THE UNIVERSITY OF CHICAGO

OPTIMAL CONGESTION PRICING IN GENERAL EQUILIBRIUM

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
THOMAS RAYMOND HIERONS

CHICAGO, ILLINOIS
JUNE 2025

Contents

List of Figures

List of Tables

Acknowledgements

I am deeply grateful to my advisors — Esteban Rossi-Hansberg, Rodrigo Adão, and Alex Torgovitsky — for their support and guidance. They shaped the project in fundamental ways, helping it become more ambitious and complete than I initially thought possible. I have learned a great deal from each of them and hope to continue to do so.

The project benefitted greatly from comments and discussion with Milena Almagro, Scott Behmer, Stéphane Bonhomme, Olivia Bordeu, Juanma Castro-Vincenzi, Jonathan Dingel, Ed Jee, Hugo Lopez, Jordan Rosenthal-Kay, Jeanne Sorin, Thomas Wiemann, and numerous other seminar participants.

The impact of fellow graduate students and faculty on me throughout my time during the PhD has led to lasting friendships and intellectual growth that are reflected throughout this thesis.

Finally, I thank my wife, Elif, for her love and support through all of it. No matter what.

Abstract

Road traffic leads to an externality: drivers do not account for the time cost they impose on others. I study optimal congestion pricing in an urban general equilibrium model in which agents choose residential and workplace locations, travel modes, and route choices with congestion. The attractiveness of workplaces and residences is also determined endogenously. I provide conditions for the uniqueness of both the competitive equilibrium and the first best planner's problem and characterize the tax instruments needed to decentralize it. I show how the model can be solved with arbitrary other taxes, including congestion toll zones. I apply these theoretical results to New York City and find that the first best tax policy would realize gains of \$0.77 per person per day with substitution between driving and public transit as a key margin of adjustment. 35% of the gains from optimal congestion pricing at the link level can be achieved by a congestion zone that covers only lower Manhattan.

1 Introduction

The average travel speed by car in Midtown Manhattan in 2023 was 5 miles per hour. Congestion affects cities globally, generating an externality as drivers do not account for the time cost they impose on others. The textbook solution to correct this externality is to impose taxes that equalize the marginal social and private cost of using every road in a city. How much do we stand to gain from such a policy? While theoretically optimal, real world congestion pricing departs from the Pigouvian ideal. Cities such as London, Singapore, Milan, and Stockholm have implemented toll zones or gates. How close can these types of policy instruments take us to the first best solution? Travel times by car in turn affect people's other choices: the route they choose to commute, whether they drive or use public transit, where they live and work. How do these margins of adjustment mediate the effects of congestion pricing?

Correcting externalities from traffic congestion is a core policy problem in urban economics but solving it in general equilibrium has remained challenging. I address this problem in an urban general equilibrium model and provide conditions for the uniqueness of both the competitive equilibrium and the first best (utilitarian) planner's problem as well as characterizing the tax instruments needed to decentralize the optimal allocation. I also show how the model can be solved with arbitrary taxes of the kind often implemented in practice, such as toll zones, so that policy relevant pricing policies can be evaluated. I develop an algorithm to solve each of these problems so that equilibrium and optimal allocations can be computed and used to calculate welfare. I apply these theoretical results to New York City to assess the welfare gains from optimal congestion pricing in practice.

The environment I study is a static urban general equilibrium model which features three levels of choices by agents: routes, modes, and locations. Roads are represented by a network over locations and, when driving, people choose a full path through the network from origin to destination. These choices determine the flow of cars through each link in the network and the resulting equilibrium travel speeds. People also choose a travel mode, accounting for the equilibrium travel times on raods as well as on other modes. Finally, agents select their residence and work location, where these choices are equilibrium outcomes jointly determined with travel times. As in many urban models, the attractiveness of a location as a home or workplace depends endogenously upon the number of people living or working there. The supply of floor-space to firms or residents provides a key dispersion force: with the supply of land fixed, floor-space becomes increasingly costly to provide. On the other hand, agglomeration economies from knowledge spillovers, labor market pooling, or input sharing may increase firm productivity in high density areas.

The complex interactions between the general equilibrium and transport parts of the model pose a challenge both theoretically and computationally. Pigouvian taxes that equate marginal private and social costs on each link are simple to state, but evaluating them in actual cities requires knowledge of the first best flows on the entire network and so requires solving the full planner's problem. I use tools from convex optimization both to establish the uniqueness of the solution and to efficiently compute it by reducing its dimension using a novel algorithm.

Turning to the competitive equilibrium problem, the presence of externalities leads to a failure of the welfare theorems. I follow an approach first suggested in Samuelson (1947), in which I characterize the equilibrium through an optimization problem: a pseudo-planner's problem. This allows the equilibrium to be handled using the same mathematical techniques as the planner's problem.

Finally, moving beyond Pigouvian taxes to consider second best policies, the problem becomes more difficult. It is no longer sufficient to offset the externality on each road locally. Instead, optimal use of *ad-hoc* policies must account for the spillovers to traffic on all other roads in the network through the full equilibrium interactions of the model. Here too, casting the equilibrium as an optimization problem is key. It allows the model to be solved with arbitrary taxes so that second best policies can be evaluated numerically. In general, this problem is computationally challenging, but in the case of a flat tax when entering a toll zone it remains tractable to solve.

The main theoretical results of the paper are summarized in three propositions. The first characterizes the planner's problem, proving existence and providing sufficient conditions for the uniqueness of the solution. The key to establishing uniqueness is to show that the problem takes the form of a strictly concave optimization problem over a convex set. The economic interpretation of these sufficient conditions follows the logic of other results in the literature: the dispersion forces in the model must be at least as large as any agglomeration forces. The second proposition establishes that the set of competitive equilibria of the model is equal to the set of turning points of the Lagrangean of a pseudo-planner's problem. This applies both to the case where the equilibrium is unique and when there is multiplicity. While the planner uses the marginal social cost in deciding how many cars to send along a particular link, private individuals will instead simply use the time it takes to cross the link (corresponding to the average cost). This wedge between marginal and average cost is key in mapping competitive equilibria to turning points of the Lagrangean of the pseudo-planner's problem. The third proposition establishes existence and gives sufficient conditions for uniqueness of the competitive equilibrium. Given the equivalence established in the second proposition, the proof proceeds entirely analogously to the first by establishing strict concavity of the problem.

Building on these three key results, I show how taxes can be incorporated, how the solution to each problem can be computed, and how different margins of adjustment can be decomposed. In a first corollary, I show that given a fixed but arbitrary set of taxes, the model has a unique solution under the appropriate restrictions on parameters. This result follows almost immediately from propositions two and three: once the equilibrium is recognized as the solution to an optimization problem, the addition of taxes poses no problem as it leaves the concavity of the objective function unchanged. A second corollary provides expressions for the taxes which decentralize the first best allocation. These correspond to the intuitive Pigouvian logic of offsetting each externality where it is produced. While the expressions for the taxes are simple, their levels depend on the optimal flows so that computing them in real world cities requires solving a full planner's problem.

The practical computation of the allocations in the planner, competitive equilibrium, and taxed problems is simplified by the fact that they all share the same mathematical structure. This means that a single algorithm can be used to solve all three. I use tools from convex optimization, in particular duality theory, and analytical simplifications of the problem, to solve a lower-dimensional, unconstrained problem. Finally, I show that it is possible to solve the model fixing some of the agents' decisions. In particular, it is possible to solve for transportation choices, fixing the location decisions of individuals, or to solve only the routing problem over the network, fixing both location and mode choice. This is important in counterfactuals, to quantify the contribution of each margin in response to policy changes, and in estimation, where it will be used as a key step in the identification of the parameters.

I then turn to an empirical application in which I use my theoretical results to study optimal congestion pricing in New York City. This provides an ideal setting to illustrate the framework for a number of reasons. Firstly, it provides a real world example in which congestion pricing has been proposed as a policy tool and is, at the time of writing, under review. Secondly, New York City has a highly developed public transit system, making the substitution pattern between transit and driving of first order importance. Thirdly, road traffic, especially in Manhattan, regularly ranks as among the worst in the US suggesting the costs from congestion are potentially high. Finally, New York City is a data rich environment, especially in terms of traffic flows. The Department of Transportation provides repeated measurement of vehicular counts on over 2000 roads throughout the city at different times of the day, which is unique among large US cities.

The estimation of the model's parameters proceeds in three steps. Firstly, I fix some parameters using estimates from the literature, most importantly the value of time, the elasticity of substitution between modes, and floor-space supply elasticities. Secondly, I use variation across hours of the day in both speed (from Google Maps API) and the flow of vehicles (from the Department of Transportation) to estimate the congestion technology using a flexible parametric form. Finally, I use data on commuting patterns between locations with a maximum likelihood estimation approach to identify the attractiveness of residential and workplace locations.

Using the fitted model parameters I undertake counterfactuals to assess the potential welfare gains from different congestion pricing policies in New York City. I find that first best taxes yield a gain of \$0.77 per person per weekday in 2019 dollars or a total of \$21.7 million per week. 12% of these gains are realized when drivers are only allowed to change their route through the network. 90% are achieved when agents are allowed to change both their transport mode and driving route. The final 10% require location changes: individuals moving their home and workplace location in response to congestion pricing. These results highlight the importance of substitution across modes, with a non-negligible role for drivers re-routing and general equilibrium effects. Finally, I assess the optimal level of a toll covering only lower Manhattan. The model implies this should be set at \$13.37 in 2019 dollars. Inflating this to 2024 dollars this gives a figure of \$16.38 which is similar in magnitude to the \$15 proposal for Manhattan rejected in June of 2024[1]. The toll yields 35% of the potential first best benefits. This suggests that substantial gains can be achieved even by relatively simple pricing policies, but there remains scope for what could be achieved by more targeted policies.

[1]The proposal was put forward in the Traffic Mobility Review Board (2023), it has subsequently been changed to a \$9 toll.

The remainder of the article is set out as follows. Section 2 discusses how the paper relates to the literature. Section 3 presents the model and key theoretical results. Section 4 shows how the parameters of the model are estimated in the empirical application to New York City. Section 5 undertakes counterfactuals using the fitted model. Section 6 concludes.

2 Related Literature

A growing literature has studied the economic effects of transportation policies (Almagro et al. (2024); Durrmeyer and Martinez (2022); Barwick et al. (2024); Kreindler (2024)). I build on this literature by retaining two key features of these models: endogenous congestion and multiple transport modes. These papers also feature rich individual level heterogeneity, both in terms of observable characteristics such as income, as well as in preferences. I abstract from many of these differences to contribute along two dimensions. Firstly, I allow for a rich set of route choices for drivers through the road network. This makes the impact of toll zones or road taxes on commuters' behavior unclear *ex-ante*: will the increased cost of traveling through a particular area induce people to change routes to avoid them or will they switch travel modes entirely? The model I develop allows these margins to be isolated separately. Secondly, the model I present features full general equilibrium interactions between agents so that location and residential decisions are determined jointly with congestion in commuting.

The inclusion of general equilibrium location decisions is motivated by a large empirical literature which finds substantial effects of transportation costs on the spatial distribution of economic activity within cities. These papers have investigated the impact of access to public transportation (Gibbons and Machin (2005); Billings (2011); Gonzalez-Navarro and Turner (2018); Tsivanidis (2023)) as well as road infrastructure (Baum-Snow (2007); Baum-Snow et al. (2017); Brinkman and Lin (2024)) on economic outcomes such as local population and housing prices. The model I develop incorporates these general equilibrium forces into the design of optimal congestion pricing policies.

A third strand of the literature to which I contribute analyzes optimal policy in models featuring rich general equilibrium interactions and endogenous congestion. Three recent papers in this group are Allen and Arkolakis (2022), Bordeu (2023), and Fajgelbaum and Schaal (2020). All are concerned with the effects of infrastructure investment whereas the present paper considers how to efficiently use existing infrastructure through congestion pricing. I build on the route choice framework for commuting with endogenous congestion from Allen and Arkolakis (2022) to incorporate multiple modes of transportation. I then use the framework for a different purpose: to study optimal congestion pricing policy in general equilibrium as well as second best tax policies. This requires solving both a planner's problem and the equilibrium problem with arbitrary taxes imposed, which are not studied in Allen and Arkolakis (2022).

Perhaps the most closely related paper is Fajgelbaum and Schaal (2020). They study optimal infrastructure investment in a network determining transportation in a neoclassical trade model. Their model is focused on the transportation of goods, whereas the model in the present paper is tailored towards commuting and so makes several different assumptions. My paper features multiple modes of transport as well as heterogeneity in consumers' preferences over where to live and work, their transport mode and route which are absent from Fajgelbaum and Schaal (2020) but a key feature of recent quantitative spatial models. Similarly to Fajgelbaum and Schaal (2020), I use techniques from convex optimization to characterize the solution to the model. However, an important distinction is that I am able to establish the global strict concavity of my problem, not only for the social planner (or with optimal taxes), but also in the equilibrium case without taxes or with arbitrary (non-Pigouvian) taxes[2]. This allows me to provide guarantees on the existence and uniqueness

[2]At at technical level, there are also differences in the way the problem is formulated. I use a route level formulation whereas Fajgelbaum and Schaal (2020) use flows along links. This allows me to avoid imposing flow conservation constraints which are automatically fulfilled by the flow decomposition theorem

of the equilibrium as well an algorithm to compute it which are necessary to quantify the potential gains from different congestion pricing policies.

The final literature I draw from is one which casts the equilibrium of models with externalities as the solution to an optimization problem. This is the key analytical device that I use throughout the paper to establish results about the competitive equilibrium with or without taxes. The welfare theorems show the connection between equilibria and Pareto optima which solve a planner's optimization problem. However, when these theorems fail, for example due to externalities, a pseudo-planner's problem may still provide a characterization of the equilibrium problem. This idea, which dates back to Samuelson (1947), has been used extensively within transportation research, beginning with the seminal work of Beckmann et al. (1956). Casting transport equilibria as the solution to optimization problems has become the dominant approach in a large literature synthesized in Sheffi (1985) and which is still active (Akamatsu (1996, 1997); Baillon and Cominetti (2008); Oyama et al. (2022)). This approach has also been widely used in macroeconomics to study models with taxes or externalities. Kehoe et al. (1992) provide examples of how this can be done as well as many references to the literature. The advantage of the approach is twofold. Firstly, it allows us to study existence and uniqueness properties of the models using properties, in particular concavity or convexity, of the associated optimization problem. Secondly, when formulated as an optimization problem, finding algorithms to compute the solution of the problem can draw from the rich literature in optimization theory. While the transportation literature models trip demands as fixed or determined in partial equilibrium, the current paper applies this approach in a novel setting to handle a general equilibrium urban model with multiple modes, route, and location choice solving planner, decentralized and taxed problems.

(see, for example, chapter 3 of Ahuja et al. (1988)). This is useful computationally. Moreover, by imposing a parametric form on the shape of agglomeration economies I am able to state sharp sufficient conditions on parameters for uniqueness in these cases.

3 Model

3.1 Consumers and Firms

Space is discrete and there is a finite set $\mathcal{N} = \{1, \ldots, N\}$ of locations. Of these, a set $\mathcal{O} \subseteq \mathcal{N}$ are home locations and $\mathcal{D} \subseteq \mathcal{N}$ are workplace locations. Note that these sets may overlap, be disjoint, or comprise of all locations. I consider a closed city equilibrium with a continuum of consumers of measure L. Consumers choose a home location $o \in \mathcal{O}$ and a work location $d \in \mathcal{D}$. They also choose how to commute between o and d. They choose a mode $m \in \mathcal{M}$ and a route $r \in \mathcal{R}_{odm}$. Consumers get utility at their home location from residential attractiveness, u_o, which they take as given but will be endogenous. They also receive a wage w_d which depends on their workplace as well as other sources of income b which do not depend on their location choices. They use their total income to consume a single freely traded consumption good which is also the numéraire. Associated with each mode $m \in \mathcal{M}$ and route $r \in \mathcal{R}_{odm}$ is a travel time t_{odmr} and consumers have a constant value of time γ which translates the time cost into units of the consumption good. The disutility from travel comes directly in the form of a utility cost of time commuting. This travel time will be the endogenous result of other agents' travel decisions because of congestion. Finally, consumers have idiosyncratic preferences ε_{odmr} over home-workplace-mode-route tuples. The consumer's problem is given by

$$
\begin{aligned}
&\max_{c \in \mathbb{R}_+, o \in \mathcal{O}, d \in \mathcal{D}, m \in \mathcal{M}, r \in \mathcal{R}_{odm}} u_o + c - \gamma t_{odmr} + \varepsilon_{odmr} \\
&\text{s.t.} \\
&c \leq w_d + b \,.
\end{aligned}
$$

I assume that the distribution of idiosyncratic preferences in the population follows a nested logit distribution with the lower nest representing choices over routes $r \in \mathcal{R}_{odm}$

conditional on location and mode choices, the second nest representing choices over modes $m \in \mathcal{M}$ and the upper nest representing the choice over home-workplace pairs $od \in \mathcal{O} \times \mathcal{D}$. For clarity I define the index set, $\mathcal{S}$, of the realizations of the shock term $\bar{\varepsilon} = (\bar{\varepsilon}_{odmr})_{odmr \in \mathcal{S}}$ as

$$\mathcal{S} := \{(o,d,m,r) : o \in \mathcal{O}, d \in \mathcal{D}, m \in \mathcal{M}, r \in \mathcal{R}_{odm}\} .$$

A realization $\bar{\varepsilon}$ takes values in $\mathbb{R}^{\mathcal{S}}$. The random vector itself will be denoted ε. Formally, the joint cumulative distribution function of ε is given by

$$\mathbb{P}\left(\varepsilon_{odmr} \leq \bar{\varepsilon}_{odmr}\ \forall odmr \in \mathcal{S}\right) = \exp\left(-\sum_{od \in \mathcal{O}\times\mathcal{D}}\left[\sum_{m \in \mathcal{M}}\left(\sum_{r \in \mathcal{R}_{odm}} \exp\left(-\bar{\varepsilon}_{odmr}/\sigma\right)\right)^{\frac{\sigma}{\nu}}\right]^{\frac{\nu}{\theta}}\right) .$$

Note that, as usual, this is a function of $\bar{\varepsilon}$. It gives the share of agents with a realization of the random vector ε such that each element is less than or equal to the corresponding element of $\bar{\varepsilon}$. Since the total mass of agents in the population is L, the mass of agents with realisation of $\varepsilon_{odmr} \leq \bar{\varepsilon}_{odmr}$ for all $odmr \in \mathcal{S}$ is then given by $L\mathbb{P}\left(\varepsilon_{odmr} \leq \bar{\varepsilon}_{odmr}\ \forall odmr \in \mathcal{S}\right)$. I denote the density of $\mathbb{P}$ with respect to the Lebesgue measure by f_ε so that $\int_{\mathbb{R}^{\mathcal{S}}} f_\varepsilon(\bar{\varepsilon}) d\bar{\varepsilon} = 1$. It is important to note that this is a deterministic model: each agent knows their own value of ε and it is a feature of their preferences. The distributions above describe how these vary over the mass L.

σ is the scale parameter governing the lower nest choice over routes, ν the scale parameter over modes and θ over locations. I assume throughout that $\theta > \nu > \sigma$ to that the model is consistent with utility maximization and belongs to the family of Generalized Extreme Value models[3]. This implies that it is easier for individuals to substitute across different

[3]See Train (2009) or Anderson et al. (1992) for details.

commuting routes than it is for them to substitute across modes which is in turn easier than changing workplace and home locations. Let ℓ_{odmr} denote the mass of consumers choosing to live in workplace $o \in \mathcal{O}$, work in $d \in \mathcal{D}$ and commute choosing mode $m \in \mathcal{M}$ along route $r \in \mathcal{R}_{od}$. Substituting the budget constraint into the consumer's problem and making use of the distributional assumption on preferences implies that the population shares will satisfy the usual formula for nested logit choice probabilities. The deterministic utility for a particular choice of $odmr$ is given by

$$v_{odmr} = b + u_o + w_d - \gamma t_{odmr} \qquad \forall odmr \in \mathcal{S}\,. \tag{1}$$

Following the literature on discrete choice, I now also define the inclusive values[4] for each origin, destination, mode. Aggregating over routes this gives

$$v_{odm} = \sigma \ln \sum_{r \in \mathcal{R}_{odm}} \exp(v_{odmr}/\sigma) \qquad \forall odm \in \mathcal{O} \times \mathcal{D} \times \mathcal{M}\,. \tag{2}$$

Similarly, aggregating over modes, the inclusive value for an origin-destination pair is given by

$$v_{od} = \nu \ln \sum_{m \in \mathcal{M}} \exp(v_{odm}/\nu) \qquad \forall od \in \mathcal{O} \times \mathcal{D}\,. \tag{3}$$

Using standard results in discrete choice, the final choice shares are given by

$$\frac{\ell_{odmr}}{L} = \frac{\exp(v_{od}/\theta)}{\sum_{o'd' \in \mathcal{O} \times \mathcal{D}} \exp(v_{o'd'}/\theta)} \frac{\exp(v_{odm}/\nu)}{\sum_{m' \in \mathcal{M}} \exp(v_{odm'}/\nu)} \frac{\exp(v_{odmr}/\sigma)}{\sum_{r' \in \mathcal{R}_{odm}} \exp(v_{odmr'}/\sigma)} \qquad \forall odmr \in \mathcal{S}. \tag{4}$$

ℓ_{odmr} denotes the mass of agents making the choice $odmr$. The first term in (4) represents

[4] This defines the expected utility over routes for an agent. See, for example, Train (2009) for further discussion.

the unconditional choice probability for a particular home-workplace pair, od. The second term represents the choice probability of choosing mode $m \in \mathcal{M}$ conditional on location choices. The third terms gives the conditional probability of route $r \in \mathcal{R}_{odm}$ given both mode and location choices. Workers take wages, w_d, residential attractiveness, u_o, and travel times, t_{odmr}, as given and their choices are fully summarized by equations (1), (2), (3), and (4).

In order to keep track of the employees at each work location and residents in each home location I define the following aggregate variables for each $o \in \mathcal{O}$ and each $d \in \mathcal{D}$:

$$\ell_o^H = \sum_{m \in \mathcal{M}} \sum_{dr \in \mathcal{D} \times \mathcal{R}_{odm}} \ell_{odmr} \qquad \forall o \in \mathcal{O}\,, \tag{5}$$

$$\ell_d^F = \sum_{m \in \mathcal{M}} \sum_{or \in \mathcal{O} \times \mathcal{R}_{odm}} \ell_{odmr} \qquad \forall d \in \mathcal{D}\,. \tag{6}$$

In each work location d, perfectly competitive fims use labor as the only input to produce the freely traded consumption good[5]. Each location has an endogenous productivity level A_d which firms take as given. The total output of firms in location d is given by

$$y_d = A_d \ell_d^F \qquad \forall d \in \mathcal{D}\,. \tag{7}$$

The free entry condition implies that wages are given by

$$w_d = A_d \qquad \forall d \in \mathcal{D}\,. \tag{8}$$

In each workplace productivities emerge endogenously from the number of workers in that location and are given by

[5]Appendix B presents a model in which firms use both labor and land in a constant returns technology to model the case where all scale effects are internalized through market interactions. The formal equations from that model are isomorphic to those presented in the main text.

$$A_d = \bar{A}_d + \alpha_d \ln \ell_d^F \qquad\qquad \forall d \in \mathcal{D} \, . \tag{9}$$

where α_d governs the semi-elasticity of productivity with respect to the number of workers.

In a home location, o, residential attractiveness results from the number of residents in that location

$$u_o = \bar{u}_o + \beta_o \ln \ell_o^H \qquad\qquad \forall o \in \mathcal{O} \, , \tag{10}$$

Whether the endogenous residential attractiveness and productivities are internalized by agents through market interactions or are fully external to their decisions will be key in comparing the equilibrium to the first best planner's problem and this will depend upon the particular micro-foundations that are used. Duranton and Puga (2004) provide a variety of micro-foundations which lead to externalities from agglomeration through matching, sharing and learning. I provide a fully internalized micro-foundation in Appendix B based on the supply of floor-space to households and firms. In this case, rents to land are rebated uniformly to consumers leading to non-wage income. In order to keep track of whether the model's general equilibrium relationships are internalized or not I introduce the following notation:

$$\mathbb{1}_{\mathrm{I}} = \begin{cases} 1 & \text{if } \alpha_d, \beta_o \text{ are fully internalized} \\ 0 & \text{if } \alpha_d, \beta_o \text{ are fully external} \end{cases} \, .$$

This leads to the following reduced form relationship for non-wage income:

$$b = \mathbb{1}_{\mathrm{I}} \left[\frac{-\sum_{d \in \mathcal{D}} \ell_d^F \alpha_d - \sum_{o \in \mathcal{O}} \ell_o^H \beta_o}{L} \right] \, . \tag{11}$$

In the case of the rental market for floorspace microfoundation $-\sum_{d \in \mathcal{D}} \ell_d^F \alpha_d - \sum_{o \in \mathcal{O}} \ell_o^H \beta_o$ corresponds exactly to the total rents that accrue to land which are divided evenly among the population of size L.

3.2 Transport

When making transportation decisions, consumers choose both a mode and a route. I assume that the set of modes is finite and given by a set $\mathcal{M}$. I assume that some of the modes are uncongested and have a fixed route and denote these by $\mathcal{M}_0 \subseteq \mathcal{M}$. Other modes require route choice and are subject to congestion, denoted $\mathcal{M}_1$. $\mathcal{M}_0$ and $\mathcal{M}_1$ partition $\mathcal{M}$ and in the empirical application $\mathcal{M}_0 = \{m_0\}$ will denote public transit and $\mathcal{M}_1 = \{m_1\}$ will denote driving. Given a particular mode $m \in \mathcal{M}_1$ where route choice is required, routes are described by a directed network over the set of locations $(\mathcal{N}, \mathcal{E}_m)$ where $\mathcal{N}$ is the same set of locations as in the previous section and $\mathcal{E}_m \subseteq \mathcal{N} \times \mathcal{N}$ is a set of directed links describing the transport connections between locations. I assume throughout that for each mode in $\mathcal{M}_1$, the network is strongly connected so that for any locations $i, j \in \mathcal{N}$ there exists a directed path[6] from i to j. Let $\mathcal{R}_{odm}$ denote the set of all paths from $o \in \mathcal{O}$ to $d \in \mathcal{D}$ using mode $m \in \mathcal{M}_1$ that have a length of no more than $K \in \mathbb{N}$. I assume that K is large and formally consider the behaviour of the model in the limit as $K \to \infty$ in Appendix B[7]. In particular, K must be large enough that the network remains connected when we restrict attention to these paths. Given a link $ij \in \mathcal{E}_m$ and a route $r \in \mathcal{R}_{odm}$ let $n^{odm}_{ij,r}$ denote the number of times that the link ij occurs on route r between origin $o \in \mathcal{O}$ and destination $d \in \mathcal{D}$ for mode $m \in \mathcal{M}$. Note that $n^{odm}_{ij,r}$ may be zero if a link does not occur on a particular route and may be greater than one if a link occurs multiple times along a route due to cycles. As in the previous section, ℓ_{odmr} denotes the mass of commuters choosing mode $m \in \mathcal{M}$ and route $r \in \mathcal{R}_{odm}$ to travel to work.

[6]Throughout I use the term path to denote any sequence of links connecting an origin to a destination allowing for cycles. This is sometimes referred to as a *walk* in the graph theory and network literature.

[7]Restricting attention to finite route sets allows all the analytical results to be presented using finite dimensional analysis, substantially simplifying the technical details without loss of economic insight. Appendix B shows that the $K \to \infty$ limit is well-behaved under explicitly stated assumptions on model primitives.

With these definitions, we can now define how journey times emerge endogenously as a result of the decisions of commuters through the transport network. For $\mathcal{M}_1$, the time along a route is given by the sum of the time it takes to cross each of its links. For $\mathcal{M}_0$ the times are fixed. This leads to

$$t_{odmr} = \begin{cases} \sum_{ij \in \mathcal{E}_m} n_{ij,r}^{odm} t_{ijm} & \text{if } m \in \mathcal{M}_1 \\ \varphi \bar{t}_{odm} & \text{if } m \in \mathcal{M}_0 \end{cases} \qquad \forall odmr \in \mathcal{S} \tag{12}$$

Note that the times for crossing each of the links are weighted by the number of times they occur on the path, $n_{ij,r}^{odm}$. The parameter φ represents the value of time spent on congested roads relative to uncongested modes. In the empirical application it will allow me to scale the costs associated with using public transit relative to those for driving to match the aggregate shares on each mode.

For a particular mode, the flow of traffic along a link ij is given by the sum of flows along all paths, weighted by the number times ij occurs on each. This gives

$$x_{ijm} = \sum_{od \in \mathcal{O} \times \mathcal{D}} \sum_{r \mathcal{R}_{odm}} n_{ij,r}^{odm} \ell_{odmr} \qquad \forall m \in \mathcal{M}_1, \forall ij \in \mathcal{E}_m \tag{13}$$

Finally, the time it takes to cross a particular link in the network is an increasing function of the amount of traffic flowing across that link for that mode, namely

$$t_{ijm} = s_{ijm}(x_{ijm}) \qquad \forall m \in \mathcal{M}_1, \forall ij \in \mathcal{E}_m \tag{14}$$

That is, $s_{ijm}(.)$ determines how traffic flows affect travel times and therefore provides congestion in the model.

The following assumption is made throughout:

Assumption 1. *s_{ijm} is differentiable and $s'_{ijm}(x) > 0$ for all $x \geq 0$ $\forall m \in \mathcal{M}_1$, $\forall ij \in \mathcal{E}_m$*

No parametric restrictions need to placed on the congestion technology for the anlaytical results in the paper. In particular, it is not assumed to constant elasticity, either within or across links and the empirical application will make use of this fact to fit a flexible congestion technology across links. Also note that the model only allows congestion to operate through links of the same mode. For example, buses and cars occupying the same lanes and being subject to cross-congestion is excluded. In the empirical application, modes are separated between driving and public transit, which is predominantly rail and subway, so that cross-congestion is a less central issue[8].

3.3 Market Clearing and Equilibrium

The final condition that must be imposed is goods market clearing which requires

$$\sum_{odmr \in \mathcal{S}} \ell_{odmr}(w_d + b) = \sum_{d \in \mathcal{D}} y_d + \mathbb{1}_{\mathrm{I}} \left[-\sum_{d \in \mathcal{D}} \ell_d^F \alpha_d - \sum_{o \in \mathcal{O}} \ell_o^H \beta_o \right] . \tag{15}$$

The term on the left hand side represents the total income of consumers which is composed of wage and non-wage income which they spend on the final good. When $\mathbb{1}_{\mathrm{I}} = 0$, corresponding to the case where α_d, β_o are externalities, $b = 0$ and total output is given by $\sum_d y_d$.

In the case of the floor-space micro-foundations presented in Appendix B, $\mathbb{1}_{\mathrm{I}} = 1$, and $b > 0$. In this model, total output is not given only by $\sum_d y_d$. The $\left[-\sum_{d \in \mathcal{D}} \ell_d^F \alpha_d - \sum_{o \in \mathcal{O}} \ell_o^H \beta_o\right]$ term represents the total amount of the final good used in payments to land by firms and households. These resources are not used up, but rather act as a transfer between agents, and contribute to the total stock of the final good available for consumption. (15) represents an equivalent condition to goods market clearing in that model. This is formally established in (B.3) but (15) is used in the main development of the text to avoid presenting the full set

[8] Cross-congestion can be handled in traffic models, and the extension would require tools of the type developed in Dafermos (1980).

of micro-foundations.

We can now state the definition of an equilibrium for the city.

Definition 1 (Competitive Equilibrium).

A competitive equilibrium is a set of quantities $(\{\ell_{odmr}\}\{\ell_d^F\}, \{\ell_o^H\}, \{y_d\}\{x_{ijm}\})$*, wages* $\{w_d\}$*, travel times* $(\{t_{odmr}\}\{t_{ij}\})$ *and inclusive values* $(\{v_{odmr}\}, \{v_{odm}\}, \{v_{od}\})$ *such that:*

1. *Consumers optimize as in (1), (2), (3), and (4)*
2. *Competitive firms optimize given their production technology as in (7) and (8).*
3. *Wages, productivities, residential attractiveness, and non-wage incomes are determined endogenously by (8), (9), (10), and (11).*
4. *Traffic flows and travel times are given endogenously by (12), (13), and (14).*
5. *Residential and workplace labor markets and final goods markets clear so that (5), (6) and (15) hold.*

The competitive equilibrium features rich interactions between location decisions and the commuting choices and travel times. These operate through residential attractiveness, u_o, productivities A_d, and travel times t_{ij} which are all endogenously determined by agents choices. When the time taken to travel between an *od* pair increases, fewer agents will wish to travel between them, endogenously affecting their attractiveness as home and workplaces through u_o and A_d. Conversely, changes to u_o and A_d change the structure of commutes which endogenously affects travel times t_{ij} through the flow of traffic on all links in the network. The simultaneous determination of all these variables is characterized by the system of equations (1)-(15).

3.4 A Planner's Problem

I now set up a utilitarian social planner's problem for the economy presented above. I will allow the social planner to observe the full set of shocks for each agent in the economy. Given a particular value of the shocks, $\bar{\varepsilon}$, the planner may choose the residence, workplace, mode, route and consumption of each agent. That is they choose a function[9] $a : \mathbb{R}^{\mathcal{S}} \to \mathcal{S} \times \mathbb{R}_+$ where $\bar{\varepsilon} \mapsto (o, d, m, r, c)$. Under the mapping $a(.)$ denote the values of (o, d, m, r, c) it gives as a function of $\bar{\varepsilon}$ by $\Big(\tilde{o}(\bar{\varepsilon}), \tilde{d}(\bar{\varepsilon}), \tilde{m}(\bar{\varepsilon}), \tilde{r}(\bar{\varepsilon}), \tilde{c}(\bar{\varepsilon})\Big)$. These choices generate aggregate level variables which much respect the same feasibility constraints as in the economy given above. In particular, the function a generates an aggregate level of consumption $c \in \mathbb{R}_+$, flows of commuters $\{\ell_{odmr}\}$, residential aggregates $\{\ell_o^H\}$ and workplace aggregates $\{\ell_d^H\}$ as well as traffic flows $\{x_{ijm}\}$ and travel times at the link and route level $\{t_{ijm}\}, \{t_{odmr}\}$.

Setting up the problem in this way gives the planner the full set of information available to agents, including knowing the exact values of all their idiosyncratic preferences. I also do not explicitly enforce a spatial mobility constraint of the kind used in Fajgelbaum and Gaubert (2020). The planner may assign all individuals where to live, work, how to commute and how much to consume, respecting only the feasibility constraints of the economy defined above. The planner's knowledge and abilities are broad and of limited practical feasibility. It is therefore crucial to establish that the results can be decentralized using a more reasonable set of tax instruments which do not depend on knowing individual shock realizations and allow agents to freely choose their location and commuting decisions. I show this in Corollary 2 below. Studying the more general problem shows that these taxes achieve the first best solution subject only to feasibility constraints. The reason such results are possible relies on several of the model's key assumptions. Most important is the additive separability of utility in time, consumption and idiosyncratic preferences. This means that the marginal

[9]The function a should be Lebesgue measurable to ensure all integrals are well defined since ε has a density over $\mathbb{R}^{\mathcal{S}}$. I denote the set of all such functions as $\mathcal{A}$

utility of consumption as well as the value of time are constant. While these assumptions are strong, the tractability they buy allows progress to be made on an otherwise difficult problem. Formally, the planner's problem can be stated as follows.

Definition 2 (Utilitarian Planner's Problem)**.**

$$\max_{\substack{a(.),\, c,\, \{\ell_{odmr}\},\\ \{\ell_d^F\},\, \{\ell_o^H\},\, \{x_{ijm}\}\\ \{t_{ijm}\},\, \{t_{odmr}\}}} \quad L\int_{\mathbb{R}^{\mathcal{S}}} \left(u_{\tilde{o}(\bar{\varepsilon})} + \tilde{c}(\bar{\varepsilon}) - \gamma t_{\tilde{o}(\bar{\varepsilon})\tilde{d}(\bar{\varepsilon})\tilde{m}(\bar{\varepsilon})\tilde{r}(\bar{\varepsilon})} + \bar{\varepsilon}_{\tilde{o}(\bar{\varepsilon})\tilde{d}(\bar{\varepsilon})\tilde{m}(\bar{\varepsilon})\tilde{r}(\bar{\varepsilon})}\right) f_{\varepsilon}(\bar{\varepsilon})d\bar{\varepsilon} \tag{16}$$

s.t.

$$L\int_{\mathbb{R}^{\mathcal{S}}} \tilde{c}(\bar{\varepsilon}) f_{\varepsilon}(\bar{\varepsilon})d\bar{\varepsilon} = c\,, \tag{17}$$

$$L\int_{\mathbb{R}^{\mathcal{S}}} \mathbb{1}\left\{\left(\tilde{o}(\bar{\varepsilon}), \tilde{m}(\bar{\varepsilon}), \tilde{d}(\bar{\varepsilon}), \tilde{r}(\bar{\varepsilon})\right) = (o,d,m,r)\right\} f_{\varepsilon}(\bar{\varepsilon})d\bar{\varepsilon} = \ell_{odmr} \qquad \forall odmr \in \mathcal{S}\,, \tag{18}$$

$$\sum_{odmr\in\mathcal{S}} \ell_{odmr} = L\,, \tag{19}$$

$$\sum_{m\in\mathcal{M}} \sum_{dr\in\mathcal{D}\times\mathcal{R}_{odm}} \ell_{odmr} = \ell_o^H \qquad \forall o \in \mathcal{O}\,, \tag{20}$$

$$\sum_{m\in\mathcal{M}} \sum_{or\in\mathcal{O}\times\mathcal{R}_{odm}} \ell_{odmr} = \ell_d^H \qquad \forall d \in \mathcal{D}\,, \tag{21}$$

$$t_{odmr} = \begin{cases} \sum_{ij\in\mathcal{E}_m} n_{ij,r}^{odm} t_{ijm} & \text{if } m \in \mathcal{M}_1 \\ \varphi \bar{t}_{odm} & \text{if } m \in \mathcal{M}_0 \end{cases} \qquad \forall odmr \in \mathcal{S}\,, \tag{22}$$

$$x_{ijm} = \sum_{od\in\mathcal{O}\times\mathcal{D}} \sum_{r\mathcal{R}_{odm}} n_{ij,r}^{odm} \ell_{odmr} \qquad \forall m \in \mathcal{M}_1, \forall ij \in \mathcal{E}_m\,, \tag{23}$$

$$t_{ijm} = s_{ijm}(x_{ijm}) \qquad \forall m \in \mathcal{M}_1, \forall ij \in \mathcal{E}_m\,, \tag{24}$$

$$u_o = \bar{u}_o + \beta_o \ln \ell_o^H \qquad \forall o \in \mathcal{O}\,, \tag{25}$$

$$c = \sum_{d \in \mathcal{D}} (\bar{A}_d + \alpha_d \ln \ell_d^F)\ell_d^F + \mathbb{1}_I \left[- \sum_{d \in \mathcal{D}} \ell_d^F \alpha_d - \sum_{o \in \mathcal{O}} \ell_o^H \beta_o \right] . \tag{26}$$

(16) integrates the utility that each agent with utility shocks $\bar{\varepsilon}$ gets according to the distribution of the shocks in the population f_ε under the choice of assignment $a \in \mathcal{A}$. (17) enforces that individual consumption assignments are consistent with the aggregate level of consumption c. Similarly, (18) ensures that the aggregate level of commuter flows ℓ_{odmr} are consistent with the individual level assignments of shocks to routes. (19) enforces that the total population equals the mass of available workers L. (20) and (21) enforce workplace and residential aggregates agree with the commuter flows. (22), (23), and (24) impose endogenous congestion. (25) requires that residential attractiveness be given endogenously. (26) requires that consumption equal total production with endogenous productivities substituted in. In the case where $\mathbb{1}_{\mathrm{I}} = 1$, the addition of the $\left[- \sum_{d \in \mathcal{D}} \ell_d^F \alpha_d - \sum_{o \in \mathcal{O}} \ell_o^H \beta_o\right]$ represents the full production of the city. Appendix B shows the details of this in the case of floor-space provision, where the term represents the fact that land is owned by the residents of the city.

The utilitarian planner's problem can be substantially simplified so that we can restrict attention to solving only for the variable $\boldsymbol{\ell} := (\ell_{odmr})_{odmr \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$. In order to do this, I first show that all other variables in the optimization problem, apart from $a(.)$ can be written purely as a function of $\boldsymbol{\ell}$. This follows by recursively applying the constraints to define functions as follows:

$$\ell_o^H(\boldsymbol{\ell}) := \sum_{m \in \mathcal{M}} \sum_{dr \in \mathcal{D} \times \mathcal{R}_{odm}} \ell_{odmr} \qquad \forall o \in \mathcal{O} \, , \tag{27}$$

$$\ell_d^H(\boldsymbol{\ell}) = \sum_{m \in \mathcal{M}} \sum_{or \in \mathcal{O} \times \mathcal{R}_{odm}} \ell_{odmr} \qquad \forall d \in \mathcal{D} \, , \tag{28}$$

$$\ell_{odm}(\boldsymbol{\ell}) := \sum_{r \in \mathcal{M}} \ell_{odmr} \qquad \forall od \in \mathcal{O} \times \mathcal{D}, \forall m \in \mathcal{M} \, , \tag{29}$$

$$\ell_{od}(\boldsymbol{\ell}) := \sum_{m\in\mathcal{M}} \sum_{r\in\mathcal{R}_{odmr}} \ell_{odmr} \qquad \forall od \in \mathcal{O}\times\mathcal{D}\,, \tag{30}$$

$$x_{ijm}(\boldsymbol{\ell}) := \sum_{od\in\mathcal{O}\times\mathcal{D}} \sum_{r\in\mathcal{R}_{odm}} n_{ij,r}^{odm}\ell_{odmr} \qquad \forall m \in \mathcal{M}_1, \forall ij \in \mathcal{E}_m\,, \tag{31}$$

$$t_{ijm}(\boldsymbol{\ell}) := s_{ijm}(x_{ijm}(\boldsymbol{\ell})) \qquad \forall m \in \mathcal{M}_1, \forall ij \in \mathcal{E}_m\,, \tag{32}$$

$$t_{odmr}(\boldsymbol{\ell}) := \begin{cases} \sum_{ij\in\mathcal{E}_m} n_{ij,r}^{odm} t_{ijm}(\boldsymbol{\ell}) & \text{if } m \in \mathcal{M}_1 \\ \varphi\bar{t}_{odm} & \text{if } m \in \mathcal{M}_0 \end{cases} \qquad \forall odmr \in \mathcal{S}\,. \tag{33}$$

With these definitions in hand, the simplification of the planner's problem is established in the following lemma:

Lemma 1 (Simplifying the Planner's Problem)**.**

$\boldsymbol{\ell} := (\ell_{odmr})_{odmr\in\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$ is a solution to the Utilitarian Planner's Problem if and only if it solves

$$\begin{aligned}
\max_{\boldsymbol{\ell}\in[0,L]^{\mathcal{S}}} & \sum_o \left(\sum_d \ell_{od}(\boldsymbol{\ell})\right)\left[\bar{u}_o - \mathbb{1}_I\beta_o + \beta_o \ln\left(\sum_d \ell_{od}(\boldsymbol{\ell})\right)\right] \\
& + \sum_d \left(\sum_o \ell_{od}(\boldsymbol{\ell})\right)\left[\bar{A}_d - \mathbb{1}_I\alpha_d + \alpha_d \ln\left(\sum_o \ell_{od}(\boldsymbol{\ell})\right)\right] \\
& - \gamma \sum_{m\in\mathcal{M}_1} \sum_{ij\in\mathcal{E}_m} x_{ijm}(\boldsymbol{\ell}) s_{ijm}(x_{ijm}(\boldsymbol{\ell})) \\
& - \gamma \sum_{od} \sum_{m\in\mathcal{M}_0} \varphi\bar{t}_{odm}\ell_{odmr} \\
& - (\theta-\nu)\sum_{od} \ell_{od}(\boldsymbol{\ell}) \ln\left(\ell_{od}(\boldsymbol{\ell})\right) \\
& - (\nu-\sigma)\sum_{odm} \ell_{odm}(\boldsymbol{\ell}) \ln\left(\ell_{odm}(\boldsymbol{\ell})\right) - \sigma \sum_{odmr} \ell_{odmr}\ln \ell_{odmr}
\end{aligned}$$

s.t.

$$\sum_{odmr \in \mathcal{S}} \ell_{odmr} = L \; .$$

where $\ell_{odm}(\boldsymbol{\ell}), \ell_{odm}(\boldsymbol{\ell})$ *and* $x_{ijm}(\boldsymbol{\ell})$ *are given by definitions (31), (32), and (33) above.*

Proof. See Appendix A

The simplification of the planner's problem makes use of three key steps. Firstly, since utility is linear in consumption and additively separable from the idiosyncratic term, the planner will not care about the distribution of consumption over individuals and will only seek to maximize total output. This means that the production technology for the economy can be plugged directly into the objective function for the problem.

The second key step is to note that fixing the aggregate variables ℓ_{odmr}, the problem of how to assign ε realisations to different *odmr* tuples is a well studied problem[10]. Intuitively, the planner will assign individuals with realizations of ε_{odmr} that are high relative to alternative $o'd'm'r'$ to *odmr*. Formally, this is an optimal transport problem, and the key implication is that the value function for the inner problem can be used to solve out for the function $a(.)$. This means that we need only consider the more tractable problem of optimizing over aggregate flows ℓ_{odmr}.

The particular form of the value function for the inner problem will depend on the distribution of ε. This can be seen in the final final three terms in Lemma 1. For intuition, consider the first of these terms: $-(\theta - \nu) \sum_{od} \ell_{od} \ln \ell_{od}$. Considered in isolation, this term reaches a maximum when ℓ_{od} is completely evenly distributed: everyone is sent to the *od* pair

[10] Variations of this problem have been studied in Anderson et al. (1992), Hofbauer and Sandholm (2002), Maher et al. (2005), Cameron et al. (2007), Galichon and Salanié (2022) and Donald et al. (2023) with similar results being rediscovered independently. I use the results proved in Galichon and Salanié (2022).

for which they have the highest shock realization. As the planner concentrates more people along a particular *od* pair, for example because productivities or residential attractiveness are high there, the planner is forced to draw from lower and lower in the shock distribution leading to less utility from idiosyncratic preferences. This acts as a dispersion force in the planner's problem, leading flows to be spread more evenly across locations, modes, and routes than they would be in the absence of idiosyncratic preferences.

The final step involves using definitions (27)-(33) and rearranging the constraints to get the desired expression. It is important to note how this simplification makes the problem tractable. In the initial formulation, we were required to choose a whole function $a(.)$ assigning each realization of the shock to a particular *odmr* and level of consumption. Lemma 1 shows that this reduces to a finite-dimensional optimization problem over a compact set with a single linear constraint. This makes the problem highly tractable and allows the use of standard techniques in optimization theory to establish both existence and uniqueness.

3.5 Existence and Uniqueness

The Planner's Problem

This section presents the main theoretical results of the paper. I prove existence and uniqueness results for both the social planner and decentralized equilibrium problems and provide an algorithm to compute them.

Proposition 1 (Existence and Uniqueness for the Planner's Problem)**.** *Under Assumption 1, a solution to the planner's problem exists. The solution for* $\{\ell_{odmr}\}, \{\ell_d^F\}, \{\ell_o^H\}, \{x_{ijm}\}, \{t_{ijm}\}, \{t_{odmr}\}$ *and total welfare is unique if*

$$\theta - \nu > \max_{o,d \in \mathcal{O} \times \mathcal{D}} \{\alpha_d + \beta_o, \alpha_d, \beta_o\}$$

$$xs_{ijm}(x) \text{ is strictly convex in } x \quad \forall m \in \mathcal{M}_1, \quad \forall ij \in \mathcal{E}_m$$

Proof. See Appendix A

Note that, as shown in Lemma 1, the distribution of consumption is not pinned down uniquely in the planner's problem since utility is linear in c and so the planner is indifferent between different distributions of the consumption good. The remaining aggregate variables are unique under the stated condition.

Existence is established by showing that the optimization problem is the maximization of a continuous function over a compact set. Uniqueness is shown by proving that the objective function is strictly concave over a convex set so that any optimum is unique. The proof proceeds by analyzing the components of the objective function piece by piece. The main terms that have the potential to cause non-concavity are the first two terms in Lemma 1. Economically, when $\alpha_d, \beta_o > 0$, these represent the agglomeration forces in the model. When more people enter a place to live or work this leads the location to become more attractive as a home or workplace. This can lead to multiplicity. When people concentrate in a particular *od* pair, that remains attractive due to the benefits of agglomeration so that no-one wishes to relocate in equilibrium. However if people had concentrated in a different pair $o'd'$ these would instead become attractive and give rise to a different equilibrium.

To offset these agglomeration forces a dispersion force across *od* pairs is required. Preference heterogeneity across *od* pairs through θ, net of the variation coming over modes through ν, provides this force. This can also be seen through the formulation of the problem as an optimization problem. The $-(\theta-\nu)\sum_{od}\ell_{od}(\boldsymbol{\ell})\ln\left(\ell_{od}(\boldsymbol{\ell})\right)$ is strictly concave in ℓ_{od} and, mathematically, makes the problem more concave. Economically, as people initially concentrate in a particular origin-destination pair *od*, only those with a high idiosyncratic preference for *od* move there. As ℓ_{od} increases however, the idiosyncratic component of their utility is drawn from lower down the shock distribution making the location less and less attractive for the marginal mover. The condition in the statement of the proposition formalizes the balance of these forces. When the dispersion force across *od* pairs are sufficiently high they will dominate. To derive this condition I make use of results on diagonally dominant matrices to sign the Hessian of these terms.

The other terms in the optimization problem do not pose a problem for concavity. Preference heterogeneity across modes and routes as well as congestion from traffic act as dispersion forces and make the problem strictly concave. The second condition in the statement of the proposition on the strict convexity of $xs_{ijm}(x)$ is relatively mild. It is satisfied whenever the marginal cost to the planner of sending a person along a link, given by $s_{ijm}(x) + xs'_{ijm}(x)$ is increasing. This is the analogue for the planner's problem of the condition in Assumption 1. It does not impose any functional form assumptions on s_{ijm} but is satisfied by many that are commonly used in practice. For example, if $s(x) = a + bx^c$, it holds whenever $b, c > 0$. It is important to note that these results allows for further flexibility in estimation than is pursued in this paper: non-parametric approaches, which impose only shape restrictions on s_{ijm} are compatible with the framework.

The Competitive Equilibrium

I now turn to the problem of characterizing the equilibrium of the model. The first step in this will be to show that the equilibria of the model correspond to turning points of the Lagrangean of a distorted planner's problem. I first define the pseudo-planner's problem and then discuss the intuition for why its turning points coincide with the equilibria of the model. I will cast the distorted planner's problem purely in terms of the vector $\boldsymbol{\ell} := (\ell_{odmr})_{odmr \in \mathcal{S}}$ as this will turn out to be useful. In order to do so it is useful to note that all other variables in the model can be viewed as a function of $\boldsymbol{\ell}$ alone.

Definition 3 (Distorted Planner's Problem)**.**

Define the distorted planner's problem as:

$$\max_{\boldsymbol{\ell}\in[0,L]^{\mathcal{S}}} \quad f(\boldsymbol{\ell}) := \sum_o \left(\sum_d \ell_{od}(\boldsymbol{\ell})\right)\left[\bar{u}_o - \beta_o + \beta_o \ln\left(\sum_d \ell_{od}(\boldsymbol{\ell})\right)\right]$$

$$+ \sum_d \left(\sum_o \ell_{od}(\boldsymbol{\ell})\right)\left[\bar{A}_d - \alpha_d + \alpha_d \ln\left(\sum_o \ell_{od}(\boldsymbol{\ell})\right)\right]$$

$$- \gamma \sum_{m\in\mathcal{M}_1}\sum_{ij\in\mathcal{E}_m}\int_0^{x_{ijm}(\boldsymbol{\ell})} s_{ijm}(z)dz$$

$$- \gamma \sum_{od}\sum_{m\in\mathcal{M}_0} \varphi \bar{t}_{odm}\ell_{odmr}$$

$$- (\theta-\nu)\sum_{od}\ell_{od}(\boldsymbol{\ell})\ln\left(\ell_{od}(\boldsymbol{\ell})\right)$$

$$- (\nu-\sigma)\sum_{odm}\ell_{odm}(\boldsymbol{\ell})\ln\left(\ell_{odm}(\boldsymbol{\ell})\right) - \sigma\sum_{odmr}\ell_{odmr}\ln \ell_{odmr}$$

s.t.

$$g(\boldsymbol{\ell}) := \sum_{odmr\in\mathcal{S}} \ell_{odmr} - L = 0$$

where $\ell_{odm}(\boldsymbol{\ell}), \ell_{odm}(\boldsymbol{\ell})$ and $x_{ijm}(\boldsymbol{\ell})$ are given by definitions (31), (32), and (33) above.

The distorted planner's problem involves three changes relative to the full planner's problem. Firstly, $\bar{u}_o - \mathbb{1}_{\mathrm{I}}\beta_o$ is replaced with $\bar{u}_o - \beta_o$. Secondly, $\bar{A}_d - \mathbb{1}_{\mathrm{I}}\alpha_d$ is replaced with $\bar{A}_d - \alpha_d$. Finally, $s_{ijm}(x_{ijm})$ is replaced with $\frac{1}{x_{ijm}}\int_0^{x_{ijm}} s_{ijm}(z)dz$. These correspond to each of the distortions in the model: externalities in residential attractiveness, productivity and congestion. Note in particular, that when $\mathbb{1}_{\mathrm{I}} = 1$ so that the endogeneity in productivities and residential attractiveness is internalized by agents, the first two sets of terms coincide. The key intuition for each of these terms in the presence of externalities is that, when making decisions agents use average costs but the planner will use marginal costs.

I focus on the case of the congestion externality. When the planner considers the cost of sending traffic along a particular link, ijm, the contribution to the objective function is given by $-\gamma x_{ijm} s_{ijm}(x_{ijm})$. The marginal cost (in terms of time) of increasing flow on that link is therefore given by $s_{ijm}(x_{ijm}) + x_{ijm} s'_{ijm}(x_{ijm})$. Each of these terms has an intuitive interpretation: $s_{ijm}(x_{ijm})$ is the utility cost paid in time for the marginal traveler. $s'_{ijm}(x_{ijm})$ is how much the marginal traveler increases journey times for all other people using the link. There are x_{ijm} people using the link so the contribution from this term is $x_{ijm} s'_{ijm}(x_{ijm})$. By contrast, in the competitive equilibrium problem, agents do not account for the $x_{ijm} s'_{ijm}(x_{ijm})$ term: the marginal driver along a link only internalizes their private time cost $s_{ijm}(x_{ijm})$ which also corresponds to the average cost along the link. The key to establishing the competitive equilibrium as an optimization problem is to distort the planner's congestion technology so that they too neglect the spillovers from the marginal driver to the travel times others. This is exactly what is achieved by taking the total contribution to the objective function of traffic flows on a link to be $\int_0^{x_{ijm}} s_{ijm}(z)dz$. When this function is differentiated it returns $s_{ijm}(x_{ijm})$ so that the pseudo-planner only accounts for the direct cost of time travel to the marginal driver. The formal verification of this intuition for congestion, as well as endogenous amenities and productivities, is given explicitly in the proof of Proposition 2.

The problem has the following Lagrangean associated with it, constructed from the objective function and constraint together with a multiplier λ. This is given in the following definition.

Definition 4 (Lagrangean of the Distorted Planner's Problem and Turning points)**.** *The Lagrangean of the Distorted Planner is a function of* $(\boldsymbol{\ell}, \lambda)$ *where* $\lambda \in \mathbb{R}, \ell \in [0, L]^{\mathcal{S}}$ *defined by:*

$$\mathcal{L}(\boldsymbol{\ell}, \lambda) := f(\boldsymbol{\ell}) - \lambda g(\boldsymbol{\ell}, \lambda)$$

A turning point of the Lagrangean of the Distorted Planner is a pair $(\boldsymbol{\ell}, \lambda)$ *satisfying:*

$$\nabla f(\boldsymbol{\ell}) - \lambda \nabla g(\boldsymbol{\ell}) = 0$$
$$g(\boldsymbol{\ell}) = 0$$

Formulating the Lagrangean is key in characterizing the competitive equilibria of the city, as is established in the following proposition.l

Proposition 2 (Characterizing the Competitive Equilibrium)**.** *All variables in the equilibrium are determined by the value of* $\boldsymbol{\ell}$*. The set of* $\boldsymbol{\ell}$*'s that are turning points of the distorted planner's Lagrangean is equal to the set of* $\boldsymbol{\ell}$*'s that are competitive equilibria of the model.*

Proof. See Appendix A

It is important to note that this result does not rely on the uniqueness of the equilibrium. If agglomeration forces are strong enough, there will be multiple equilibria. Fujita et al. (2001) show how this issue can be approached in models with a few locations by enumerating the full set of equilibria using direct methods. Proposition 2 suggests an alternative: by casting the equilibrium as an optimization problem, enumerating equilibria of the model is reposed as the problem of enumerating the turning points of a Lagrangian. While not pursued in the current paper, it is possible this approach could shed light on issues of multiplicity in modern quantitative spatial models featuring many locations of the kind reviewed in Redding and Rossi-Hansberg (2017).

Given that the equilibria of the model can be cast as an optimization problem with a very similar structure to the planner's problem, establishing existence and uniqueness of the competitive equilibrium in the model will also proceed in a similar way, namely by showing that the program is strictly concave. Under standard regularity conditions, the turning point of the Lagrangean will give us the unique optimum of the program provided the objective function is strictly concave. This leads to the following proposition:

Proposition 3 (Existence and Uniqueness for the Competitive Equilibrium Problem)**.** *Under Assumption 1, a solution to the competitive equilibrium problem exists. It is unique if*

$$\theta - \nu > \max_{o,d \in \mathcal{O} \times \mathcal{D}} \{\alpha_d + \beta_o, \alpha_d, \beta_o\}$$

Proof. See Appendix A

The intuition for this result is similar to the case of the planner's problem: the agglomeration forces in the model must be offset by significantly strong dispersion forces to maintain the overall concavity of the problem and resulting uniqueness of the equilibrium. It is important to note that we only require $s'_{ijm}(x) > 0$ as given in Assumption 1 for the equilibrium case. This is a weaker requirement than that $xs_{ijm}(x)$ be strictly convex. The reason for this is that the replacement of $xs_{ijm}(x)$ by $\int_0^{x_{ijm}} s_{ijm}(z)dz$ acts as a further convexifying force on the objective function. The planner's problem requires that the total cost along a link be convex in x_{ijm} whereas the pseudo-planner requires only that the marginal effect of traffic flows on times be positive. The latter is enough to ensure that, in a competitive equilibrium, multiplicity is avoided: when more people travel along a road, the costs to private individuals increase. The planner must also account for spillovers: as flows on a road increase, the marginal increases in costs to other drivers must also increase.

3.6 Decentralization

I now turn to how the optimal allocation can be decentralized as a competitive equilibrium with taxes. I assume that there are three tax instruments available: a flat tax (or subsidy[11]) on individuals at their workplace τ_d^F, a flat tax at their residence τ_o^H and a flat tax imposed each time they cross a link on each congested mode τ_{ijm}. I assume that the tax revenues from these taxes are rebated lump-sum to individuals and included in their non-wage income b. The three taxes are chosen to offset the three potential externalities in the model. Intuitively, this should give us strong enough tools to correct each of them, a point that is formally shown below. I first show the taxes change the equilibrium equations (1)-(15) to define the concept of an equilibrium with taxes.

With this definition in hand I establish two corollaries to the main propositions of the paper. Firstly, I show that given an arbitrary set of taxes $\left(\left(\tau_d^F\right)_{d\in\mathcal{D}}, \left(\tau_o^H\right)_{o\in\mathcal{O}}, \left(\tau_{ijm}\right)_{m\in\mathcal{M}_1, ij\in\mathcal{E}_m}\right)$ the model can be solved for the equilibrium with taxes in an analogue of Proposition 3. Secondly, I find expressions for the taxes that decentralize the planner's solution.

Equations (9) and (10) are adjusted in a straightforward way to incorporate the taxes

$$A_d = \bar{A}_d - \tau_d^F + \alpha_d \ln \ell_d^F \qquad \forall d \in \mathcal{D}\,, \tag{$9^\#$}$$

$$u_o = \bar{u}_o - \tau_o^H + \beta_o \ln \ell_o^H \qquad \forall o \in \mathcal{O}\,. \tag{$10^\#$}$$

The total revenues from the taxes are rebated lump sum to consumers so that we have

$$b = \frac{\sum_{d\in\mathcal{D}} \ell_d^F (\tau_d^F - \mathbb{1}_{\mathrm{I}}\alpha_d) + \sum_{o\in\mathcal{O}} \ell_o^H (\tau_o^H - \mathbb{1}_{\mathrm{I}}\beta_o) + \sum_{m\in\mathcal{M}_1}\sum_{ij\in\mathcal{E}_m} \tau_{ijm} x_{ijm}}{L} \tag{$11^\#$}$$

[11] I represent subsidies as negative values of the τ variables. I refer to them as taxes throughout but they should be understood to allow for negative values.

It is important for what follows that the rebated tax revenues are equal for all agents. In particular, they must not depend upon any aspect of the choices over *odmr*. For example, if the taxes were rebated locally, so that the revenue from each area's taxes was given to the agents who lived or worked there, this would distort locations. In turn this would lead to complicated interactions with the congestion externality and the endogenous attractiveness of workplaces and residences. The fact that lump-sum transfers that are constant across agents do not affect location decisions can be seen mathematically in equation (4). Note that the b term will cancel in all numerators and denominators of the expression: as in all (additive) discrete choice models, shifting the value of each option by a constant does not affect the choice probabilities.

Turning to the tax on links for congested modes we have

$$t_{ijm} = \sum_{ij \in \mathcal{E}_m} s_{ijm}(x_{ijm}) + \tau_{ij}/\gamma \quad \forall m \in \mathcal{M}_1, ; \forall ij \in \mathcal{E}_m \tag{14$^{\#}$}$$

τ_{ijm} is divided by γ since the tax is paid in units of the consumption good rather than time. The taxes are paid each time the link is crossed terms of the consumption good and so are equivalent, in utility terms, to a time cost of τ_{ij}/γ

The market clearing condition in (15) must also be adjusted to account for the presence of taxes too, namely,

$$\sum_{odmr \in \mathcal{S}} \ell_{odmr}(w_d + b) = \sum_{d \in \mathcal{D}} y_d + \sum_{d \in \mathcal{D}} \ell_d^F(\tau_d^F - \mathbb{1}_{\mathrm{I}}\alpha_d) + \sum_{o \in \mathcal{O}} \ell_o^H(\tau_o^H - \mathbb{1}_{\mathrm{I}}\beta_o) + \sum_{m \in \mathcal{M}_1} \sum_{ij \in \mathcal{E}_m} \tau_{ijm} x_{ijm} \tag{15$^{\#}$}$$

Note that the taxes appear on both sides of equation ($15^{\dagger}$) through the b term on the left hand side and that these will cancel. As in the case of the equilibrium the $\mathbb{1}_{\mathrm{I}}$ adjusts the condition for whether or not the endogenous components of residential attractiveness and productivities are internalized by agents. I now define an equilibrium with taxes.

Definition 5 (Competitive Equilibrium with Taxes)**.**
Given $\left(\left(\tau_d^F\right)_{d\in\mathcal{D}}, \left(\tau_o^H\right)_{o\in\mathcal{O}}, \left(\tau_{ijm}\right)_{m\in\mathcal{M}_1, ij\in\mathcal{E}_m}\right)$ *a competitive equilibrium with taxes is set of quantities* $(\{\ell_{odmr}\}\{\ell_d^F\}, \{\ell_o^H\}, \{y_d\}\{x_{ijm}\})$*, wages* $\{w_d\}$*, travel times* $(\{t_{odmr}\}\{t_{ij}\})$ *and inclusive values* $(\{v_{odmr}\}, \{v_{odm}\}, \{v_{od}\})$ *such that:*

1. *Consumers optimize as in (1), (2), (3), and (4)*
2. *Competitive firms optimize given their production technology as in (7) and (8).*
3. *Wages, productivities, residential attractiveness, and non-wage incomes are determined endogenously by (8), (9#), (10#), and (11#).*
4. *Traffic flows and travel times are given endogenously by (12), (13), and (14#).*
5. *Residential and workplace labor markets and final goods markets clear so that (5), (6) and (15#) hold.*

These conditions imply that all agents take the taxes as given and optimize conditional on them. The total revenue from the taxes is rebated lump sum to all individuals in the economy. The problem is otherwise the same as the competitive equilibrium problem.

Corollary 1 (Existence and Uniqueness of Competitive Equilibrium with Taxes)**.** *Given a set of taxes* $\left(\left(\tau_d^F\right)_{d\in\mathcal{D}}, \left(\tau_o^H\right)_{o\in\mathcal{O}}, \left(\tau_{ijm}\right)_{m\in\mathcal{M}_1, ij\in\mathcal{E}_m}\right)$*, and under Assumption 1, a solution to the competitive equilibrium with taxes problem exists. It is unique if*

$$\theta - \nu > \max_{o,d\in\mathcal{O}\times\mathcal{D}} \{\alpha_d + \beta_o, \alpha_d, \beta_o\}$$

Proof. See Appendix A

Given the similarity of the problem with taxes to the competitive equilibrium, establishing existence and uniqueness proceeds along similar lines. The problem is first cast as an optimization problem which is isomorphic to the distorted planner's problem for the equilibrium problem. Since the taxes are lump sum, they do not affect the concavity properties of the program. This result is key in investigating second best policy counterfactuals. The set of taxes available can be constrained arbitrarily and the model can be solved at any fixed value of the taxes to evaluate the welfare consequences. For example, a toll zone which imposes a fixed tax on a subset of links can be modeled as setting $\tau_{ijm} = \bar{\tau}$ for all links entering the zone and zero elsewhere. The optimal level of the toll can be investigated numerically by solving the model for different values of $\bar{\tau}$ and comparing the resulting utility.

I now show that the tax instruments given above are sufficient to achieve the first best allocation from the planner's problem and to find the taxes that do so. This is established in the following proposition.

Corollary 2 (Optimal Taxes). *Let $\left(x^*_{ijm}\right)_{m\in\mathcal{M}_1, ij\in\mathcal{E}_m}$ be the link level flows that solve the planner's problem. In an equilibrium with taxes, the taxes that reproduce the allocation and welfare from the planner's problem are:*

$$\begin{aligned}
\tau^H_o &= -(1-\mathbb{1}_I)\beta_o && \forall o \in \mathcal{O} \\
\tau^F_d &= -(1-\mathbb{1}_I)\alpha_d && \forall d \in \mathcal{D} \\
\tau_{ijm} &= \gamma x^*_{ijm} s'_{ijm}(x^*_{ijm}) && \forall m \in \mathcal{M}_1,\ \forall ij \in \mathcal{E}_m
\end{aligned}$$

Proof. See Appendix A

The proof of the proposition proceeds by comparing the conditions for optimality of the planner's problem with those for the pseudo-planner's problem for a competitive equilibrium with taxes. It is verified that the proposed taxes lead to the same allocation. The taxes follow the intuitive Pigouvian logic of offsetting each externality where it occurs. Taxes by residence and home location offset any externalities that occur there. Note in particular that they are zero when $\mathbb{1}_{\mathrm{I}} = 1$ and endogenous residential attractiveness and productivities are fully internalized by agents. Turning to the congestion taxes, these offset the time cost to others of traveling on each road. As discussed above, the marginal traveler on link ijm imposes a total cost of $x_{ijm} s'_{ijm}(x_{ijm})$ on other travelers using that link. τ_{ijm} corrects for this by aligning marginal private and social costs of using each link.

While the formulae for taxes are easy to state, evaluating them numerically requires the full structure of the model. In particular, x^*_{ijm} comes from the solution to the first best planner's problem and accounts for the full set of equilibrium interactions. This means that solving for the optimal tax policy requires computing the solution to a full general equilibrium social planner's problem. Doing so will be a key part of the empirical application to evaluate the potential gains from congestion pricing. Note also, the fact that τ^H_o and τ^F_d do not depend on ℓ^H_o and ℓ^F_d is a consequence of the functional form assumptions that have been made in the model. In particular the log-linear form of equations (9) and (10) ensures that any externalities from residence and workplace location decisions are constant in the number of people. More general functional forms would also lead τ^H_o and τ^F_d to be dependent on population[12].

3.7 Computation

I now turn to the computation of the first best, competitive equilibrium and taxed equilibrium problems. Since the empirical case I will consider involves $\alpha_d, \beta_o < 0$, $\mathcal{M}_1 = \{m_1\}$, and

[12] In principle, more general functional forms for residential and productive externalities could be accommodated. For example if $A_d = f_d(\ell^F_d)$ then the $\ell^F_d \left[\bar{A}_d - \alpha_d + \alpha_d \ln \ell^F_d\right]$ term in the distorted planner's problem would need to be replaced with $\int_0^{\ell^F_d} f_d(z)dz$, analogous to the case of congestion. This makes sufficient conditions for strict concavity of the distorted planner and planner's programs more abstract, so I have imposed functional forms throughout.

$\mathcal{M}_0 = \{m_0\}$ I design an algorithm specifically tailored to this case. If α_d, β_o are composed of multiple forces, the condition that they are negative states that the balance of these forces must be negative in both cases. It achieves computational efficiency by greatly reducing the dimensionality of the problem. In order to address all three problems within the same framework I make the following definitions:

$$\tilde{u}_o := \begin{cases} \bar{u}_o - \mathbb{1}_{\mathrm{I}}\beta_o & \text{for the first best problem} \\ \bar{u}_o - \beta_o & \text{for the competitive equilibrium problem} \\ \bar{u}_o - \beta_o - \tau_o^H & \text{for the taxed equilibrium problem} \end{cases}$$

$$\tilde{A}_d := \begin{cases} \bar{A}_d - \mathbb{1}_{\mathrm{I}}\alpha_d & \text{for the first best problem} \\ \bar{A}_d - \alpha_d & \text{for the competitive equilibrium problem} \\ \bar{A}_o - \beta_d - \tau_d^F & \text{for the taxed equilibrium problem} \end{cases}$$

$$\tilde{s}_{ijm}(x) := \begin{cases} s_{ijm}(x) & \text{for the first best problem} \\ \frac{1}{x}\int_0^x s_{ijm}(z)dz & \text{for the competitive equilibrium problem} \\ \frac{1}{x}\int_0^x s_{ijm}(z)dz + \tau_{ijm}/\gamma & \text{for the taxed equilibrium problem} \end{cases}$$

Proposition 4 (Computable Program)**.** *Suppose that $\alpha_d < 0$ for all $d \in \mathcal{D}$ and $\beta_o < 0$ for all $o \in \mathcal{O}$, $\mathcal{M}_1 = \{m_1\}$, and $\mathcal{M}_0 = \{m_0\}$. Suppose also that Assumption 1 holds. Let $\tilde{u}_o, \tilde{A}_d, \tilde{s}_{ijm}$ be given by the definitions above according to which problem we are solving. The solution to the problem can be obtained by solving the following unconstrained minimization problem in dual variables:*

$$\min_{\{\mu_{ij}\},\{\lambda_o^H\},\{\lambda_d^F\}} \quad -\sum_{o\in\mathcal{O}} \beta_o \exp\left(\frac{\lambda_o^H - \tilde{u}_o - \beta_o}{\beta_o}\right) - \sum_{d\in\mathcal{D}} \alpha_d \exp\left(\frac{\lambda_d^F - \tilde{A}_d - \alpha_d}{\alpha_d}\right)$$
$$+ \sum_{ij\in\mathcal{E}_{m_1}} \tilde{s}^{*-1}_{ijm_1}(\mu_{ij}/\gamma)\left[\mu_{ij}/\gamma - \tilde{s}_{ijm_1}\left(\tilde{s}^{*-1}_{ijm_1}(\mu_{ij}/\gamma)\right)\right]$$

$$+ \theta L \ln \left(\sum_{od} \exp\left(\frac{\lambda_o^H + \lambda_d^F}{\theta}\right) \left[\exp\left(\frac{-\varphi\gamma\bar{t}_{od}}{\nu}\right) + \left(\sum_r \exp\left(- \sum_{ij} \mu_{ij} n^{od}_{ijmr_1,r} / \sigma \right) \right)^{\frac{\sigma}{\nu}} \right]^{\frac{\nu}{\theta}} \right)$$

where

$$\tilde{s}^*_{ijm_1}(x) := \tilde{s}_{ijm_1}(x) + x\tilde{s}'_{ijm_1}(x) \ ,$$

and the primal variables $\{x_{ijm_1}\}, \{\ell_o^H\}, \{\ell_d^F\}, \{\ell_{odm_0}\}, \{\ell_{odm_1}\}, \{\ell_{od}\}$ *can be recovered through duality relationships as a function of the optimal* $\{\mu_{ij}\}, \{\lambda_o^H\}, \{\lambda_d^F\}$*.*

Proof. See Appendix A

Formally, this result comes from convex duality. However, it also provides economic intuition about the problem. The proof proceeds by defining slack variables that represent the aggregates $x_{ijm_1}, \ell_o^H, \ell_d^F$. These are constrained to be equal to their definitions in terms of $\boldsymbol{\ell}$ and $\mu_{ij}, \lambda_o^H, \lambda_d^F$ represent the multipliers on $x_{ijm_1}, \ell_o^H, \ell_d^F$ respectively. As usual, these have an interpretation as the value of relaxing each of the constraints marginally. Consider the planner's problem. The proposition shows that all the planner needs to know to pin down a solution is the optimal marginal value of an extra worker in each workplace, λ_d^F, an extra resident in each home location, λ_o^H, and an extra car flowing along each link in the network, μ_{ij}.

3.8 Solving Sub-models

I now consider the problem of fixing some of agents' choices within the model. I consider the case with taxes as this nests the case without taxes by setting all taxes equal to zero. Firstly, I consider fixing $\{\ell_{od}\}$ so that people are no longer allowed to change their locations *od* but are still able to adjust their mode choice m and their route choice r. This may be thought of as corresponding to the medium term where people can switch how they commute to work and the route they choose to drive but cannot move their residence or workplace.

Given a fixed origin and destination, the only choice agents face is to maximize their utility over modes and routes. They take into account the shocks they receive ε_{mr} as well as the time it takes to travel t_{odmr}. ε_{mr} now has a GEV distribution with parameters (ν, σ) governing the scale over modes and routes. The problem facing a consumer travelling from o to d is therefore:

$$\max_{m\in\mathcal{M}, r\in\mathcal{R}_{odm}} -\gamma t_{odmr} + \varepsilon_{mr}$$

Following the same logic as for the general model, the choice shares will satisfy:

$$v^{\text{tr}}_{odm} = \sigma \ln \sum_{r\in\mathcal{R}_{odm}} \exp\left(-\gamma t_{odmr}/\sigma\right) \qquad \forall od \in \mathcal{O}\times\mathcal{D},\ \forall m \in \mathcal{M} \tag{34}$$

$$\frac{\ell_{odmr}}{\ell_{od}} = \frac{\exp\left(v^{\text{tr}}_{odm}/\nu\right)}{\sum_{m'\in\mathcal{M}} \exp\left(v^{\text{tr}}_{odm'}/\nu\right)} \frac{\exp\left(-\gamma t_{odmr}/\sigma\right)}{\sum_{r'\in\mathcal{R}_{odm}} \exp\left(-\gamma t_{odmr'}/\sigma\right)} \qquad \forall odmr \in \mathcal{S} \tag{35}$$

Note that the transportation choices made by individuals still give rise to endogenous travel times through the definitions in (12), (13) and (14†). Equation (14†) also allows for the incorporation of taxes in the form of $\{\tau_{ijm}\}$ which may be equal to zero but will allow us to assess the effects of congestion policies through different sub-models. We can now formulate the following definition for an equilibrium over the transport part of the model.

Definition 6 (Transport Equilibrium with Taxes)**.**
Given a set of origin-destination flows $\{\ell_{od}\}$ and a set of taxes over links $\{\tau_{ijm}\}$, a transport equilibrium with taxes is a set of quantities $(\{\ell_{odmr}\}, \{x_{ijm}\})$, travel times $(\{t_{odmr}\}, \{t_{ij}\})$ and inclusive values $\{v^{tr}_{odm}\}$ such that

1. *Consumers optimize over routes and modes as in* (23) *and* (24)

2. *Traffic flows and travel times are given endogenously by* (12), (13), (14†)

In the transport equilibrium $\{\ell_{od}\}$ pins down workplace and residential aggregates through equations (5), (6), (7), (8), (9†), (10†), (11†) and of the model. It also determines the total output of the city. We can therefore evaluate welfare in these models and the incorporation of taxes is entirely analogous to the full general equilibrium case. In general, individuals would want to change locations or modes in these models if they could: they do not maximize the full utility function from the general model. However, if the $\{\ell_{od}\}$ are themselves the equilibrium values that result from the larger competitive equilibrium model then the solutions to the transport equilibrium will coincide with the $\{\ell_{odmr}\}$ flows that are consistent with them. This can be seen directly by appropriately summing and rearranging equations (1)-(4) and comparing the results with equations (16) and (17). The details of these results as well as the computational algorithm used to solve these problems are presented in Appendix B. They follow a similar logic to the general case.

It is possible to constrain agents choices even further. Fixing both the location and mode choices is equivalent to fixing $\{\ell_{odm}\}$ for each $od \in \mathcal{O} \times \mathcal{D}$ and each $m \in \mathcal{M}$. This fixes the mass of agents along taking all modes between all origin and destination pairs. Agents with $m \in \mathcal{M}_0$) now face no choice: their is no route choice in their problem and their origin, destination and mode have been specified. However, agents with $m \in \mathcal{M}_1$ still face a choice over the route they take to commute. This can be thought of as the very short term: drivers can alter their travel route daily, but permanently changing travel mode generally takes longer, for example requiring buying or selling a car. Drivers now account for the shocks they receive only over routes ε_r which are now simply Gumbel distributed with scale parameter σ as well as travel times. A consumer traveling from o to d by mode $m \in \mathcal{M}_1$ solves

$$\max_{r \in \mathcal{R}_{odm}} -\gamma t_{odmr} + \varepsilon_r$$

In this case, the choice shares are simply given by the usual logit formula:

$$\frac{\ell_{odmr}}{\ell_{odm}} = \frac{\exp(-\gamma t_{odmr}/\sigma)}{\sum_{r'}\exp(-\gamma t_{odmr'}/\sigma)} \qquad \forall odmr \in \mathcal{S} \tag{36}$$

Note that this equation holds trivially for $m \in \mathcal{M}_0$ but that for $m \in \mathcal{M}_1$ it provides an optimality condition and the t_{odmr} terms on the right hand must agree with the endogenous flows of drivers on the left. This leads us to the following definition of a driving equilibrium:

Definition 7 (Driving Equilibrium with Taxes)**.**
Given a set of origin-destination-mode flows $\{\ell_{odm}\}$ and a set of taxes over links $\{\tau_{ijm}\}$, a driving equilibrium is a set of quantities $(\{\ell_{odmr}\}, \{x_{ijm}\})$ and travel times $(\{t_{odmr}\}, \{t_{ij}\})$ such that

1. *Consumers optimize over routes as in (25)*
2. *Traffic flows and travel times are given endogenously by* (12), (13), ($14^\dagger$)

Again, the aggregate production and residential attractiveness of locations in the model will be pinned down by $\{\ell_{odm}\}$, allowing us to evaluate welfare in the driving equilibrium. Individuals may wish to re-optimize either over modes or locations or both. However, if $\{\ell_{odm}\}$ are generated from a transport equilibrium then the sub-choice of routes from a driving equilibrium taking these flows as given will generate the same $\{\ell_{odmr}\}$ flows. Similarly $\{\ell_{odm}\}$ generated from the full competitive equilibrium problem will also have $\{\ell_{odmr}\}$ flows which coincide with those from the driving sub-problem. With the two problems defined, I provide a final existence and uniqueness result.

Proposition 5 (Existence and Uniqueness for Transport and Driving Equilibria with Taxes)**.**
Under assumption 1, there exists a unique solution for the the transport equilibrium with taxes and there exists a unique solution for the driving equilibrium with taxes.

Proof. See Appendix A

The proof closely parallels that of propositions (2) and (3) but is simplified considerably by the absence of location choices and the possibility for agglomeration forces through α_d, β_o. In these models, the conditions for uniqueness are mild: there are only congestion forces so there is no scope for multiplicity provided that assumption 1 holds and increased traffic leads to increasing journey times. These propositions are closely related to results in the transport literature as reviewed in Sheffi (1985). I state them here for clarity. I also use the results for different purposes to what has been done in that literature. Firstly, in counterfactuals, I will use both the transport and driving equilibrium to assess how welfare changes in response to policy before the margins of mode and route choice have adjusted. Secondly, solving the transport equilibrium problem will be a key step in the identification of the model's parameters.

4 An Application to New York City Congestion Pricing

This section shows how the model is estimated in an application to New York City. I begin by describing the sources of data I use and the the parameters that I take from the literature. Estimation of the set of parameters proceeds in two steps. Firstly, I estimate the congestion technology from time of day variation in the flow of cars on roads. Secondly, I use these congestion estimates together with the structure of the model to identify residential attractiveness and productivities in each location using commuting data and a maximum likelihood estimation strategy.

In what follows I make the following simplifications to the more general framework outlined above.

$$\mathcal{M}_1 = \{m_1\} \tag{37}$$

$$\mathcal{M}_0 = \{m_0\} \tag{38}$$

$$\mathbb{1}_{\mathrm{I}} = 1 \tag{39}$$

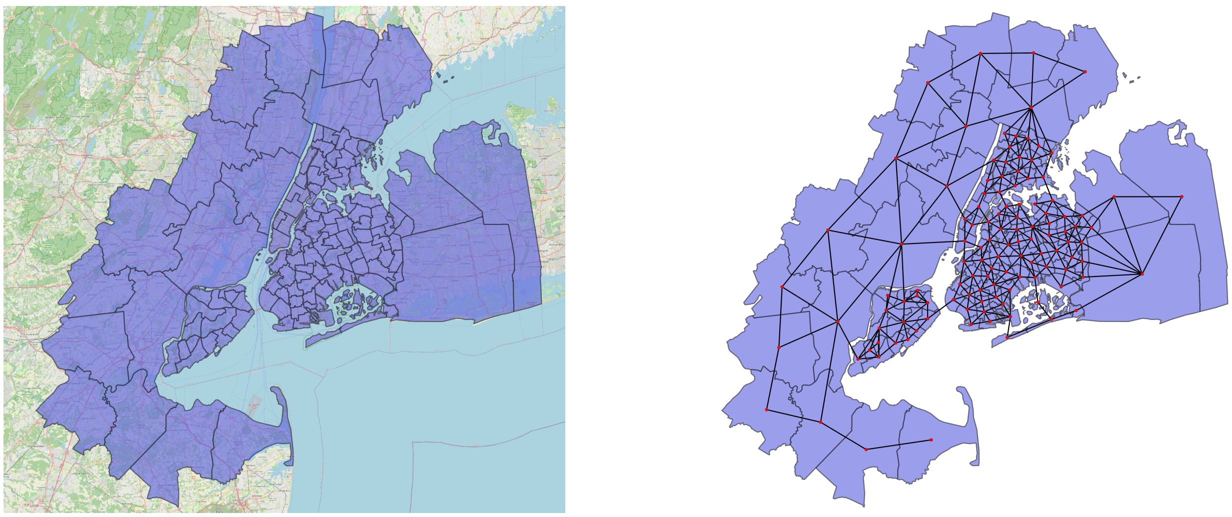
Equation (16) states that there is a single congested mode, which will represent driving with equation (17) representing the single uncongested mode: public transit. (18) implies that we will consider a model without externalities in home and work locations, focusing on the case where they are given by floor-space supply elasticities and internalized by agents. This allows me to isolate the effects of the key externality of the paper: traffic congestion. It also allows the results from Proposition 4 to be directly applied.

4.1 Data Sources

Taking the model to data requires mapping the discrete locations in the model to areas in and around New York City. Within the city itself, I take 263 taxi zones used by New York's Taxi and Limousine Company to describe neighborhoods. I aggregate these into 100 approximately equally sized areas which provide a level of spatial granularity that I can solve computationally. Outside of New York City I include all sub-counties within a 20km radius of the edge of the city. These capture areas in New York State and New Jersey in which a large number of commuters into New York City live. As these vary considerably in size, I again aggregate them to a total of 20 approximately equal zones. The resulting 120 locations are illustrated in panel (a) of Figure 1. The area accounts for the home locations of 89% of the workforce in New York City. Further details on the construction of these zones is provided in Appendix C.

To represent the roads in the area as a network, I use Open Street Maps data (OSM).

This provides the individual roads and junctions over the study area together with a range of road characteristics such as the direction, number of lanes, length and road and junction type. I extract a directed network from this data by placing a link from a source zone to a target if any of the OSM roads go in that direction between the two zones. The resulting network is displayed in panel (b) of Figure 1. Road and junction characteristics will also be used to estimate the congestion technology below.



(a) Zones (b) Road Network

Figure 1: New York City Zones and Road Network

To estimate the relationship between travel times by car and traffic flows I make use of two data sources. For travel speeds I use Google Maps API to provide the speeds at different hours of the day along different links in the network. I take the minimum speed observed at any hour of the day to be the free-flow speed along that link. For traffic flows I use the New York Department of Transportation's automated traffic volume counts between 2017 and 2019. These provide a measure of the flow of vehicles on over 2,000 road segments within New York City by hour of the day and across different dates in the year. I aggregate these to find the average volume per lane kilometer by hour of the day for each link in the network, which I match to the travel time data. Journey times by public transit are also obtained from Google Maps API. I query each origin-destination pair to find the fixed travel time between locations allowing for the use of all public transport options (including subway, buses and rail) as well as walking. These queries are made to achieve a 9am arrival time on Monday morning, corresponding to a typical morning rush-hour commute.

Finally, I make use of two commonly used data sources. For commuting I use the Longitudinal Employer-Household Dynamics, Origin-Destination Employment Statistics (LODES) from 2019. This provides commuting flows between home and work census tracts which I aggregate to the zones defined above. I also use the American Community Survey (ACS) data for 2019 to provide information on wages, hours worked per day, and transportation modes used when commuting. Further details on all aspects of the data construction process are provided in Appendix C.

4.2 Fixed and Calibrated Parameters

Translating time savings into dollars requires taking a stance on the value of time (VoT) for individuals: γ in the model. To obtain this quantity I use recent estimates of the value of time from Almagro et al. (2024). They provide estimates that vary across the income distribution for Chicago in 2010. In order to make these estimates applicable to New York City in 2019 and find a single value of time, I use their estimated parameters together with the mean income per capita for New York[13]. I deflate this to 2010 dollars to compute the implied value of time in 2010 dollars and then reinflate it to find its value in 2019[14]. I find a value of time of $25.92. I also draw on Almagro et al. (2024) to find the mode shock scale: ν. Using their estimates for peak-hour travel and converting into units of dollars using the

[13]Estimates from the transportation economics literature consistently find the value of time traveling to be substantially below the wage rate. Those reviewed in Small and Verhoef (2007) find it to be around half the wage. The estimates I use are close to this for New York.

[14]In the notation of Almagro et al. (2024) The value of time is given by $\alpha_T y^{1-\alpha_{py}}/\alpha_p$ where y is income and $\alpha_T, \alpha_{py}, \alpha_p$ are estimated parameters. This is the relationship I use, plugging in the mean value of y for New York City.

value of time estimate I find a value of $\nu = 8.869$.

To obtain values for α_d, β_o I make use of the floor-space supply micro-foundations presented in Appendix B, together with floor-space supply elasticity estimates that vary over space taken from Baum-Snow and Han (2024). The equations from the micro-foundations imply expressions for the parameters of the model as follows

$$\beta_o = \frac{-\delta^H}{1 + \left(\frac{\psi_o}{1-\psi_o}\right)}$$
$$\alpha_d = \frac{-\delta^F}{1 + \left(\frac{\psi_d}{1-\psi_d}\right)}$$

Where δ^H is a preference parameter determining household expenditure on floor-space, δ_F is a production parameter of firms determining their use of floor-space in production, and $\frac{\psi_o}{1-\psi_o}$ is the floor-space supply elasticity which I take from Baum-Snow and Han (2024) in each location. I calibrate δ^H to match the average expenditure share on housing from the BLS for New York City and New Jersey which is 0.391. For firms, I match the average expenditure share on floor-space to 0.2. These then recover α_d and β_o as given parameters.

The final set of parameters that I calibrate and fix are φ and σ. I calibrate φ so that the aggregate share of people who use public transport in the equilibrium of the transport model matches the data. I find a value of $\varphi = 1.047$ implying a higher utility cost of time spend on public transit relative to driving. This parameter ensures the model matches aggregate mode shares and therefore captures many effects that exist in reality but that are abstracted from in the model. For example, the user cost of owning and maintaining a car versus the monetary cost of using public transit are not directly modeled. Equally, the difficulty of parking in New York or the fact that driving may be more or less pleasant than riding the subway are not explicitly modeled. To the extent that these parameters affect the attractiveness of taking public transit relative to driving they will be reflected in, and partially captured by, φ. If these costs are proportional to time spent traveling φ is the correct way to model them. φ therefore provides a reduced form way to capture many omitted factors[15]. Finally, I fix $\sigma = 0.648$ which is small in magnitude relative to the other elasticities in the model. This leads to routing choices that approximate the case where travelers take the shortest time path through the network given the equilibrium traffic flows.

4.3 Congestion Technology Estimation

To estimate the congestion technology I make use of a commonly used functional form within the transport literature developed by the Bureau of Public Roads (BPR)[16]. On a particular link level ij, it relates the time taken at an hour of the day h, t_{ijh}, to the flow on the link at that time x_{ijh} through the following formula:

$$t_{ijh} = t_{ij}^{0} \left[1 + a_{ij} x_{ijh}^{b_{ij}}\right]$$

where t_{ij}^{0} represents the free-flow travel time across the link which I obtain directly as the minimum observed travel time across hours of the day from Google Maps API. The free-flow speed is scaled by an increasing function of the flow of traffic on the link so that journey times increase as flows increase. a_{ij} and b_{ij} are parameters governing the slope and curvature of the relationship and depend on features of the road network. The function is convex whenever $b_{ij} > 1$ and for large values of b_{ij} journey times can rise rapidly corresponding to bottlenecks in the network. Note that I only observe traffic flows, x_{ijh}, on a subset of links in the network where the Department of Transport has sensors measuring the counts. In particular, I do

[15] With disaggregated data on mode shares by origin and destination, it would be possible to be even more flexible, fitting a separate φ_{od} parameter for each pair. Calculating costs directly is an even more data intensive alternative and has been developed in great detail recently by Almagro et al. (2024).

[16] See Sheffi (1985) for a discussion of this and other parametric congestion functions.

not observe any flow measures outside of New York City. To make progress on this issue, I assume that the parameters are a function of link level observable variables, z_{ij} so that

$$
\begin{aligned}
a_{ij} &= \exp\left(g_0(z_{ij})\right) \ , \\
b_{ij} &= \exp\left(g_1(z_{ij})\right) \ .
\end{aligned}
$$

I use over 90 covariates including the number of intersections in each zone, the total length of roads, the latitude and longitude, and the types of roads and junctions in each zone. A full list is provided in Appendix C. Applying these to the BPR formula with an error term, η_{ij}, included gives

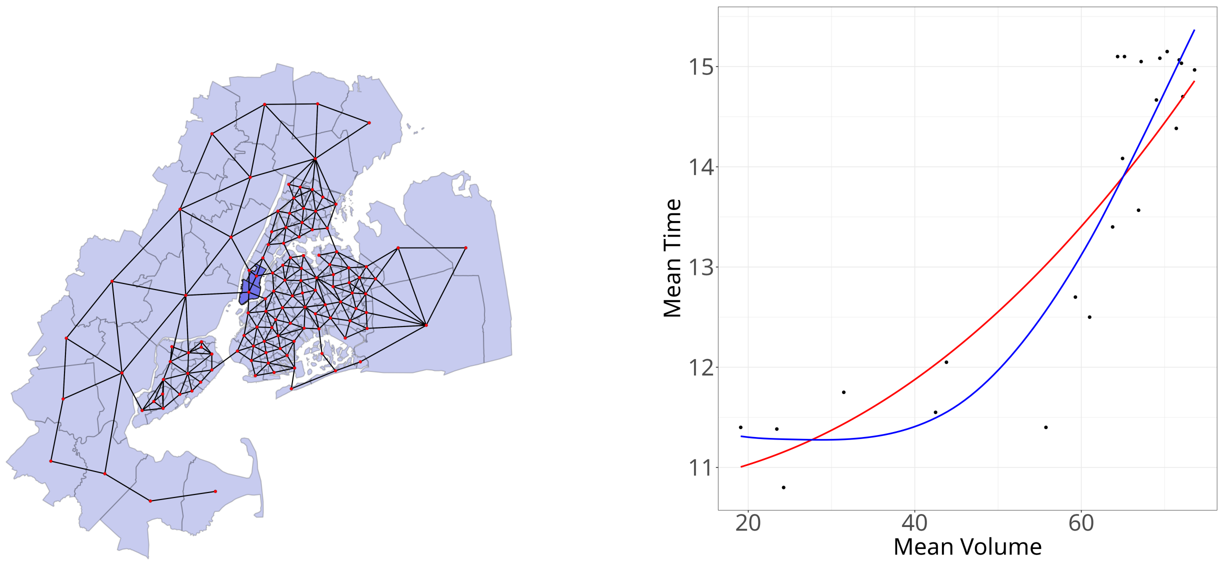
$$
t_{ijh} = t_{ij}^0 [1 + e^{g_0(z_{ij}) \eta_{ijh}} x_{ijh}^{g_1(z_{ij})}] \ . \tag{40}
$$

Rearranging this by noting that t_{ij}^0 is observable yields the estimating equation

$$
\ln\left(\frac{t_{ijh}}{t_{ij}^0} - 1\right) = g_0(z_{ij}) + g_1(z_{ij}) \ln x_{ijh} + \eta_{ijh} \ . \tag{41}
$$

This equation is of a form analyzed by Athey et al. (2019) and I use machine learning techniques, in particular generalized random forests to deal both with the high dimensional set of covariates to avoid over-fitting and to obtain a more precise prediction of the relationship between travel speeds and traffic flows. I take point estimates for $g_0(z_{ij})$ and $g_1(z_{ij})$ for all links in the network by extrapolating the fitted functions from links in which I observe traffic flows to those in which I do not. I do not consider inference for these quantities which is challenging in such settings.

Figure 2 shows the goodness of fit of the relationship between traffic flows and travel times on the particular link in lower Manhattan highlighted in panel (a). Panel (b) shows



(a) Link in Lower Manhattan (b) Goodness of Fit

Figure 2: Fit on a Single Link

the raw data as 24 points, one for each hour of the day for which the aggregated flows and speeds are observed. In panel (b), the red curve shows the estimates from the BPR formula using generalized random forests and the blue curve shows a non-parametric kernel fit using only the data on the link itself. The BPR formula captures the relationship well and as is further illustrated in Figure 3 which shows the predicted versus actual travel times for all links and times at which both flows and speeds are observed. The correlation coefficient between the two is 0.97.

The estimated values of $g_0(z_{ij})$ and $g_1(z_{ij})$ provide a mapping between the average flow per lane and travel times. The output of the model is not directly comparable to flows per lanes. In order to address this issue I scale the model implied flows x_{ij} to match the mean flow per lane on observed links at the morning rush hour time of 8-9:00am. This parameter is calibrated jointly with φ at a fixed value of σ by repeatedly solving the transportation sub-problem as defined in proposition 5. This calibration procedures means that I am able to match both the aggregate share of people using public transit (versus driving) as well as the mean flow per lane in the data.

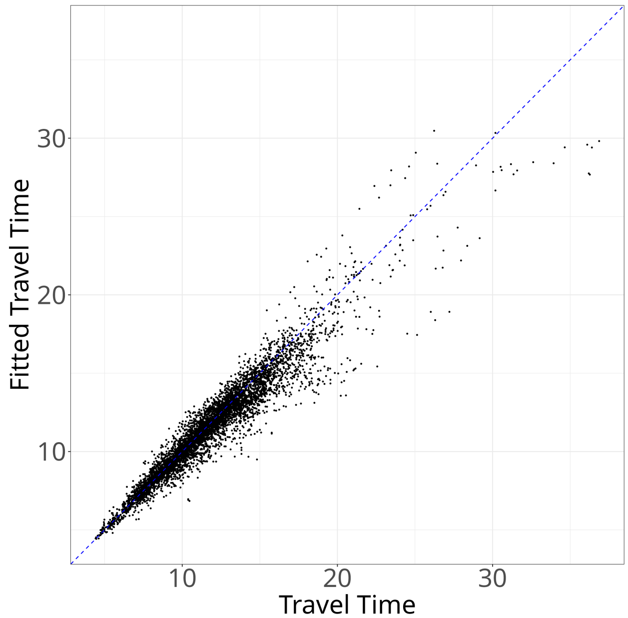


Figure 3: Overall Goodness of Fit for Congestion

There are at least two potential concerns with the estimation strategy developed above. Firstly, if η_{ij} is interpreted as measurement error in x_{ijh} then the estimates for $g_0()$ and $g_1(.)$ will generally be biased. There are reasons to suspect that measurement error may be a concern: only a subset of the roads within a zone are observed, and while the sample appears representative there may be issues with both selection and missing data. To gain some intuition on this issue, consider the case of classical measurement error in $\ln x_{ijh}$ and where equation (20) is estimated on a single link using OLS where g_0, g_1 are parameters and no longer a function of observables. In this case, g_1 will be biased towards zero leading to less curvature in the estimated congestion relationship than in reality. A second potential concern is that η_{ij} represents shocks to the congestion technology at different hours of the day which are correlated with the demand for traffic on the road $\ln x_{ijh}$ and observed by individuals leading to simultaneity bias. Considering again the case where equation (20) is estimated by OLS on a single link, it is reasonable to expect this correlation to be negative if drivers observe the congestion technology. As an example, suppose that it is slower to drive at night. It is reasonable to expect people to adjust to this by taking fewer trips during night-time hours than they otherwise would. Again, this would lead the estimated g_1 to be biased downwards.

The case where ϵ_{ijh} represents a shock to travel times that is uncorrelated with $\ln x_{ijh}$ leads to consistent estimates in (20). As argued above, the likely sign of bias in estimation is downwards. This intuition is supported by the IV strategies employed in Almagro et al. (2024) and Akbar and Duranton (2017) who both find higher elasticities of congestion when using IV compared to OLS.

4.4 Commuting: Maximum Likelihood Estimation

In this section I show how $\{A_d\}_{d\in\mathcal{D}}$, $\{u_o\}_{o\in\mathcal{O}}$ and θ can be recovered from the LODES commuting data using the structure of the model and the parameters $\sigma, \nu, \varphi, \gamma$ and $\{s_{ijm_1}(.)\}_{ij\in\mathcal{E}_{m_1}}$ obtained in the previous sections.

In the notation of the model, the number of people commuting between o and d is given by $\ell_{od} := \sum_{m\in\mathcal{M}}\sum_{r\in\mathcal{R}_{odm}}\ell_{odmr}$. Equations (1)-(4) and (12) imply that

$$\frac{\ell_{od}}{L} = \frac{\exp\left((u_o + A_d + \kappa_{od})/\theta\right)}{\sum_{o'd'\in\mathcal{O}\times\mathcal{D}}\left((u'_o + A'_d + \kappa_{o'd'})/\theta\right)} \quad \forall od \in \mathcal{O}\times\mathcal{D} \tag{42}$$

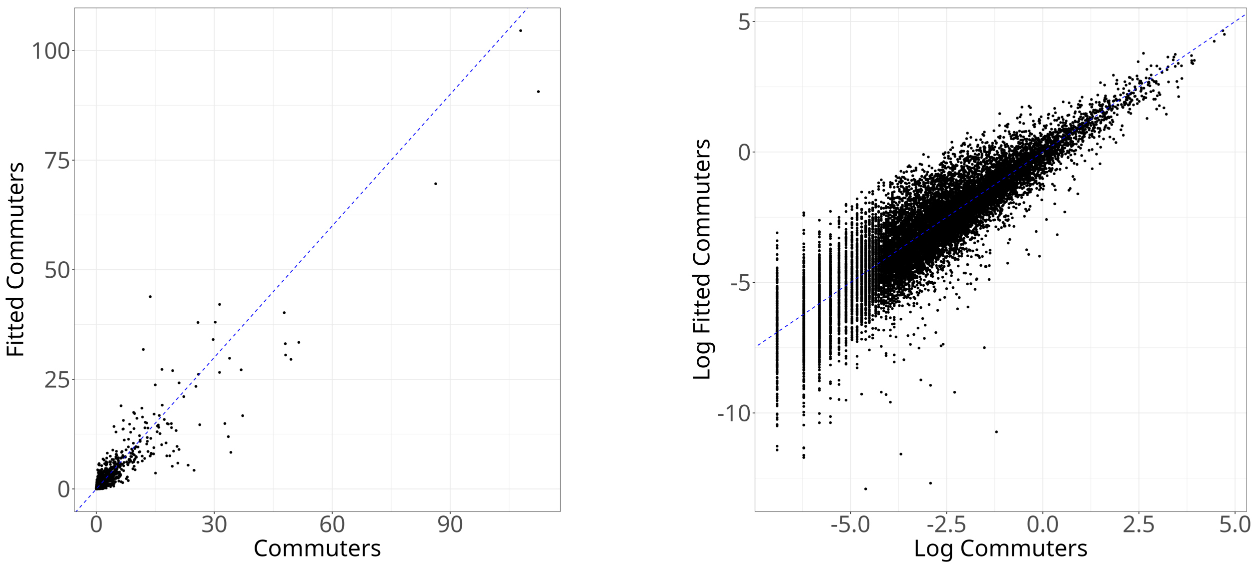
$$\kappa_{od} := \nu \ln\left[\exp\left(\frac{\phi\gamma}{\nu}\bar{t}_{od}\right) + \exp\left(\frac{\sigma}{\nu}\ln\sum_{r\in\mathcal{R}_{odm_1}}\exp\left(-\gamma t_{odm_1 r}\right)\right)\right] \tag{43}$$

Equation (42) gives the usual logit formula for choice probabilities over locations given deterministic utilities $u_o + A_d + \kappa_{od}$. The bilateral cost term κ_{od} in (43) accounts for the nesting over routes and modes using the usual log-sum-exp formulas. In particular, it gives the expected total utility inclusive of all preference heterogeneity over agents, of traveling between o and d. Importantly each term in this expression, apart from u_o, A_d, is known or can be computed from the data. $\nu, \sigma, \varphi, \gamma$ have already been fixed and $\bar{t}_{od}$ is data. $\sum_{r \in \mathcal{R}_{odm_1}} \exp\left(-\gamma t_{odm_1 r}\right)$ can be computed provided we know the equilibrium link level speeds t_{ijm_1} using an expansion in terms of powers of a weighted adjacency matrix over the network. This is formally described in Appendix B. I now turn to how u_o, A_d, and θ can be recovered from the data by using Proposition 5.

The LODES data provides a measure of commuting flows between origins and destinations which I denote $\tilde{\ell}_{od}$. I interpret as a finite, i.i.d., sample of size $N = \sum_{od \in \mathcal{O} \times \mathcal{D}} \tilde{\ell}_{od}$ from the measures $\{\ell_{od}\}_{od \in \mathcal{O} \times \mathcal{D}}$ of people commuting in the continuum model which has a fixed total mass L. In an asymptotic regime in which $N \to \infty$ while L remains fixed, the weak law of large numbers then implies that $\frac{\tilde{\ell}_{od}}{N} \xrightarrow{p} \frac{\ell_{od}}{L}$.

Given Proposition 5, fixing the values of ℓ_{od} at the observed levels from the LODES data $\tilde{\ell_{od}}$, I solve the transport sub-model given the parameters $\sigma, \nu, \varphi, \gamma, \{s_{ijm_1}(.)\}_{ij \in \mathcal{E}_{m_1}}$. This provides me with an estimate of the traffic flows on each driving link, x_{ijm_1}, together with the required travel times along particular driving links t_{ijm_1}. As all functions in the model are smooth, the continuous mapping theorem suggests that these estimates should be consistent as N grows large given the asymptotic regime outlined in the previous paragraph. I do not provide a formal proof of this. Equipped with a measure of t_{ijm_1}, equations (42) and (43) can be estimated from the data using maximum likelihood and replacing u_o/θ and A_d/θ with fixed effects. In practice, this can be implemented by a Poisson regression, once we have computed κ_{od} as is standard in the urban-spatial literature. The coefficient on κ_{od} will give us $1/\theta$. The fixed effects, A_d and u_o, are then identified up to an additive constant as is the case in general in discrete choice models: the level of deterministic utility of all options can be shifted up or down by any constant without changing the choice probabilities. The resulting estimate is $\theta = 21.603$ and so the parameters in the model satisfy the requirement that $\theta > \nu > \sigma$ as required for GEV.

In order to pin down the levels of A_d and u_o from the estimated fixed effects I make use of outside data from the ACS. I set the level of wages $w_d = A_d$ so that the mean daily wages for the study area in the model match those observed for the same area in the ACS data. In particular I add a constant to across locations to u_o and A_d to ensure that $\sum_{d \in \mathcal{D}} A_d \ell_d^F / L$ matches the values observed in the data.



(a) Commuting Fit in Levels

(b) Commuting Fit in Logs

Figure 4: Goodness of Fit for Commuting

Figure 4 shows the fit of the commuting model. Panel (a) shows the fit in levels and the correlation between fitted and actual commuting flows is 0.926. Panel (b) shows the fit in logs for ease of visualization given the large range of the data, dropping zeros. It is worth noting that at the level of aggregation considered in this paper, 96.9% of origin-destination pairs have non-zero flows.

5 Conclusion

In this paper, I develop a general equilibrium urban model with location, transport mode and route choice to study the welfare effects of congestion pricing. Travel times and the attractiveness of locations as home or workplaces are determined endogenously as a result of agents' decisions. I study the competitive equilibrium for the economy with or without taxes as well as a utilitarian planner's problem whose solution can be fully decentralized and provide conditions for the existence and uniqueness of solutions to these problems within a unified framework. I also develop an algorithm to solve the model numerically.

I apply the framework to study the welfare effects of congestion pricing in New York, fitting model parameters using the rich data available there. I find large gains from optimal congestion pricing, with the first best taxes yielding gains from time saved equivalent to $21.7 million dollars per week. These gains are mediated through individuals' choices of location, mode and route in the model. The results suggest that in New York, where there is a highly developed system of public transit, substitution across modes is of first order importance for welfare with location decisions and drivers' re-routing behavior playing a more modest but non-negligible role.

The model suggests that toll zones of the kind proposed in Manhattan have the potential to substantially increase welfare over the status quo. Over a third of the gains from optimal pricing can be achieved by a toll zone covering only lower Manhattan. However, this policy leaves much scope for improvement. The space of potential second best policy instruments is large and it would be of practical policy relevance to find alternative policies that do better.

A number of extensions to the model would be interesting to consider from both a practical and theoretical perspective. Firstly, incorporating different types of agents, with different income levels and values of time, would allow the study of distributional implications of congestion pricing which may interact in novel ways with the sorting of individuals across space through their home and workplace choices. Secondly, while the model has focused on commuting, non-commuting trips account for a large share of journeys made within New York and their importance has recently been highlighted by Miyauchi et al. (2021). How these trips interact with endogenous congestion may yield further insights into the impacts and potential benefits from congestion pricing. Thirdly, the model is fully static. The substitution of trips over different hours of the day, as studied by Kreindler (2024), provides a further margin through which congestion pricing affects welfare. How this interacts with location, mode and route choice in full general equilibrium has yet to be fully studied.

Finally, the work has relevance beyond the case of traffic congestion. Models with distortions that occur through a network structure are pervasive in economics. For example, congestion and agglomeration externalities operating through connections in space are central to much of the urban and spatial economics literature. Equally, models of supply chain networks often feature externalities between different producers. Casting the equilibria of these economies as the solution to an optimization problem offers a different point of view to study them, one which may generate novel insights. The tools developed in this paper could therefore have scope to be applied more broadly and, I believe, offer a promising avenue for future research.

Bibliography

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1988). Network flows.

Akamatsu, T. (1996). Cyclic flows, markov process and stochastic traffic assignment. *Transportation Research Part B: Methodological*, 30(5):369–386.

Akamatsu, T. (1997). Decomposition of path choice entropy in general transport networks. *Transportation Science*, 31(4):349–362.

Akbar, P. and Duranton, G. (2017). Measuring the cost of congestion in highly congested city: Bogotá.

Allen, T. and Arkolakis, C. (2022). The welfare effects of transportation infrastructure improvements. *The Review of Economic Studies*.

Almagro, M., Barbieri, F., Castillo, J. C., Hickok, N. G., and Salz, T. (2024). Optimal urban transportation policy: Evidence from chicago. Technical report, National Bureau of Economic Research.

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1992). *Discrete choice theory of product differentiation*. MIT press.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests.

Baillon, J.-B. and Cominetti, R. (2008). Markovian traffic equilibrium. *Mathematical Programming*, 111(1-2):33–56.

Barwick, P. J., Li, S., Waxman, A., Wu, J., and Xia, T. (2024). Efficiency and equity impacts of urban transportation policies with equilibrium sorting. *American Economic Review*, 114(10):3161–3205.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The quarterly journal of economics*, 122(2):775–805.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017). Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448.

Baum-Snow, N. and Han, L. (2024). The microgeography of housing supply. *Journal of Political Economy*, 132(6):1897–1946.

Beckmann, M., McGuire, C. B., and Winsten, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press.

Billings, S. B. (2011). Estimating the value of a new transit option. *Regional Science and Urban Economics*, 41(6):525–536.

Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, environment and urban systems*, 65:126–139.

Bordeu, O. (2023). Commuting infrastructure in fragmented cities. *Job Market Paper, University of Chicago Booth School of Business*, 6.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Brinkman, J. and Lin, J. (2024). Freeway revolts! the quality of life effects of highways. *Review of Economics and Statistics*, 106(5):1268–1284.

Cameron, S., Chaudhuri, S., and McLaren, J. (2007). Trade shocks and labor adjustment: theory.

Dafermos, S. (1980). Traffic equilibrium and variational inequalities. *Transportation science*, 14(1):42–54.

Dingel, J. I. and Tintelnot, F. (2020). Spatial economics for granular settings. Technical report, National Bureau of Economic Research.

Donald, E., Fukui, M., and Miyauchi, Y. (2023). Unpacking aggregate welfare in a spatial economy.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, volume 4, pages 2063–2117. Elsevier.

Durrmeyer, I. and Martinez, N. (2022). The welfare consequences of urban traffic regulations.

Fajgelbaum, P. D. and Gaubert, C. (2020). Optimal spatial policies, geography, and sorting. *The Quarterly Journal of Economics*, 135(2):959–1036.

Fajgelbaum, P. D. and Schaal, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, 88(4):1411–1452.

Fujita, M., Krugman, P. R., and Venables, A. (2001). *The spatial economy: Cities, regions, and international trade.* MIT press.

Galichon, A. and Salanié, B. (2022). Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629.

Gibbons, S. and Machin, S. (2005). Valuing rail access using transport innovations. *Journal of urban Economics*, 57(1):148–169.

Gonzalez-Navarro, M. and Turner, M. A. (2018). Subways and urban growth: Evidence from earth. *Journal of Urban Economics*, 108:85–106.

Graham, M. R., Kutzbach, M. J., and McKenzie, B. (2014). Design comparison of lodes and acs commuting data products. Technical report.

Hofbauer, J. and Sandholm, W. H. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294.

Kehoe, T. J., Levine, D. K., and Romer, P. M. (1992). On characterizing equilibria of economies with externalities and taxes as solutions to optimization problems. *Economic Theory*, 2:43–68.

Kreindler, G. (2024). Peak-hour road congestion pricing: Experimental evidence and equilibrium implications. *Econometrica*, 92(4):1233–1268.

Maher, M., Stewart, K., and Rosa, A. (2005). Stochastic social optimum traffic assignment. *Transportation Research Part B: Methodological*, 39(8):753–767.

Miyauchi, Y., Nakajima, K., and Redding, S. J. (2021). Consumption access and agglomeration: evidence from smartphone data. Technical report, National Bureau of Economic Research.

Oyama, Y., Hara, Y., and Akamatsu, T. (2022). Markovian traffic equilibrium assignment based on network generalized extreme value model. *Transportation Research Part B: Methodological*, 155:135–159.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9(1):21–58.

Rudin, W. (1976). Principles of mathematical analysis. *3rd ed.*

Samuelson, P. A. (1947). Foundations of economic analysis.

Sheffi, Y. (1985). *Urban transportation networks*, volume 6. Prentice-Hall, Englewood Cliffs, NJ.

Small, K. and Verhoef, E. T. (2007). *The economics of urban transportation*. Routledge.

Takayama, A. (1985). *Mathematical Economics*. Cambridge university press.

Traffic Mobility Review Board (2023). Congestion pricing in new york: A toll structure recommendation from the traffic mobility review board. Technical report, Traffic Mobility Review Board. Available at `https://new.mta.info/document/127761`.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Tsivanidis, N. (2023). Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio. *University of California, Berkeley*.

A Proofs

Throughout the proofs I freely use basic facts about concave functions. For background material on convex optimization see Boyd and Vandenberghe (2004).

Lemma 1.

The proof proceeds in two steps. Firstly, I show that the planner is indifferent between different allocations of the consumption good, c, and I note how this simplifies the problem. Secondly, I fix ℓ_{odmr} and use the results from Galichon and Salanié (2022) to show how this simplifies the problem to the desired form by solving out for the function $a(.)$. Thirdly, I show that the problem can be expressed purely as a function of $\boldsymbol{\ell} := (\ell_{odmr})_{odmr\in\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$.

First note that independently of the choice of the assignment $a(.)$, given the constraints of the model we will have that:

$$
\begin{aligned}
&L\int_{\mathbb{R}^{\mathcal{S}}} \tilde{c}(\bar{\varepsilon}) f_{\varepsilon}(\bar{\varepsilon}) d\bar{\varepsilon} = c \\
&\qquad\qquad\qquad = \sum_{d\in\mathcal{D}} (\bar{A}_d + \alpha_d \ln \ell_d^F)\ell_d^F + \mathbb{1}_{\mathrm{I}}\left[-\sum_{d\in\mathcal{D}} \ell_d^F \alpha_d - \sum_{o\in\mathcal{O}} \ell_o^H \beta_o\right]
\end{aligned}
$$

This shows that the impact on the objective of the allocation of c is independent of $a()$ provided that the constraints are satisfied. This holds because the planner is utilitarian with utility linear in consumption: they care only about productive efficiency, not how utility is divided.

Secondly, consider fixing $\{\ell_{odmr}\}$ at a particular set of values. Note that, the constraints of the model pin down all other variables apart from the assignment function $a(.)$. In particular $u_o = u_o(\{\ell_{odmr}\})$ and $t_{odmr} = t_{odmr}(\{\ell_{odmr}\})$. The objective function for the planner, ignoring the fixed $L \int_{\mathbb{R}^{\mathcal{S}}} \tilde{c}(\bar{\varepsilon}) f_\varepsilon(\bar{\varepsilon}) d\bar{\varepsilon}$ term which is unaffected by $a(.)$ becomes:

$$\max_{a \in \mathcal{A}} \quad L \int_{\mathbb{R}^{\mathcal{S}}} \left(u_{\tilde{o}(\bar{\varepsilon})} - \gamma t_{\tilde{o}(\bar{\varepsilon})\tilde{d}(\bar{\varepsilon})\tilde{m}(\bar{\varepsilon})\tilde{r}(\bar{\varepsilon})} + \bar{\varepsilon}_{\tilde{o}(\bar{\varepsilon})\tilde{d}(\bar{\varepsilon})\tilde{m}(\bar{\varepsilon})\tilde{r}(\bar{\varepsilon})} \right) f_\varepsilon(\bar{\varepsilon}) d\bar{\varepsilon}$$

s.t.

$$\int_{\mathbb{R}^{\mathcal{S}}} \mathbb{1} \left\{ \left(\tilde{o}(\bar{\varepsilon}), \tilde{m}(\bar{\varepsilon}), \tilde{d}(\bar{\varepsilon}), \tilde{r}(\bar{\varepsilon}) \right) = (o, d, m, r) \right\} f_\varepsilon(\bar{\varepsilon}) d\bar{\varepsilon} = \ell_{odmr}/L \qquad \forall odmr \in \mathcal{S}$$

Where u_o and t_{odmr} are understood to be functions of the fixed value of ℓ_{odmr}. This is an optimal transport problem. For this proof only, redefine $v_{odmr} := u_o + -\gamma t_{odmr}$, again understood as a function of ℓ_{odmr} and collect the full vector in $\mathbf{v}$.

The value of the problem above is equivalent to maximizing $v_{odmr} + \varepsilon_{odmr}$ point-wise for each realization of ε. That is, it is equivalent to

$$L \int_{\mathbb{R}^{\mathcal{S}}} \left(\max_{odmr} \{ v_{odmr} + \bar{\varepsilon}_{odmr} \} \right) f(\bar{\varepsilon}) d\bar{e}$$

Following Galichon and Salanié (2022) define

$$G(\mathbf{v}) = \int_{\mathbb{R}^{\mathcal{S}}} \left(\max_{odmr} \{ v_{odmr} + \bar{\varepsilon}_{odmr} \} \right) f(\bar{\varepsilon}) d\bar{\varepsilon}$$

By well known results in discrete choice for GEV distributions we know that, up to an additive constant[17]:

$$G(\mathbf{v}) = \theta \ln \sum_{od} \exp \left(\left[\nu \sum_{m} \exp \left(\left(\sigma \ln \sum_{r} \exp(v_{odmr}/\sigma) \right) / \nu \right) \right] / \theta \right)$$

Again following Galichon and Salanié (2022) define the Legendre-Fenchel transform of G

[17] Throughout I neglect constant terms involving the Euler-Mascheroni constant as they do not affect any of the optimization problems or changes across counterfactuals

for $\mathbf{q} \in \mathbb{R}^{\mathcal{S}}$

$$G^*(\mathbf{q}) = \begin{cases} \sup_{\tilde{\mathbf{v}}} \sum_{odmr} \tilde{v}_{odmr} q_{odmr} - G(\tilde{\mathbf{v}}) & \text{if } \sum_{odmr} q_{odmr} \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

Following the duality arguments in section 3 and Appendix B.1[18] of Galichon and Salanié (2022) we know that for any vector $\mathbf{q} \in \mathbb{R}_+^{\mathcal{S}}$ such that $\sum_{odmr} q_{odmr} = 1$:

$$-G^*(\mathbf{q}) = \theta \ln \sum_{od} \exp\left(\left[\nu \sum_m \exp\left(\left(\sigma \ln \sum_r \exp(\tilde{v}_{odmr}(\mathbf{q})/\sigma)\right)/\nu\right)\right]/\theta\right) - \sum_{odmr} q_{odmr} \tilde{v}_{odmr}(\mathbf{q})$$

Where $\tilde{v}_{odmr}(\mathbf{q})$ solves the system of equations:

$$q_{odmr} = \frac{\partial}{\partial \tilde{v}_{odmr}} \theta \ln \sum_{od} \exp\left(\left[\nu \sum_m \exp\left(\left(\sigma \ln \sum_r \exp(\tilde{v}_{odmr}/\sigma)\right)/\nu\right)\right]/\theta\right)$$

This characterizes $G^*(\mathbf{q})$ as an implicit function of $\mathbf{q}$ I now solve for this function in closed form. I make the following definitions:

$$q_{odm} = \sum_r q_{odmr}$$

$$q_{od} = \sum_m q_{odm}$$

$$\tilde{v}_{odm} = \sigma \ln \sum_r \exp(\tilde{v}_{odmr}/\sigma)$$

$$\tilde{v}_{od} = \nu \ln \sum_m \exp(\tilde{v}_{odm}/\nu)$$

[18] Note that there is a sign error in equation B.3 of their Appendix B

$$\bar{v} = \theta \ln \sum_{od} \exp(v_{od}/\theta)$$

Now, taking the derivative above we have the usual nested logit probability formula:

$$q_{odmr} = \frac{\exp(\tilde{v}_{od}/\theta)}{\sum_{o'd'} \exp(\tilde{v}_{o'd'}/\theta)} \frac{\exp(\tilde{v}_{odm}/\nu)}{\sum_{m'} \exp(\tilde{v}_{odm'}/\nu)} \frac{\exp(\tilde{v}_{odmr}/\sigma)}{\sum_{r'} \exp(\tilde{v}_{odmr'}/\sigma)}$$

From this, after some algebra, we obtain:

$$\begin{aligned}
\tilde{v}_{od} &= \theta \ln q_{od} + \bar{v} \\
\tilde{v}_{odm} &= \nu \ln q_{odm} + (\theta - \nu) \ln q_{od} + \bar{v} \\
\tilde{v}_{odmr} &= \sigma \ln q_{odmr} + (\nu - \sigma) \ln q_{odm} + (\theta - \nu) \ln q_{od} + \bar{v}
\end{aligned}$$

Plugging these expressions back into the expression for $-G^*(\mathbf{q})$ and simplifying, the terms involving $\bar{v}$ cancel and we obtain an expression purely in terms of $\mathbf{q}$:

$$-G^*(\mathbf{q}) = -(\theta - \nu) \sum_{od} q_{od} \ln q_{od} - (\nu - \sigma) \sum_{odm} q_{odm} \ln q_{odm} - \sigma \sum_{odmr} q_{odmr} \ln q_{odmr}$$

At the optimum, equations (3.2) and (3.5) from Galichon and Salanié (2022) imply that:

$$G(\mathbf{v}) = \sum_{odmr} v_{odmr} p_{odmr} - G^*(\mathbf{p})$$

Where $p_{odmr} := \ell_{odmr}/L$ is the share implied by the fixed value of ℓ_{odmr}. Together with the closed form expression for $G^*(\mathbf{p})$, this shows that we have reduced the problem of finding an assignment function $a(.)$ by solving the inner optimization problem first. Plugging the definitions back in and simplifying gives[19]:

$$LG(\mathbf{v}(\{\ell_{odmr}\})) = \sum_d \ell_o^H u_o - \gamma \sum_{odmr} t_{odmr}\ell_{odmr}$$
$$- (\theta - \nu) \sum_{od} \ell_{od} \ln \ell_{od} - (\nu - \sigma) \sum_{odm} \ell_{odm} \ln \ell_{odm} - \sigma \sum_{odmr} \ell_{odmr} \ln \ell_{odmr}$$

Combining this with the expression for $L \int_{\mathbb{R}^{\mathcal{S}}} \tilde{c}(\bar{\varepsilon}) f_\varepsilon(\bar{\varepsilon}) d\bar{\varepsilon}$ derived above and substituting in the constraints for u_o, A_d gives the following formulation of the planner's problem:

$$\max_{\substack{\{\ell_{odmr}\},\{\ell_d^F\},\{\ell_o^H\},\\ \{x_{ijm}\},\{t_{ijm}\},\{t_{odmr}\}}} \sum_{o\in\mathcal{O}} \ell_o^H [\bar{u}_o - \mathbb{1}_{\mathrm{I}}\beta_o + \beta_o \ln \ell_o^H] + \sum_{d\in\mathcal{D}} \ell_d^F [\bar{A}_d - \mathbb{1}_{\mathrm{I}}\alpha_d + \alpha_d \ln \ell_d^F]$$
$$- \gamma \sum_{odmr\in\mathcal{S}} \ell_{odmr} t_{odmr}$$
$$- \theta \sum_{od\in\mathcal{O}\times\mathcal{D}} \ell_{od} \ln \ell_{od} - \nu \left[\sum_{m\in\mathcal{M}} \sum_{od\in\mathcal{O}\times\mathcal{D}} \ell_{odm} \ln \ell_{odm} - \sum_{od} \ell_{od} \ln \ell_{od} \right]$$
$$- \sigma \left[\sum_{od\in\mathcal{O}\times\mathcal{D}} \sum_{m\in\mathcal{M}} \sum_{r\in\mathcal{R}_{odm}} \ell_{odmr} \ln \ell_{odm_1r} - \sum_{od\in\mathcal{O}\times\mathcal{D}} \sum_{m\in\mathcal{M}} \ell_{odm} \ln \ell_{odm} \right]$$

s.t.

$$\sum_{odr\in\mathcal{S}} \ell_{odmr} = L$$
$$\sum_{m\in\mathcal{M}} \sum_{dr\in\mathcal{D}\times\mathcal{R}_{odm}} \ell_{odmr} = \ell_o^H \quad \forall o \in \mathcal{O}$$
$$\sum_{m\in\mathcal{M}} \sum_{or\in\mathcal{O}\times\mathcal{R}_{odm}} \ell_{odmr} = \ell_d^F \quad \forall d \in \mathcal{D}$$

[19] Again neglecting additive constants.

$$\sum_{m\in\mathcal{M}}\sum_{r\in\mathcal{R}_{odm}}\ell_{odmr}=\ell_{od}\quad\forall od\in\mathcal{O}\times\mathcal{D}$$

$$\sum_{m\in\mathcal{M}}\sum_{r\in\mathcal{R}_{odm}}\ell_{odmr}=\ell_{odm}$$

$$t_{odmr}=\begin{cases}\sum_{ij\in\mathcal{E}_m}n^{odm}_{ij,r}t_{ijm} & \text{if } m\in\mathcal{M}_1\\ \varphi\bar{t}_{odm} & \text{if } m\in\mathcal{M}_0\end{cases}\quad\forall odmr\in\mathcal{S}$$

$$x_{ijm}=\sum_{od\in\mathcal{O}\times\mathcal{D}}\sum_{r\mathcal{R}_{odm}}n^{odm}_{ij,r}\ell_{odmr}\qquad\forall m\in\mathcal{M}_1,\forall ij\in\mathcal{E}_m$$

$$t_{ijm}=s_{ijm}(x_{ijm})\qquad\forall m\in\mathcal{M}_1,\forall ij\in\mathcal{E}_m$$

I now turn to the third step of the proof: showing that this program can be expressed only as a function of $\boldsymbol{\ell}\in\mathbb{R}^{\mathcal{S}}$.

First note that by the constraints:

$$\begin{aligned}\sum_{od}\sum_{m\in\mathcal{M}_1}\sum_{r}t_{odmr}\ell_{odmr}&=\sum_{od}\sum_{m\in\mathcal{M}_1}\sum_{r}\left(\sum_{ij\in\mathcal{E}_m}n^{odm}_{ij,r}t_{ijm}\right)\ell_{odmr}\\&=\sum_{m\in\mathcal{M}_1}\sum_{ij\in\mathcal{E}_m}\left(\sum_{o}d\sum_{r}n^{odm}_{ij,r}\ell_{odmr}\right)t_{ijm}\\&=\sum_{m\in\mathcal{M}_1}\sum_{ij\in\mathcal{E}_m}x_{ijm}s_{ijm}(x_{ijm})\end{aligned}$$

This result simply states that the total time spent traveling through the network can be computed either by summing flows multiplied by times along routes or along links. The latter is a more useful representation.

For uncongested routes we simply have:

$$\sum_{od} \sum_{m \in \mathcal{M}_0} \sum_{r} t_{odmr} \ell_{odmr} = \sum_{od} \sum_{m \in \mathcal{M}_0} \ell_{odm} \varphi \bar{t}_{odm}$$

With these results in hand we are able to restate the problem as an optimization problem over only $\boldsymbol{\ell}$ with a single constraint[20]. Using definitions (16)-(22) from the main text and substituting in to the program above gives the problem in the desired form. □

Proposition 1.

Existence The feasible set is compact since $[0, L]^{\mathcal{S}}$ is the product of closed and bounded intervals and therefore compact. The intersection of this with a linear constraint is also compact. All functions in the objective function are continuous. By Weierstrass' theorem (see for example, Rudin (1976) p.89-90) the objective function attains a maximum.

Uniqueness Uniqueness proceeds by considering each of the pieces of the objective function in turn and arguing that they are concave or strictly concave in $\boldsymbol{\ell}$ so that the resulting function is strictly concave over a convex set and any solution is unique. The convexity of the feasible set follows immediately from the linearity of the constraint and the fact that the bounds on $\boldsymbol{\ell}$ are the product of closed and bounded intervals. Recall again that concave transformations of a concave function are concave.

First consider the term $-\gamma \sum_{m \in \mathcal{M}_1} \sum_{ij \in \mathcal{E}_m} x_{ijm}(\boldsymbol{\ell}) s_{ijm}(x_{ijm}(\boldsymbol{\ell}))$. Since $x_{ijm} s_{ijm}(x_{ijm})$ is everywhere convex by assumption, we have the sum of convex functions which is convex.

[20] I define $x \ln x := \lim_{x \to 0} x \ln x = 0$ so that the function is defined over the entire domain. Note however that a corner will never be the optimum as the marginal benefit from any $x \ln x$ term approaches infinity as x approaches 0. This in turn implies that the other corner at L will never be optimal.

The negative sign makes it concave and the linear aggregation of $\boldsymbol{\ell}$ preserves concavity so this term is concave in $\boldsymbol{\ell}$

Now turn to $-\gamma \sum_{od} \sum_{m \in \mathcal{M}_0} \varphi \bar{t}_{odm} \ell_{odmr}$ which is linear and therefore concave.

Next consider the term $-(\nu - \sigma) \sum_{odm} \ell_{odm}(\boldsymbol{\ell}) \ln (\ell_{odm}(\boldsymbol{\ell})) - \sigma \sum_{odmr} \ell_{odmr} \ln \ell_{odmr}$. First splitting this term up in terms of $\mathcal{M}_1, \mathcal{M}_0$ and noting that for $m \in \mathcal{M}_0$ there is no route choice so that $\ell_{odmr} = \ell_{odm}$ we find that the term simplifies to:

$$-(\nu - \sigma) \sum_{od} \sum_{m \in \mathcal{M}_1} \ell_{odm}(\boldsymbol{\ell}) \ln \ell_{odm}(\boldsymbol{\ell}) - \nu \sum_{od} \sum_{m \in \mathcal{M}_o} \ell_{odmr} \ln \ell_{odmr} - \sigma \sum_{m \in \mathcal{M}_1} \sum_{r} \ell_{odmr} \ln \ell_{odmr}$$

Consider the first term in this expression. $-(\nu - \sigma)\ell_{odm} \ln \ell_{odm}$ is concave in ℓ_{odm} for all $m \in \mathcal{M}_1$ and all od. The first term is therefore concave. I now show that the second term is *strictly concave in the full vector* $\boldsymbol{\ell}$. I show this by a second derivative test considering the Hessian of the function. Firstly note that all cross partial terms of the Hessian are zero: it is a diagonal matrix. It suffices to show that the diagonal entries are strictly negative. For $m \in \mathcal{M}_1$ the second derivative with respect to ℓ_{odmr} is given by $\frac{-\sigma}{\ell_{odmr}} < 0$ for $\ell_{odmr} > 0$. For $m \in \mathcal{M}_0$ the second derivative with respect to ℓ_{odmr} is given by $\frac{-\nu}{\ell_{odmr}} < 0$ for $\ell_{odmr} > 0$. The function is therefore strictly concave on the interior of the domain and we have already noted that the optimum cannot be at a corner since the marginal benefit of increasing ℓ_{odmr} approaches infinity as ℓ_{odmr} approaches zero.

The final terms to consider are:

$$\sum_{o} \left(\sum_{d} \ell_{od} \right) \left[\bar{u}_o - \mathbb{1}_{\mathrm{I}} \beta_o + \beta_o \ln \left(\sum_{d} \ell_{od} \right) \right] + \sum_{d} \left(\sum_{o} \ell_{od} \right) \left[\bar{A}_d - \mathbb{1}_{\mathrm{I}} \alpha_d + \alpha_d \ln \left(\sum_{o} \ell_{od} \right) \right]$$
$$- (\theta - \nu) \sum_{od} \ell_{od} \ln (\ell_{od})$$

Note that the terms involving $\bar{u}_o, \mathbb{1}_{\mathrm{I}}\beta_0, \bar{A}_d, \mathbb{1}_{\mathrm{I}}\alpha_d$ are linear in ℓ_{od} and so concave and may be ignored. it suffices to show that the following function is concave:

$$\sum_o \beta_o \left(\sum_d \ell_{od} \right) \ln \left(\sum_d \ell_{od} \right) + \sum_d \alpha_d \left(\sum_o \ell_{od} \right) \ln \left(\sum_o \ell_{od} \right)$$
$$- (\theta - \nu) \sum_{od} \ell_{od} \ln (\ell_{od})$$

Now note that if $\alpha_d \leq 0$ or $\beta_o \leq 0$ any of the terms involving those values will be concave, since, for example $\beta_o (\sum_d \ell_{od}) \ln (\sum_d \ell_{od})$ is concave in $\{\ell_{od}\}$ whenever $\beta_o \leq 0$. Such terms may be ignored from the function since they are already concave. Keeping only terms where $\alpha_d > 0$ and $\beta_o > 0$ by using indicator functions, the function that we have to show is concave is:

$$f(\{\ell_{od}\}) := \sum_o \mathbb{1}_{\{\beta_o > 0\}} \beta_o \left(\sum_d \ell_{od} \right) \ln \left(\sum_d \ell_{od} \right) + \sum_d \mathbb{1}_{\{\alpha_d > 0\}} \alpha_d \left(\sum_o \ell_{od} \right) \ln \left(\sum_o \ell_{od} \right)$$
$$- (\theta - \nu) \sum_{od} \ell_{od} \ln (\ell_{od})$$

Again I proceed with a second derivative test based on the Hessian. First evaluate the partial derivatives that make up the gradient vector:

$$\frac{\partial f}{\partial \ell_{od}} (\{\ell_{\tilde{o}\tilde{d}}\}) = \mathbb{1}_{\{\beta_o > 0\}} \beta_o + \mathbb{1}_{\{\beta_o > 0\}} \beta_o \ln \left(\sum_d \ell_{od} \right)$$
$$+ \mathbb{1}_{\{\alpha_d > 0\}} \alpha_d + \mathbb{1}_{\{\alpha_d > 0\}} \alpha_d \ln \left(\sum_o \ell_{od} \right) - (\theta - \nu) (1 + \ln \ell_{od})$$

Now turn to the Hessian:

$$\frac{\partial^2 f}{\partial \ell_{od} \partial \ell_{o'd'}}(\{\ell_{\tilde{o}\tilde{d}}\}) = \begin{cases} \frac{\mathbb{1}_{\{\beta_o>0\}}\beta_o}{\sum_{\tilde{d}} \ell_{o\tilde{d}}} + \frac{\mathbb{1}_{\{\alpha_d>0\}}\alpha_d}{\sum_{\tilde{o}} \ell_{\tilde{o}d}} - \frac{\theta-\nu}{\ell_{od}} & \text{if } o = o',\ d = d' \\ \frac{\mathbb{1}_{\{\beta_o>0\}}\beta_o}{\sum_{\tilde{d}} \ell_{o\tilde{d}}} & \text{if } o = o',\ d \neq d' \\ \frac{\mathbb{1}_{\{\alpha_d>0\}}\alpha_d}{\sum_{\tilde{o}} \ell_{\tilde{o}d}} & \text{if } o \neq o',\ d = d' \\ 0 & \text{if } o \neq o',\ d \neq d' \end{cases}$$

Note that the Hessian is an $|\mathcal{O}||\mathcal{D}| \times |\mathcal{O}||\mathcal{D}|$ matrix with rows and columns each index by an od pair[21]. I will show that this is negative definite which entails that the function is strictly concave in $\{\ell_{od}\}$. Fix the Hessian at a particular value of $\{\ell_{\tilde{o}\tilde{d}}\}$. Note that since $(\theta - \nu) > \max_{o,d}\{\alpha_d, \beta_o, \alpha_d + \beta_d\}$ and that $\ell_{od} \leq \min\{\sum_{\tilde{o}} \ell_{\tilde{o}d}, \sum_{\tilde{d}} \ell_{o\tilde{d}}\}$ we have that

$$\frac{\partial^2 f}{\partial \ell_{od}^2}(\{\ell_{\tilde{o}\tilde{d}}\}) < 0$$

By Theorem 4.C.2 (in particular the remark following it) in Takayama (1985) this means that if the Hessian is diagonally dominant, then the matrix is negative definite as required. Using the structure of the Hessian, the condition for diagonal dominance from the definition on page 381 of Takayama (1985) requires that there exist positive numbers $\{a_{od}\}$ such that for any od:

$$a_{od} \left| \frac{\partial^2 f}{\partial \ell_{od}^2}(\{\ell_{\tilde{o}\tilde{d}}\}) \right| > \sum_{d' \neq d} a_{od'} \left| \frac{\partial^2 f}{\partial \ell_{od} \partial \ell_{od'}}(\{\ell_{\tilde{o}\tilde{d}}\}) \right| + \sum_{o' \neq o} a_{o'd} \left| \frac{\partial^2 f}{\partial \ell_{od} \partial \ell_{o'd}}(\{\ell_{\tilde{o}\tilde{d}}\}) \right|$$

Where I have made use of the fact that many of the terms in the Hessian are zero and so are dropped from the summations on the right hand side. Taking $a_{od} := \ell_{od}$ and applying the computed values for the Hessian we find that the expression above evaluates to:

$$\theta - \nu > \beta_o \mathbb{1}_{\{\beta_o>0\}} + \alpha_d \mathbb{1}_{\{\alpha_d>0\}}$$

[21] We may take an arbitrary ordering as long as it is the same for rows and columns.

Now note that the assumption from the proposition was that $\theta - \nu > \max\{\beta_o, \alpha_d, \beta_o + \alpha_d\}$. This ensures that the above inequality holds and so the hessian of f is negative definite. This shows that $f(\ell_{od})$ is strictly concave in $\{\ell_{od}\}$ and therefore concave when viewed as a function of $\{\ell_{odmr}\}$. This completes the final piece of the objective function.

I have shown that the full objective function is globally strictly concave on the interior of a convex domain and that the optimum cannot occur on the boundary of $[0, L]^{\mathcal{S}}$. Conclude that any solution to the optimization problem is interior and unique. □

Proposition 2.

The proof proceeds in three steps. First, I show that the competitive equilibrium can be reduced to a set of equations defining choice shares as a fixed point in the variable $\boldsymbol{\ell}$. Second, I show that any $\boldsymbol{\ell}$ satisfying the equilibrium equations allows the construction of a λ such that $(\boldsymbol{\ell}, \lambda)$ is a turning point of the distorted planner's Lagrangean. Third, I show that any $(\boldsymbol{\ell}, \lambda)$ that is a turning point of the distorted planner's Lagrangean will have an $\boldsymbol{\ell}$ that satisfies the equilibrium equations.

Equilibrium as a fixed point in $\boldsymbol{\ell}$

First note that all the variables in the model can be represented as a function of $\boldsymbol{\ell}$. Equations (16)-(22) in the main text can be augmented by making the following recursive definitions:

$$u_o(\boldsymbol{\ell}) := \bar{u}_o + \beta_o \ln\left(\ell_o^H(\boldsymbol{\ell})\right) \qquad \forall o \in \mathcal{O}$$

$$A_d(\boldsymbol{\ell}) := \bar{A}_d + \alpha_d \ln\left(\ell_d^F(\boldsymbol{\ell})\right) \qquad \forall d \in \mathcal{D}$$

$$
\begin{aligned}
b(\boldsymbol{\ell}) &:= \mathbb{1}_{\mathrm{I}}\left[\frac{-\sum_{d\in\mathcal{D}}\ell_d^F(\boldsymbol{\ell})\alpha_d - \sum_{o\in\mathcal{O}}\ell_o^H(\boldsymbol{\ell})\beta_o}{L}\right] \\
v_{odmr}(\boldsymbol{\ell}) &:= b(\boldsymbol{\ell}) + u_o(\boldsymbol{\ell}) + A_d(\boldsymbol{\ell}) - \gamma t_{odmr}(\boldsymbol{\ell}) && \forall odmr \in \mathcal{S} \\
v_{odm}(\boldsymbol{\ell}) &:= \sigma \ln \sum_r \exp(v_{odmr}(\boldsymbol{\ell})/\sigma) && \forall odm \in \mathcal{O}\times\mathcal{D}\times\mathcal{M} \\
v_{od}(\boldsymbol{\ell}) &:= \nu \ln \sum_m \exp(v_{odm}(\boldsymbol{\ell})/\nu) && \forall od \in \mathcal{O}\times\mathcal{D}
\end{aligned}
$$

This means that we can state equation (4) from the set of equilibrium equations as a fixed point problem in the vector $\boldsymbol{\ell}$:

$$
\frac{\ell_{odmr}}{L} = \frac{\exp(v_{od}(\boldsymbol{\ell})/\theta)}{\sum_{o'd'}\exp(v_{o'd'}(\boldsymbol{\ell})/\theta)}\frac{\exp(v_{odm}(\boldsymbol{\ell})/\nu)}{\sum_{m'}\exp(v_{odm'}(\boldsymbol{\ell})/\nu)}\frac{\exp(v_{odmr}(\boldsymbol{\ell})/\sigma)}{\sum_{r'}\exp(v_{odmr'}(\boldsymbol{\ell})/\sigma)} \tag{†}
$$

If (†) holds then equations (1)-(14) are pinned down as functions of $\boldsymbol{\ell}$ through the definitions given above. Equation (15) will hold by Walras' law. Noting that since $b(\boldsymbol{\ell})$ is constant over *odmr* it does not affect the choice probabilities as it drops out from all numerators and denominators. we may therefore redefine:

$$
\begin{aligned}
\tilde{v}_{odmr}(\boldsymbol{\ell}) &:= u_o(\boldsymbol{\ell}) + A_d(\boldsymbol{\ell}) - \gamma t_{odmr}(\boldsymbol{\ell}) && \forall odmr \in \mathcal{S} \\
\tilde{v}_{odm}(\boldsymbol{\ell}) &:= \sigma \ln \sum_r \exp(\tilde{v}_{odmr}(\boldsymbol{\ell})/\sigma) && \forall odm \in \mathcal{O}\times\mathcal{D}\times\mathcal{M} \\
\tilde{v}_{od}(\boldsymbol{\ell}) &:= \nu \ln \sum_m \exp(\tilde{v}_{odm}(\boldsymbol{\ell})/\nu) && \forall od \in \mathcal{O}\times\mathcal{D}
\end{aligned}
$$

It will also be useful to define:

$$\bar{v}(\boldsymbol{\ell}) := \theta \ln \sum_{od} \exp(\tilde{v}_{od}(\boldsymbol{\ell})/\theta)$$

Which gives the expected utility in the model.

Given b does not affect choices over locations, modes or routes, we can instead consider the fixed point problem:

$$\frac{\ell_{odmr}}{L} = \frac{\exp(\tilde{v}_{od}(\boldsymbol{\ell})/\theta)}{\sum_{o'd'} \exp(\tilde{v}_{o'd'}(\boldsymbol{\ell})/\theta)} \frac{\exp(\tilde{v}_{odm}(\boldsymbol{\ell})/\nu)}{\sum_{m'} \exp(\tilde{v}_{odm'}(\boldsymbol{\ell})/\nu)} \frac{\exp(\tilde{v}_{odmr}(\boldsymbol{\ell})/\sigma)}{\sum_{r'} \exp(\tilde{v}_{odmr'}(\boldsymbol{\ell})/\sigma)} \tag{*}$$

Noting that the solutions to $(*)$ and $(\dagger)$ coincide. We have now shown that the competitive equilibrium problem reduces to the problem of finding a fixed point to a set of equations in $\boldsymbol{\ell}$ only.

If $\boldsymbol{\ell}$ is a turning point of the Lagrangean for the Distorted Planner then it is a solution to $(*)$

In what follows I drop the explicit dependence on $\boldsymbol{\ell}$, but all variables are to be understood as functions of $\boldsymbol{\ell}$. The turning points of the Lagrangean are characterized by the first order conditions:

$$\tilde{v}_{odmr} - (\theta - \nu) \ln \ell_{od} - (\nu - \sigma) \ln \ell_{odm} - \sigma \ln \ell_{odmr} = \lambda + \theta$$
$$\sum_{odmr} \ell_{odmr} = L$$

Let $(\boldsymbol{\ell}, \lambda)$ satisfy these equations. Manipulating the equations and using the definitions of variables in terms of $\boldsymbol{\ell}$ above, we obtain the following expressions:

$$\begin{aligned}
\ell_{odm} &= \exp\left(-\left(\frac{\theta-\nu}{\nu}\right)\ln \ell_{od} - \left(\frac{\lambda+\theta}{\nu}\right) + \frac{\tilde{v}_{odm}}{\nu}\right) \\
\ell_{od} &= \exp\left(-\left(\frac{\lambda+\theta}{\theta}\right) + \frac{\tilde{v}_{od}}{\theta}\right) \\
\lambda &= \bar{v} - \theta - \theta \ln L
\end{aligned}$$

We will now plug these expressions into the first order condition to obtain an expression for ℓ_{odmr} and verify that it agrees with $(*)$.

$$\begin{aligned}
\ell_{odmr} &= \exp\left(\frac{\tilde{v}_{odmr}}{\sigma} - \left(\frac{\theta-\nu}{\sigma}\right)\ln \ell_{od} - \left(\frac{\nu-\sigma}{\sigma}\right)\ln \ell_{odm} - \left(\frac{\lambda+\theta}{\sigma}\right)\right) \\
&= \exp\left(\frac{\tilde{v}_{odmr}}{\sigma} - \left(\frac{\lambda+\theta}{\theta}\right) + \left(\frac{\nu-\theta}{\nu\theta}\right)\tilde{v}_{od} + \left(\frac{\sigma-\nu}{\sigma\nu}\right)\tilde{v}_{odm}\right) \\
&= \exp\left(\frac{\tilde{v}odmr}{\sigma} - \frac{\bar{v}}{\theta} + \ln L + \left(\frac{\nu-\theta}{\nu\theta}\right)\tilde{v}_{od} + \left(\frac{\sigma-\nu}{\sigma\nu}\right)\tilde{v}_{odm}\right) \\
&= \frac{\exp\left(\frac{\tilde{v}_{odmr}}{\sigma}\right)\exp\left(\left(\frac{\nu-\theta}{\nu\theta}\right)\tilde{v}_{od}\right)\exp\left(\left(\frac{\sigma-\nu}{\sigma\nu}\right)\tilde{v}_{odm}\right)}{\exp\left(\frac{\bar{v}}{\theta}\right)} L \\
&= \frac{\exp(\tilde{v}_{od}/\theta)}{\sum_{o'd'}\exp(\tilde{v}_{o'd'}/\theta)} \frac{\exp(\tilde{v}_{odm}/\nu)}{\sum_{m'}\exp(\tilde{v}_{odm'}} \frac{\exp(\tilde{v}_{odmr}/\sigma)}{\sum_{r'}\exp(\tilde{v}_{odmr'}/\sigma)} L
\end{aligned}$$

Which is equivalent to $(*)$ as desired.

If $\boldsymbol{\ell}$ is a solution to $(*)$ then it defines a turning point of the Lagrangean for the Distorted Planner

Suppose $\boldsymbol{\ell}$ satisfies $(*)$. Summing $(*)$ over $odmr$ we obtain $\sum_{\ell_{odmr}} = L$ immediately. It therefore remains to show that there exists a value of λ such that the first condition for a turning point of the Lagrangean is satisfied for all ℓ_{odmr}.

Taking logs of $(*)$ and using the definitions we obtain:

$$\ln \ell_{odmr} - \ln L = \left(\frac{\nu - \theta}{\nu\theta}\right) \tilde{v}_{od} + \left(\frac{\sigma - \nu}{\sigma\nu}\right) \tilde{v}_{odm} + \frac{\tilde{v}_{odmr}}{\sigma} - \frac{\bar{v}}{\theta}$$

Now note that by taking the appropriate summations $(*)$ also implies that:

$$\tilde{v}_{od} = \theta \ln \frac{\ell_{od}}{L} + \bar{v}$$
$$\tilde{v}_{odm} = \nu \ln \frac{\ell_{odm}}{L} + (\theta - \nu) \ln \frac{\ell_{od}}{L} + \bar{v}$$

Applying these to the formula above and simplifying yields:

$$\tilde{v}_{odmr} - (\theta - \nu) \ln \ell_{od} - (\nu - \sigma) \ln \ell_{odm} - \sigma \ln \ell_{odmr} = \bar{v} - \theta \ln L$$

Taking $\lambda := \bar{v} - \theta \ln L - \theta$ ensures that the first order conditions for a turning point hold for all *odmr* as required. □

Proposition 3.

Existence Note that by the same argument as in proposition 1, an optimum to the distorted planner exists. By Theorem 1.D.6 in Takayama (1985) this optimum will also be a turning point[22] of the distorted planner's Lagrangean since gradient of $g(\boldsymbol{\ell})$ is a vector of ones so the constraint qualification holds. By proposition 2, turning points of the Lagrangean are equilibria. Conclude that an equilibrium exists.

[22] Note that Takayama uses the term quasi-saddlepoint (QSP), see the definitions in chapter 1.D

Uniqueness Given proposition 2, the problem is isomorphic to the simplified problem in proposition 1 where $\bar{u}_o - \mathbb{1}_{\mathrm{I}}\beta_o$ is replaced with $\bar{u}_o - \beta_o$, $\bar{A}_d - \mathbb{1}_{\mathrm{I}}\alpha_d$ with $\bar{A}_d - \alpha_d$ and $s_{ijm}(x_{ijm})$ with $\frac{1}{x_{ijm}} \int_0^{x_{ijm}} s_{ijm}(z)dz$. To show the problem is strictly concave under the stated assumptions, the only condition which remains to be verified is that the following function is globally convex:

$$x_{ijm} \left(\frac{1}{x_{ijm}} \int_0^{x_{ijm}} s_{ijm}(x)dx \right)$$

Taking its second derivative with respect to x_{ijm} shows that this holds whenever $s'_{ijm}(x) > 0$ for any x, which is the assumption. Note that unlike the case of the planner's problem, all variables in the equilibrium (including consumption) are pinned down as a function of $\boldsymbol{\ell}$ through the definitions above and equations (16)-(22) in the main text. Conclude that the equilibrium is unique.

$\square$

Corollary 1.

The equilibria of a competitive equilibrium with taxes are characterized as the turning points of the Lagrangean for the following program:

$$\begin{aligned}
\max_{\boldsymbol{\ell} \in [0,L]^{\mathcal{S}}} \quad & \sum_o \left(\sum_d \ell_{od}(\boldsymbol{\ell}) \right) \left[\bar{u}_o - \tau_o^H - \beta_o + \beta_o \ln \left(\sum_d \ell_{od}(\boldsymbol{\ell}) \right) \right] \\
& + \sum_d \left(\sum_o \ell_{od}(\boldsymbol{\ell}) \right) \left[\bar{A}_d - \tau_d^F - \alpha_d + \alpha_d \ln \left(\sum_o \ell_{od}(\boldsymbol{\ell}) \right) \right] \\
& - \gamma \sum_{m \in \mathcal{M}_1} \left(\sum_{ij \in \mathcal{E}_m} \int_0^{x_{ijm}(\boldsymbol{\ell})} s_{ijm}(z)dz + \frac{x_{ijm}(\boldsymbol{\ell})\tau_{ijm}}{\gamma} \right)
\end{aligned}$$

$$
\begin{aligned}
&- \gamma \sum_{od} \sum_{m \in \mathcal{M}_0} \varphi \bar{t}_{odm} \ell_{odmr} \\
&- (\theta - \nu) \sum_{od} \ell_{od}(\boldsymbol{\ell}) \ln \left(\ell_{od}(\boldsymbol{\ell})\right) \\
&- (\nu - \sigma) \sum_{odm} \ell_{odm}(\boldsymbol{\ell}) \ln \left(\ell_{odm}(\boldsymbol{\ell})\right) - \sigma \sum_{odmr} \ell_{odmr} \ln \ell_{odmr}
\end{aligned}
$$

s.t.

$$
\sum_{odmr \in \mathcal{S}} \ell_{odmr} - L = 0
$$

Where $\ell_{odm}(\boldsymbol{\ell}), \ell_{odm}(\boldsymbol{\ell})$ and $x_{ijm}(\boldsymbol{\ell})$ are given by definitions (18), (19) and (20). The proof parallels that in proposition 2 exactly. The only changes are the replacement of $\bar{u}_o$ with $\bar{u}_o - \tau_o^H$, $\bar{A}_d$ with $\bar{A}_d - \tau_d^F$ and $\int_0^{x_{ijm}(\boldsymbol{\ell})} s_{ijm}(z)dz$ with $\int_0^{x_{ijm}(\boldsymbol{\ell})} s_{ijm}(z)dz + \frac{x_{ijm}(\boldsymbol{\ell})\tau_{ijm}}{\gamma}$. Note that none of these changes alter the concavity properties of the optimization problem. By the arguments in propositions 1 and 3 conclude that an equilibrium exists under assumption 1 and is unique provided that $\theta - \nu > \max_{od \in \mathcal{O} \times \mathcal{D}} \{\alpha_d + \beta_o, \alpha_d, \beta_o\}$. □

Corollary 2.

The proof proceeds by considering the turning points of the Lagrangean for the planner's problem and the problem with taxes. By proposition 1 and corollary 1 we know that both of these uniquely identify an $\boldsymbol{\ell}$. It therefore remains to show that the two systems of equations are equivalent under the stated taxes so that the solutions for $\boldsymbol{\ell}$ coincide. Note that the derivative with respect to λ is the same in both cases, simply restating the constraint that $\ell_{odmr} - L = 0$. I turn to the first order conditions with respect to ℓ_{odmr} in each case.

In the case of the planner's problem we have:

$$\begin{aligned}
&\bar{u}_o + (1 - \mathbb{1}_{\mathrm{I}})\beta_o + \beta_o \ln \ell_o^{H^*} + \bar{A}_d + (1 - \mathbb{1}_{\mathrm{I}})\alpha_d + \alpha_d \ln \ell_d^{F^*} \\
&- \mathbb{1}_{\{m \in \mathcal{M}_1\}} \gamma \sum_{ij} s_{ijm}(x^*_{ijm}) + x^*_{ijm} s'_{ijm}(x^*_{ijm}) - \mathbb{1}_{\{m \in \mathcal{M}_o\}} \gamma \phi \bar{t}_{odm} \\
&- (\theta - \nu)(1 + \ln \ell^*_{od}) - (\nu - \sigma)(1 + \ell^*_{odm}) - \sigma(1 + \ell^*_{odmr}) - \lambda^* = 0
\end{aligned}$$

In the case of the program for the equilibrium with taxes defined in corollary 1 evaluated at the taxes from the statement of corollary 2 above we have:

$$\begin{aligned}
&\bar{u}_o + (1 - \mathbb{1}_{\mathrm{I}})\beta_o + \beta_o \ln \ell_o^{H} + \bar{A}_d + (1 - \mathbb{1}_{\mathrm{I}})\alpha_d + \alpha_d \ln \ell_d^{F} \\
&- \mathbb{1}_{\{m \in \mathcal{M}_1\}} \gamma \sum_{ij} \left(s_{ijm}(x_{ijm}) + x^*_{ijm} s'_{ijm}(x^*_{ijm}) \right) - \mathbb{1}_{\{m \in \mathcal{M}_o\}} \gamma \phi \bar{t}_{odm} \\
&- (\theta - \nu)(1 + \ln \ell_{od}) - (\nu - \sigma)(1 + \ell_{odm}) - \sigma(1 + \ell_{odmr}) - \lambda = 0
\end{aligned}$$

Note that the first system of equations has a unique solution when combined the constraint by proposition 1. Now note that since $*$ variables solve the first system of equations, $\ell_o^H = \ell_o^{H^*}$, $\ell_d^F = \ell_d^{F^*}$, $\ell_{od} = \ell^*_{od}$, $\ell_{odm} = \ell^*_{odm}$, $\ell_{odmr} = \ell^*_{odmr}$, $x_{ijm} = x^*_{ijm}$ and $\lambda = \lambda^*$ must also solve the second system. By corollary 1, this is the unique solution. Conclude that the proposed taxes decentralize the first best solution to $\boldsymbol{\ell}$.

□

Proposition 4.

The proof proceeds in three steps. Firstly, each optimization problem is cast as the saddle point to a Lagrangean problem in slack variables. Secondly, by using duality, we consider the dual of the optimization problem. Finally, considering the inner optimization problem over primal variables, first order conditions are used to simplify the program and obtain a problem stated purely in terms of multipliers which is simplified further by taking first order conditions over those.

Using the notation for $\tilde{u}_o$, $\tilde{A}_d$, and $\tilde{s}_{ijm}$ from the text the mathematical program defining the solution to the planner, competitive equilibrium, and equilibrium with taxes problems can be expressed in a unified framework. Under the assumptions of the proposition we have:

$$\begin{aligned}
\max_{\boldsymbol{\ell}\in[0,L]^{\mathcal{S}}} & \sum_o \left(\sum_d \ell_{od}(\boldsymbol{\ell})\right)\left[\tilde{u}_o - \mathbb{1}_{\mathrm{I}}\beta_o + \beta_o \ln\left(\sum_d \ell_{od}(\boldsymbol{\ell})\right)\right] \\
& + \sum_d \left(\sum_o \ell_{od}(\boldsymbol{\ell})\right)\left[\tilde{A}_d - \mathbb{1}_{\mathrm{I}}\alpha_d + \alpha_d \ln\left(\sum_o \ell_{od}(\boldsymbol{\ell})\right)\right] \\
& - \gamma \sum_{ij\in\mathcal{E}_m} x_{ijm_1}(\boldsymbol{\ell})\tilde{s}_{ijm_1}(x_{ijm_1}(\boldsymbol{\ell})) \\
& - \gamma\varphi\bar{t}_{odm_0}\ell_{odm_0r} \\
& - (\theta-\nu)\sum_{od}\ell_{od}(\boldsymbol{\ell})\ln\left(\ell_{od}(\boldsymbol{\ell})\right) \\
& - (\nu-\sigma)\sum_{odm}\ell_{odm}(\boldsymbol{\ell})\ln\left(\ell_{odm}(\boldsymbol{\ell})\right) - \sigma\sum_{odmr}\ell_{odmr}\ln\ell_{odmr}
\end{aligned}$$

s.t.

$$\sum_{odmr\in\mathcal{S}} \ell_{odmr} = L$$

This problem has been shown to be strictly concave. I now introduce slack variables for ℓ_o^H, ℓ_d^F , ℓ_{odm} and x_{ijm} and corresponding multipliers so that they satisfy the constraints defined by the functions in equations (16)-(22) of the main text. With a slight abuse of notation I let $\boldsymbol{\ell}$ denote all of $\{\ell_{odmr}\}$, $\{\ell_o^H\}$, $\{\ell_d^F\}$, $\{\ell_{odm}\}$ stacked as a single vector. I let $\boldsymbol{x}$ denote the stacked traffic flows and $\boldsymbol{\lambda}, \boldsymbol{\mu}$ denote the stacked vectors of multipliers. This leads to the following Lagrangean formulation of the problem[23]:

$$
\begin{aligned}
\max_{\boldsymbol{\ell},\boldsymbol{x}} \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} & \sum_o \ell_d^H \left[\tilde{u}_o + \beta_o \ln \ell_o^H\right] + \sum_d \ell_d^F \left[\tilde{A}_d + \alpha_d \ln \ell_d^F\right] - \sum_{ij} x_{ijm_1} \tilde{s}_{ijm_1}(x_{ijm_1}) - \varphi \sum_{od} \ell_{odm_0r} \bar{t}_{odm_o} \\
& - (\theta - \nu) \sum_{od} \ell_{od} \ln \ell_{od} - (\nu - \sigma) \sum_{odm} \ell_{odm} \ln \ell_{odm} - \sigma \sum_{odmr} \ell_{odmr} \ln \ell_{odmr} \\
& - \sum_o \lambda_o^H \left[\ell_o^H - \sum_{dmr} \ell_{odmr}\right] - \sum_d \lambda_d^F \left[\ell_d^F - \sum_{omr} \ell_{odmr}\right] - \sum_{od} \lambda_{odm_1}^M \left[\ell_{odm_r} - \sum_r \ell_{odm_1r}\right] \\
& \lambda_{od}^C \left[\ell_{od} - \sum_{mr} \ell_{odmr}\right] - \sum_{ij} \mu_{ij} \left[\sum_{odr} \ell_{odm_1r} n_{ijm_1r}^{od} - x_{ijm}\right] - \lambda \left[L - \sum_{odmr} \ell_{odmr}\right]
\end{aligned}
$$

Note that since all constraints are linear, the Lagrange multipliers are not constrained in sign. Not also that with $\alpha_d, \beta_o < 0$ the function is concave in its arguments. $\boldsymbol{x}, \boldsymbol{\ell}$ need only be positive but, as argued above, the solution will always be interior so that non-negativity constraints are omitted. This completes the first step in the proof.

Now note that since the constraints are linear and satisfy the constraint qualification, strong duality holds[24], and we may instead solve the problem

$$
\min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \max_{\boldsymbol{\ell},\boldsymbol{x}} \sum_o \ell_d^H \left[\tilde{u}_o + \beta_o \ln \ell_o^H\right] + \sum_d \ell_d^F \left[\tilde{A}_d + \alpha_d \ln \ell_d^F\right] - \sum_{ij} x_{ijm_1} \tilde{s}_{ijm_1}(x_{ijm_1}) - \varphi \sum_{od} \ell_{odm_0r} \bar{t}_{odm_o}
$$

[23]Note that since existence has been established, $\sup, \inf$ are replaced with $\max, \min$

[24]See the results in section 5.3.2 of Boyd and Vandenberghe (2004)

$$
\begin{aligned}
&-(\theta-\nu)\sum_{od}\ell_{od}\ln\ell_{od}-(\nu-\sigma)\sum_{odm}\ell_{odm}\ln\ell_{odm}-\sigma\sum_{odmr}\ell_{odmr}\ln\ell_{odmr}\\
&-\sum_{o}\lambda_o^H\left[\ell_o^H-\sum_{dmr}\ell_{odmr}\right]-\sum_{d}\lambda_d^F\left[\ell_d^F-\sum_{omr}\ell_{odmr}\right]-\sum_{od}\lambda_{odm_1}^M\left[\ell_{odm_r}-\sum_{r}\ell_{odm_1r}\right]\\
&\lambda_{od}^C\left[\ell_{od}-\sum_{mr}\ell_{odmr}\right]-\sum_{ij}\mu_{ij}\left[\sum_{odr}\ell_{odm_1r}n_{ijm_1r}^{od}-x_{ijm}\right]-\lambda\left[L-\sum_{odmr}\ell_{odmr}\right]
\end{aligned}
$$

Where the order of max and min have been interchanged: the dual problem. This completes the second step of the proof.

Now consider the inner problem of optimizing over $\boldsymbol{\ell},\boldsymbol{x}$ for fixed values of $\boldsymbol{\lambda},\boldsymbol{\mu}$. This problem is concave so that first order conditions will be sufficient for an optimum. It will be useful to define the following:

$$
\tilde{s}^*_{ijm}(x):=\tilde{s}_{ijm}(x)+x\tilde{s}'_{ijm}(x)
$$

We will take first order conditions with respect to $\boldsymbol{x},\boldsymbol{\ell}$.

The first order conditions give:

$$
\begin{aligned}
\ln\ell_o^H&=\frac{\lambda_o^H-\tilde{u}_o-\beta_o}{\beta_o}\\
\ln\ell_d^F&=\frac{\lambda_d^F-\tilde{A}_d-\alpha_d}{\alpha_d}\\
x_{ijm}&=\tilde{s}^{*^{-1}}_{ijm}(-\mu_{ij}/\gamma)\\
\ln\ell_{od}&=\frac{-\lambda_{od}^C}{\theta-\nu}-1\\
\ln\ell_{odm_or}&=\frac{-\varphi\gamma\bar{t}_{odm_o}+\lambda_o^H+\lambda_d^F+\lambda_{od}^c+\lambda}{\nu}\\
\ln\ell_{odm_1}&=\frac{-\lambda_{odm_1}^M}{\nu-\sigma}-1\\
\ln\ell_{odm_1r}&=\frac{\lambda_o^H+\lambda_d^F+\lambda_{odm_1}^M+\lambda_{od}^C+\lambda-\sum_{ij}\mu_{ij}n_{ijm_1r}^{od}/\gamma}{\sigma}-1
\end{aligned}
$$

Note that if we can solve for the multipliers, these equations recover all primal variables as a function of them, which is final part of the statement of the proposition. Using these first order conditions, algebraic manipulations lead to the following program which involves only Lagrange multipliers.

$$
\begin{aligned}
\min_{\boldsymbol{\lambda},\boldsymbol{\mu}} & -\sum_o \beta_o \exp\left(\frac{\lambda_o^H - \tilde{u}_o - \beta_o}{\beta_o}\right) - \sum_d \alpha_d \exp\left(\frac{\lambda_d^F - \tilde{A}_d - \alpha_d}{\alpha_d}\right) \\
& + \sum_{ij\in\mathcal{E}_{m_1}} \tilde{s}^{*-1}_{ijm_1}(\mu_{ij}/\gamma)\left[\mu_{ij}/\gamma - \tilde{s}_{ijm_1}\left(\tilde{s}^{*-1}_{ijm_1}(\mu_{ij}/\gamma)\right)\right] \\
& + (\theta-\nu)\sum_{od}\exp\left(\frac{-\lambda^C_{od}}{\theta-\nu} - 1\right) + \nu\sum_{od}\exp\left(\frac{-\varphi\gamma\bar{t}_{odm_o} + \lambda_o^H + \lambda_d^F + \lambda^c_{od} + \lambda}{\nu}\right) \\
& + \sigma\sum_{od}\exp\left(\frac{\lambda_o^H + \lambda_d^F + \lambda^M_{odm_1} + \lambda^C_{od} + \lambda}{\sigma} - 1\right)\sum_r \exp\left(\frac{-\sum_{ij}\mu_{ij}n^{od}_{ijm_1r}/\gamma}{\sigma}\right)
\end{aligned}
$$

Note that by duality, this problem will be convex. First order conditions are therefore sufficient for an optimum. We will now simplify this program even further by removing all multipliers except for μ_{ij}, λ_o^H and λ_d^F by taking first order conditions with respect to the other variables. Doing so, and after some further algebraic manipulations, we get:

$$
\begin{aligned}
\min_{\lambda,\mu} \quad & -\sum_{o\in\mathcal{O}} \beta_o \exp\left(\frac{\lambda_o^H - \tilde{u}_o - \beta_o}{\beta_o}\right) - \sum_{d\in\mathcal{D}} \alpha_d \exp\left(\frac{\lambda_d^F - \tilde{A}_d - \alpha_d}{\alpha_d}\right) \\
& + \sum_{ij\in\mathcal{E}_{m_1}} \tilde{s}^{*-1}_{ijm_1}(\mu_{ij}/\gamma)\left[\mu_{ij}/\gamma - \tilde{s}_{ijm_1}\left(\tilde{s}^{*-1}_{ijm_1}(\mu_{ij}/\gamma)\right)\right] \\
& + \theta L \ln\left(\sum_{od}\exp\left(\frac{\lambda_o^H + \lambda_d^F}{\theta}\right)\left[\exp\left(\frac{-\varphi\gamma\bar{t}_{od}}{\nu}\right) + \left(\sum_r \exp\left(-\sum_{ij}\mu_{ij}n^{od}_{ijm_1,r}/\sigma\right)\right)^{\frac{\sigma}{\nu}}\right]^{\frac{\nu}{\theta}}\right)
\end{aligned}
$$

Which is the program from the statement of the proposition. $\square$

Proposition 5.

The strategy for both proofs parallels that of the general equilibrium case. The equilibrium equations of the models are cast as fixed points whose solution is shown to coincide with the solution to an optimization problem.

Transport Equilibrium with Taxes

Let the fixed origin destination flows be given by $\{\bar{\ell}_{od}\}$ and the link level taxes by $\{\tau_{ijm}\}$.. The equilibrium for the transport model with taxes can be cast as the following fixed point problem in $\boldsymbol{\ell}$:

$$\frac{\ell_{odmr}}{\bar{\ell}_{od}} = \frac{\exp\left(v^{\mathrm{tr}}_{odm}(\boldsymbol{\ell})/\nu\right)}{\sum_{m'\in\mathcal{M}}\exp\left(v^{\mathrm{tr}}_{odm'}(\boldsymbol{\ell})/\nu\right)} \frac{\exp\left(\left(-\gamma t_{odmr}(\boldsymbol{\ell})\right)/\sigma\right)}{\sum_{r'\in\mathcal{R}_{odm}}\exp\left(\left(-\gamma t_{odmr'}(\boldsymbol{\ell})\right)/\sigma\right)} \tag{44}$$

Where $t_{odmr}(\boldsymbol{\ell})$ is given as in equations $(20), (14^{\dagger}), (22)$ and

$$v^{\mathrm{tr}}_{odm}(\boldsymbol{\ell}) := \sigma \ln \sum_{r\in\mathcal{R}_{odm}} \exp\left(\left(-\gamma t_{odmr}(\boldsymbol{\ell})\right)/\sigma\right)$$

The optimization problem corresponding to this problem is:

$$\begin{aligned}
&\max_{\boldsymbol{\ell}} -\gamma\sum_{ijm}\int_0^{x_{ijm}(\boldsymbol{\ell})} s_{ijm}(z)dz - \sum_{ijm}\tau_{ijm}x_{ijm}(\boldsymbol{\ell}) - (\nu-\sigma)\sum_{odm}\ell_{odm}(\boldsymbol{\ell})\ln\left(\ell_{odm}(\boldsymbol{\ell})\right) \\
&\qquad -\sigma\sum_{odmr}\ell_{odmr}\ln\ell_{odmr} \\
&\quad \text{s.t.} \\
&\quad \sum_{mr}\ell_{odmr} = \bar{\ell}_{od} \quad \forall od \in \mathcal{O}\times\mathcal{D}
\end{aligned}$$

Verifying that the turning points of the Lagrangean coincide with the equilibria defined by (44) proceeds analagously to the general case. Moreover it is clear the objective function is strictly concave given the arguments made above. Conclude that there exists a unique transport equilibrium with taxes.

Diving Equilibrium with Taxes

Let the fixed origin-destination-mode flows be given by $\{\bar{\ell}_{odm}\}$ and the link level taxes be given by $\{\tau_{ijm}\}$. The equilibrium for this model can be case as a fixed point problem in $\boldsymbol{\ell}$

$$\frac{\ell_{odmr}}{\bar{\ell}_{odm}} = \frac{\exp\left(-\gamma t_{odmr}(\boldsymbol{\ell})/\sigma\right)}{\sum_{r' \in \mathcal{R}_{odm}} \exp\left(-\gamma t_{odmr'}(\boldsymbol{\ell})/\sigma\right)} \tag{45}$$

Where again $t_{odmr}(\boldsymbol{\ell})$ is given as in equations (12), (13), and ($14^{\#}$).

The optimization problem corresponding to this problem is:

$$\max_{\boldsymbol{\ell}} -\gamma \sum_{ijm} \int_0^{x_{ijm}(\boldsymbol{\ell})} s_{ijm}(z)dz - \sum_{ijm} \tau_{ijm} x_{ijm}(\boldsymbol{\ell}) - \sigma \sum_{odmr} \ell_{odmr} \ln \ell_{odmr}$$

s.t.

$$\sum_r \ell_{odmr} = \bar{\ell}_{odm} \quad \forall od \in \mathcal{O} \times \mathcal{D},\ \forall m \in \mathcal{M}$$

That the turning points of the Lagrangean coincide with the equilibria defined by (45) again proceeds as in the general case. Moreover the objective function is strictly concave. Conclude that there exists a unique driving equilibrium with taxes.

□

B Model Extensions and Micro-foundations

B.1 The Limit as $K \to \infty$

Recall that we constrained the set of routes to have a length no longer than K in the main text. I now show formally the implications of the limit as $K \to \infty$. In equation (4) the term which causes an issue is

$$\sum_r \exp(v_{odmr}/\sigma) \ .$$

In order for the model to be well-behaved, and converge to a definite limit as $K \to \infty$, we must ensure that this series does not diverge. If this term diverges, expected utilities will become unbounded and the model is not well-defined in the limit. I proceed by providing conditions that guarantee that this term is bounded, for any K.

First note that using the definitions from the main text we have

$$\begin{aligned}\sum_r \exp(v_{odmr}/\sigma) &= \exp\left(\frac{b+u_o+w_d}{\sigma}\right)\sum_r \exp\left(-\frac{\gamma}{\sigma}\sum_{ij} n^{od}_{ijmr} t_{ijm}\right)\\ &\leq \exp\left(\frac{b+u_o+w_d}{\sigma}\right)\sum_r \exp\left(-\frac{\gamma}{\sigma}\sum_{ij} n^{od}_{ijmr} s_{ijm}(0)\right)\end{aligned}$$

The first equality applies definitions from the main text and the inequality follows from Assumption 1: $t_{ijm} \geq s_{ijm}(0)$ since $s'_{ijm} > 0$ at any level of traffic. Intuitively, this states that the journey time along a link in equilibrium is at least as large as the free-flow speed with no traffic. It remains to bound the final term. First note that $\exp\left(\frac{b+u_o+w_d}{\sigma}\right)$ is bounded since the population is of finite mass L. This follows from equations (8), (9), (10), and (11). So we must bound $\sum_r \exp\left(-\frac{\gamma}{\sigma}\sum_{ij} n^{od}_{ijmr} s_{ijm}(0)\right)$. This involves a sum over routes $r \in \mathcal{R}_{odm}$ which may be of arbitrary length. It will be helpful to define the following $|\mathcal{N}| \times |\mathcal{N}|$ matrix for each $m \in \mathcal{M}_1$

$$[W_m]_{ij} = \begin{cases} \exp\left(-\frac{\gamma}{\sigma} s_{ijm}(0)\right) & \text{if } ij \in \mathcal{E}_m \\ 0 & \text{otherwise} \end{cases}$$

This is a weighted adjacency matrix for the graph $\mathcal{E}_m$ over the set of all possible locations $\mathcal{N}$. The weights are given as a function of the free-flow speeds across links. Recall that the ijth entry of the kth power of the weighted adjacency matrix, W_m^k, gives the sum of the products of weights on all routes from node i to node j. This means that when paths are limited to those of length K, $\sum_r \exp\left(-\frac{\gamma}{\sigma} \sum_{ij} n_{ijmr}^{od} s_{ijm}(0)\right)$ is given by the odth entry of the following matrix sum:

$$W_m + W_m^2 + \ldots + W_m^K$$

Showing that all such sums converge in the limit as $K \to \infty$ is equivalent to ensuring that $(I - W_m)$ is invertible (see, for example Theorem 4.C.6 in Takayama (1985)). That is

$$(I - W_m)^{-1} - I = \sum_{k=1}^{\infty} W_m^k \, .$$

This invertibility issue arises in many models of traffic equilibria in which agents have a distribution of idiosyncratic preferences over routes as in Akamatsu (1996), Akamatsu (1997), and Allen and Arkolakis (2022). Here I provide a condition in terms of model primitives which is, to the best of my knowledge, new to the literature in both economics and transportation. I provide sufficient conditions for the matrix to be dominant diagonal and therefore invertible by Theorem 4.C.1 in Takayama (1985). A sufficient condition for this is that

$$\sum_{j:ij\in\mathcal{E}_m} \exp\left(-\frac{\gamma}{\sigma}s_{ijm}(0)\right) < 1 \qquad \forall i \in \mathcal{N},\, \forall m \in \mathcal{M}_1\,. \tag{46}$$

Since all terms are positive, this simply restates the definition of dominant diagonal matrices on page 381 of Takayama (1985). Note that it is stated purely in terms of model primitives and depends on three things. Firstly, the condition requires $s_{ijm}(0)$ to be sufficiently large: when the cost of crossing each link is too low, the benefits from taking long paths with cycles do not decline rapidly enough and they may remain attractive as the route set becomes larger. Secondly it depends on $\frac{\gamma}{\sigma}$. When the value of time crossing links, γ is small, this makes the condition harder to fulfill as people care less about the time they spend crossing links. Similarly, when the shock variance is large, captured by σ the condition is harder to fulfill. There is greater heterogeneity in preferences over routes and so longer routes become more attractive for some. Finally, it depends on the network structure through $\mathcal{E}_m$. When the network is densely connected, there are more j's to sum over for each i, making the condition harder to fulfill. Importantly, given parameter estimates for σ, γ, and s_{ijm}, condition (46) can be numerically checked before attempting to solve a model with arbitrary route lengths. This ensures that the defined objects will be well behaved in the limit.

B.2 Computing the Objective Function

I now note how the results of the previous section can be used in the computational results in Section 3.7 of the paper. In particular in evaluating the term $\sum_r \exp\left(-\sum_{ij}\mu_{ij}n^{od}_{ijmr_1,r}/\sigma\right)$ in the objective function of the program. Analogous to the above define

$$\left[\tilde{W}_m\right]_{ij} = \begin{cases} \exp\left(-\frac{\mu_{ij}}{\sigma}\right) & \text{if } ij \in \mathcal{E}_m \\ 0 & \text{otherwise} \end{cases}.$$

Then when paths are limited to length K, $\sum_r \exp\left(-\sum_{ij}\mu_{ij}n^{od}_{ijmr_1,r}/\sigma\right)$ is given by the *od*th entry of

$$\tilde{W}_m + \tilde{W}_m^2 + \ldots + \tilde{W}_m^K \ .$$

When K is large, we have that the infinite series converges and so

$$(I - \tilde{W}_m)^{-1} - I \simeq \tilde{W}_m + \tilde{W}_m^2 + \ldots + \tilde{W}_m^K$$

The left hand side is what is used in the algorithm that I implement. In practice, the approximation becomes good rapidly and approaches numerical precision even for modest values of K around 50.

B.3 The Supply and Demand of Floor-space

This section provides a micro-foundation for α_d, β_o in terms of floor-space supply to firms and households. I assume that the markets are segmented: firms and households consume floor-space in separate markets that are priced and clear individually and by location. I show how this leads to the same reduced form expressions as developed in the main text when $\mathbb{1}_{\mathrm{I}} = 1$ for both the competitive equilibrium problem and in the case of the planner. This also provides an explicit expression for α_d, β_o which I use to calibrate the parameters using supply elasticities from Baum-Snow and Han (2024).

Floor-space Supply

Floor-space is supplied by competitive firms in each location $i \in \mathcal{N}$. The market for floor-space is divided by sector $K \in \{H, F\}$ representing the supply for floor-space to households and firms. A representative firm in sector K, location i uses material inputs M and land T in a Cobb-Douglas technology to produce floor-space. They take as given the prices of floor-space, r_i^K and land, p_i^T, in that location as well as the price of materials which are assumed to be constant across space, p^M, and determined exogenously, outside the city's economy. The firm's problem is:

$$\max_{Q,M,T} r_i^K Q - p^M M - p_i^T T \quad \text{s.t.} \quad Q = M^{\psi_i} T^{1-\psi_i}$$

Land is fixed for each sector in each location. This could be interpreted as zoning: certain areas are marked for firms and others for residents.

$$T_i^K = \bar{T}_i^K$$

The firm's first order conditions imply

$$\begin{aligned} p^M M_i^K &= \psi_i r_i^K Q_i^K \\ p_i^T T_i^K &= (1-\psi) r_i^K Q_i^K \end{aligned}$$

The free entry condition together with the fixed supply of land then imply that the inverse housing supply function in a given location and sector is

$$r_i^K = \left(\frac{p^M}{\psi_i \bar{T}^{(1-\psi_i)/\psi_i}} \right) \left(Q_i^K \right)^{(1-\psi_i)/\psi_i}$$

Note that the elasticity of the rental price with respect to housing quantity is $(1-\psi_i)/\psi_i$ which will be fixed in each location to the values from Baum-Snow and Han (2024). The housing supply function is given by simple rearrangement as:

$$Q_i^K = \left(\frac{\psi_i r_i^K}{p^M} \right)^{\frac{\psi_i}{1-\psi_i}} \bar{T}_i^K$$

Household Floor-space Demand

Consider a consumer who has already made their location decisions over a given workplace d and home o and focus on their decision on how much housing to consume. I fix all the levels of population in each home and work location ℓ_d^F, ℓ_o^H and their wage income w_d at arbitrary levels. I will assume that their non-wage income b comes from evenly rebated land rents in the model and provide an explicit expression for this below. Each consumer takes b, w_d and r_o^H as given and choose only how to allocate their income between consumption, c, and floor-space, q. I also assume that there are fundamental amenities in each location $u^\dagger$ that are exogenous. The consumer solves the following problem:

$$\max_{c,q} \quad u_o^\dagger + c + \delta \ln q \quad \text{s.t.} \quad c + r_o^H q \leq w_d + b$$

The solution for the consumer's Marshallian demand at an interior solution (which is assumed throughout) is

$$q_o^H = \frac{\delta^H}{r_o^H}$$

Their indirect utility function is therefore:

$$v(r_o^H, w_d + b) = w_d + b - \delta^H + \delta^H \ln \delta^H - \delta^H \ln r_o^H + u_o^\dagger$$

Firm Floor-space Demand

Firms in combine labor and floor-space to produce the final good in a constant returns to scale technology. As usual, the scale and number of firms will be arbitrary in the competitive equilibrium with free entry and a zero profit condition. The representative firm solves:

$$\max_{\ell,q} \ell \left(A^{\dagger} + \delta^F \ln \frac{q}{\ell} \right) - w_d \ell - r_d^F q$$

The production function is intuitive. The average product of labor is an increasing but concave function of the amount of floor-space per worker and the function is constant returns in both floor-space and labor but decreasing returns when floor-space is fixed. This will create aggregate decreasing returns to scale in each location when the supply of land to produce floor-space is fixed. Solving the firm's first order conditions and imposing zero profits implies:

$$q_d^F = \frac{\delta^F \ell_d^F}{r_d^F}$$
$$w_d = A_d^{\dagger} + \delta^F \ln \frac{q_d^F}{\ell_d^F} - \delta^F$$

Floor-space Market Clearing

Floor-space markets clear by location and sector. For residential floor-space noting the fixed levels of ℓ_d^F, ℓ_o^H we have:

$$\frac{\ell_o^H \delta^H}{r_o^H} = \left(\frac{\psi_o r_o^H}{p^M} \right)^{\frac{\psi_o}{1-\psi_o}} \bar{T}_o^H$$

This gives

$$u_o^{\dagger} - \delta^H \ln r_o^H = u_o^{\dagger} \delta^H + \delta^H \psi_o \left(\ln \left(\frac{\delta^H \psi_o}{p^M} \right) - 1 \right) + \delta^H (1-\psi_o) \ln \bar{T}_o^H + \delta^H (\psi_o - 1) \ln \ell_o^H + \delta^H (\psi_o - 1)$$

Defining :

$$
\begin{aligned}
\bar{u}_o &:= u_o^\dagger \delta^H + \delta^H \psi_o \left(\ln \left(\frac{\delta^H \psi_o}{p^M} \right) - 1 \right) + \delta^H (1-\psi_o) \ln \bar{T}_o^H + \delta^H (\psi_o - 1) \\
\beta_o &:= \delta^H (\psi_o - 1)
\end{aligned}
$$

$$
u_o^\dagger - \delta^H \ln r_o^H = \bar{u}_d + \alpha_d \ln \ell_o^H
$$

Now using Workplace market clearing we have:

$$
\frac{\delta^F \ell_d^F}{r_d^F} = \left(\frac{\psi_d r_d^F}{p^M} \right)^{\frac{\psi_d}{1-\psi_d}} \bar{T}_d^F
$$

Combining this with the firm's optimality conditions and zero profit condition we find that:

$$
w_d = A_d^\dagger + \delta^F \psi_d \left(\ln \left(\frac{\delta^F \psi_d}{p^M} \right) - 1 \right) + \delta^F (1-\psi_d) \ln \bar{T}_d + \delta^F (\psi_d - 1) \ln \ell_d^F + \delta^F (\psi_d - 1)
$$

Defining:

$$
\begin{aligned}
\bar{A}_d &:= A_d^\dagger + \delta^F \psi_d \left(\ln \left(\frac{\delta^F \psi_d}{p^M} \right) - 1 \right) + \delta^F (1-\psi_d) \ln \bar{T}_d + \delta^F (\psi_d - 1) \\
\alpha_d &:= \delta^F (\psi_d - 1)
\end{aligned}
$$

We get that:

$$
w_d = \bar{A}_d + \alpha_d \ln \ell_d^F
$$

Finally note that total land rents in the model are given by:

$$\sum_o p_o^H \bar{T}_o^H + \sum_d p_d^F \bar{T}_d^F = \sum_o (1-\psi_o) r_o^H Q_o^H + \sum_d (1-\psi_d) \ell_d^F \delta^F$$
$$= -\sum_o \alpha_d \ell_d^F - \sum_o \beta_o \ell_o^H$$

When land rents are rebated uniformly across consumers we get that their non-wage income is given by

$$b = \frac{-\sum_d \ell_d^F \alpha_d - \sum_o \ell_o^H \beta_o}{L}$$

We can re-express the consumer's indirect utility, now as a function of ℓ_d^F, ℓ_o^H as:

$$v_{od} = \bar{u}_o + \beta_o \ln \ell_o^H + \bar{A}_d + \alpha_d \ln \ell_d + b$$

Note that all the expressions here agree with those in the main text, with the only terms missing being those that concern travel and the idiosyncratic preference terms. Since conditional on their location and commuting problems, consumers will always choose housing and consumption in the way described above. We have therefore shown that a decentralized market for floor-space for firms and households provides a micro-foundation for the reduced form expressions provided in the main body of the paper.

Implications for Output Market Clearing

Market clearing for the final good in the economy described above requires total output of the final good net of payments for materials to equal consumer expenditure on the final good. That is:

$$\sum_d \ell_d^F(A^\dagger + \delta^F \ln \frac{q^F}{\ell^F}) - \sum_d p^M M_d^F - \sum_o p^M M_o^H = \sum_d w_d \ell_d^F + \sum_{od} \ell_{od} b - \sum_o r_o^H q_o^H \ell_o^H \quad (47)$$

The first term on the left hand side gives total production, the second two express payments to materials. The terms on the write hand side express consumer income from wage and non-wage income, net of their payments to housing. This section shows that this expression is equivalent to expression (15) in the main text.

The total output remaining net of costs paid out of the city for materials is therefore

$$\sum_d \ell_d^F(A^\dagger + \delta^F \ln \frac{q^F}{\ell^F}) - \sum_d p^M M_i^F + \sum_o p^M M_i^H = \sum_d A_d \ell_d^F - \sum_d \alpha_d \ell_d^F - \sum_o \psi_o \delta^H \ell_o^H$$

Noting that $r_o^H q_o^H = \delta^H$ and defining $y_d = A_d \ell_d^F$ as in the main text, and rearranging (47) gives

$$\sum_{od} \ell_{od}(w_d + b) = \sum_d y_d - \sum_d \ell_d^F \alpha_d - \sum_o \ell_o^H \beta_o$$

Which is equivalent to (15) when $\mathbb{1}_{\mathrm{I}} = 1$ as desired.

Implications for the Planner's Problem

I consider the sub-problem of a planner choosing only the amount of floor-space to supply to firms and households. That is, I suppose that the aggregate variables c, $\{\ell_{odmr}\}$, $\{\ell_{od}\}$, $\{\ell_d^F\}$, $\{\ell_o^H\}$ have been fixed at an arbitrary level and solve for the optimal choices for floor-space production given that. I show that the resulting value function is isomorphic to the planner's problem in the main text when $\mathbb{1}_{\mathrm{I}} = 1$.

I assume that the planner must pay materials costs, as in the economy, and faces the same supply function for floor-space for firms and households. The planner will choose these quantities to maximize the total utility produced by production and utility from housing given the fixed values of ℓ_d^F, ℓ_o^H. Dividing the optimization problem into two stages will lead to the same solution as the fixed values of the parameters are arbitrary. Such a planner's problem, neglecting constant terms, is:

$$\max_{\{q_o^H,\},\{q_d^F\},\{M_o^H\},\{M_d^F\}} \sum_d \ell_d^F \left[A_d^\dagger + \delta^F \ln \frac{q_d^F}{\ell_d^F}\right] - \sum_o p_m M_o^H - \sum_d p_m M_d^H + \sum_o \ell_o^H \left[\delta^H \ln q_o^H + u_o^\dagger\right]$$

$$\begin{aligned}
&\text{s.t.}\\
&\ell_o^H q_o^H = \left(M_o^F\right)^{\psi_o} \left(\bar{T}_o^H\right)^{1-\psi_o}\\
&q_d^F = \left(M_o^F\right)^{\psi_o} \left(\bar{T}_o^H\right)^{1-\psi_o}
\end{aligned}$$

The first term in the objective function represents total production with the second and third giving materials costs. The final term is the total utility produced from housing and residential amenities. At fixed ℓ_d^F, ℓ_o^H, the planner will always choose to maximize this objective function. Eliminating q_o^H, q_d^F from the problem by using the constraints we obtain the following unconstrained, concave problem:

$$\begin{aligned}
\max_{\{M_o^H\},\{M_d^F\}} &\sum_d \ell_d^F \left[A_d^\dagger + \delta^F \psi_d \ln M_d^F + \delta^F(1-\psi_d) \ln \bar{T}_d^F - \delta^F \ln \ell_d^F\right]\\
&- \sum_o p^M M_o^H - \sum_d p^M M_d^H\\
&+ \sum_o \ell_o^H \left[u_o^\dagger + \delta^h \ln \psi_o \ln M_o^H + \delta^H(1-\psi_o) \ln \bar{T}_o^H - \delta^H \ln \ell_o^H\right]
\end{aligned}$$

The first order conditions for the optimum imply that

$$M_o^H = \frac{\ell_o^H \delta^H \psi_o}{p^M}$$
$$M_d^F = \frac{\ell_d^F \delta^F \psi_d}{p^M}$$

Plugging in the solution gives the value function for this problem as a function of the fixed variables:

$$\sum_d \ell_d^F \left[A_d^\dagger + \delta^F \psi_d \left(\ln \frac{\delta^F \psi_d}{p^M} - 1\right) + \delta^F (1-\psi_d) \ln \bar{T}_d + \delta^F (\psi_d - 1) \ln \ell_d^F\right] +$$
$$\sum_o \ell_o^H \left[u_o^\dagger + \delta^H \psi_o \left(\ln \frac{\delta^H \psi_o}{p^M} - 1\right) + \delta^H (1-\psi_o) \ln \bar{T}_o^H + \delta^H (\psi_o - 1) \ln \ell_o^H\right]$$

Using the same definitions for $\alpha_d, \bar{A}_d$ and $\beta_o, \bar{u}_o$ as we did above we find that this yields

$$\sum_d \ell_d^F \left[\bar{A}_d - \alpha_d + \alpha_d \ln \ell_d^F\right] + \sum_o \ell_o^H \left[\bar{u}_o - \beta_o + \beta_o \ln \ell_o^H\right]$$

When $\mathbb{1}_{\mathrm{I}} = 1$, this coincides with the expression for the planner's problem in the text at a given level of ℓ_d^F, ℓ_o^H completing the demonstration of the isomorphism.

Note that the expression for total consumption in the main text includes the utility benefits stemming from the fact that land is owned by the city. This is captured in the $-\sum_d \ell_d^F \alpha_d - \sum_o \ell_o^H \beta_o$ term.

C Data

C.1 Construction of Zones

In order to obtain a set of locations for the model I use two sources of data. Firstly, I make use of county subdivisions from the US census for New York State and New Jersey. I take a buffer of 20km around the five boroughs in New York City and include all county subdivisions which intersect with the buffer. I also take taxi zones from New York's Taxi and Limousine Company. These describe neighborhoods and provide a more granular view of locations within New York City itself to help capture a more detailed view of traffic routing. I aggregate both of these sets of locations for two reasons. Firstly, the regions differ in geographic size and so I create aggregate zones that are more similar in terms of area. Secondly, for computational reasons, it is useful to slightly reduce the number of regions considered for the analysis.

In order to aggregate these areas I proceed by considering the areas outside New York City and inside separately. I sequentially merge zones by taking the smallest zone and merging it with its nearest neighbor in terms of centroid distance, repeating this process until the desired number of locations is achieved. For the areas outside New York City, I merge zones until 20 locations remain. For the areas within New York City I merge zones until 100 locations remain. This enables me to capture granular commuting flows within New York City while also allowing for in-traffic from the surrounding areas. The locations under consideration account for 89% of all commuters into New York City.

C.2 Department of Transportation's Automated Vehicle Counts

In order to measure traffic flows within the city I make use of the New York Department of Transportation's Automated Vehicle Counts. These provide automated counts of the number of vehicles passing through a given road segment by hour of the day on different days. The number of observations per segment, and the dates it is recorded, vary by segment. I match these counts to the Department of City Planning's LION file for road segments to find their geographical location. This results in 2,344 road segments that are displayed in Figure 5. They cover a large section of New York City and are relatively well dispersed throughout the five boroughs.

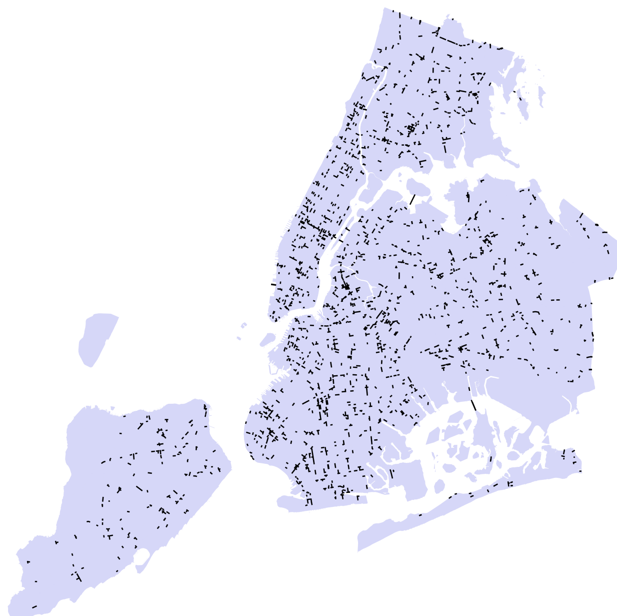


Figure 5: Traffic Count Locations

These counts are then aggregated to the zones in the previous section to find the bilateral mean flow of traffic per lane between pairs of zones in the study area. Lane information comes from the LION data and I take means first by road segment and hour of the day for weekdays and then across roads within each pair of zones. This is used in the estimation of the congestion technology.

C.3 Open Street Map Data and Covariates

For detailed data on roads throughout the whole study area I make use of Open Street Maps (OSM). This provides the exact layout of roads and junctions in the area. I use Boeing

(2017) to process the raw shapefiles from OSM into a set of nodes and edges. An example of part of this from lower Manhattan and Brooklyn is presented in figure 6

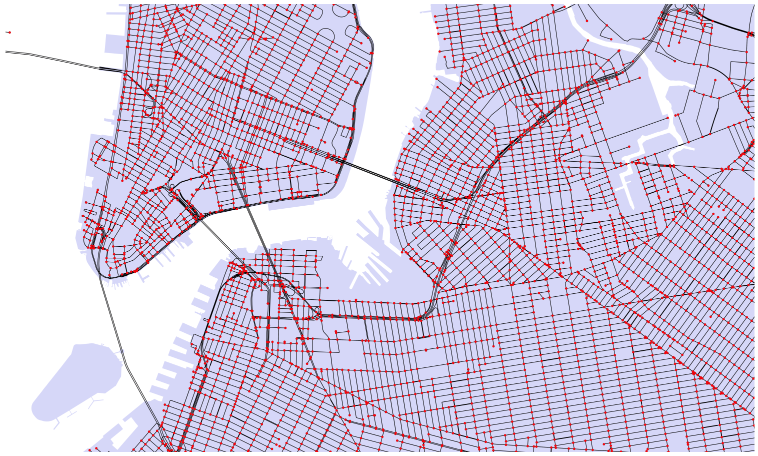


Figure 6: Open Street Maps Data

To obtain the network obtained in the main text, I create a directed link between two zones whenever there is a road that passes in that direction between them. I also use the data from open street maps to create covariates to parameterize the congestion function across links. The full set of covariates is presented in the table below. All variables are included separately for the source and target zone of a link so that the covariates represent the directed nature of the network.

This leads to a total of 115 covariates that are used in the estimation of the congestion function. Given, the large number included, the machine learning techniques discussed in the main text are used to avoid over-fitting and return a good estimate of the congestion technology.

C.4 LODES and ACS Data

I use the Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics from 2019 for commuting flows. I use primary jobs (JT01) for 2019 as the main source of data. Note that, as highlighted by Dingel and Tintelnot (2020) and detailed in Graham et al.

Variable Description	Further Details	Number of Variables
Length of edges		2
Length of lanes		2
Quantiles of lane distribution	At {0, 0.25, 0.5, 0.7, 1.0}	10
Quantiles of speed limit	At {0, 0.25, 0.5, 0.7, 1.0}	10
Quantiles of freeflow time by road	At {0, 0.25, 0.5, 0.7, 1.0}	10
Freeflow time of the link		1
Counts of road segments by type	Types: {motorway, motorway link, primary, primary link, residential, secondary, secondary link, tertiary, tertiary link, trunk, trunk link, unclassified, living street, crossing, road, busway}	32
Quantiles of node degrees	At {0, 0.25, 0.5, 0.7, 1.0}	10
Counts of junction types	Types: {crossing, junction, stop, signals, turning circle, turning loop, crossing, motorway junction, traffic signals-crossing, gantry, mini-roundabout, trailhead, give way, priority, bus stop, disused}	32
Distance to the CBD		2
Longitude and Latitude		4

Table 1: Included Covariates Table

(2014), the LODES counts have noise added to preserve anonymity. The data is provided at the tract level and issues with noise become increasingly important at more granular levels. Since I aggregate a large number of census tracts into zones, this is less of a concern in the present study. I also find that 96% of commuting pairs have non-zero flows at the level of aggregation I consider. Secondly, I use the American Commuting Survey (ACS) 5-year data for 2019 to find data on wage income, hours worked, and commuting mode shares. Again, I aggregate the data to the level of zones defined in C.1.