_i \approx \widehat{L}_i$; 30% above average
Ω	1.5105×10^{-10}	To make $\widehat{L}_i = 1$
D	10	10% of average y_i across locations
α	0.9849	To match z_i and τ_i to data
λ	0.2047	To match z_i and τ_i to data
δ	1.5714	To match z_i and τ_i to data
Endogenous variables		
Variable	**Value**	**Comment**
τ_i^*	0.0011	Transit tax rate when drivers majority
τ_i^{**}	0.0031	Transit tax rate when non-drivers majority
z_i^*	0.9213	Share driving when public transit becomes operational ($L_i = \widehat{L}_i$)
$\widehat{L}_i$	1	Normalization
$\widehat{\widehat{L}}_i$	3.0862	Population density where non-drivers become majority

Table 1: Variable values used for benchmark simulations.

Exogenous variables		
Variable	**Value**	**Comment**
y_i	NA	Set as in data
Ω	3.9008×10^{-12}	To make $\widehat{L}_i = 1$ when $y_i = 130$
D	10	10% of average y_i across locations; same as benchmark
α	0.9062	To match z_i and τ_i to data
λ	0.0266	To match z_i and τ_i to data
δ	1.0769	To match z_i and τ_i to data
Endogenous variables		
Variable	**Value**	**Comment**
τ_i^*	0.0013	Transit tax rate when drivers majority
τ_i^{**}	0.0023	Transit tax rate when non-drivers majority
z_i^*	0.9868	Share driving when public transit becomes operational ($L_i = \widehat{L}_i$)
$\widehat{L}_i$	1	Normalization
$\widehat{\widehat{L}}_i$	3.8738	Population density where non-drivers become majority

Table 2: Variable values used for simulations where per-capita incomes are allowed to vary as in data.

County	Population Density (Normalized)	Share driving (z_i)		
		Data	Model (Equil.)	Model (Optimal)
Pinellas, FL	0.71	0.98	1.00	0.75
Richmond City, VA	0.73	0.92	1.00	0.75
Orange, CA	0.81	0.97	1.00	0.69
Bergen, NJ	0.83	0.86	1.00	0.69
Denver, CO	0.84	0.91	1.00	0.68
Milwaukee, WI	0.84	0.94	1.00	0.68
Norfolk, VA	0.96	0.95	1.00	0.63
Nassau, NY	1.00	0.83	0.92	0.62
St. Louis City, MO	1.10	0.88	0.88	0.59
Union, NJ	1.11	0.89	0.88	0.58
Cook, IL	1.17	0.80	0.85	0.57
Essex, NJ	1.33	0.78	0.81	0.53
Baltimore City, MD	1.64	0.79	0.73	0.48
Arlington, VA	1.71	0.69	0.72	0.47
Richmond, NY	1.71	0.67	0.72	0.47
Alexandria, VA	1.99	0.76	0.67	0.43
District of Columbia, DC	2.10	0.53	0.65	0.42
Philadelphia, PA	2.43	0.70	0.61	0.39
Suffolk, MA	2.65	0.61	0.58	0.37
Hudson, NJ	2.93	0.56	0.56	0.36
San Francisco, CA	3.67	0.59	0.46	0.32
Queens, NY	4.39	0.44	0.42	0.29
Bronx, NY	7.03	0.35	0.34	0.23
Kings, NY	7.55	0.29	0.33	0.23
New York, NY	14.83	0.14	0.24	0.16

Table 3: Simulation results for the 25 most densely populated counties in the benchmark setting, where per-capita incomes are kept constant across locations.

County	Population Density (Normalized)	Share driving (z_i)		
		Data	Model (Equil.)	Model (Optimal)
Pinellas, FL	0.71	0.98	1.00	0.64
Richmond City, VA	0.73	0.92	1.00	0.64
Orange, CA	0.81	0.97	1.00	0.64
Bergen, NJ	0.83	0.86	1.00	0.69
Denver, CO	0.84	0.91	1.00	0.63
Milwaukee, WI	0.84	0.94	1.00	0.56
Norfolk, VA	0.96	0.95	0.88	0.50
Nassau, NY	1.00	0.83	1.00	0.63
St. Louis City, MO	1.10	0.88	0.84	0.48
Union, NJ	1.11	0.89	0.97	0.55
Cook, IL	1.17	0.80	0.90	0.51
Essex, NJ	1.33	0.78	0.89	0.51
Baltimore City, MD	1.64	0.79	0.70	0.40
Arlington, VA	1.71	0.69	0.88	0.50
Richmond, NY	1.71	0.67	0.74	0.42
Alexandria, VA	1.99	0.76	0.81	0.46
District of Columbia, DC	2.10	0.53	0.76	0.43
Philadelphia, PA	2.43	0.70	0.60	0.34
Suffolk, MA	2.65	0.61	0.68	0.38
Hudson, NJ	2.93	0.56	0.57	0.32
San Francisco, CA	3.67	0.59	0.60	0.34
Queens, NY	4.39	0.44	0.43	0.24
Bronx, NY	7.03	0.35	0.30	0.17
Kings, NY	7.55	0.29	0.32	0.18
New York, NY	14.83	0.14	0.32	0.18

Table 4: Simulation results for the 25 most densely populated counties when letting per-capita incomes vary.

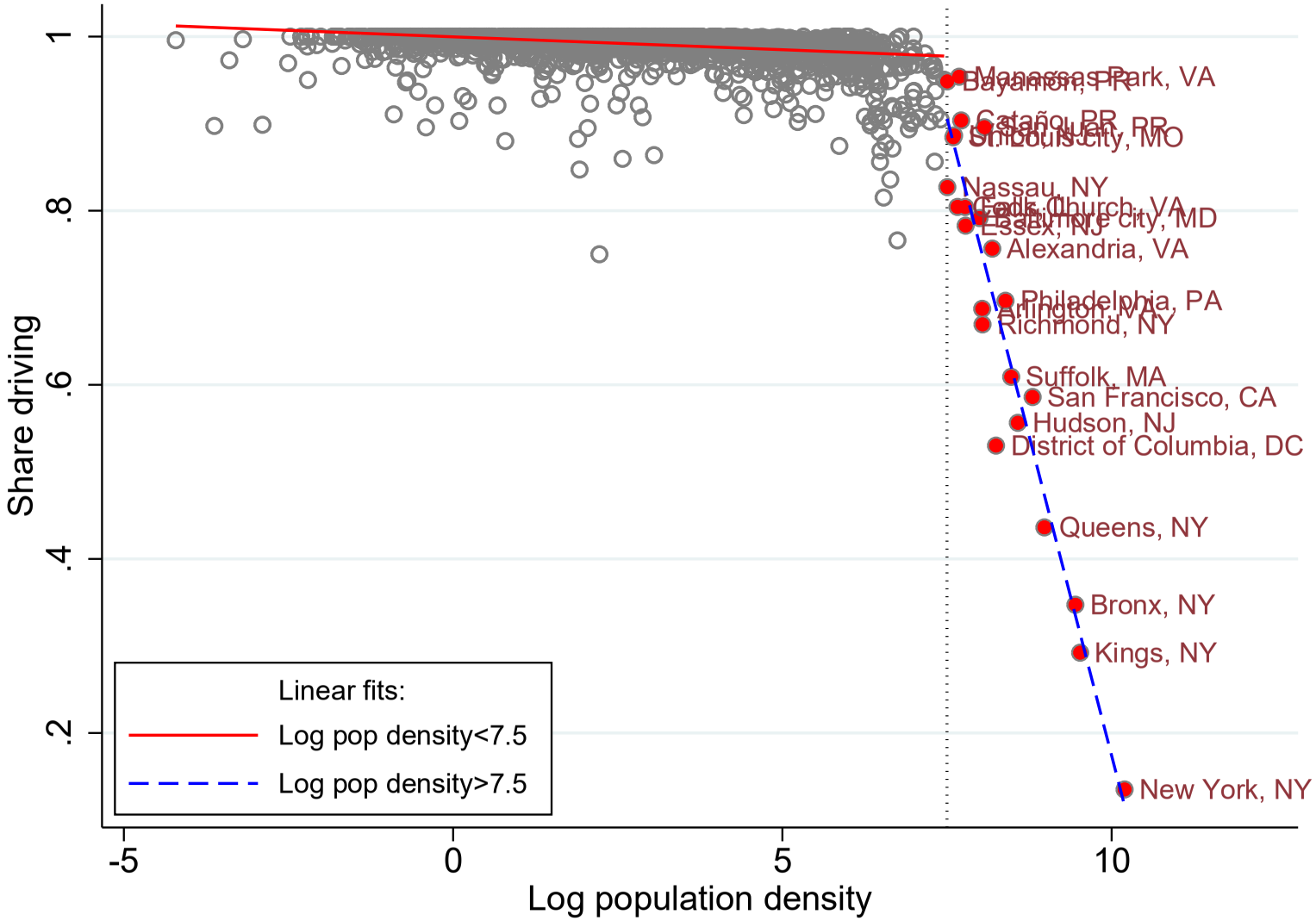


Figure 1: Driving and population density across U.S. counties.

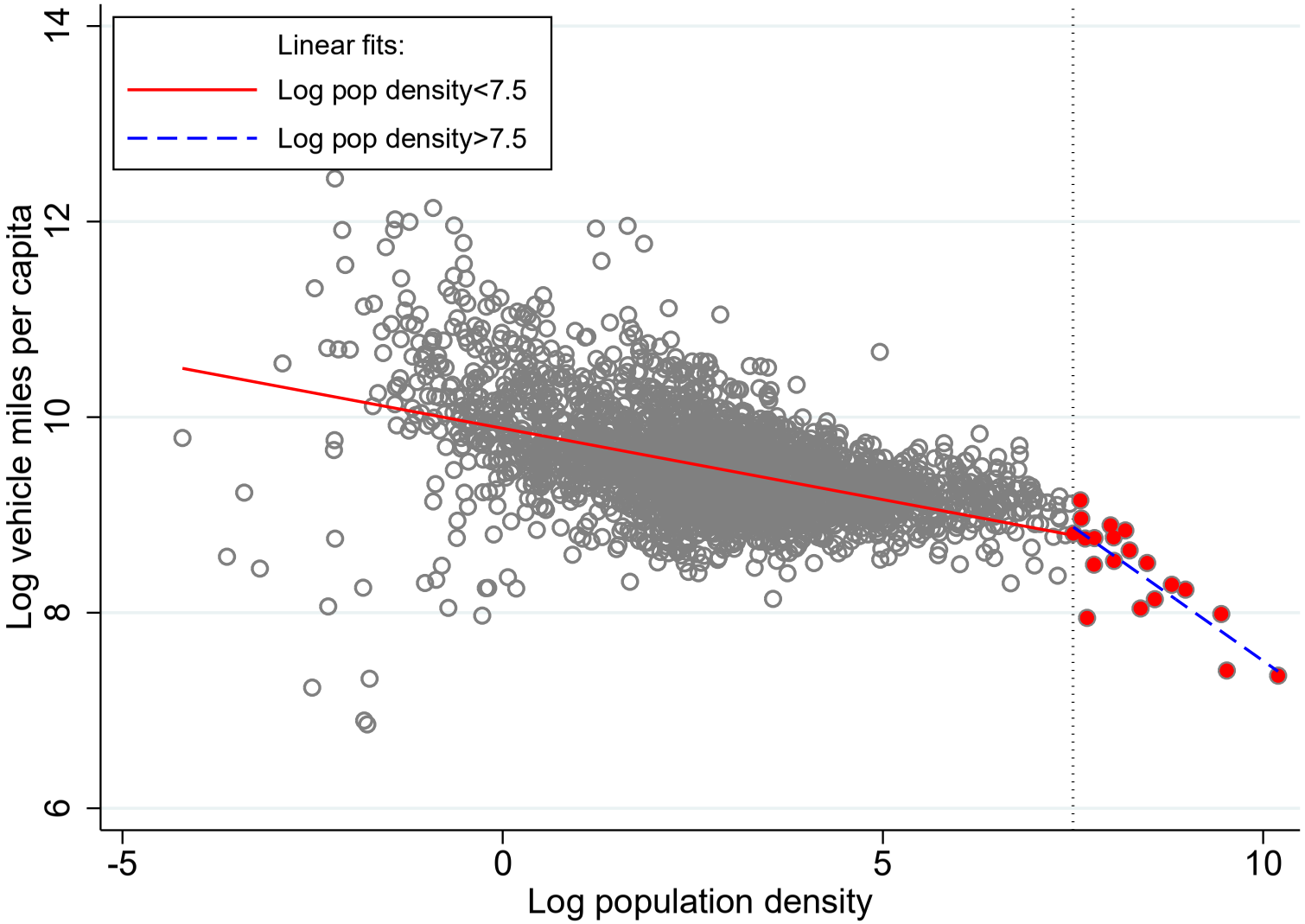


Figure 2: Vehicle miles per capita and population density across U.S. counties.

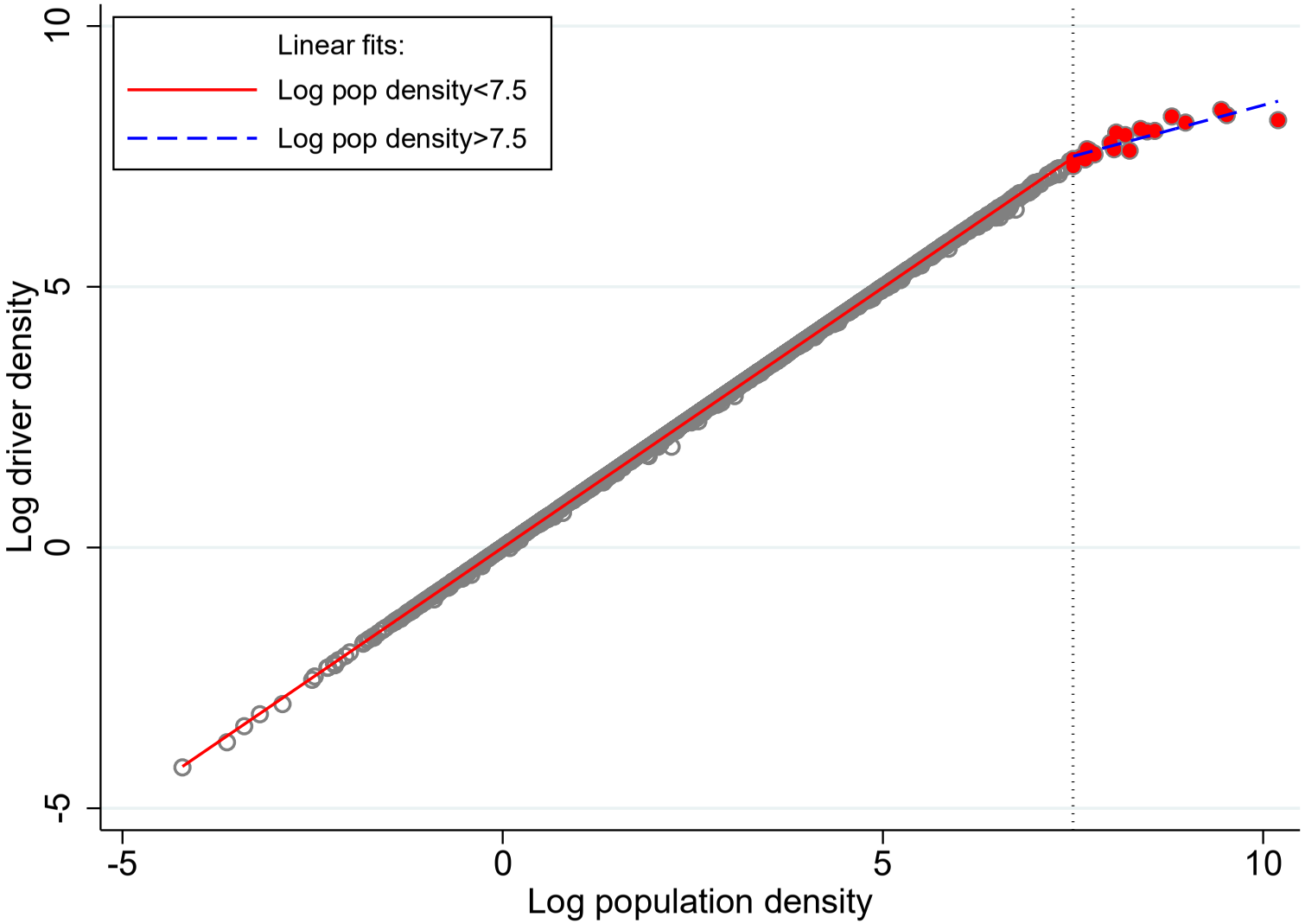


Figure 3: Driver density and population density across U.S. counties.

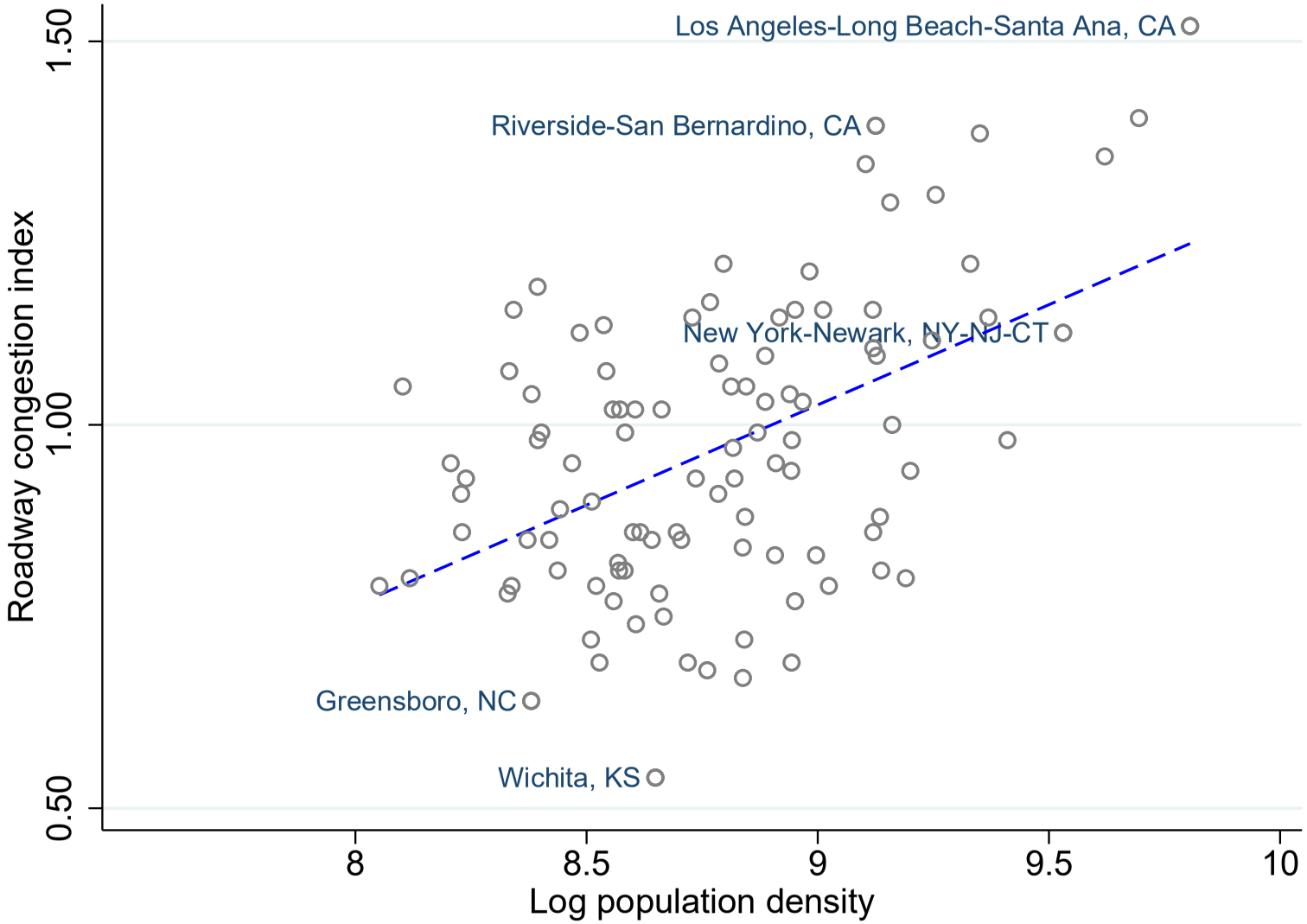


Figure 4: The congestion index and population density across U.S. urban areas.

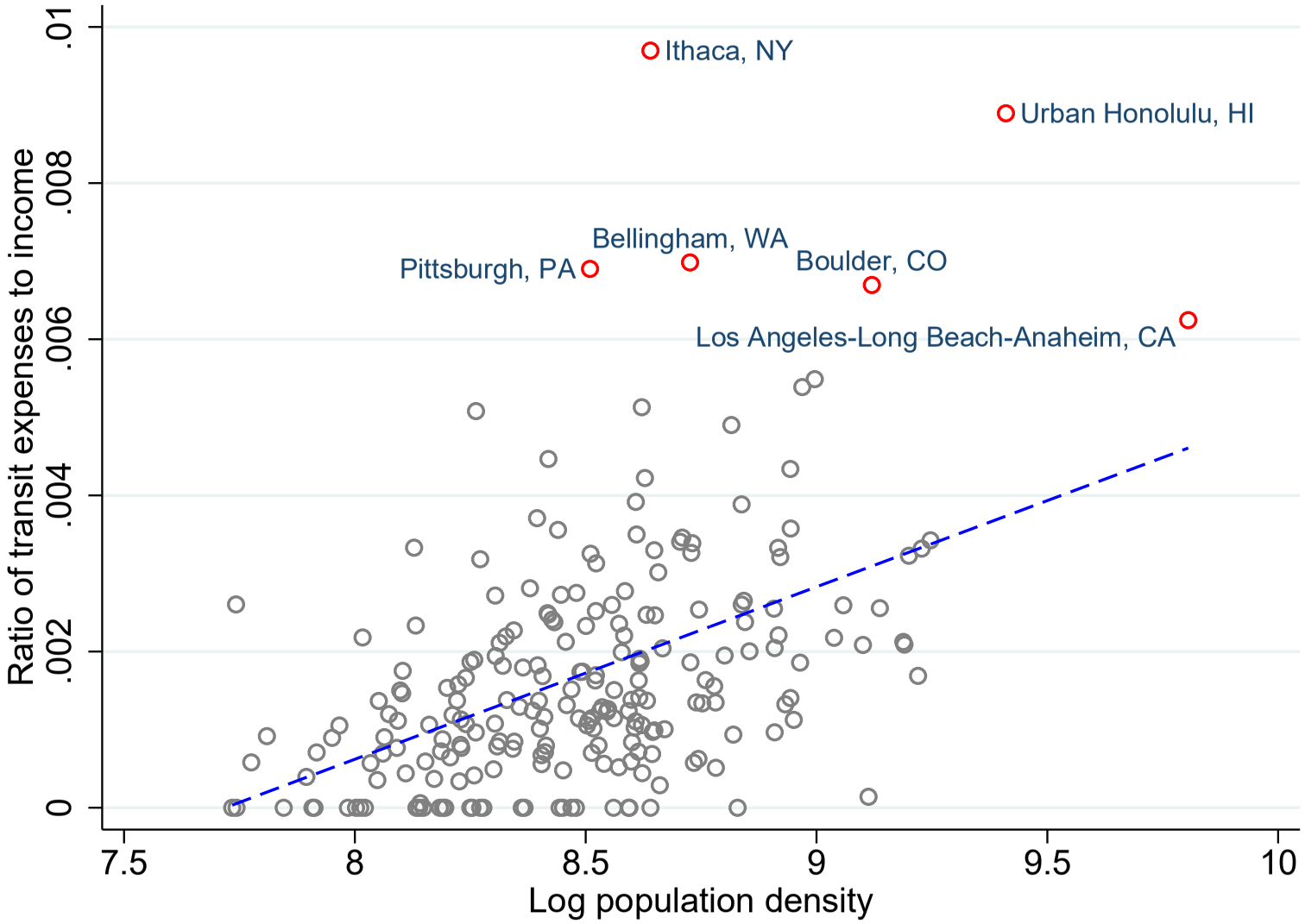


Figure 5: The ratio of transit system expenditures to income plotted against population density across urbanized areas (UZAs).

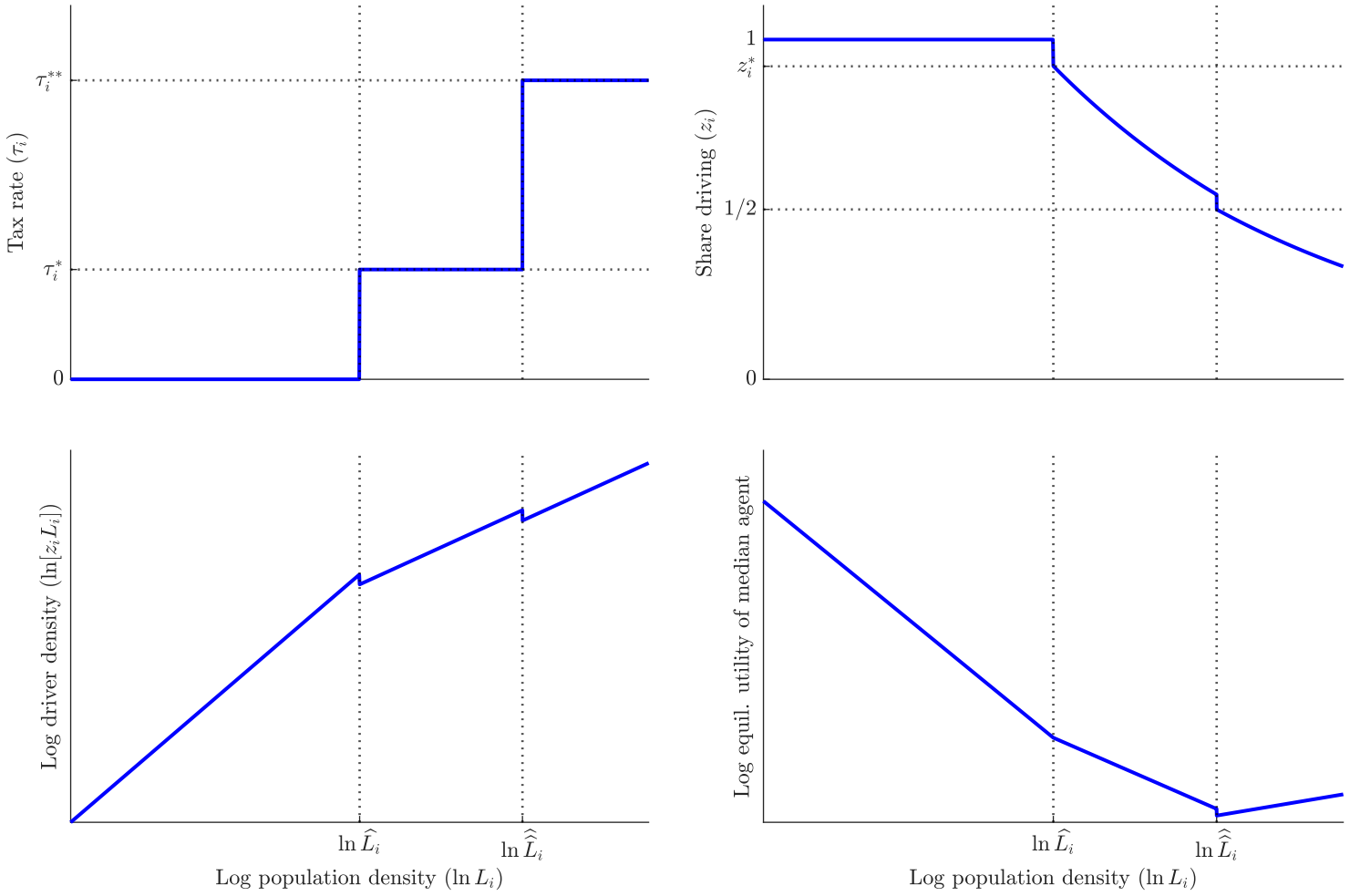


Figure 6: How the main variables depend on population density.

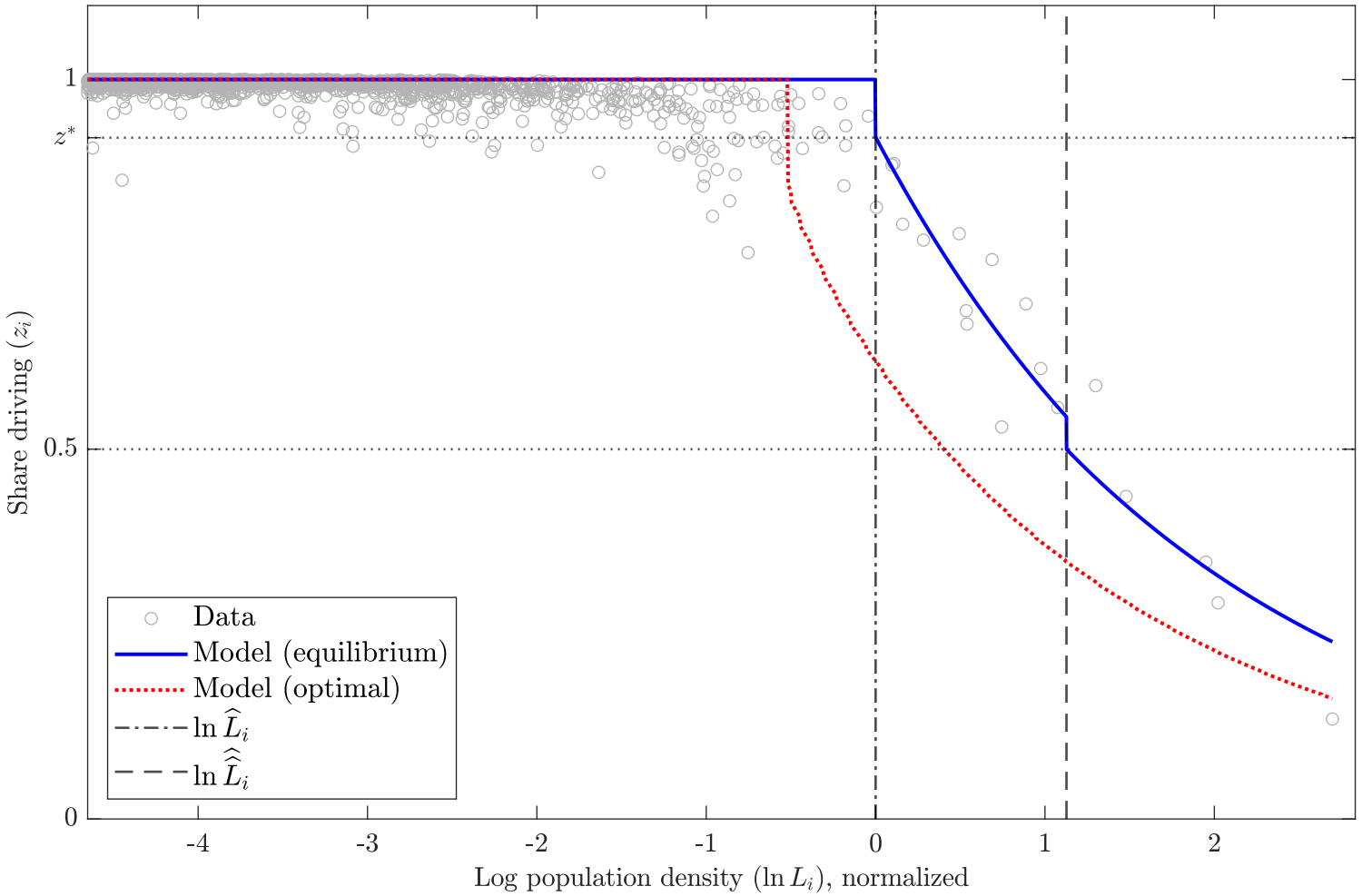


Figure 7: The share drivers in model and data, benchmark simulation.

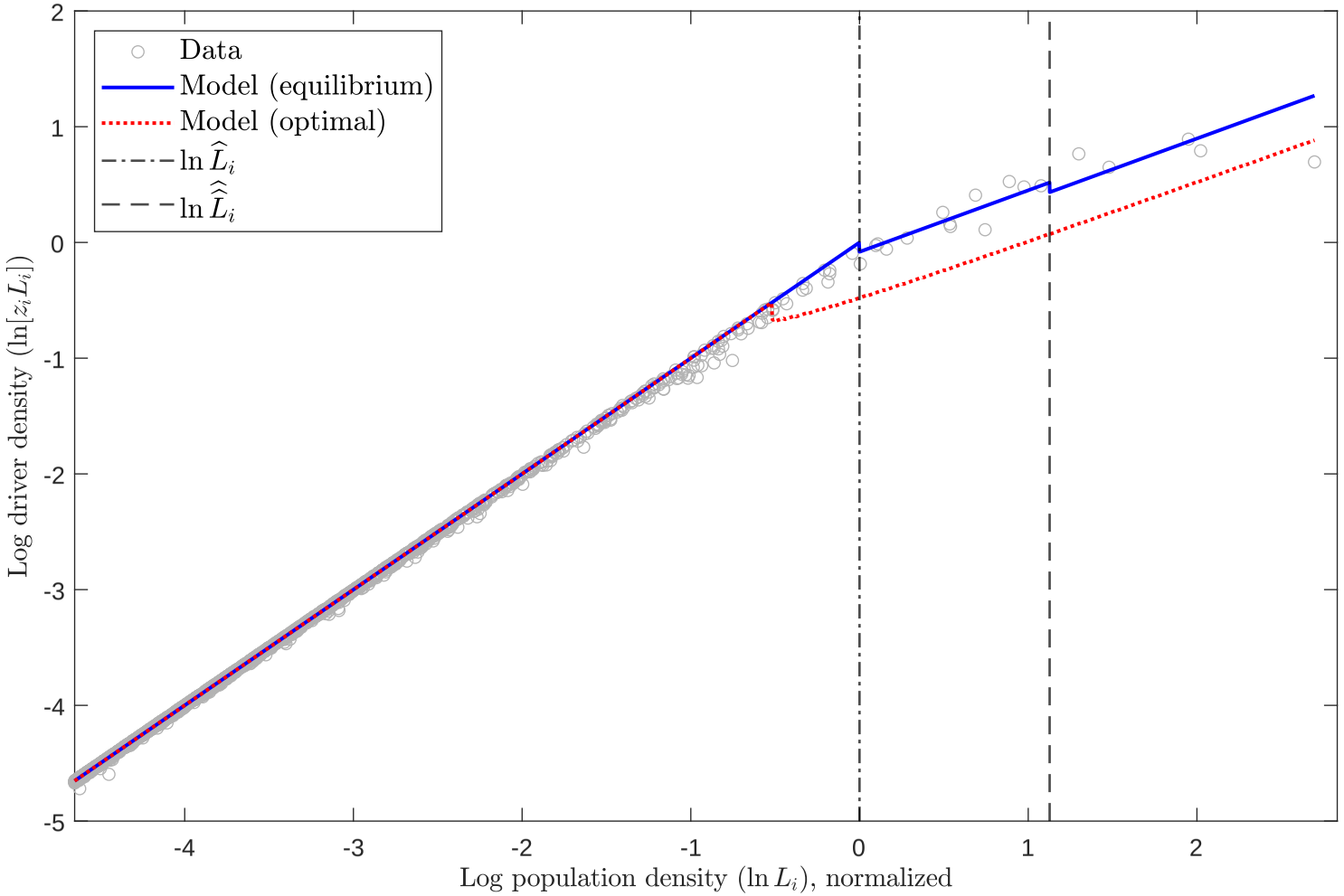


Figure 8: Driver density in model and data, benchmark simulation.

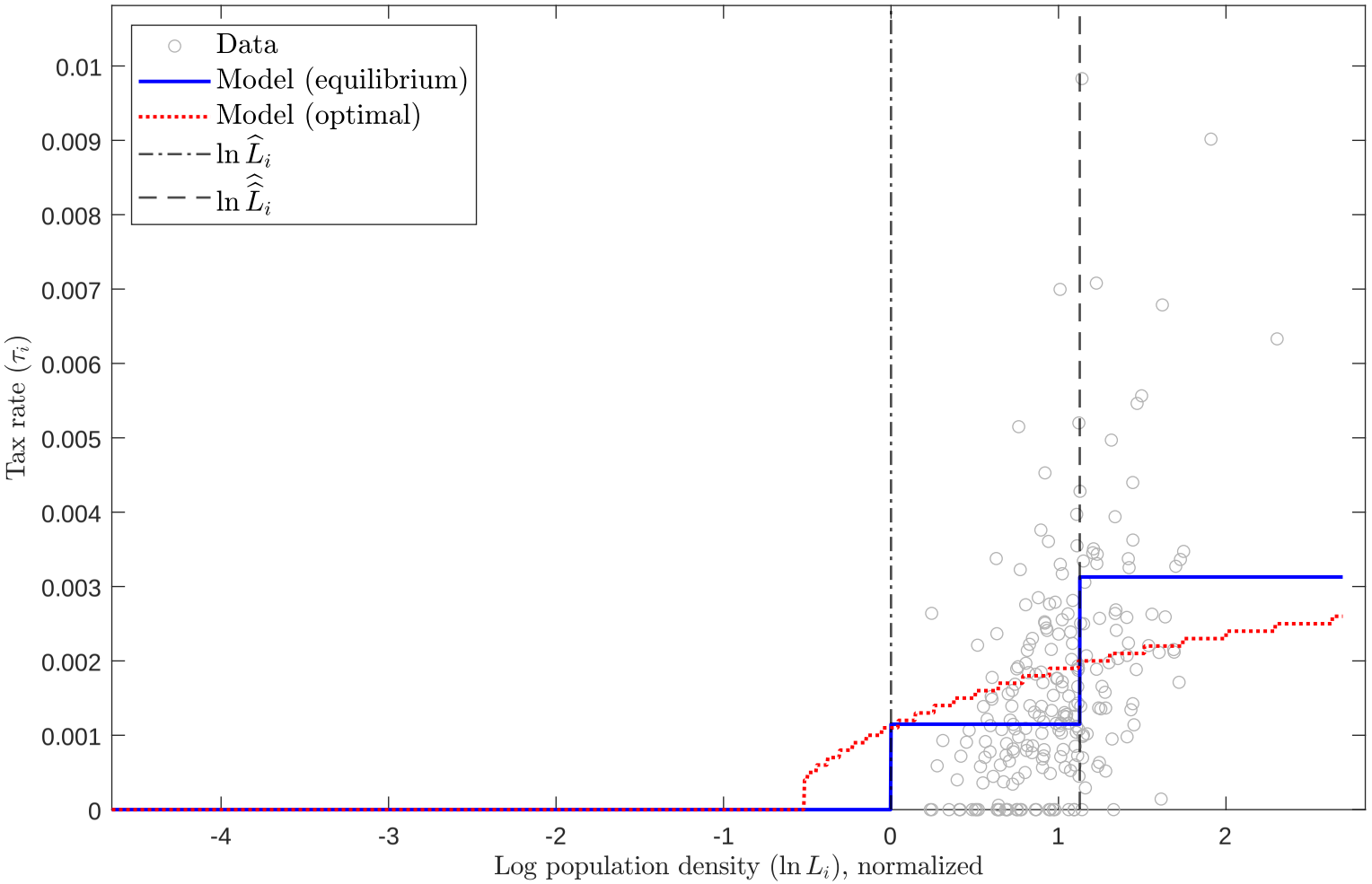


Figure 9: Tax rate in the model and data, benchmark simulation.

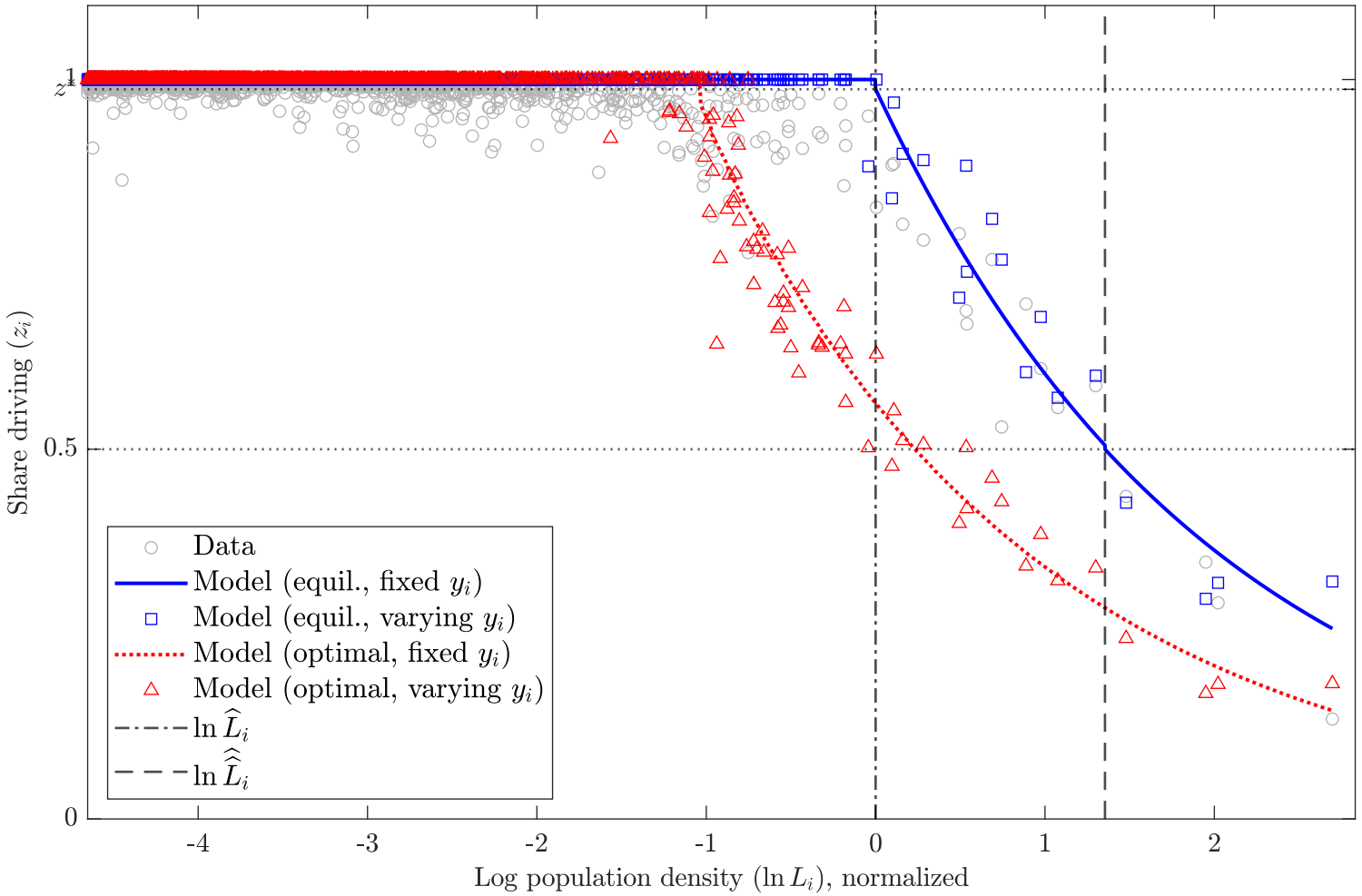


Figure 10: The share drivers in model and data, letting per-capita incomes vary as in the data.

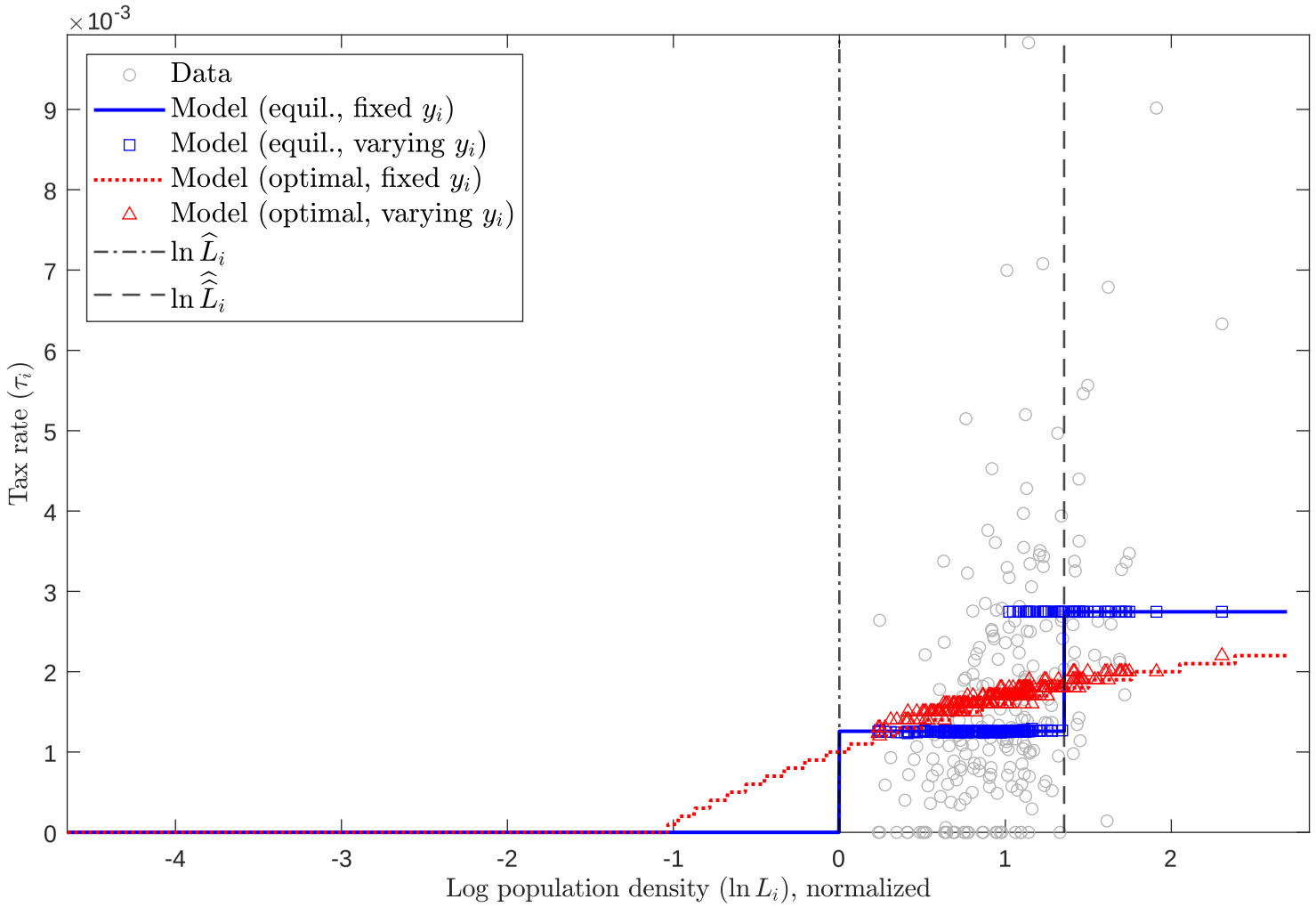


Figure 11: Tax rate in the model, letting per-capita incomes vary as in the data.

A Online Appendix

A.1 Equilibrium

A.1.1 Finding $Z^{\mathbf{eq}}(y_i, \tau_i, L_i)$

Consider an equilibrium in which some drive and some use public transit, so that $z_i < 1$. In this equilibrium there must exist some agent who is indifferent between driving and not driving. Recall that we have denoted that agent's driving preference by $\widetilde{\eta}$. Equalizing the expressions for V_{ij}^D and V_{ij}^{ND} in (7) and (8), with $\eta_j = \widetilde{\eta}$, gives

$$\left[\frac{(1-\tau_i)y_i - D}{(1-\tau_i)y_i}\right]^{\alpha} = \left[\frac{(\Omega\tau_i y_i)^{\lambda} L_i^{1+\lambda} z_i}{\widetilde{\eta}}\right]^{1-\alpha}. \tag{A.1}$$

Recall from (14) that the marginal driver's preference equals $\widetilde{\eta} = z_i^{-\delta}$. Substituting this into (A.1), we can now solve for z_i to get the expression for $Z^{\text{eq}}(y_i, \tau_i, L_i)$ in (23). Thus, if an equilibrium exists where some agents are not driving, then it must be such that $z_i = Z^{\text{eq}}(y_i, \tau_i, L_i)$.

A.1.2 Finding τ_i^*

Next consider τ_i^*. Recall that this is the tax rate in an equilibrium where some use public transit and drivers are in majority, $z_i \in [1/2, 1)$. Let $\widetilde{V}_i^D$ denote the utility of the agent who is indifferent between driving and not driving. Setting $\eta_j = \widetilde{\eta}$ in (7) (where, recall, $\widetilde{\eta}$ is the η_j of the marginal driver) gives

$$\ln\left(\widetilde{V}_i^D\right) = \alpha \ln\left[(1-\tau_i)y_i - D\right] + (1-\alpha)\ln\left(\frac{1}{z_i L_i}\right) + (1-\alpha)\ln(\widetilde{\eta}). \tag{A.2}$$

The utility of any other agent who drives (i.e., with $\eta_j > \widetilde{\eta}$) is also given by (7). Using (14) and (A.2), we can write

$$\begin{aligned}
\ln\left(V_{ij}^D\right) &= \alpha \ln\left[(1-\tau_i)y_i - D\right] + (1-\alpha)\ln\left(\tfrac{1}{z_i L_i}\right) + (1-\alpha)\ln(\eta_j) \\
&= \ln\left(\widetilde{V}_i^D\right) + (1-\alpha)\ln(\eta_j) - (1-\alpha)\ln(\widetilde{\eta}) \\
&= \ln\left(\widetilde{V}_i^D\right) + (1-\alpha)\ln(\eta_j) + (1-\alpha)\delta\ln(z_i) \\
&= \alpha \ln\left[(1-\tau_i)y_i\right] + (1-\alpha)\lambda\ln\left(\Omega\tau_i y_i L_i\right) + (1-\alpha)\ln(\eta_j) + (1-\alpha)\delta\ln(z_i),
\end{aligned} \tag{A.3}$$

where the last equality uses the fact $\widetilde{V}_i^D = V_{ij}^{ND}$, which must hold by the definition of $\widetilde{V}_i^D$, and the expression for V_{ij}^{ND} in (8).

The tax rate in an equilibrium where drivers are in a majority maximizes (A.3), subject to $z_i = Z^{\text{eq}}(y_i, \tau_i, L_i)$. Substituting $z_i = Z^{\text{eq}}(y_i, \tau_i, L_i)$ in (23) into (A.3), and ignoring terms that do not involve τ_i (containing the logs of Ω, y_i, L_i, and η_j) gives this objective function:

$$\begin{aligned} R_i^D = \; & \alpha \ln(1-\tau_i) + (1-\alpha)\lambda \ln(\tau_i) \\ & + (1-\alpha)\delta \left\{ \left[\tfrac{\alpha}{(1-\alpha)(1+\delta)} \right] \ln \left[\tfrac{(1-\tau_i)y_i - D}{1-\tau_i} \right] - \left(\tfrac{\lambda}{1+\delta} \right) \ln(\tau_i) \right\}. \end{aligned} \tag{A.4}$$

Rearranging, and factoring out $1/(1+\delta)$, we can write

$$R_i^D = \frac{1}{1+\delta} \left\{ \alpha \ln(1-\tau_i) + (1-\alpha)\lambda \ln(\tau_i) + \alpha\delta \ln\left([1-\tau_i]\, y_i - D \right) \right\}. \tag{A.5}$$

Maximizing (A.5) with respect to τ_i, the first-order condition can be written

$$-\frac{\alpha}{1-\tau_i} + \frac{(1-\alpha)\lambda}{\tau_i} - \frac{\alpha\delta y_i}{(1-\tau_i)\, y_i - D} = 0. \tag{A.6}$$

The left-hand side of (A.6) is simply the function $G(\tau_i)$ in (15). Thus, the equilibrium tax rate when drivers are in a majority, τ_i^*, is defined from $G(\tau_i^*) \equiv 0$.

A.1.3 Finding τ_i^{**}

Consider next τ_i^{**}, which is the tax rate in an equilibrium where non-drivers are in majority, $z_i < 1/2$. This is found by simply maximizing V_{ij}^{ND} in (8) with respect to τ_i, which gives τ_i^{**} as in (9). It can also be found by setting $\delta = 0$ in (A.6). That is, when heterogeneity is removed all agents are indifferent between driving and not driving, and thus agree on the same tax rate.

A.1.4 Finding $\widehat{L}_i$

Let $V_{ij}^{D,0}$ denote the level of V_{ij}^D associated with zero taxes, and thus no public transit and all agents driving. Setting $\tau_i = 0$ and $z_i = 1$ in (7) gives

$$V_{ij}^{D,0} = \alpha \ln(y_i - D) + (1-\alpha) \ln\left(\frac{\eta_j}{L_i} \right). \tag{A.7}$$

Similarly, let V_{ij}^{D,τ^*} denote the level of V_{ij}^{D} associated with the optimal tax rate when drivers are in majority and some use public transit, i.e., $\tau_i = \tau_i^*$ and $z_i < 1$. From (7), this can be written

$$V_{ij}^{D,\tau^*} = \alpha \ln\left[(1-\tau_i^*)y_i - D\right] + (1-\alpha)\ln\left(\frac{\eta_j}{z_i L_i}\right). \tag{A.8}$$

Setting $V_{ij}^{D,\tau^*} = V_{ij}^{D,0}$ in (A.7) and (A.8) we can solve for what z_i must equal for agents to be indifferent between the two tax rates ($\tau_i = \tau_i^*$ and $\tau_i = 0$, respectively). Denote that level of z_i by z_i^*. Equalizing (A.7) and (A.8), setting $z_i = z_i^*$, gives the expression in (18).

The next task is to find what level of population implements $z_i = z_i^*$. Using (23), with $\tau_i = \tau_i^*$, gives

$$z_i^* = Z^{\text{eq}}(y_i, \tau_i^*, L_i) = \left[\frac{(1-\tau_i^*)y_i - D}{(1-\tau_i^*)y_i}\right]^{\frac{1}{(1-\alpha)(1+\delta)}} \left(\frac{1}{\tau_i^* y_i}\right)^{\frac{\lambda}{1+\delta}} L_i^{-\frac{1+\delta}{1+\lambda}} = B_i^* L_i^{-\frac{1+\delta}{1+\lambda}}, \tag{A.9}$$

where $B_i^* = Z^{\text{eq}}(y_i, \tau_i^*, 1)$ is defined in (17). Solving (A.9) for L_i gives $\widehat{L}_i$ in (16).

A.1.5 Finding $\widehat{\widehat{L}}_i$

Recall that $\widehat{\widehat{L}}_i$ is the population density threshold above which non-drivers become a majority, implementing their preferred tax rate, τ_i^{**}. Thus, $\widehat{\widehat{L}}_i$ is defined from $Z^{\text{eq}}(y_i, \tau_i^{**}, \widehat{\widehat{L}}_i) = 1/2$. Using (23), with $\tau_i = \tau_i^{**}$, gives

$$\widehat{\widehat{L}}_i = \left\{2\left[\frac{(1-\tau_i^{**})y_i - D}{(1-\tau_i^{**})y_i}\right]^{\frac{1}{(1-\alpha)(1+\delta)}} \left(\frac{1}{\tau_i^{**} y_i}\right)^{\frac{\lambda}{1+\delta}}\right\}^{\frac{1+\delta}{1+\lambda}}, \tag{A.10}$$

which can be rewritten as in (19) and (20).

A.2 Social Optimum

A.2.1 Deriving $V_i^{D,\text{tot}}$

Moving factors that do not depend on η_j outside of the integral in (24), and using (25), we can write

$$V_i^{D,\text{tot}} = \left[(1-\tau_i)y_i - D\right]^{\alpha} \left(\frac{1}{z_i L_i}\right)^{1-\alpha} \left(\frac{1}{\delta}\right) \int_{\widetilde{\eta}}^{\infty} \eta_j^{1-\alpha-\frac{1+\delta}{\delta}} d\eta_j. \tag{A.11}$$

Solving the integral in (A.11) gives

$$\begin{aligned}\int_{\widetilde{\eta}}^{\infty} \eta_j^{1-\alpha-\frac{1+\delta}{\delta}} d\eta_j &= \int_{\widetilde{\eta}}^{\infty} \eta_j^{-\left(\alpha+\frac{1}{\delta}\right)} d\eta_j \\ &= \frac{1}{1-\left(\alpha+\frac{1}{\delta}\right)} \left[\lim_{\eta\to\infty} \eta^{1-\left(\alpha+\frac{1}{\delta}\right)} - \widetilde{\eta}^{1-\left(\alpha+\frac{1}{\delta}\right)}\right] \\ &= \left(\frac{1}{\left[\alpha+\frac{1}{\delta}\right]-1}\right) \widetilde{\eta}^{1-\left(\alpha+\frac{1}{\delta}\right)} \\ &= \left(\frac{1}{\left[\alpha+\frac{1}{\delta}\right]-1}\right) z_i^{\delta\left[\left(\alpha+\frac{1}{\delta}\right)-1\right]} \\ &= \left[\frac{\delta}{1-\delta(1-\alpha)}\right] z_i^{1-\delta(1-\alpha)},\end{aligned} \tag{A.12}$$

where the third equality notes that $\alpha + \frac{1}{\delta} > 1$ (since $\delta < 1$), implying that $\lim_{\eta\to\infty} \eta^{1-\left(\alpha+\frac{1}{\delta}\right)} = 0$, and where the fourth equality uses $\widetilde{\eta} = z_i^{-\delta}$ from (14).

Substutiting (A.12) into (A.11), and simplifying, gives

$$V_i^{D,\text{tot}} = [(1-\tau_i)y_i - D]^{\alpha} \left(\frac{1}{L_i}\right)^{1-\alpha} \left[\frac{1}{1-\delta(1-\alpha)}\right] z_i^{\alpha-\delta(1-\alpha)}, \tag{A.13}$$

where we let δ cancel and note that the exponent on z_i equals $1-\delta(1-\alpha)-(1-\alpha) = \alpha-\delta(1-\alpha)$. Rearranging (A.13) gives (26).

A.2.2 Deriving $Z^{\mathbf{opt}}(y_i, \tau_i, L_i)$

The first-order condition for an interior solution with respect to z_i says that

$$\begin{aligned}\frac{\partial W_i}{\partial z_i} = &\left[\frac{\alpha-\delta(1-\alpha)}{1-\delta(1-\alpha)}\right] z_i^{\alpha-\delta(1-\alpha)-1} [(1-\tau_i)y_i - D]^{\alpha} \left(\frac{1}{L_i}\right)^{1-\alpha} \\ &- [(1-\tau_i)y_i]^{\alpha} (\tau_i y_i L_i)^{\lambda(1-\alpha)} = 0.\end{aligned} \tag{A.14}$$

From the first-order conditon in (A.14), it helps to first isolate the only factor involding z_i to write

$$z_i^{-(1-\alpha)(1+\delta)} = \left[\frac{1-\delta(1-\alpha)}{\alpha-\delta(1-\alpha)}\right] \left(\frac{(1-\tau_i)y_i}{(1-\tau_i)y_i - D}\right)^{\alpha} (\tau_i y_i)^{\lambda(1-\alpha)} L_i^{(1+\lambda)(1-\alpha)}, \tag{A.15}$$

where we have also simplified the exponent on z_i. Inverting and solving for z_i gives

$$z_i = \left[\frac{\alpha-\delta(1-\alpha)}{1-\delta(1-\alpha)}\right]^{\frac{1}{(1-\alpha)(1+\delta)}} \left(\frac{(1-\tau_i)y_i - D}{(1-\tau_i)y_i}\right)^{\frac{\alpha}{(1-\alpha)(1+\delta)}} \left(\frac{1}{\tau_i y_i}\right)^{\frac{\lambda}{1+\delta}} L_i^{-\frac{1+\lambda}{1+\delta}}, \tag{A.16}$$

which can be written as in (29) and (30).

A.3 Data

A.3.1 Counties

Population density. Data on population density come from the website 2010 Census Urban and Rural Classification and Urban Area Criteria, hosted by the U.S. Census Bureau. The data used here are in the Excel file "Percent Urban and Rural In 2010 by State and County [$<$ 1.0 MB]," downloadable from the website. To calculate population density, I divide total population (POP_COU) by area (AREA_COU), and then multiply by $1,000,000$ to convert from square meters to square kilometers.

Per-capita income: Per-capita income data are from the Bureau of Economic Analysis and refer to per capita personal income (in dollars) for the year 2010. The data were retrieved by first selecting "County and MSA personal income summary: personal income, population, per capita personal income" under the tab Table. I then make the following selections: County under Major Area; "All counties is the US" under State; "Per capita personal income" under Statistic; Levels under Unit of Measure; and 2010 under Time Period.

The fraction commuters driving: This variable is based on numbers from the American Community Survey, conducted by the U.S. Census Bureau and referring to the year 2010. The table is called B08130–MEANS OF TRANSPORTATION TO WORK BY PLACE OF WORK–STATE AND COUNTY LEVEL, which can be downloaded as a zipped file, and unzipped as a .csv file. I first calculated the total number of commuters as the sum of all people using public transit, those driving alone, and those car pooling. The fraction using public transit is then calculated as the total number of people using public transit over the total number of commuters.

A.3.2 Urban Areas (UZAs) and Metropolitian Statistical Areas (MSAs)

Operating expenses by UZA: Data on operating expenses were retrieved through the National Transit Database (NTD) portal. Here we use an Excel file labelled "2011_UZA_Allocation.xlsx"

(containing information for the census year 2010), which can be found by searching under Data Product Year 2011 and selecting Expenses under Data Categories.

These data refer to operating expenses. To get an estimate of total expenses for public transit, including capital expenditures, we use data from Zhao et al. (2021). Across 471 UZAs in 2014, they report average levels of operating and capital transit expenditures of $88 million and $39.5 million, respectively, implying capital expenditures of about 41% of operating expenditures. Based on this, we estimate total transit expenditures as 1.41 times operating expenditures.

Population by UZA: Data on population density by UZA in 2010 is provided in a crosswalking file called "UZA_CHANGES_1990-2020_2_5.xlsx," available here for downloading. The file can also be found by entering "UZA CHANGES 1990 2020" (or similar terms) in the search field at the Federal Transit Administration website.

Population density is reported in the column labelled "2010 Density (Population per square mile)." To get density in units per square kilometer we multiply by 2.59. Total population is in the column "2010 Population."

These population data can be matched to the file with operating expenses using the UZA code in the column labelled "2010 NTD UZA Code (Deprecated)."

Incomes by MSA: We do not know per-capita or total income by UZA, so we instead use per-capita income numbers by MSA, which we then merge by name to UZAs, and use total population to infer total income.

We first turn to the Bureau of Economic Analysis, which we used for the county level data above, specifically the BEA's Regional Economic Accounts. Under Choose a category, we select "Personal income (state and local)," and then MARPP under Table. This downloads a zipped file, which contains an Excel file labelled "MARPP_MSA_2008_2023.xlsx," with per-capita income data by MSA (possibly with a different file name, depending on the time of download). From this file we select the column for year 2010, and the variable " Real per capita personal income (constant (2017) dollars)" (keeping rows with Linecode=2).

Price adjustments: We convert the numbers for personal income to 2010 (current) prices by dividing by 1.1048, implying a 10.48% increase in prices between 2010 and 2017. This is consistent with the change in the Implicit regional price deflator from the BEA,

referring to the whole United States: the ratio of the index values for 2017 and 2010 comes to ($100/90.514 \approx 1.1048$).

These price data are available by first going to Regional Price Parities by State and Metro Area at the BEA, and selecting "Interactive Data" and "Interactive Tables: Regional Price Parities," which leads to a new page titled "Regional Data GDP and Personal Income." Select "Summary table for GDP, personal income, and related data," and then "SASUMMARY State annual summary statistics: personal income, GDP, consumer spending, price indexes, and employment." Then select United States under AREA, "Regional price parities (RPPs)" under Statistic, and Levels under Unit of Measure. Then click Next Step and All Years under Time Period, which opens the data in the browser, and allows for subsequent download.

We merge the MSA income data with the UZA expenditure and population data by the name of the UZA or MSA. This results in 224 matches, out of a total of 481 UZAs with population and expenditure data.

Total income of a UZA is calculated as its population times the per-capita personal income in the matched MSA. The transit tax rate is then defined as total expenditures (which, recall, is computed as 1.41 time operating expenditures) divided by total income.

Online Appendix Tables and Figures

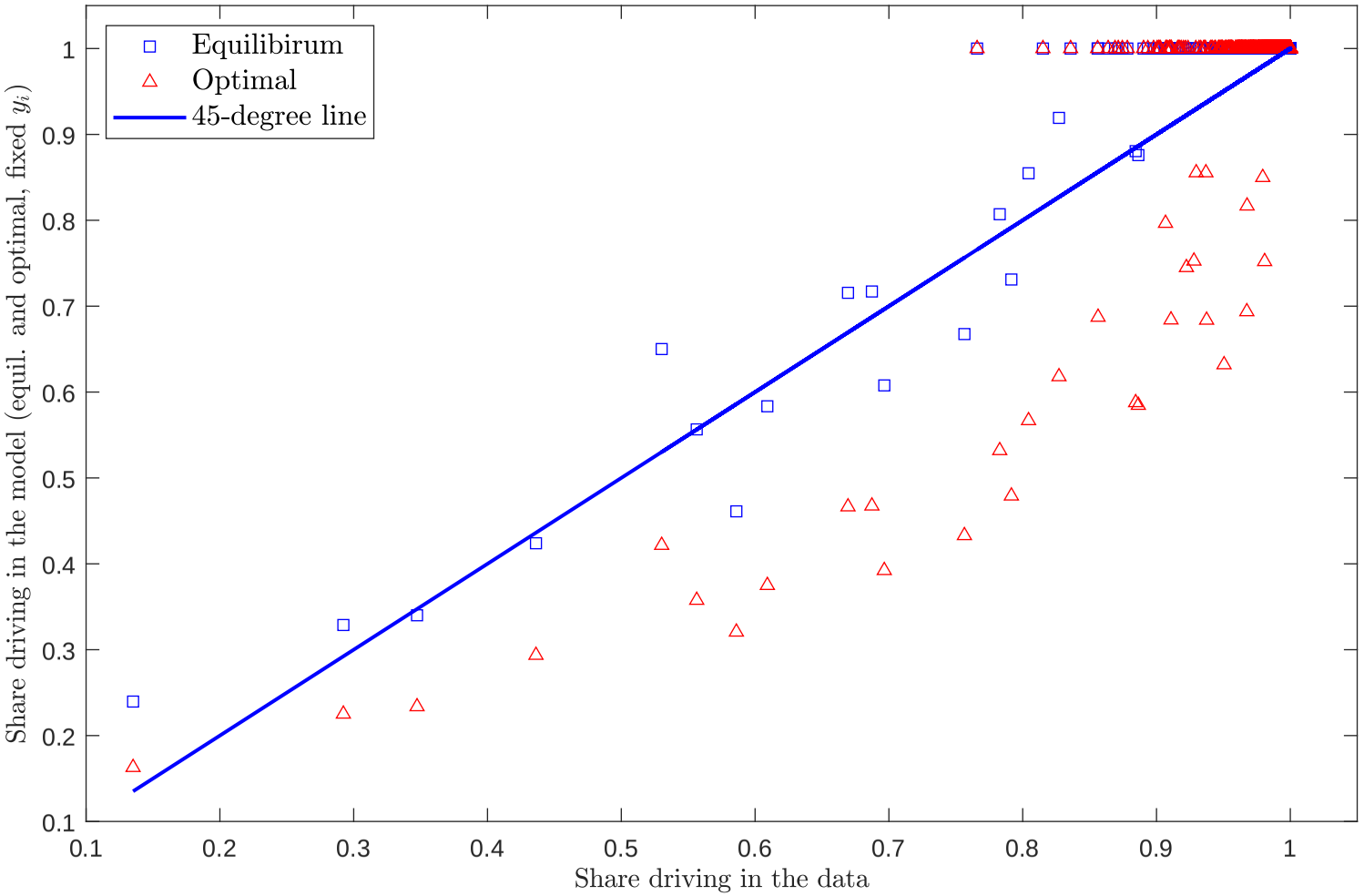


Figure A.1: Comparing the share driving in model and data, benchmark simulation.

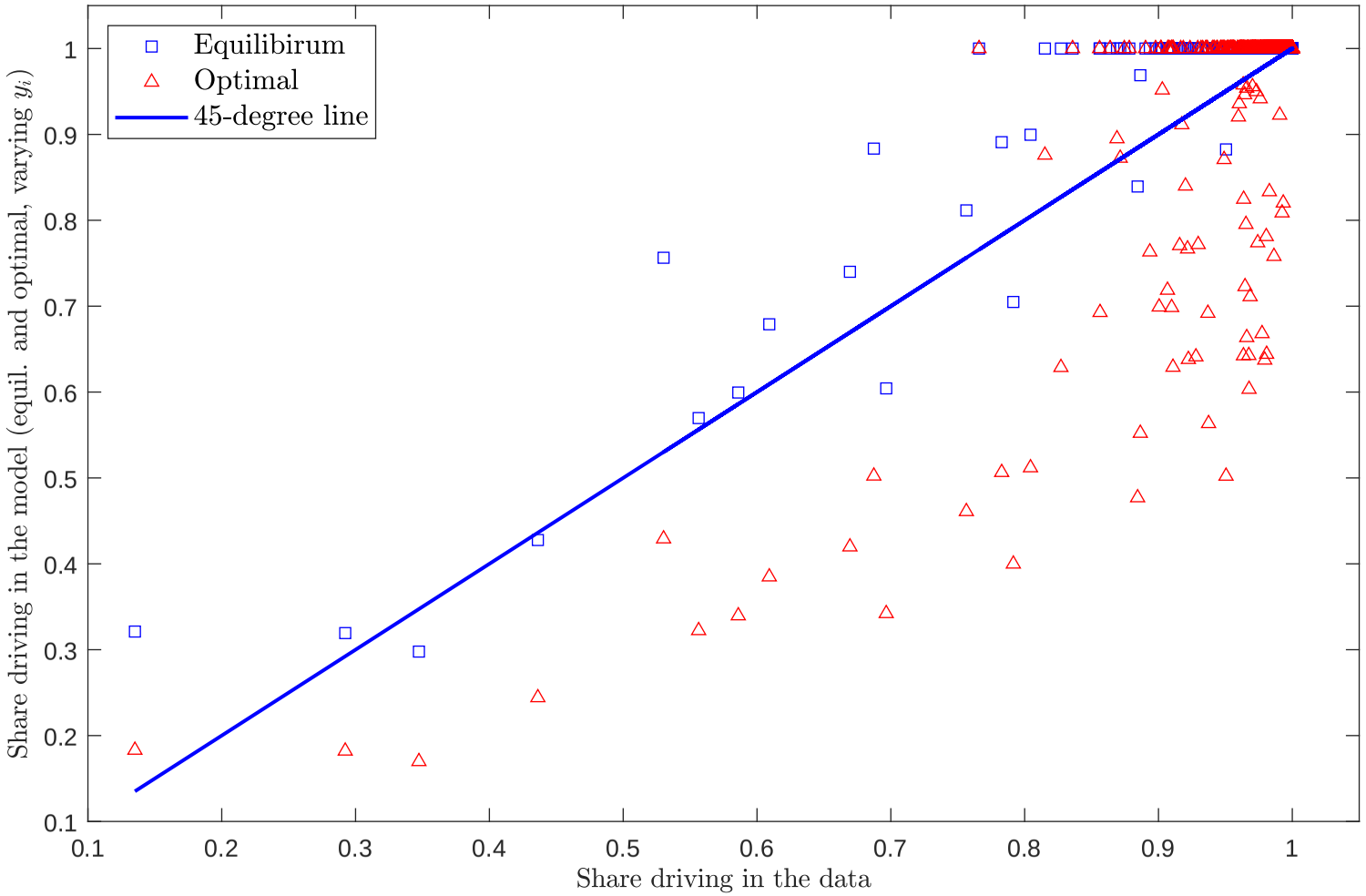


Figure A.2: Comparing the share driving in model and data, letting per-capita incomes vary as in the data.