unibz
Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management
Paper Series

NO 67/ 2020

Congestion and Incentives in the Age of Driverless Fleets

Federico Boffa, Alessandro Fedele,
Alberto Iozzi

Congestion and Incentives in the Age of Driverless Fleets*

Federico Boffa† Alessandro Fedele‡ Alberto Iozzi§

November 16, 2020

Abstract

Fleets of ridesharing companies, such as Uber and Lyft, are providing an increasing share of passenger travel. Autonomous vehicles will likely only magnify this trend. We analyze the welfare effects of the transition from a decentralized regime, in which travelers are atomistic and do not internalize the congestion externality, to a centralized regime, where travelers are supplied by a monopolist's fleet. The monopolist can sort travelers across routes based on their congestion disutility, and ration them. A centralized regime is always welfare-reducing when the monopolist chooses not to ration travelers. We then characterize optimal road taxes throughout the transition.

Keywords: congestion externality, fleets, autonomous vehicles, sorting, rationing.

JEL Codes: R41, R11.

*We are grateful to Paul Belleflamme, Toomas Hinnosaar, Jakub Kastl, Ryan Kellogg, Ignacio Monzon, Giacomo Ponzetto, Aleksey Tetenov, Matthew Turner and Tommaso Valletti for useful discussions and helpful comments. We thank participants to the NBER Workshop on "The economics of autonomous and electric vehicles" in Stanford, to seminars at Collegio Carlo Alberto, Free U. of Bolzano, Hong Kong Polytechnic, IMT Lucca, U. of Siena, SOAS U. of London, U. Catholique de Louvain, U. of Rome 'Tor Vergata', U. of Naples 'Federico II'/CSEF, to workshops at Católica Porto Business School, U. Cattolica (NERI), U. of Sassari (IBEO), and U. of Bergamo, and to session conferences at AISRE, ASSET, EARIE, ERSA, IAERE, ITEA, JEI, SIE, SIEPI, and SIET. All errors are ours.

†Free University of Bolzano and Collegio Carlo Alberto: federico.boffa@unibz.it.

‡Free University of Bolzano: alessandro.fedele@unibz.it.

§Università di Roma 'Tor Vergata' and SOAS University of London: alberto.iozzi@uniroma2.it.

1 Introduction

Technological advancement is rapidly changing the mobility industry. As a result of the diffusion of smart phones and of geo-localization systems, an increasing share of passenger travel is provided by the fleets of ridesharing companies, such as Uber and Lyft. In 2016, fleets accounted for 15% of all intra San Francisco vehicle trips. In New York, the share of vehicle trips covered by fleets doubled annually between 2014 and 2016 (Erhard et al. 2019). This trend will likely only be magnified when fleets will become cheaper as fully autonomous vehicle, which do not require a driver, will be deployed. Some smaller scale experimentation with fleets of robotaxi, i.e. taxis operated by ride service companies through autonomous vehicles (AVs, henceforth), is currently ongoing in selected urban areas. Examples include Baidu's fleet in Changsha City in China (Liu et al., 2020), Voyage's in Florida (Bloomberg, 2020) and Waymo's - an Alphabet's subsidiary - in Phoenix, in partnership with Lyft. In addition, Tesla has recently announced that it will soon stop selling cars to private owners, and use them for its own - comparatively more profitable - robotaxis fleet instead.[1] This is in line with the widespread prediction that consumers' investment into private cars is bound to shrink and that urban traffic will be organized around fleets (Fagnant and Kockelman, 2015, and Ward et al., 2019).

The increased importance of fleets managed by ridesharing companies is entaling a process of centralization of urban traffic. This centralization already involves the pricing as an increasing share of urban traffic is being priced, and pricing strategies are determined by ridesharing companies. This process is likely to continue insofar ridesharing companies may gain control also of location and routing decision of vehicles.[2] This last step, albeit possible with traditional vehicles, will be a natural development with fleets of AVs. This will have dramatic consequences on congestion. Congestion in the mobility industry not only derives from transport infrastructures being inadequate relative to demand, but is also the result of a standard externality. In a conventional decentralized setting, with no fleets, drivers do not factor in their travel decisions the external effect in terms of congestion they impose on fellow travelers.

The paper studies how the ongoing transition to a centralized mobility market, organized around ridesharing companies that manage their own fleets, affects congestion, and, as a result, welfare. Centralization, in our model, affects both the pricing and the routing decisions. It may seem intuitive that firms in a centralized setting internalize (at least partly) the congestion externality inherent in the decentralized setting, thereby contributing to an increase in welfare. We show that this intuition is incomplete, and we emphasize the distortions occurring in a centralized market. We then explore the taxation schemes that allow to restore first best. We show that congestion charges are not optimal in a setting with ridesharing

[1] https://www.wsj.com/articles/a-tale-of-two-teslas-elon-musk-to-tout-robot-cars-amid-sales-slump-11555848000 (last accessed November 6, 2020).

[2] Currently, ridesharing companies set the trip fares. However, they cannot directly set their drivers' supply and their drivers' location, since drivers are independent contractors, as opposed to employees. As a result, ridesharing companies have up to now been acting as two-sided platforms, and using location-based peak-load pricing (the so-called surge pricing) to incentivize drivers to be available when demand is high, and to show up in the location that the company finds optimal (Guda and Subramanian, 2019). Drivers' contractual position is bound to change as a result of some state court rulings that require Uber and Lyft to classify their drivers as employees (https://www.npr.org/2020/10/22/926916925/uber-and-lyft-must-make-drivers-employees-california-appeals-court-rules?t=1604069805691, last accessed November 6, 2020). If this ruling will end up being implemented, ridesharing companies will then be able to directly control the availability and the location of their drivers.

fleets.

Understanding the impact of traffic centralization on congestion is of utmost importance because congestion costs, while hidden and hard to measure because of their nature of opportunity costs, represent a major component of the traveling costs. Congestion not only increases the average travel time, but it also raises its variance. The yearly congestion cost has been estimated to amount to more than one hundred billion dollars in the US, and to be steadily increasing over time (Schrank, Lomax and Eisele, 2011). Fleets managed by ridesharing companies, mainly Uber and Lyft, have increased congestion in San Francisco (Erhardt et al., 2020). In the absence of appropriate policies, congestion costs are not bound to disappear when AVs will be deployed. On the one hand, with AVs, consumers may spend more productively their time on vehicles, thereby reducing the disutility of congestion when one holds the level of congestion constant. On the other hand, however, the ability to use time in the car more fruitfully will increase congestion, by inducing more travel demand (Gucwa, 2014). The overall effect on disutility of congestion will depend on the relative strength of these two effects, something hard to predict now. It is well possible that congestion costs will ultimately increase, reproducing an effect similar to the positive association between new infrastructure and kilometers traveled expressed in the fundamental law of road congestion (Downs, 1962; Duranton and Turner, 2011).

A proper analysis of congestion requires to consider travelers' heterogeneity in the disutility from congestion. This is substantial (see, for instance, Small, 2012) and reflects heterogeneity in individuals' value of time, as well as in value of reliability. Small, Winston and Yan (2005) find that the difference in value of time between the 25th and the 75th percentile is about \$10 per hour, with the median being about \$21. Even more starkly, the difference in value of reliability between the 25th and the 75h percentile is \$13 per hour and higher than the median of \$12. Also, as intuitive, disutility from congestion is known to be tightly positively associated to income. Estimates of the elasticity of time value to income range from about 0.5 to 1, and they are increasing at higher levels of income (Börjesson, Fosgerau and Algers, 2012). The current trends of increasing inequality in many Western countries, in the context of a rising average income, should only exacerbate such heterogeneity.

With heterogenous travelers, the reduction in aggregate congestion costs, as well as welfare maximization, requires to act not only on the margin of the total number of vehicles that travel, but also on the efficient sorting of travelers. To see this, consider a highway with two lanes. With heterogeneous disutility from congestion, efficiency may require to differentiate the speed across the two lanes, thereby creating a faster lane with fewer vehicles for travelers that dislike congestion a lot, and a slower lane with more vehicles for drivers who are less bothered by congestion.

Efficient sorting of travelers is economically very relevant. The combined evidence in Small, Winston and Yan (2005, 2006) shows that a little less than the average hourly disutility from congestion can be mitigated through an efficient sorting of travelers. They simulate the effects of several alternative lane management schemes through tolls and/or high occupancy vehicles lanes, which give rise to different allocations of heterogenous travelers on a California State highway. They find that the welfare-maximizing scheme yields about \$2.50 welfare gain in a 15 minute trip over the alternative with no tolls and no lane utilization rules. This is equivalent to $10 per hour per passenger, which is more than half of the average hourly wage in the United States, and just less than half of the median value of time that they estimate.

In addition, sorting vehicles across different routes will become increasingly easy for a central decision-maker, be it a government or a ridesharing company, when AVs will be deployed. The software that drives the AVs may itself be used to sort vehicles across different routes, even based on other vehicles' behavior, without the possibility of interference by travelers. Also, GPS and mobile technologies will contribute to the diffusion of road pricing schemes. These schemes will be increasingly sophisticated, as prices may be conditioned on a variety of dimensions (route, time of the day, occupancy, etc.), and cheaper, as a result of the reduction in the costs of the required infrastructure (Ostrovsky and Schwarz, 2018). AVs will also reduce the extent of the matching externalities that taxi and ridesharing companies are currently facing (Frechette, Lizzeri and Salz, 2019).

In this paper, we consider a stylized framework with individuals using AVs to travel on a road, segmented into two separate parallel lanes, both congested. The lanes are ex-ante identical, but can ex-post differ in the amount of congestion. Individuals are assumed to be heterogenous as to the utility they derive from the trip and to the disutility they derive from congestion. Consistent with evidence pointing to a positive relation between income and value of time, we assume that individuals with a larger utility from the trip suffer from a larger cost of congestion. We look at the equilibrium assignment of individuals to one of the two lanes or to not traveling.

We first show that welfare maximization requires to differentiate the congestion level in the two lanes, reflecting the heterogeneity in individuals' value of time. Individuals with low disutility of congestion travel in a slow lane, while those with a high disutility of congestion travel in a fast lane. Furthermore, an individual travels as long as her benefit from traveling exceeds the increase in aggregate congestion costs she imposes on fellow travelers. Thus, if the congestion cost is sufficiently large, efficiency requires to prevent some low-value individuals from traveling.

We then study the welfare effects of moving from a decentralized to a centralized regime. In a decentralized regime, all individuals are atomistic and do not factor in their travel decisions the external congestion effect they impose on fellow travelers. In a centralized regime, all the mobility services are provided by a monopolist through its fleet. We also look at a mixed regime, where a portion of individuals is atomistic, while the others are supplied by the monopolist. This reflects the fact that the transition towards an economy fully organized around fleets, which has already started, will likely be gradual. We first focus on how travelers are sorted across the two lanes in a situation in which parameter values are such that, in all regimes, all individuals travel. We then study parameter values for which the aggregate amount of travelers varies and allow for rationing.

When parameter values are such that, in each regime, all individuals travel, the two lanes have exactly the same mass of travelers under a decentralized regime. When the market is centralized, there is too much differentiation in congestion across lanes with respect to the social optimum. From the welfare standpoint, this excess differentiation turns out to be worse than the no differentiation prevailing in the decentralized setting. An additional welfare-reducing inefficiency may emerge under the mixed regime, in that all atomistic travelers, including those with a relative low value of traveling and disutility from congestion, travel in the fast lane. Overall, we find that welfare remains unchanged as long as the share of travelers using the ridesharing fleet is small enough. When the share of travelers using the ridesharing fleet is large enough, welfare monotonically declines as this share increases. In reality, we have no evidence of ridesharing companies strategically trying to manipulate congestion so as to offer some travelers a more expensive, but quicker travel option. This squares with the ridesharing company still having too small a share of total travel to be able to affect the congestion patterns. Under these circumstances, our model shows that the presence of ridesharing companies does not affect welfare.

Our results change for parameter values under which the monopolist excludes some low-value individuals from traveling – in order to increase prices on, and extract more value from, travelers – while all individuals still travel in the decentralized setting. If congestion costs are sufficiently severe that the social planner would efficiently exclude some low-value agents from traveling, the quantity reduction operated by the monopolist with respect to a decentralized regime may be efficient. If, to the contrary, congestion costs are not so large in the first place, and the planner would not ration any individual, the monopolist's screening is welfare-reducing over decentralized travel.

Part of our results parallel those obtained in the airline economics literature, under monopolistic air carriers, in terms of the congestion levels: see, for instance, Brueckner (2002) and Basso (2008), and the empirical counterparts estimating the relation between airport concentration and congestion (Mayer and Sinai, 2003; Rupp, 2009; Daniel and Harback, 2008; and Molnar, 2013). However, we crucially add individuals' heterogeneity in the disutility from congestion, and, as a result, we examine how the sorting of travelers across lanes can be used to mitigate the negative effects of congestion on welfare. Our findings also relate to Czerny and Zhang (2015), who study third degree price discrimination in the presence of congestion externalities.

We then analyze the case of a tax authority able to impose taxes that restore social optimality. In a decentralized regime, a traditional congestion charge, i.e., a Pigouvian tax, equal to the marginal (external) cost imposed on the other vehicles, restores optimality. This mirrors the finding obtained in the bottleneck model (see Vickrey, 1969; and Arnott, de Palma and Lindsey, 1990). Instead, under a centralized regime, the tax that restores social optimum is very different. Since the monopolist considers in its pricing policy the effects of the congestion costs on all travelers, there is no scope for a congestion charge. This result aligns with those obtained in the literature on airports when carries have market power (Daniel, 1995; Brueckner, 2002; Pels and Verhoef, 2004; Brueckner 2005; Basso and Zhang, 2007; and Silva and Verhoef, 2013). We then characterize a simple tax/subsidy scheme that restores the incentives to optimality both in the degree of differentiation across lanes and in the aggregate number of travelers. We show that, when the congestion problem is particularly severe, this scheme may involve a net subsidy for the monopolist. This can prove politically challenging, even more than traditional congestion pricing schemes, and may require some countermeasures by the tax authority that improve its political feasibility. We also find that, in the mixed regime, atomistic travelers should pay the congestion charge, while travelers using the monopolist's service should pay a tax similar to that prevailing in the centralized regime. This result casts doubts on the welfare effects of the congestion charge imposed in Manhattan, since 2019, to taxis and ridesharing companies (NYC Taxi and Limousine Commission and Department of Transportation, 2019).

To the best of our knowledge, there are only a few papers that consider congestion with reference to AVs. Lamotte, De Palma and Geroliminis (2017) develop a bottleneck model to investigate the commuters' choice between conventional and autonomous vehicles, while van den Berg and Verhoef (2016) focus on the impact of AVs on road capacity, studying the deployment of infrastructures resulting from the transition to AVs.[3] Finally, our paper is close to Ostrovsky and Schwarz (2018), who investigate the interplay between autonomous transportation, carpooling, and road pricing to achieve socially efficient outcomes.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 illustrates the first best. Section 4 and Section 5 characterize the equilibrium for parameter values under which the total number of travelers is fixed across regimes and varies across regimes, respectively. Section 6 examines taxation to restore social optimality. Section 7 concludes. Derivations and proofs of all Propositions are relegated to an appendix.

2 The model

Lanes and individuals' utility. There is a unit mass of individuals, each with unit demand for a trip from a common origin to a common destination. Trips occur along a single road connecting the origin and the destination. The road is divided into two ex-ante identical lanes that are congested at any positive mass of travelers. The lanes may, however, differ ex-post because of a different mass of travelers, leading to different levels of congestion. We refer to the (weakly) more congested lane as the *slow* lane and denote its mass of traveler by s. Similarly, the (weakly) less congested lane is referred to as *fast* lane and its mass of traveler is denoted by f.

Individuals are heterogeneous. For simplicity, their type θ is assumed to be uniformly distributed in the $[0, 1]$ interval.[4] A type-θ individual has the following utility function:

$$U(\theta) = \begin{cases} 0 & \text{when not traveling;} \\ B(\theta) - \theta g s & \text{when traveling in the slow lane;} \\ B(\theta) - \theta g f & \text{when traveling in the fast lane.} \end{cases} \tag{1}$$

The term $B(\theta)$, with $B'(\theta) > 0$ and $B''(\theta) \leq 0$, is the gross benefit from traveling. The terms $\theta g s$ and $\theta g f$ denote the disutility from congestion. They depend on: *i*) the mass of travelers in the same lane, either s or f; *ii*) a type-independent parameter $g > 0$ representing the common component of congestion disutility across travelers; *iii*) the type θ representing the idiosyncratic component of congestion disutility. Note that θ determines the travelers' value of travel and, at the same time, affects their cost of congestion. The assumption that both increase with θ is consistent with evidence that points at a positive relation

[3] Two other papers analyze equilibria when drivers are non-atomistic (Silva et al. 2016, Lindsey, de Palma and Silva, 2019), while Simoni et al. (2019) use agent-based simulations to evaluate the impact of different congestion pricing and tolling strategies in the presence of AVs.

[4] See, e.g., Brueckner (2002) for the uniform distribution assumption in the case of airline passengers.

between wage and value of time (see, for instance, Small, 2012).

We assume that the common component of congestion disutility, g, is sufficiently low so that the net utility of travelers is increasing in θ for any s and f, i.e., $U'(\theta) > 0$. A necessary and sufficient condition for this to occur is

Assumption 1. $g < B'(1)$.

This relatively low value of g may be consistent with the likely reduction in the disutility from congestion entailed by the use of AVs.

To avoid the uninteresting case of some low θ-types never wanting to travel, we posit that the type-0 individual's utility from traveling is nonnegative,

Assumption 2. $U(0) = B(0) \geq 0$.

Assumptions 1 and 2 together imply that all individuals get nonnegative utility from traveling.

Individuals' identity. We assume that the mass of individuals is potentially composed of two different groups:

- *atomistic individuals*: when traveling, they use AVs that do not belong to a fleet, and do not factor in their travel decisions the congestion cost they impose on fellow travelers;
- *corporate individuals*: when traveling, they use the services of a fleet of AVs.

We let the mass of corporate individuals be equal to $\mu \in [0, 1]$ and the remaining mass $1 - \mu$ be composed of atomistic individuals. These two proportions are assumed to be exogenous. This is because belonging to either group may depend on individual preferences or on long-run decisions, such as the choice to use an owned car or not, which are not modeled here. Moreover, we let the distribution of the two groups of individuals be independent of the type θ. That is, in any subinterval $[\theta, \theta + \epsilon]$ of the unit line, with $\epsilon > 0$, there is a fraction μ of corporate individuals and a fraction $1 - \mu$ of atomistic individuals.

Depending on the value of μ, we will explore three possible compositions of the total mass of individuals, meant to illustrate the stages of the transition from a decentralized to a centralized regime:

- *atomistic individuals only:* $\mu = 0$. All individuals use their own car. This is referred to as the decentralized regime;[5]
- *atomistic and corporate individuals:* $\mu \in (0, 1)$. A share $1 - \mu$ of individuals is atomistic and the remaining share μ of individuals use the fleet of AVs. This is referred to as the mixed regime. The larger is μ, the more advanced the process of transition towards a centralized regime is;
- *corporate individuals only:* $\mu = 1$. All individuals use the fleet of AVs. This is referred to as the centralized regime.

[5]This scenario resembles the current traffic organization with traditional vehicles. The diffusion of navigation systems using real time information provides individuals with an easy solution to the informational and computation problem of choosing the individually optimal route.

The fleet manager. We assume that the fleet of AVs providing transportation services to corporate individuals is centrally managed by a monopolistic firm. Trips have zero marginal costs. The monopolist charges a different price in each lane, but uniform within each lane. It can price discriminate across customers based on the lane used only, but not on the individual's type θ, possibly because of a privacy protection regulation (Montes, Sand-Zantman and Valletti, 2018).

The game. We initially look at a two stage game. In the first stage, the monopolist chooses the fares paid by corporate individuals; in the second stage, all individuals, both corporate and atomistic, simultaneously make their travel decisions. All players have full information on the entire game and the equilibrium concept is subgame perfect Nash equilibrium. When all individuals are atomistic, the first stage is irrelevant and we ignore it.

In Section 6, we add an initial stage to this game, in which a tax authority chooses a tax scheme. After this initial stage, the rest of the game unfolds as described before. In this game, we restrict the tax authority to charge unit (per travel) taxes, possibly different by lane but not by individual's identity.

Individuals' incentives. Both atomistic and corporate individuals choose whether or not to travel, and in which lane to do so, based on private incentives. For a type θ, the decision depends on her utility $U(\theta)$ and on the fares and/or taxes (if any) she pays when traveling in the slow or the fast lane, denoted as $\sigma \geq 0$ and $\phi \geq 0$, respectively. A type-θ individual travels in the slow lane if and only her individual rationality (IR) constraint holds, that is

$$B(\theta) - \theta g s - \sigma \geq 0, \tag{2}$$

and if and only if her incentive compatibility (IC) constraint holds, that is

$$B(\theta) - \theta g s - \sigma \geq B(\theta) - \theta g f - \phi. \tag{3}$$

Similarly, she travels in the fast lane if and only if

$$B(\theta) - \theta g f - \phi \geq 0, \tag{4}$$

$$B(\theta) - \theta g f - \phi \geq B(\theta) - \theta g s - \sigma. \tag{5}$$

When at least one of the IR constraints holds for all individuals, they all travel so the market is fully covered (i.e., $s + f = 1$). Instead, if for some individuals neither IR constraints hold, such individuals prefer not to travel so the market is only partially covered (i.e., $s + f < 1$).

3 First best

We consider an utilitarian welfare maximizing social planner, who is perfectly informed. It can decide which individuals travel and directly allocate those traveling to the two lanes. The social welfare function is given by

$$W \equiv \int_0^1 U(\theta) d\theta. \tag{6}$$

Welfare maximization requires to partition travelers in (at most) three groups. Some low θ-types may not travel, while all the other θ's are partitioned into the two lanes, with the highest θ's traveling in the fast lane.[6] Then, (6) may be rewritten as:

$$W' \equiv \int_{1-s-f}^{1-f} (B(\theta) - \theta g s)\, d\theta + \int_{1-f}^{1} (B(\theta) - \theta g f)\, d\theta \tag{7}$$

where the first integral gives the aggregate utility of types traveling in the slow lane, and the second integral gives the aggregate utility of types traveling in the fast lane. The planner's problem may be written as follows:

$$\max_{s \geq 0, f \geq 0} \quad W' \tag{8}$$
$$\text{s.t.} \quad s + f \leq 1.$$

An interior solution satisfies the following first order conditions (FOCs, henceforth):

$$B(1 - s_{FB} - f_{FB}) - 2 g s_{FB} \left(1 - f_{FB} - \frac{3}{4} s_{FB}\right) = 0, \tag{9}$$

$$B(1 - s_{FB} - f_{FB}) + g \left(s_{FB}^2 + \frac{3}{2} f_{FB}^2 - 2 f_{FB}\right) = 0, \tag{10}$$

where the subscript FB is a mnemonic for equilibrium variables referred to First Best. The solution to the planner's problem is characterized in the following Proposition.

Proposition 1. *Let s_{FB} and f_{FB} denote the solutions to problem (8) when the market is partially covered, i.e., $s_{FB} + f_{FB} < 1$. Also, let $\bar{s}_{FB}$ and $\bar{f}_{FB}$ denote the solutions to problem (8) when the market is fully covered, i.e., $\bar{s}_{FB} + \bar{f}_{FB} = 1$. Finally, let*

$$g_{FB} \equiv \frac{36\, B(0)}{4 + \sqrt{7}} \cong 5.4179 \times B(0). \tag{11}$$

Then, in equilibrium

- *when $g > g_{FB}$, the market is partially covered. The solutions s_{FB} and f_{FB} exist and are unique. Travelers with $\theta \in [0, 1 - s_{FB} - f_{FB}]$ do not travel, those with $\theta \in [1 - s_{FB} - f_{FB}, 1 - f_{FB}]$ are allocated to the slow lane and those with $\theta \in [1 - f_{FB}, 1]$ are allocated to the fast lane.*

- *when $g \leq g_{FB}$, the market is fully covered. The solutions $\bar{s}_{FB}$ and $\bar{f}_{FB}$ are equal to*

$$\bar{s}_{FB} = \frac{1}{2} + \frac{\sqrt{7} - 2}{6} \cong 0.6076 \quad \textit{and} \quad \bar{f}_{FB} = \frac{1}{2} - \frac{\sqrt{7} - 2}{6} \cong 0.3924. \tag{12}$$

[6] See the appendix for the derivation of this result.

Travelers with $\theta \in [0, \bar{s}_{FB}]$ are allocated to the slow lane and those with $\theta \in [\bar{s}_{FB}, 1]$ are allocated to the fast lane.

This Proposition characterizes the socially optimal allocation of individuals. The planner may exclude the individuals with the lowest benefit from traveling, so that the market is not fully covered. This occurs when the utility from traveling in the slow lane enjoyed by the type-0 individual, $B(0)$, is lower than the increase in the aggregate congestion costs this individual imposes on all fellow travelers in the slow lane.[7] All the remaining travelers are sorted in the two lanes. A mass of travelers equal to s_{FB} (or $\bar{s}_{FB}$, in the case of full coverage) is allocated to the slow lane and a mass of travelers f_{FB} (or $\bar{f}_{FB}$, in the case of full coverage) to the fast lane. Intuitively, travelers allocated to the fast lane are those with the highest θ.

In the case of partial coverage, the solutions to the planner's problem exist and are unique. However, their explicit characterization requires to specify the gross benefit function $B(\cdot)$. While keeping the general form of $B(\cdot)$, useful insights can be obtained by simply expressing the optimal choice of f_{FB} as a function of s_{FB}. This is obtained by simply manipulating the FOCs in (9) and (10) and reads as follows

$$f_{FB}(s_{FB}) = \frac{1}{3}\left(2\left(1 + s_{FB}\right) - \sqrt{7s_{FB}^2 - 4s_{FB} + 4}\right). \tag{13}$$

This relationship is illustrated in Figure 1 by the green increasing line, and allows us to discuss some features of the first-best equilibrium. First, the planner finds it optimal to differentiate across lanes. Travelers with relatively high θ, who suffer the most from congestion, are assigned to the fast, less congested, lane. This is why the green line in Figure 1 always lies below the 45° line. Second, by implicit differentiation of the FOCs in (9) and (10), we find that $\frac{\partial s_{FB}}{\partial g} < 0$ and $\frac{\partial f_{FB}}{\partial g} < 0$. This reflects the intuition that a larger common component of the disutility from congestion, g, is associated to a lower mass of travelers in each lane, and, therefore, to a lower market coverage. Full market coverage is illustrated in Figure 1 by the green solid circle located at the intersection between the green line illustrating equation (13) and the constraint $s + f = 1$.

4 Full coverage: equilibrium analysis

In our framework, the optimal management of congestion may require both to optimally ration travel and to optimally sort travelers across lanes. As discussed in the introduction, given the heterogeneity in individuals' disutility from congestion, efficient sorting may have a very significant impact on reducing the aggregate congestion costs. In many markets, the introduction of road pricing involves sorting consumers across routes with different speed, but little or no rationing (as can be inferred, for instance, by the analysis in Small, Winston and Yan, 2006). In addition, AVs, along with other recent developments in urban transport, are bound to dramatically reduce the cost of the sorting technology, thereby making sorting a viable and increasingly important tool to manage congestion.

[7] Condition $g > g_{FB}$ can be rearranged as $B(0) < g\frac{(\bar{s}_{FB})^2}{2}$, where the RHS is indeed the aggregate marginal congestion cost for those traveling in the slow lane.

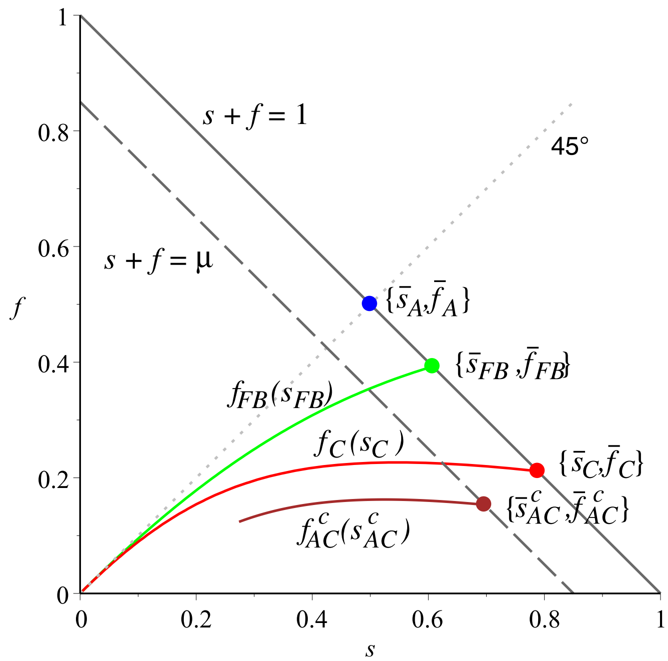

Figure 1: Equilibrium outcomes.
The Figure illustrates: *i*) the equilibrium solutions with decentralization (blue dot, $\{\bar{s}_A, \bar{f}_A\}$); *ii*) the full coverage solutions in the first best (green dot, $\{\bar{s}_{FB}, \bar{f}_{FB}\}$), *iii*) the full coverage solutions with centralization (red dot, $\{\bar{s}_C, \bar{f}_C\}$); *iv*) the full coverage solutions in the mixed regime (brown dot, $\{\bar{s}^c_{AC}, \bar{f}^c_{AC}\}$); *v*) the relationship between equilibrium s and f in the first best (green line, $f_{FB}(s_{FB})$); *vi*) the relationship between equilibrium s^c_{AC} and f^c_{AC} with centralization (red line, $f_C(s_C)$); *vii*) the relationship between equilibrium s^c_{AC} and f^c_{AC} in the mixed regime (brown line, $f^c_{AC}(s^c_{AC})$): this is defined for values of s^c_{AC} above a threshold, since when s^c_{AC} is too low the equilibrium involves no differentiation, assuming a mass of atomistic travelers equal to $\mu = 0.85$).

We thus start by considering the set of parameters for which the market is fully covered (thereby ruling out, for now, rationing), so as to focus on the allocative problem of sorting travelers between the fast and the slow lane, and on the resulting differentiation in the level of congestion across lanes. To this aim, we introduce the following assumption that gives a sufficient condition for full coverage to occur in the first best and in equilibrium for all the regimes:[8]

Assumption 3. $\frac{B'(0)}{B(0)} \leq 1$.

Assumption 3 states that at $\theta = 0$ the semielasticity of the gross benefit function $B(\theta)$ is weakly lower than 1. In other words, the gross benefit from traveling for type-0 individual is large relative to its increase as θ rises.

4.1 Decentralized regime

We first study a decentralized regime populated by atomistic individuals only. Each individual uses an AV that does not belong to a fleet, does not pay any fare and chooses the lane giving her the highest

[8]See the appendix for the derivation of this result.

utility, ignoring the effect of this choice on other individuals. Since the monopolist has no role in this regime, the stage of the game in which it sets its fares is muted.

The next Proposition shows that in any equilibrium of this game travelers split equally across the two lanes, so that, differently than the social optimum, there is no differentiation across lanes. Here and in the rest of the paper, equilibrium variables referred to the case of a market with atomistic individuals only have the subscript A, a mnemonic for Atomistic.

Proposition 2. *Assume Assumption 3 holds and all individuals are atomistic. Let $\bar{s}_A$ and $\bar{f}_A$ be the equilibrium masses of travelers in the two lanes in full coverage. Then, any allocation of travelers such that*

$$\bar{s}_A = \bar{f}_A = \frac{1}{2} \tag{14}$$

is an equilibrium.

This Proposition illustrates that in equilibrium the two lanes feature the same level of congestion. To get the intuition behind this result, argue by contradiction and suppose that, for instance, $s_A > f_A$. This cannot be an equilibrium outcome because any traveler in the more congested lane, irrespective of her type θ, would prefer to switch to the less congested lane where she would enjoy a higher net utility. We remark that there are infinite allocations that satisfy condition (14), and they all are equilibria of the game. This is because travelers, independently of their type, get the same nonnegative utility across lanes.

4.2 Mixed regime

We now analyze the case of a mixed regime, where a mass $\mu \in (0, 1)$ of individuals use the fleet of AVs managed by the monopolist, while the remaining mass $1 - \mu$ are atomistic travelers. Corporate travelers pay a lane-specific fare to the monopolist in exchange for the service. The monopolist sets uniform fares within lanes: a fare p for the slow lane and a fare P for the fast lane. Fares are used by the monopolist as the only instrument to direct corporate individuals to the two lanes. Here and in the rest of the paper, equilibrium variables referred to this market have a subscript AC, a mnemonic for Atomistic and Corporate.

We first establish some conditions on the equilibrium allocations of travelers. Denote with s^c and f^c the mass of corporate travelers in the slow and fast lane. Similarly, denote with s^a and f^a the mass of atomistic travelers in the slow and fast lane. Thus, the total mass of travelers in the two lanes is given by $s = s^a + s^c$ and $f = f^a + f^c$.

When μ is relatively low, $1 - \mu \geq s^c - f^c$, only equilibrium allocations of travelers in which there is no differentiation across lanes may exist, so that in equilibrium it must be that

$$s = f = \frac{s^c + f^c + s^a + f^a}{2}. \tag{15}$$

Indeed, if some differentiation across lanes existed, atomistic travelers in the slow lane would prefer to switch lane and enjoy the lower congestion of the faster lane. Because of their relatively large mass, this would make the fast lane more congested, contradicting its nature of faster lane. Instead, when μ is relatively high, $1-\mu < s^c - f^c$, an equilibrium allocation of travelers is such that there is differentiation across lanes and all atomistic individuals travel in the fast lane, so that

$$\begin{aligned} s &= s^c, \\ f &= f^c + f^a = f^c + (1-\mu). \end{aligned} \tag{16}$$

In this case, the mass of atomistic travelers is relatively small. For any allocation of corporate travelers, when all atomistic travelers choose the fast lane, congestion is still less than in the slow lane.

We turn now to analyze our two-stage game, starting from the case where μ is relatively low. The equilibrium allocation of travelers must satisfy condition (15). Since $s = f$, the monopolist must charge the same fares for the two lanes, i.e., $p = P$. The IR constraints (2) and (4) collapse to a single one, which, because of full coverage, can be written simply as $p \leq B(0)$. The monopolist problem may thus be written as

$$\begin{aligned} &\max_{s^c \geq 0, f^c \geq 0} B(0)(s^c + f^c), \\ &\text{s.t.} \quad s^c + f^c = \mu, \end{aligned} \tag{17}$$

where we take constraint $p \leq B(0)$ to be binding, as it must be in equilibrium. This maximization problem has multiple solutions in that the monopolist's profits are affected by the total mass of corporate travelers only, which in turn is uniquely determined by the constraint. As it will be clear in what follows, these multiple solutions are all equivalent, in terms of profits, congestion level of the two lanes and welfare.

Next, focus on the case where μ is relatively large, so that the equilibrium allocation of travelers must satisfy (16). In setting the fares to allocate corporate individuals – p for the slow lane and P for the fast lane – the monopolist faces constraints from (2) to (5). By a standard argument, only the IR constraint (2) for the marginal traveler in the slow lane – type $\theta = 0$ given our focus on full coverage – and the IC constraint (5) for the (corporate) traveler indifferent between the slow and the fast lane – type $\theta = 1 - \frac{f^c}{\mu}$ – are binding. These constraints can be written as $p = B(0)$ and $P = p + g\left(1 - \frac{f^c}{\mu}\right)(s-f)$. The fare p is set to make type 0 corporate individual just willing to travel in the slow lane. Instead, P clearly illustrates the trade-off the monopolist faces in choosing the profit maximizing degree of differentiation between lanes. On the one hand, an increasing difference in the mass of travelers, $s - f$, entails an increasing extra-fare paid by travelers in the fast lane, hence an increasing mark-up. On the other hand, a large differentiation implies that the mass of travelers in the fast lane is small.

Incorporating the two fares in the problem faced by the monopolist, this is given by

$$\begin{aligned} &\max_{s^c \geq 0, f^c \geq 0} B(0)(s^c + f^c) + g\left(1 - \frac{f^c}{\mu}\right)(s-f) f^c, \\ &\text{s.t.} \quad s^c + f^c = \mu. \end{aligned} \tag{18}$$

The next Proposition characterizes the subgame perfect Nash equilibrium of the two stage game described in Section 2.

Proposition 3. *Assume Assumption 3 holds and a mixed regime. Let* $\bar{s}^c_{AC}$ *and* $\bar{f}^c_{AC}$ *denote the equilibrium mass of corporate travelers in full coverage. Similarly, let* $\bar{s}^a_{AC}$ *and* $\bar{f}^a_{AC}$ *denote the equilibrium mass of atomistic travelers in full coverage. Then,*

- *when* $\mu \in \left(0, \frac{1}{2}\right]$*, an equilibrium is any allocation of travelers such that*

$$\begin{aligned} \bar{s}^c_{AC} + \bar{s}^a_{AC} = \bar{f}^c_{AC} + \bar{f}^a_{AC} = \tfrac{1}{2}, \\ \bar{s}^c_{AC} + \bar{f}^c_{AC} = \mu, \\ \bar{s}^a_{AC} + \bar{f}^a_{AC} = 1 - \mu; \end{aligned} \tag{19}$$

- *when* $\mu \in \left(\frac{1}{2}, 1\right)$*, the unique equilibrium is such that corporate travelers with* $\theta \in \left[0, \frac{\bar{s}^c_{AC}}{\mu}\right]$ *are in the slow lane and corporate travelers with* $\theta \in \left[\frac{\bar{s}^c_{AC}}{\mu}, 1\right]$ *are in the fast lane, together with all* $1-\mu$ *atomistic travelers, where*

$$\begin{aligned} \bar{s}^c_{AC} &= \tfrac{1}{2} + \tfrac{\sqrt{4\mu^2 - 2\mu + 1} - 2(1-\mu)}{6}, \\ \bar{f}^c_{AC} &= \tfrac{1}{2} - \tfrac{\sqrt{4\mu^2 - 2\mu + 1} + 4(1-\mu)}{6}, \\ \bar{s}^a_{AC} &= 0, \\ \bar{f}^a_{AC} &= 1 - \mu. \end{aligned} \tag{20}$$

First, we remark that the threshold level of μ that distinguishes between the two types of equilibria, with and without lane differentiation, is equal to $\frac{1}{2}$.[9]

When the mass of corporate individuals is not higher than $\frac{1}{2}$, multiple equilibrium allocations exist. In all of these, the two lanes are equally congested. No matter how the monopolist sorts corporate individuals, atomistic travelers allocate themselves across the two lanes, so as to make them equally congested. This induces the monopolist to charge the same fare across lanes. The equilibrium price is uniquely determined by the IR of the lowest θ, $p = B(0)$ because of full coverage.

When, instead, the mass μ of corporate individuals is higher than $\frac{1}{2}$, the equilibrium allocation of travelers is unique. The mass of atomistic travelers is not sufficiently large to bridge the congestion gap across lanes as determined by the prices set by the monopolist to sort corporate travelers. This means that the monopolist can price discriminate across lanes. Taking into account that all atomistic individuals will use the fast lane, the monopolist allocates a large enough mass of corporate travelers to the slow lane and a relatively small mass to the fast lane. This solution is illustrated by the brown dot in Figure 1.

4.3 Centralized regime

We now consider a centralized regime, $\mu = 1$, in which all travelers are corporate and use AVs that are part of a fleet managed by the monopolist. As in the mixed regime with $\mu \in \left(\frac{1}{2}, 1\right)$, the monopolist faces constraints from (2) to (5) in setting the two fares. Under full coverage, the binding IR constraint (2) for

[9] To see this, we plug $\bar{s}^c_{AC}$ and $\bar{f}^c_{AC}$ as in (20) into conditions $1 - \mu \geq (<)\, s^c - f^c$ and solve by μ; this yields $\mu \leq (>)\, \frac{1}{2}$.

the marginal traveler in the slow lane and the binding IC constraint (5) for the traveler indifferent between the slow and the fast lane read as $p = B(0)$ and $P = p + g(1-f)(s-f)$. The monopolist problem can therefore be written as

$$\max_{s \geq 0, f \geq 0} \; B(0)(s+f) + [g(1-f)(s-f)] f, \tag{21}$$
$$\text{s.t.} \quad s + f = 1.$$

In the next Proposition, we show how the monopolist allocates travelers when it fully covers the market. Here and in the rest of the paper, equilibrium variables referred to the centralized market have a subscript C, a mnemonic for Corporate.

Proposition 4. *Assume Assumption 3 holds and a centralized regime. Let $\bar{s}_C$ and $\bar{f}_C = 1 - \bar{s}_C$ denote the solutions to the monopolist problem (21) under full coverage. Then, the unique equilibrium is such that corporate travelers with $\theta \in [0, \bar{s}_C]$ are in the slow lane and those with $\theta \in [\bar{s}_C, 1]$ are in the fast lane, where*

$$\begin{aligned} \bar{s}_C &= \tfrac{1}{2} + \tfrac{\sqrt{3}}{6} \cong 0.7887, \\ \bar{f}_C &= \tfrac{1}{2} - \tfrac{\sqrt{3}}{6} \cong 0.2113. \end{aligned} \tag{22}$$

The Proposition characterizes the profit maximizing allocation of travelers. High-θ travelers, in a mass equal to $\bar{f}_C = 1 - \bar{s}_C$, are sorted into the fast lane and low-θ travelers, in a mass equal to $\bar{s}_C$, into the slow lane. This solution is illustrated in Figure 1 by the red solid circle. The figure shows that the outcome of a centralized market is overdifferentiation across lanes. Too few travelers travel in the fast lane as compared to the socially optimal level, $\bar{f}_C < \bar{f}_{FB}$, and too many in the slow lane, $\bar{s}_C > \bar{s}_{FB}$. This result, due to the IC constraint that allows to charge an increasingly high P the larger the congestion differential across lanes, is reminiscent of Mussa and Rosen (1978).

4.4 Welfare analysis

In this section, we investigate the welfare effects of the transition from a decentralized to a centralized regime when parameter values are such that all individuals travel in all regimes. At full coverage, where $s = 1 - f$, social welfare as in (6) may be rewritten as

$$\begin{aligned} \bar{W} = \int_0^1 B(\theta) d\theta - \mu \left[\int_0^{1-\frac{f^c}{\mu}} \theta g(s^c + s^a) d\theta + \int_{1-\frac{f^c}{\mu}}^1 \theta g(f^c + f^a) d\theta \right] + \\ - (1-\mu) \left[\int_0^{1-\frac{f^a}{1-\mu}} \theta g(s^a + s^c) d\theta + \int_{1-\frac{f^a}{1-\mu}}^1 \theta g(f^a + f^c) d\theta \right] \end{aligned} \tag{23}$$

where we use the "bar" notation as in the rest of this section to denote full coverage.[10] The first term illustrates the aggregate gross benefit from traveling, enjoyed by all travelers because of full coverage. The

[10] As in (6), fares do not appear in the welfare function because they are simply transfers from travelers to the monopolist. This same argument will apply also in the case of taxes, in which case they are transfers from travelers to the society.

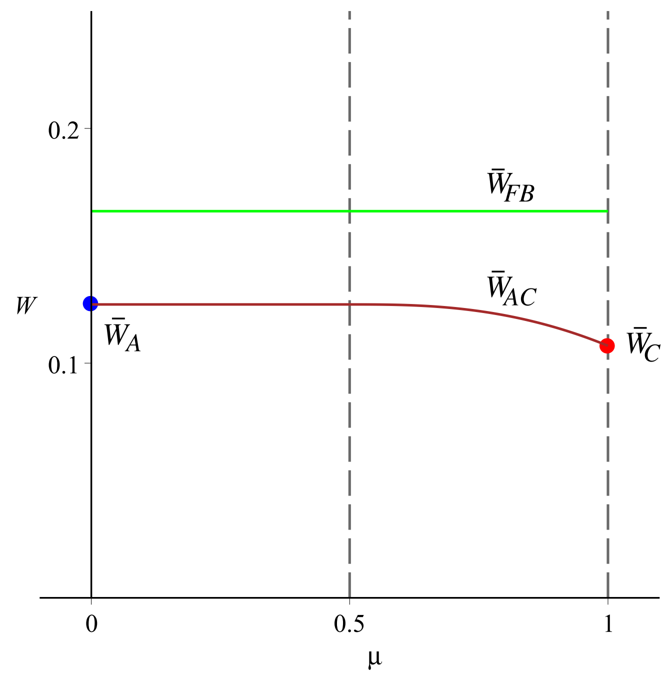

Figure 2: Equilibrium welfare.
Equilibrium welfare under full coverage in the first best ($\bar{W}_{FB}$, green line), with decentralization ($\bar{W}_A$, blue dot), in the mixed regime ($\bar{W}_{AC}$, brown line) and with centralization ($\bar{W}_C$, red dot); $g = \frac{1}{2}$ and $\int_0^1 B(\theta)d\theta = \frac{1}{4}$.

expressions in square brackets for each group of travelers (corporate and atomistic) are the sum of the congestion disutility suffered by those traveling in the slow (first term) and fast (second term) lane.

We evaluate (23) at the equilibrium for different values of μ. It may seem intuitive that, as the share of corporate travelers μ increases, welfare is positively affected by the increasing ability of the monopolist to internalize the congestion externality. We will show that this intuition is incomplete, as it ignores two welfare-reducing distortions that emerge under the centralized regime. Our results are summarized in the following Proposition.

Proposition 5. *Assume Assumption 3 holds. Then, in equilibrium, $\bar{W}$ is strictly below first best for all values of μ, does not vary with μ when $\mu \in \left[0, \frac{1}{2}\right]$, and is strictly decreasing in μ when $\mu \in \left(\frac{1}{2}, 1\right]$.*

The Proposition, also illustrated in Figure 2, shows that, for any share of corporate individuals μ, welfare is always lower than that achieved, in the first best, by the social planner (which does not depend on μ). It also illustrates that the increasing ability of the monopolist to internalize the congestion externality when the share of corporate individuals μ increases does not reflect into a welfare improvement. To the contrary, welfare is monotonically decreasing in μ (and strictly decreasing for μ above $\frac{1}{2}$).

Welfare is higher in the first-best because of two distortions that emerge when the share of corporate individuals μ is strictly positive. The first distortion involves the level of differentiation across lanes. Under the decentralized regime, the two lanes have exactly the same number of travelers, leading to suboptimal differentiation vis-à-vis the socially optimal level. Under the centralized regime, instead, there is excess differentiation in congestion across lanes with respect to social optimum. The excess differentiation in monopoly turns out to reduce welfare more than the suboptimal differentiation under the decentralized regime because it involves more travelers (i.e., $\bar{s}_C - \bar{s}_{FB} > \bar{s}_{FB} - \bar{s}_A$) and travelers with higher θ's, who are more bothered by congestion.

The second distortion relates to the identity of individuals traveling in each lane, rather than to their mass. When $\mu \in \left(\frac{1}{2}, 1\right)$, Proposition 3 states that all the atomistic individuals travel in the fast lane, including those with a relative low value of traveling and disutility from congestion. By contrast, only high-θ corporate individuals travel in the fast lane. As a result, travelers are misallocated in that some corporate travelers with a relatively high θ travel in the slow lane, while some atomistic travelers with lower θ travel in the fast lane. In particular, there exists a level of μ (i.e., $\mu = \frac{2\sqrt{7}+1}{9} \cong 0.6991$), such that the mass of travelers in the slow and in the fast lane exactly replicate those chosen by the social planner under full coverage. That notwithstanding, welfare is below that achieved in the first best, because of the misallocation of travelers, in terms of their θ type, described above. This distortion emphasizes that, given individuals' heterogeneity, welfare is maximized not only by generating the appropriate level of congestion in the two lanes, but also by ensuring that each traveler is correctly allocated.

We conclude this discussion with some remarks on the distributional effects in terms of individuals' utility net of fares across different regimes. As compared to the social optimum, a decentralized regime makes individuals with relatively low θ (i.e., $\theta \in [0, \bar{s}_{FB}]$) better off and individuals with relatively high θ (i.e., $\theta \in [\bar{s}_{FB}, 1]$) worse off. On the one hand, travelers with low θ's, who would travel in the slow lane in the first best, travel in a less congested lane in the decentralized regime. On the other hand, travelers with high θ's, who would travel in the fast lane in the first best, end up traveling in a more congested lane in a decentralized regime.

It is also possible to assess the distributional effects, net of fares, of an increase in the share of corporate individuals. When $\mu \leq \frac{1}{2}$, lanes have the same level of differentiation in equilibrium and, therefore, there is no distributional effects. When $\mu \geq \frac{1}{2}$, a higher μ increases the level of differentiation across lanes. This makes all atomistic travelers and corporate travelers traveling in the fast lane increasingly better off, on the one hand, and corporate travelers traveling in the slow lane increasingly worse off, on the other hand.

5 Partial coverage: equilibrium analysis

In Section 4, we focused on sorting. We showed the welfare effects of moving from a centralized to a decentralized regime when parameter values are such that the aggregate amount of travelers does not change across regimes, but their allocation across lanes does, with an effect on aggregate congestion costs, and, as a result, on welfare. In this section, we focus on parameter values under which in all regimes the equilibrium aggregate amount of travelers may vary. We thus abandon Assumption 3. As a result, we consider the perhaps more intuitive strategy to deal with congestion, namely restricting aggregate output.

5.1 Decentralized regime

Full coverage in the decentralized regime is implied by Assumptions 1 and 2 only and does not depend on Assumption 3 being met. Hence, the analysis contained in Section 4.1 applies here too. All atomistic individuals travel and they split equally across the two lanes, so that $s_A = f_A = \bar{s}_A = \bar{f}_A = \frac{1}{2}$, where s_A and f_A denote the equilibrium mass of atomistic travelers when Assumption 3 is not met.

5.2 Mixed regime

We now look at the mixed regime. We start our analysis by noting that conditions $1 - \mu \geq s^c - f^c$ and $1 - \mu < s^c - f^c$ determining the equilibrium allocation of travelers and giving rise to the results of no differentiation, (15), and differentiation, (16), do not depend on Assumption 3. Therefore, they apply also when Assumption 3 is relaxed.

We first focus on the case of a small share of corporate individuals. When μ is sufficiently small, then $1 - \mu \geq s^c - f^c$ holds in equilibrium. The equilibrium allocations of travelers must satisfy $s = f = \frac{s^c+f^c+s^a+f^a}{2}$ as in (15), implying that the mass of travelers in the two lanes is always identical. The only constraint faced by the monopolist is the IR constraint for the corporate traveler that is indifferent between not traveling and traveling in either lane, that is the θ-type with $\theta = 1 - \frac{s^c+f^c}{\mu}$. When binding, this constraint may be written as

$$p = B\left(1 - \frac{s^c + f^c}{\mu}\right) - \left(1 - \frac{s^c + f^c}{\mu}\right) g\,\frac{s^c + f^c + s^a + f^a}{2}, \tag{24}$$

Since the monopolist charges the same fare to all corporate travelers, its profit is affected by the total mass of corporate travelers only, that we denote by $c \equiv s^c + f^c$. This implies that $\pi = p(s^c + f^c) = p\,c$. Using (24), the monopolist problem can then be written as

$$\max_{c\geq 0} \left[B\left(1 - \frac{c}{\mu}\right) - \left(1 - \frac{c}{\mu}\right) g\,\frac{c + s^a + f^a}{2}\right] c, \quad \text{s.t.} \quad c \leq \mu. \tag{25}$$

We denote by c_{AC} the solution to this problem.

Next, we turn our attention to the case of a large share of corporate individuals. When μ is sufficiently large, then $1 - \mu < s^c - f^c$ holds in equilibrium. The equilibrium allocation of travelers must then satisfy $s = s^c$ and $f = f^c + (1 - \mu) < s$, as in (16), implying that there is differentiation across lanes and all atomistic individuals travel in the fast lane. In this case, using the same arguments as in Section 4.2, the constraints faced by the monopolist are[11]

$$p = B\left(1 - \frac{s^c + f^c}{\mu}\right) - \left(1 - \frac{s^c + f^c}{\mu}\right) g\left(s^c + s^a\right), \tag{26}$$

[11] The FOCs to this problem are given in the appendix.

$$P = p + g\left(1 - \frac{f^c}{\mu}\right)\left[s^c + s^a - \left(f^c + f^a\right)\right], \tag{27}$$

where $\theta = 1 - \frac{f^c}{\mu}$ is the type of the corporate traveler that is indifferent between traveling in the fast and the slow lane. Using (26) and (27), the monopolist's problem may then be written as

$$\begin{aligned} \max_{s^c \geq 0, f^c \geq 0} & \left[B\left(1 - \frac{s^c + f^c}{\mu}\right) - \left(1 - \frac{s^c + f^c}{\mu}\right) g\left(s^c + s^a\right)\right]\left(s^c + f^c\right) + \\ & + g\left(1 - \frac{f^c}{\mu}\right)\left[s^c + s^a - \left(f^c + f^a\right)\right] f^c, \\ \text{s.t.} \; & s^c + f^c \leq \mu. \end{aligned} \tag{28}$$

The next Proposition illustrates the subgame perfect Nash equilibrium of our two stage game for the different values of μ. The Proposition illustrates that, as in the case of full coverage, whether or not the equilibrium of the game implies differentiation across lanes depends on the share of corporate individuals. The critical value of this share is given by $\mu' \in \left(\frac{1}{2}, 1\right)$ and it is discussed in the appendix.

Proposition 6. *Assume a mixed regime. Let s^a_{AC} and f^a_{AC} denote the equilibrium mass of atomistic travelers under partial coverage. Let s^c_{AC} and f^c_{AC} denote the equilibrium mass of corporate travelers under partial coverage. Then,*

- *partial coverage occurs when*

$$g < g_{AC} \equiv K\left(\mu\right)\left[B'\left(0\right) - B\left(0\right)\right], \tag{29}$$

where

$$K\left(\mu\right) \equiv \begin{cases} 2 & \text{when } \mu \in \left(0, \mu'\right), \\ \frac{18\mu}{8\mu^2 + 5\mu - 1 + (4\mu - 1)\sqrt{4\mu^2 - 2\mu + 1}} & \text{when } \mu \in \left[\mu', 1\right); \end{cases} \tag{30}$$

- *when $\mu \in (0, \mu']$ and $g < g_{AC}$, c_{AC} exists and it is unique. Then, there are multiple equilibrium values of s^c_{AC}, s^a_{AC}, f^c_{AC} and f^a_{AC}. An equilibrium is any allocation of travelers such that*

$$\begin{aligned} s^c_{AC} + s^a_{AC} &= f^c_{AC} + f^a_{AC}, \\ s^c_{AC} + f^c_{AC} &= c_{AC} < \mu, \\ s^a_{AC} + f^a_{AC} &= 1 - \mu; \end{aligned} \tag{31}$$

- *when $\mu \in (\mu', 1)$ and $g < g_{AC}$, th equilibrium values of s^c_{AC}, s^a_{AC}, f^c_{AC} and f^a_{AC} exist and are unique. In equilibrium, corporate individuals with $\theta \in \left[0, 1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right]$ do not travel, corporate travelers with $\theta \in \left[1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}, 1 - \frac{f^c_{AC}}{\mu}\right]$ travel in the slow lane, all atomistic travelers and corporate travelers* *with* $\theta \in \left[1 - \frac{f^c_{AC}}{\mu}, 1\right]$ *travel in the fast lane, so that*

$$\begin{aligned} s^a_{AC} &= 0, \\ f^a_{AC} &= 1-\mu, \\ s^c_{AC} + f^c_{AC} &< \mu. \end{aligned} \tag{32}$$

The Proposition first illustrates the condition under which partial coverage occurs. Condition $g < g_{AC}$ in (29) shows a stark contrast with the first best partial coverage condition $g > g_{FB}$. Differently than a social planner, the monopolist rations corporate individuals when the congestion cost g is sufficiently low. To get the intuition for this perhaps counterintuitive result, consider that, as g gets larger, the travelers' utility function $U(\theta)$ gets flatter in θ, because the higher willingness to pay of higher θ types is increasingly compensated by their congestion disutility. This affects the traditional price/quantity tradeoff by providing the monopolist with a greater incentive to increase market coverage.[12]

The Proposition also characterizes the equilibrium allocation of travelers, which depends on the relative shares of atomistic and corporate travelers. These allocations mirror closely those in the case of full coverage. When the share of corporate travelers is relatively small, i.e., $\mu \in (0, \mu']$, there are multiple equilibria. Corporate travelers are indifferent between the two lanes, because they are charged the same price. Any allocation of the two types of travelers that results in identical lanes is an equilibrium. When instead the share of corporate travelers is relatively large, i.e., $\mu \in (\mu', 1)$, the monopolist differentiates the masses of corporate travelers allocated in the two lanes enough, that atomistic travelers cannot equalize them. Atomistic travelers then all travel in the less congested lane. As in the case of full coverage, an explicit characterization of s^c_{AC} and f^c_{AC} would require to specify the gross beneift function $B(\cdot)$. While keeping the general form of $B(\cdot)$, by simply manipulating the FOCs of the monopolist' problem, we can derive the optimal choice of f^c_{FB} as a function of c^c_{FB}. This expression, which is the analogous of (13), is illustrated by the brown line in Figure 1, and reads as follows

$$f^c_{AC}(s^c_{AC}) = \frac{1}{3}\left(2\mu - 1 + 2s^c_{AC} - \sqrt{7\left(s^c_{AC}\right)^2 - 2s^c_{AC}(2-\mu) + 1 - \mu + \mu^2}\right). \tag{33}$$

The Proposition also illustrates that the critical value of the share of corporate travelers for the emergence of an equilibrium with or without differentiation across lanes is given by μ'. Differently from the full coverage case, this threshold level of μ is strictly above $\frac{1}{2}$ under partial coverage. In fact, it takes a smaller mass of atomistic travelers to equalize the total mass of travelers in each lane, $s = f$, when the monopolist prefers to ration a portion of its customers.

[12] The monopolist's full coverage condition (29) depends also on $B'(0)$, with the same logic as above; when $B'(0)$ is small, $U(\theta)$ gets flatter in θ, which makes quantity more sensitive to price.

5.3 Centralized regime

We now consider the centralized regime, $\mu = 1$. The monopolist's problem is as in (28) when setting $\mu = 1$ and is omitted.[13] Here, we illustrate the subgame perfect Nash equilibrium of our two-stage game in the following Proposition.

Proposition 7. *Assume a centralized regime. Let s_C and f_C denote the solutions to the monopolist profit maximization problem under partial coverage. Then,*

- *partial coverage occurs when*

$$g < g_C \equiv \frac{6}{4+\sqrt{3}}\left[B'(0) - B(0)\right]; \tag{34}$$

- *when $g < g_C$, the solutions s_C and f_C exist and are unique. In equilibrium, corporate individuals with $\theta \in [0, 1 - s_C - f_C]$ do not travel, corporate travelers with $\theta \in [1 - s_C - f_C, 1 - f_C]$ are in the slow lane and corporate travelers with $\theta \in [1 - f_C, 1]$ are in the fast lane.*

To compare this equilibrium with those previously derived, it is useful to express f_C as a function of s_C. As before, this is obtained by simply manipulating the FOCs of the monopolist's problem and results in the following expression, which is also illustrated by the red curve in Figure 1,

$$f_C(s_C) = \frac{1}{3}\left(1 + 2s_C - \sqrt{7s_C^2 - 2s_C + 1}\right). \tag{35}$$

An interesting feature of this equilibrium is that full coverage may occur under the centralized regime but not at the social optimum. This implies that the monopolist may dispatch more travelers than the social planner. Intuitively, this may happen when g is relatively large. This result is clearly at odds with the standard outcome that a monopolist reduces total output. However, our result echoes the possibility that a monopolist underprovides quality (increases the level of congestion, in our model) relatively to the social optimum (Spence, 1975).

5.4 Welfare analysis

In Section 4.4, we considered the welfare effects for parameter values for which the market is fully covered under all regimes. In that setting, welfare changed only as a result of the different sorting of travelers across lanes. We observed that two distortions may arise as we move from a decentralized to a centralized regime, one related to the masses of travelers and the other one to the identity of travelers. Overall, we showed that welfare monotonically decreases as the share of corporate travelers increases.

That analysis ruled out another, fundamental, potential source of distortion, namely the total mass of travelers, which may depart from social optimality. While in a decentralized regime the market is always covered, we indeed proved that both a social planner and a monopoly may want to ration travel, albeit in

[13] The FOCs of this problem are given in the appendix.

a different manner. In this section, we show that the interplay between rationing and the two distortions already observed when sorting is the only issue produces a welfare comparison that is not clear cut.

To investigate the welfare effects of the transition from a decentralized to a centralized regime, we compare the welfare under the decentralized regime ($\mu = 0$), given by $W_A = \int_0^1 \left(B(\theta) - \theta g \frac{1}{2}\right) d\theta$, with the welfare under a mixed regime and a centralized regime ($\mu > 0$). However, to avoid duplicating the analysis in Section 4.4, we restrict to cases in which partial coverage occurs when $\mu > 0$. On top of that, we first focus on mixed regimes with $\mu \in (0, \mu']$, so that lanes are not differentiated in equilibrium, and then move to regimes with $\mu \in (\mu', 1]$ so that differentiation occurs.[14] Finally, we derive our welfare results by making use of a specific (linear) functional form for the gross benefit function, $B(\theta) = b_0 + b\theta$.[15]

No differentiation case $\mu \in (0, \mu']$. We denote as W_{AC} the welfare when $\mu \in (0, \mu']$, so that lanes are not differentiated in equilibrium, and let $g < g_{AC}$, so that partial coverage occurs according to Proposition 6. In symbols,

$$W_{AC} = \mu \int_{1-\frac{c_{AC}}{\mu}}^{1} \left(B(\theta) - \theta g \frac{c_{AC} + 1 - \mu}{2}\right) d\theta + (1-\mu) \int_0^1 \left(B(\theta) - \theta g \frac{c_{AC} + 1 - \mu}{2}\right) d\theta. \tag{36}$$

The first (second) term of this expression is the aggregate net utility of corporate (atomistic) travelers. The level of congestion is identical across all travelers, since both lanes feature the same mass of travelers, $\frac{c_{AC}+1-\mu}{2}$. However, while the entire mass $(1-\mu)$ of atomistic individuals travels, only a fraction $c_{AC} \in (0, \mu)$ of corporate individuals does so. We denote as ΔW the difference between W_{AC} and the welfare under a decentralized regime W_A,

$$\begin{aligned} \Delta W = {} & (1-\mu) \left[\int_0^1 \left(\theta g \frac{1}{2} - \theta g \frac{c_{AC} + 1 - \mu}{2} \right) d\theta \right] + \\ & + \mu \left[\int_{1-\frac{c_{AC}}{\mu}}^{1} \left(\theta g \frac{1}{2} - \theta g \frac{c_{AC} + 1 - \mu}{2} \right) d\theta - \int_0^{1-\frac{c_{AC}}{\mu}} \left(B(\theta) - \theta g \frac{1}{2} \right) d\theta \right]. \end{aligned} \tag{37}$$

This expression clearly illustrates the welfare pros and cons of a mixed regime, when the monopoly excludes some individuals. The first line illustrates the positive effect on the aggregate net utility of atomistic travelers. They all travel, as under the decentralized regime, but they now face a lower congestion level because of the rationing imposed by the monopolist on corporate travelers. The second line expresses the effect on the aggregate net utility of corporate travelers. The first term is the gain for those with $\theta \in \left[1 - \frac{c_{AC}}{\mu}, 1\right]$ who are still traveling, but now face lower congestion. The second term illustrates the loss for those with $\theta \in \left[0, 1 - \frac{c_{AC}}{\mu}\right]$ who are no longer traveling.

We state the following result.

[14] When $\mu \in (\mu', 1]$, no analytical results can be obtained. We therefore run numerical simulations.

[15] When $B(\theta) = b_0 + b\theta$, Assumption 1 becomes $g < b$, Assumption 2 becomes $b_0 \geq 0$, and relaxing Assumption 3 yields $b_0 < b$. Also, it is easy to check the second-order conditions are fully satisfied in all the problems we analyze.

Proposition 8. *Assume $g < g_{AC}$. Let*

$$\underline{c} \equiv \frac{2b + g(2\mu - 1) - \sqrt{(2b-g)^2 - 4g(4b_0 - g)\mu}}{2g}. \tag{38}$$

For any $\mu \in (0, \mu']$, welfare under a mixed regime is larger than welfare under a decentralized regime (i.e., $\Delta W > 0$) if and only if both following conditions hold: i) $b_0 < \frac{g}{4}$; ii) $c_{AC} \in (\underline{c}, \mu)$. Conditions i) and ii) are satisfied at least for $\mu \to 0$ and $b_0 < \frac{g}{2} - \frac{1}{3}b$.

The Proposition illustrates that mixed regimes with $\mu \in (0, \mu']$ enhance welfare provided that two rather intuitive conditions apply. First, condition $b_0 < \frac{g}{4}$ implies that it is socially optimal to ration travel. Indeed, the benefit to the type-0 traveler, b_0, must be smaller than the increase in the aggregate congestion costs this individual would impose on fellow travelers in the same lane if she traveled, $\frac{g}{4}$. Second, condition $c_{AC} \in (\underline{c}, \mu)$ shows that not too many corporate individuals have to be excluded. Recall indeed that corporate individuals may, on aggregate, lose from a mixed regime, since, as a result, some of them are excluded. Condition $c_{AC} \in (\underline{c}, \mu)$ requires this rationing to be limited.

Differentiation case $\mu \in (\mu', 1]$. We extend our analysis to regimes with a relatively high share of corporate individuals $\mu \in (\mu', 1]$ so that lanes are differentiated in equilibrium. As mentioned, we resort to numerical methods, due to the difficulty in characterizing the analytical solutions to the monopolist problem. A graphical illustration of our results is given in Figure 3. In each panel of this Figure, we illustrate, for a given value of μ, the sign of the difference between the welfare under a mixed or a centralized regime and the welfare under a decentralized regime for a grid of admissible values of the parameters b_0 and g, having normalized b to 1.

Figure 3 shows that the results in Proposition 8 extend nicely also to the case of relatively high values of μ. We include in the Figure the results of numerical simulations for relatively low values of μ, for which we have full analytical results, to emphasize that our analytical results carry through to larger values of μ. A further interesting feature that can be observed from comparing the results in the different panels of Figure 3 is that the set of parameters under which the mixed regime has a positive welfare effect shrinks as μ increase. An intuition behind this result is that, as the mass of corporate individuals rises, there is more room for the negative excess rationing effect clearly identified in Proposition 8 in the case of a small μ. On top of this, as μ increases, the two distortions due to sorting gain prominence, as proved in Section 4.4.

6 Taxes: equilibrium analysis

Up to now, we have excluded government intervention. This laissez-faire approach is often implemented in practice. While economists advocate road pricing and congestion taxes as tools to improve upon market outcomes, these are rarely implemented in practice (notable exceptions include London, Manhattan, Singapore and Stockholm). This is likely to occur for political economy reasons (Oberholzer-Gee and

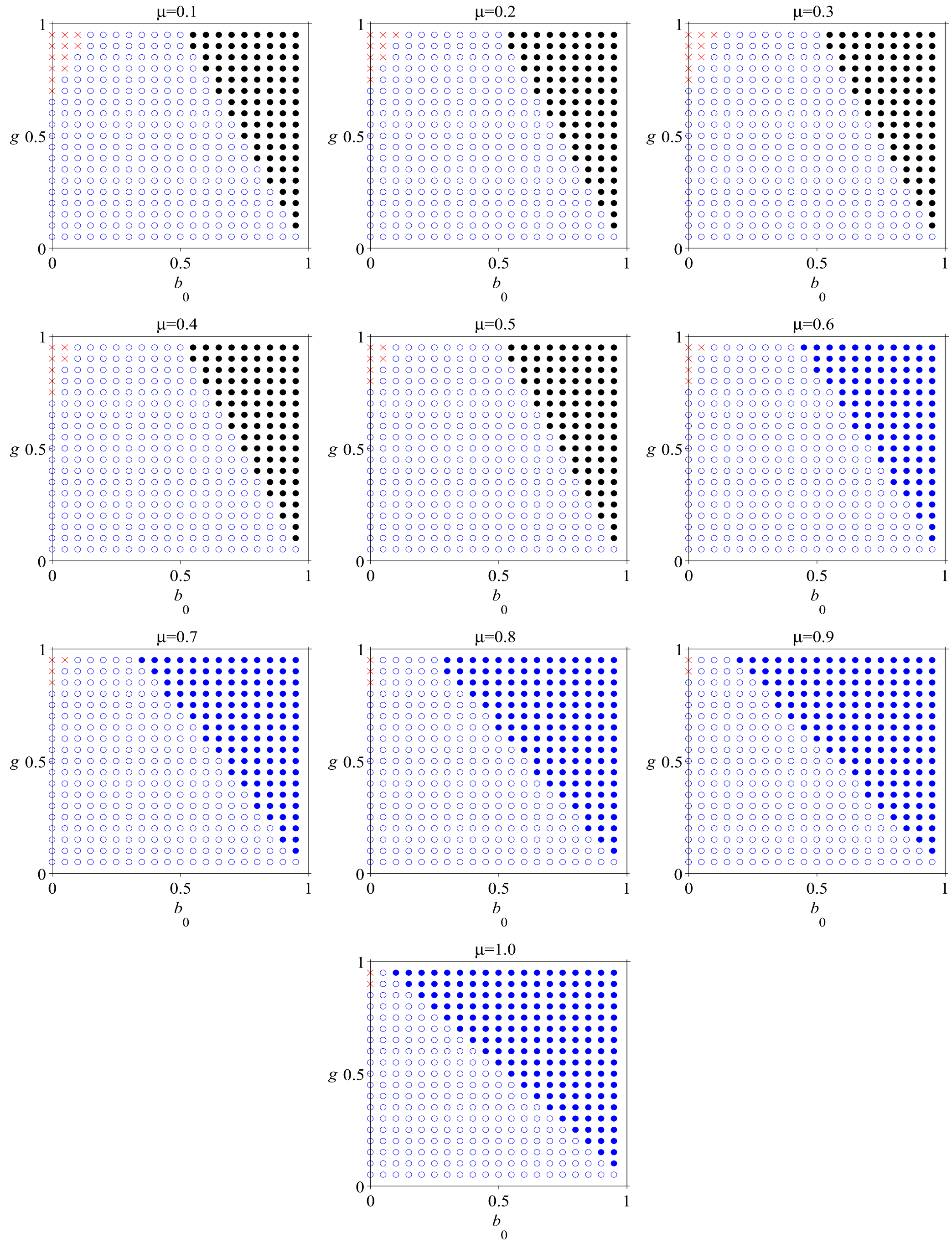

Figure 3: Welfare comparisons across different market equilibria when $b = 1$.
From left to right and top to bottom, μ varies from 0.1 to 1 by 0.1 increases. Blue dots illustrate combinations of parameters such that a decentralized regime delivers higher welfare (i.e., $W_{AC} < W_A$, or $W_C < W_A$ when $\mu = 1$), red crosses illustrate combinations of parameters such that a centralized regime delivers higher welfare (i.e., $W_A < W_{AC}$, or $W_A < W_C$ when $\mu = 1$); black dots illustrate combinations of parameters such that welfare is identical in the two regimes (i.e., $W_A = W_{AC}$). Full (empty) dots denote full (partial) coverage under centralization.

Weck-Hannemann, 2002), motivated by the fact that they may penalize low-income travelers that are inflexible in their arrival time (Hall, 2018, 2020). This section characterizes the tax/subsidy schemes that restore optimality under the two polar cases of a decentralized and of a centralized regime, $\mu = 0$ and $\mu = 1$. We show that the structure of the optimal scheme under the centralized regime differs sharply from that under the decentralized regime, and we illustrate that these features of the tax system carry over to the mixed regime. We then discuss the political feasibility and distributional effects of the first-best restoring tax/subsidy schemes.

To this end, we move to a three-stage game in which the tax authority sets the taxes in the first stage, while the two subsequent stages are identical to those analyzed in the previous sections. We restrict to per-traveler unit taxes, differentiated by lane but not by the travelers' identity. We denote by t and T the unit tax levied on individuals traveling in the slow and fast lane.

6.1 Decentralized regime

In an economy with atomistic individuals and externalities, it is well known that optimality is restored by Pigouvian taxes, which impose on each individual the non internalized social cost. In our framework, optimal road taxes have two goals: *i*) they need to induce to travel those and only those individuals whose private benefit from a trip is larger than the social cost they impose on fellow travelers; *ii*) they need to restore the appropriate allocation of travelers across lanes, to minimize the aggregate congestion cost.

Let t_A and T_A denote the unit taxes levied on atomistic individuals traveling in the slow and fast lane that restore social optimum. These taxes are described in the following:

Proposition 9. *Assume a decentralized regime. The pairs of taxes that replicate the social optimum are*

- *when* $g \leq g_{FB} \cong 5.4179 \times B(0)$ *as in (11),*

$$\begin{gathered} t_A \leq B\left(0\right), \\ T_A = t_A + g\tfrac{5-\sqrt{7}}{18}; \end{gathered} \tag{39}$$

- *when* $g > g_{FB} \cong 5.4179 \times B(0)$ *as in (11),*

$$\begin{gathered} t_A = g\, s_{FB}\left(1 - f_{FB} - \tfrac{s_{FB}}{2}\right), \\ T_A = t_A + g\left(1 - f_{FB}\right)\left(s_{FB} - f_{FB}\right). \end{gathered} \tag{40}$$

Taxes modify the after-tax net utility that atomistic individuals enjoy from traveling, thereby affecting their choices as to whether to travel and in which lane. In particular, when $g \leq g_{FB}$, misallocation of travelers across lanes is the only distortion to be solved, since full coverage would occur not only with atomistic travelers, as always, but also in the first best. Hence, the optimal pair of taxes - as in (39) - should not restrict market coverage. A multiplicity of low enough t, including $t_A = 0$, delivers this. On the other hand, optimal differentiation across lanes is obtained through an appropriate difference $T_A - t_A$, which ensures that the location of the traveler indifferent across the two lanes is identical to that of the social planner. Notice that the indeterminacy of t_A provides a flexible set of alternatives to the tax authority, ranging from solutions that minimize tax burden, when only those who travel in the fast lanes are taxed, to others associated to a larger tax burden, borne by travelers in both lanes.

When instead g is larger than g_{FB}, a social planner would exclude atomistic individuals with a low value for the travel, while these travelers would travel without taxation. Hence, both the total number of travelers and their allocation across lanes have to be corrected. The tax in the slow lane, t_A in (40), is then uniquely determined and ensures that the market coverage replicates the first best. On the other hand, the tax in the fast lane, T_A, induces an optimal degree of differentiation across lanes.

Taxes in (39) and (40) are congestion charges, reflecting the external cost imposed by the marginal travelers on fellow travelers. They align the incentives of the marginal travelers to social optimality. The distributional consequences, hence the political feasibility, of such congestion charges depend on how the revenue from them is used (Small, 1992). Since congestion charges are welfare-improving, an appropriate redistribution scheme that fully compensates losers could be Pareto-improving. However, probably as a result of imperfect or unclear compensations, congestion charges are typically unpopular, and therefore rarely implemented (Oberholzer-Gee and Weck-Hannemann, 2002). Without compensation, low θ atomistic individuals stand to lose from the tax scheme, net of the tax payment; if, on the one hand, the market remains fully covered after the tax scheme is implemented, all θ's traveling in the slow lane face more congestion, thus enjoying a lower level of utility; if, on the other hand, the tax scheme excludes the lower portion of θ's, those excluded are worse-off.

6.2 Centralized regime

We now look at the welfare maximizing tax/subsidy scheme to be imposed to the monopolist managing the entire AVs fleet.[16] We show that it is remarkably different than the tax scheme to be applied to atomistic travelers. Indeed, a monopolist already internalizes the congestion externality, so congestion charges are not appropriate. The tax/subsidy scheme has instead to correct the distortions in terms of sorting and rationing typical of the centralize regime, illustrated in the previous sections.

We consider a per-traveler tax/subsidy, potentially differing by lane, imposed on the monopolist. We allow taxes to depend on the mass of travelers in the two lanes, so that $t = t(s, f)$ and $T = T(s, f)$ denote the per-traveler tax/subsidy in the slow and fast lane. In the appendix, we solve the maximization problem faced by the monopolist in the presence of taxes and then establish the following result.

Proposition 10. *Assume a centralized regime. The social optimum is restored by a system of per-traveler taxes/subsidies of the following form*

$$\begin{aligned} t_C &= g\,s - z_C, \\ T_C &= g\,f - z_C, \end{aligned} \tag{41}$$

[16] Our analysis would be identical if taxes, instead, were imposed directly on travelers.

where

$$z_C \equiv \begin{cases} 0 & \textit{if } g \leq \frac{18\left(B(0)-B'(0)\right)}{4+\sqrt{7}} \\ & \quad \cong 2.7085 \times \left(B(0) - B'(0)\right); \\ B'\left(0\right) - B\left(0\right) + g\,\frac{4+\sqrt{7}}{18} & \textit{if } \frac{18\left(B(0)-B'(0)\right)}{4+\sqrt{7}} \leq g \leq \frac{36\,B(0)}{4+\sqrt{7}}; \\ B'(1 - s_{FB} - f_{FB})(s_{FB} + f_{FB}) + & \\ \quad +g\left(2f_{FB} - (s_{FB}^2 - \frac{3}{2}f_{FB}^2\right) & \textit{if } g \geq \frac{36\,B(0)}{4+\sqrt{7}} \quad \cong 5.4179 \times B(0). \end{cases} \tag{42}$$

The Proposition illustrates that social optimality is restored by imposing on the monopolist a per-traveler tax/subsidy scheme based on the mass of travelers, differentiated by lane. Hence, t_C and T_C consist of a tax component, gs and gf, positively affected by the mass of travelers in that lane, and a subsidy component, z_C, exogenously determined by the tax authority, based on the socially optimal number of travelers and equal across the two lanes.

The tax components of t_C and T_C differ dramatically vis-à-vis those in case of a decentralized regime, as they are not congestion charges. Instead, they induce the monopolist to mitigate the excess differentiation typical of a centralized regime, by shifting passengers from the slow to the fast lane, as long as $s > f$. The logic of this tax is similar, for instance, to that of the tax on quality in Lambertini and Mosca (1999), and Cremer and Thisse (1994) in the context of a vertically differentiated oligopoly, as long as in our model congestion is interpreted as a quality level.

An additional instrument is, however, needed to eliminate the distortion caused by the monopolist when it excludes travelers in a socially inefficient manner. Without taxes, as discussed in Section 5.3, the total number of travelers may be below or above the social optimum. After introducing the tax components of t_C and T_C, instead, the monopolist always reduces total travel below its efficient level. To ensure a socially efficient number of travelers, the monopolist must be granted a subsidy on the total mass of travelers, as in (42).

The two components of the taxes serve very different purposes, and also have remarkably different features. The levels of the tax components only depend on the monopolist's choices, being contingent on s and f. In this respect, it is a very simple tax to set, since it does not require any specific knowledge by the tax authority, except for the value of g. On the other hand, the subsidy component requires a deeper knowledge of the market, being based on a perfect knowledge of the travelers' benefit function and of the solution to the first best problem.

Overall, in equilibrium, the subsidy component may exceed the tax component, so that the monopolist receives a net subsidy from the tax authority. This situation always occurs when g is sufficiently high so that efficiency commands to partially cover the market. The tax/subsidy scheme therefore requires to absorb some funding from general taxation. Pels and Verhoef (2004) emphasize the political difficulties in implementing a negative tax scheme on airline companies. The same logic might well apply to the AVs market. The political feasibility of this tax/subsidy scheme appears very dubious, even more so than the congestion charges. A potential solution to improve political feasibility is to associate the scheme to an upfront fixed license that preserves the budget neutrality.

In the Appendix, we show that our first-best restoring taxes for the two cases of a centralized and decentralized regime carry over to the case of a mixed regime. In a mixed regime, atomistic and corporate travelers should pay, at the optimum, two different taxes. Atomistic travelers should pay a congestion charge, while corporate travelers should pay a similar tax to that charged under the centralized regime. This result has relevant policy implications, as it casts doubts on the policy, discussed in several states, and adopted in New York State, to charge a congestion charge to taxi and ridesharing companies, only. Uber and Lyft have lobbied in favor of it.[17] However, a uniform (within a city) congestion charge will not solve the misalignment of ridesharing companies' incentives vis-à-vis social welfare, in particular once their increased market share will allow them to offer menus of prices depending on the speed of the service. Quite paradoxically, such a policy misses the target. It charges a congestion charge only to companies, which have an incentive to internalize the congestion externality already (but, as we have shown, face several other distortions), while not charging indipendent travelers, who have no economic incentives to internalize the congestion externality.

7 Conclusions

The transition to AVs will open a variety of important issues in several domains, including technology, law and ethics.

The technological trajectory of AVs is linked to the progress of artificial intelligence. The ability of artificial intelligence to learn quickly, and, in particular, to adapt switfly to new circumstances will determine how fast the level of automation will progress. The standards set by the Society of Automotive Engineers International identify five different levels of automation, ranging from level 0 of no automation, to level 5 of a fully automated vehicle, able to move autonomously in all terrains and under all circumstances in which an experienced human driver would drive. The most advanced currently manufactured vehicles, produced by Waymo, stand at level 4, defined as highly automatized vehicles that can run without a driver in selected (usually urban) areas. There is some debate on the time horizon for the emergence and commercialization of level 5 vehicles.[18] However, level 4 vehicles are enough for urban traffic to be organized around robotaxis and fleets, with all the resulting welfare effects illustrated in our analysis.

From the legal standpoint, the introduction of AVs, both at level 4 and at level 5, will likely shift the responsibility for accidents from drivers to manufacturers, which will require the design of a new liability regime in the urban transport context (Abraham and Rabin, 2019).

From the moral perspective, Awad et al. (2018) discuss the tradeoffs involved in distributing the wellbeing created by machines, as well as the harm they cannot eliminate. An important moral question has to do, for instance, with how to solve the problem of the division of road risk between the different parties, including the occupants of the cars and the other stakeholders, for example pedestrians.

The transition to AVs, however, will also raise many important economic questions. Key for our analysis, a crucial effect will come from the process of traffic centralization, resulting from the organization

[17] https://www.ft.com/content/bb89ecd0-558a-11e9-91f9-b6515a54c5b1

[18] https://www.economist.com/leaders/2019/10/10/driverless-cars-are-stuck-in-a-jam (last accessed August 3, 2020).

of urban traffic around fleets. This is currently underway already, with fleets of ride-sharing companies, for example, accounting for 20% of total vehicle trips in New York's Central Business District in 2018. AVs will further reduce the cost of running a fleet, by eliminating drivers' costs, and will therefore provide an extra boost to this process.

Our paper investigated the welfare effects of moving from a decentralized regime, with atomistic vehicles only, to a centralized regime, in which all travelers use vehicles that are part of a fleet managed by a monopolist, through a transition period where some atomistic vehicles share the road with others that are part of a fleet. We analyzed an environment with heterogenous travelers who are sorted in one of two lanes, with potentially different levels of congestion. In this setting, a reduction in aggregate congestion costs potentially arises not only as a result of a reduction in overall travel, but also as a result of an optimal sorting of travelers across different lanes with different speeds. While ridesharing companies have not yet reached a sufficient market share to be able to manipulate congestion by sorting travelers across routes, this will likely come in the very near future. Recent developments in urban transport are decreasing the cost of the sorting technology, thereby making it a viable and important alternative to manage congestion. We proved that, while a centralized regime internalizes the congestion externality, it introduces additional distortions. We examined sorting and rationing as tools to manage congestion. We found that, when sorting is the most relevant, welfare decreases when moving to a centralized a regime. When, instead, rationing is also required, a centralized regime may be welfare-superior.

The self driving technology will likely affect both travel demand, and the welfare-maximizing level of congestion. With AVs, consumers may spend more productively their time on vehicles. This will arguably increase the welfare-maximizing level of congestion. At the same time, however, travel demand is expected to increase as well as a result (Gucwa, 2014). Whether the combination of the two effects will induce more or less travel rationing under the social planner is debatable. Additional factors, including the improved vehicles coordination, as well as changes in the cost of infrastructural expansions, could contribute to changing both travel demand and the welfare-maximizing level, with an unclear effect on the required rationing. As a result, it may well possible that sorting will become, in several urban contexts, more important than reducing the aggregate amount of cars circulating, to curb congestion cost. Our model showed that, in such environments, the centralization process brought about by the emergence of fleets of AVs is likely to have adverse welfare effects.

For sorting to be feasible and, as a result, welfare-reducing, ridesharing companies must be sufficiently large. Their current share of vehicle trips, albeit remarkable, is below that threshold, so we currently do not observe any forms of manipulation of congestion that allows them to offer differentiated speeds of service, and, as result, pricing menus. However, we are likely not far from that level, which requires to take appropriate and immediate policy actions.

Notice, finally, that a monopolistic company that manages the infrastructure (rather than a fleet of AVs) and can charge different prices across lanes behaves exactly like our monopolist in a centralized setting. Our results suggest that, when congestion is dealt with by sorting, and not by rationing, a monopolist that manages the infrastructure reduces welfare with respect to the decentralized case, with atomistic drivers only. This should raise a word of caution on infrastructural projects for AVs managed by private unregulated entities.

We then explored how to restore first best with road taxes. The familiar congestion charges are optimal under a decentralized regime. However, they fail to restore optimal welfare when vehicles are part of a fleet. In this case, the welfare-maximizing tax/subsidy scheme is very different, and may require subsidizing the company – something likely to be politically very unappealing. Our analysis shows that certain policies implemented in major urban contexts to tackle congestion are ill-conceived. For example, New York City applies a congestion charge only to ridesharing companies and to taxi, and not to private vehicles. Our model shows that congestion charge should be applied, quite to the contrary, to private vehicles, which do not internalize the congestion externality. Ridesharing companies, instead, internalize the congestion externality, so the tax levied on them should be designed differently, to tackle a different set of distortions.

A couple of extensions of our model seem natural. One would analyze welfare when multiple companies, each managing a fleet of AVs, compete for passengers. While we believe the qualitative results would not be altered, the analysis might provide some additional insights. A second possible extension would involve the analysis of the interaction between a welfare-oriented public transit company and a profit-maximizing service provider. This could be illustrative of the effects of government direct involvement in the urban transit business in a world of fleets.

References

Abraham, K. and R. Rabin, 2019. Automated Vehicles and Manufacturer Responsibility for Accidents: a New Legal Regime for a New Era. Virginia Law Review, 105(1), 127-171.

Arnott, R., A. de Palma and R. Lindsey, 1990. Economics of a bottleneck. Journal of Urban Economics, 27, 111-130.

Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon and I. Rahwan, 2018. The Moral Machine experiment. Nature, 563, 59-64.

Basso, L., 2008. Airport deregulation: Effects on prices and capacity. International Journal of Industrial Organization, 26, 1015-1031.

Basso, L. and A. Zhang, 2007. Congestible facility rivalry in vertical structures. Journal of Urban Economics, 61, 218-237.

Börjesson, M., M. Fosgerau and S. Algers, 2012. On the income elasticity of the value of travel time. Transportation Research Part A: Policy and Practice, 46, 368-377.

Brueckner, J. K., 2002. Airport congestion when carriers have market power. American Economic Review, 92(5), 1357-1375.

Brueckner, J. K., 2005. Internalization of airport congestion: A network analysis. International Journal of Industrial Organization, 23(7-8), 599-614.

Cremer, H. and J.-F. Thisse, 1994. Commodity taxation in a differentiated oligopoly. International Economic Review, 35(3), 613-633.

Czerny, A. and A. Zhang, 2015. Third-degree price discrimination in the presence of congestion externality. Canadian Journal of Economics, 48(4), 1430-1455.

Daniel, J. I., 1995. Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues. Econometrica, 63(2), 327-370.

Daniel, J. I. and K. T. Harback, 2008. (When) Do hub airlines internalize their self-imposed congestion delays? Journal of Urban Economics, 63(2), 583-612.

Downs, A., 1962. The law of peak-hour express-way congestion. Traffic Quarterly, 16(3), 393-409.

Duranton, G. and M. Turner, 2011. The Fundamental Law of Road Congestion: Evidence from US Cities. American Economic Review, 101(6), 2616-2652.

Erhardt, G. D., S. Roy, D. Cooper, B. Sana, M. Chen and J. Castiglione, 2019. Do transportation network companies decrease or increase congestion?. Science Advances, 5(5), eaau2670.

Fagnant, D. and K. Kockelman, 2015. Preparing a nation for autonomous vehicles. Transportation Research Part A: Policy and Practice, 77(4), 167-181.

Fréchette, G. R., A. Lizzeri and T. Salz, 2019. Frictions in a Competitive, Regulated Market: Evidence from Taxis, American Economic Review, 109(8), 2954-2992.

Gucwa, M., 2014. Mobility and Energy Impacts of Automated Cars. Presentation to the Automated Vehicles Symposium. Burlingame, CA, July 16.

Guda, H. and U. Subramanian, 2019. Your Uber Is Arriving: Managing On-Demand Workers Through Surge Pricing, Forecast Communication, and Worker Incentives. Management Science, 65(5), 1995-2014.

Hall, J. D., 2018. Pareto Improvements from Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways. Journal of Public Economics, 158, 113-125.

Hall, J. D. 2020. Can Tolling Help Everyone? Estimating the Aggregate and Distributional Consequences of Congestion Pricing. Journal of the European Economic Association, jvz082, https://doi.org/10.1093/jeea/jvz082.

Lamotte, R., A. de Palma and N. Geroliminis, 2017. On the use of reservation-based autonomous vehicles for demand management. Transportation Research Part B: Methodological, 99, 205-227.

Lindsey, R., A. de Palma and H. Silva, 2019. Equilibrium in the Bottleneck Model with Atomic and non Atomic Users. Transportation Research Part B: Methodological, 124, 82-107.

Mayer, C. and T. Sinai, 2003. Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil. American Economic Review, 93(4), 1194-1215.

Molnar, A., 2013. Congesting the commons: A test for strategic congestion externalities in the airline industry, mimeo.

Montes, R., W. Sand-Zantman and T. Valletti, 2018. The Value of Personal Information in Online Markets with Endogenous Privacy. Management Science, 65(3), 1342-1362.

Mussa, M. and S. Rosen, 1978. Monopoly and Product Quality. Journal of Economic Theory, 18, 301-317.

NYC Department of Transportation, 2019. New York City Mobility Report. New York: NYC Department of Transportation.

NYC Taxi and Limousine Commission and Department of Transportation, 2019. Improving Efficiency and Managing Growth in New York's For-Hire Vehicle Sector: Final Report. New York: NYC Taxi and Limousine Commission.

Lambertini, L. and M. Mosca, 1999. On the Regulation of a Vertically Differentiated Market. Australian Economic Papers, 38(4), 354-366.

Oberholzer-Gee, F. and H. Weck-Hannemann, 2002. Pricing road use: politico-economic and fairness considerations. Transportation Research Part D: Transport and Environment, 7(2), 357-371.

Ostrovsky, M. and M. Schwarz, 2018. Carpooling and the Economics of Self-Driving Cars. NBER Working Paper No. 24349.

Pels, E. and E. Verhoef, 2004. The economics of airport congestion pricing. Journal of Urban Economics, 55(2), 257-277.

Rupp, N. G., 2009. Do carriers internalize congestion costs? Empirical evidence on the internalization question. Journal of Urban Economics, 65(1), 24-37.

Schrank, D., T. Lomax and B. Eisele, 2011. TTI's 2011 Urban Mobility Report. Texas Transportation Institute, The Texas A&M University System.

Silva, H. and E. Verhoef, 2013. Optimal pricing of flights and passengers at congested airports and the efficiency of atomistic charges. Journal of Public Economics, 106, 1-13.

Silva H., R. Lindsey, A. de Palma and V. van den Berg, 2016. On the Existence and Uniqueness of Equilibrium in the Bottleneck Model with Atomic Users. Transportation Science, 51 (3), 863-881.

Simoni M., K. Kockelmana, K. Gurumurthya and J. Bischoff, 2019. Congestion pricing in a world of self-driving vehicles: An analysis of different strategies in alternative future scenarios. Transportation Research Part C: Emerging technologies, 98, 167-185.

Small, K., 1992. Using the revenues from congestion pricing. Transportation, 19, 359-381.

Small, K., 2012. Valuation of travel time. Economics of Transportation, 1, 2-14.

Small, K., C. Winston and J. Yan, 2005. Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. Econometrica, 73 (4), 1367-1382.

Small, K., C. Winston and J. Yan, 2006. Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design. Brookings-Wharton Papers on Urban Affairs, 53-96.

Spence, A., 1975. Monopoly, quality, and regulation. Bell Journal of Economics. 6(2), 417-429.

Van den Berg, V. and E. Verhoef, 2016. Autonomous cars and dynamic bottleneck congestion: The effects on capacity, value of time and preference heterogeneity. Transportation Research Part B: Methodological, 94, 43-60.

Vickrey, W. S., 1969. Congestion theory and transport investment. American Economic Review, 59(2), 251-260.

Ward, J. W., J. J. Michalek, I. L. Azevedo, C. Samarasa and P. Ferreira, 2019. Effects of on-demand ridesourcing on vehicle ownership, fuel consumption, vehicle miles traveled, and emissions per capita in U.S. States. Transportation Research Part C: Emerging Technologies, 108, 289-301.

Appendix: derivations and proofs

This Appendix contains the proofs of all Propositions, together with the derivations of additional results contained in the paper.

Derivation of (7). We show here that the maximization of (6) implies the maximization of (7). First, we show that, if the planner optimally excludes some travelers, it would exclude travelers with the lowest θ's. By contradiction, suppose that it dispatches a θ'-type, and it does not dispatch a θ''-type, with $\theta' < \theta''$. Then, a switch between the θ''-type not traveling and the θ'-type traveling would leave congestion unaffected while increasing the aggregate net benefit from traveling because $B'(\theta) > 0$, thereby increasing social welfare.

Second, we show that it is never optimal to assign to the same lane travelers in non-contiguous partitions of the unit line. Let types θ' and θ'' be traveling in a lane with a mass of travelers equal to $l' + l''$. Assume, by contradiction, that $\theta' \in [\bar{\theta}', \bar{\theta}' + l']$ and $\theta'' \in [\bar{\theta}'', \bar{\theta}'' + l'']$, with $\bar{\theta}' + l' < \bar{\theta}''$. Let also type θ''' be traveling in a lane with a mass of travelers equal to l''' and that $\theta''' \in [\bar{\theta}' + l', \bar{\theta}'']$. Next, assume that $l' + l'' < l'''$ so that the lane where types θ' and θ'' travel is less congested than the one where type θ''' travels. A switch between types θ' and θ''' would leave congestion in both lanes unaltered, while increasing the aggregate net benefit from traveling because $\theta' < \theta'''$ and $B'(\theta) > 0$. Similarly, assume that $l''' < l' + l''$ so that the lane where types θ' and θ'' travel is more congested than the one where type θ''' travels. A switch between types θ'' and θ''' would leave congestion in both lanes unaltered, while increasing the aggregate net benefit from traveling because $\theta'' > \theta'''$ and $B'(\theta) > 0$.

As a result, the welfare function can be rewritten as follows

$$W = \int_{1-l_1-l_2}^{1-l_2} (B(\theta) - \theta g l_1)\, d\theta + \int_{1-l_2}^{1} (B(\theta) - \theta g l_2)\, d\theta,$$

where l_1 is the mass of travelers in lane 1 and l_2 is the mass of travelers in lane 2. It is easy to prove that $l_1 > l_2$, so that travelers with lower θ's are placed in lane 1, the more congested lane. Consider two travelers, θ' and θ'', with $\theta' < \theta''$ and first suppose θ' uses lane 1, while θ'' uses lane 2. In equilibrium, the aggregate net benefit for the two travelers is $w \equiv B(\theta') - \theta' g l_1 + B(\theta'') - \theta'' g l_2$. Suppose now that θ'' uses lane 1 and θ' uses lane 2. The aggregate net benefit for the two travelers is $w' \equiv B(\theta'') - \theta'' g l_1 + B(\theta') - \theta' g l_2$. Then, $w - w' = B(\theta') - \theta' g l_1 + B(\theta'') - \theta'' g l_2 - (B(\theta'') - \theta'' g l_1 + B(\theta') - \theta' g l_2) = g(\theta'' - \theta')(l_1 + l_2) > 0$, which shows that social welfare is higher in the first case. The planner problem may then be written as in (7). □

Proof of Proposition 1. The Lagrangean of problem (8) is

$$\mathcal{L}_{FB} \equiv \int_{1-s-f}^{1-f} (B(\theta) - \theta g s)\, d\theta + \int_{1-f}^{1} (B(\theta) - \theta g f)\, d\theta - \lambda(s + f - 1).$$

At the solutions to this problem, denoted by s_{FB}, f_{FB} and λ_{FB}, Kuhn-Tucker conditions require

$$\frac{\partial \mathcal{L}_{FB}}{\partial s} = B(1 - s_{FB} - f_{FB}) - 2g s_{FB}\left(1 - f_{FB} - \frac{3}{4} s_{FB}\right) - \lambda_{FB} = 0, \tag{A-1}$$

$$\frac{\partial \mathcal{L}_{FB}}{\partial f} = B(1 - s_{FB} - f_{FB}) + g\left(s_{FB}^2 + \frac{3}{2} f_{FB}^2 - 2 f_{FB}\right) - \lambda_{FB} = 0, \tag{A-2}$$

$$\frac{\partial \mathcal{L}_{FB}}{\partial \lambda} = s_{FB} + f_{FB} - 1 \le 0, \qquad \lambda_{FB} \ge 0 \text{ and } \qquad \frac{\partial \mathcal{L}_{FB}}{\partial \lambda}\lambda_{FB} = 0.$$

Assume that $\lambda_{FB} = 0$ and $s_{FB} + f_{FB} - 1 < 0$, so that the solution is interior. Substitute $\lambda_{FB} = 0$ in (A-1) and (A-2), equate them and solve w.r.to f_{FB} to obtain (13).

Assume instead that $\lambda_{FB} \ge 0$ and $s_{FB} + f_{FB} = 1$. Substitute $s_{FB} = 1 - f_{FB}$ in (A-1) and (A-2), equate them and solve w.r.to f_{FB} to obtain $f_{FB} = \frac{1}{2} - \frac{\sqrt{7}-2}{6}$. Use again $s_{FB} = 1 - f_{FB}$ to get $s_{FB} = \frac{1}{2} + \frac{\sqrt{7}-2}{6}$. Plug f_{FB} and s_{FB} thus obtained into (A-1) or (A-2) and solve w.r.to λ_{FB} to obtain $\lambda_{FB} = B(0) - g\frac{4+\sqrt{7}}{36}$. Solve $\lambda_{FB} \ge 0$ w.r.to g to get $g \le g_{FB}$. ■

Derivation of the comparative static results in the first best. We derive here the comparative statics results mentioned in Section 3, i.e., $\frac{\partial s_{FB}}{\partial g} < 0$ and $\frac{\partial f_{FB}}{\partial g} < 0$.

Denote the FOCs of the maximization problem for the planner in (8), given in (9) and (10), as $h_s(s_{FB}, f_{FB}, g) = 0$ and $h_f(s_{FB}, f_{FB}, g) = 0$, respectively. Implicit differentiation of the FOCs w.r.to g gives

$$\frac{ds_{FB}}{dg} = \frac{\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial f_{FB}} - \frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial g}}{\frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial f_{FB}}};$$

$$\frac{df_{FB}}{dg} = \frac{\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial g}}{\frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial f_{FB}}}.$$

In problem (8), SOCs require $\frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial f_{FB}} > 0$. As a result,

$$\operatorname{sgn}\left(\frac{ds_{FB}}{dg}\right) = \operatorname{sgn}\left(\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial f_{FB}} - \frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial g}\right)$$

and

$$\operatorname{sgn}\left(\frac{df_{FB}}{dg}\right) = \operatorname{sgn}\left(\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial g}\right).$$

Using (9) and (10) yields

$$\frac{\partial h_s}{\partial s_{FB}} = -B'(1 - s_{FB} - f_{FB}) - 2g\left(1 - f_{FB} - \frac{3}{2} s_{FB}\right),$$

$$\frac{\partial h_s}{\partial f_{FB}} = -B'(1 - s_{FB} - f_{FB}) + 2g s_{FB},$$

$$\frac{\partial h_f}{\partial s_{FB}} = -B'(1 - s_{FB} - f_{FB}) + 2g s_{FB},$$

$$\frac{\partial h_f}{\partial f_{FB}} = -B'(1 - s_{FB} - f_{FB}) + g\,(3f_{FB} - 2)\,,$$

$$\frac{\partial h_s}{\partial g} = -2s_{FB}\left(1 - f_{FB} - \frac{3}{4}s_{FB}\right),$$

$$\frac{\partial h_f}{\partial g} = (s_{FB})^2 + \frac{3}{2}(f_{FB})^2 - 2f_{FB}.$$

When evaluated at the equilibrium relationship (13), the numerator of $\frac{ds_{FB}}{dg}$ is negative for any $s_{FB} \le \bar{s}_{FB}$, where $\bar{s}_{FB}$ is the full coverage and maximum value of s_{FB}, computed in Proposition 1. Hence, $\frac{ds_{FB}}{dg} < 0$. Also, from (13), $\frac{df_{FB}}{ds_{FB}} > 0$, hence $\frac{df_{FB}}{dg} = \frac{df_{FB}}{ds_{FB}}\frac{ds_{FB}}{dg} < 0$. □

Proof of Proposition 2. We first prove that any equilibrium must be such that $s = f$. Assume, by contradiction, that $s > f$; any traveler in the slow lane has an incentive to switch to the other lane, proving that this cannot be an equilibrium. When $s = f$ and in the absence of fares, the two IC constraints (3) and (5) are always trivially satisfied and the two IR constraints (2) and (4) become identical and equal to $B(\theta) - \theta g s = B(\theta) - \theta g f$. This value is nonnegative for any θ, showing that all atomistic travelers travel. ■

Derivation of (15) and (16). We proved in Proposition 2 that all atomistic travelers travel. It follows that

$$s^a + f^a = 1 - \mu. \tag{A-3}$$

First, assume $1 - \mu \ge s^c - f^c$. We want to prove that s^a and f^a are such that $s = s^a + s^c = f^a + f^c = f$. Assume, by contradiction, s^a and f^a are such that $s > f$. For all θ-type atomistic travelers in the slow lane, it must be the case that

$$B(\theta) - \theta g(s^a + s^c) \ge B(\theta) - \theta g(f^a + f^c),$$

which reduces to $s^a + s^c \le f^a + f^c$ or, equivalently, using $s = s^a + s^c$ and $f = f^a + f^c$, to $s < f$. This contradicts our initial hypothesis that $s > f$. A similar argument rules out the possibility that $s < f$. Hence, when $1 - \mu \ge s^c - f^c$, s^a and f^a are such that $s = f$.

Next, assume $1 - \mu < s^c - f^c$. We want to prove that $s^a = 0$. Assume, by contradiction, that $s^a > 0$. For all θ-type atomistic travelers in the slow lane, it must be the case that

$$B(\theta) - \theta g(s^a + s^c) \ge B(\theta) - \theta g(f^a + f^c),$$

which reduces to $s^a + s^c \le f^a + f^c$ or, using (A-3), to $1 - \mu \ge s^c - f^c + 2s^a$. This contradicts $1 - \mu < s^c - f^c$. Hence, when this inequality holds, $s^a = 0$ and, substituting $s^a = 0$ into (A-3), $f^a = 1 - \mu$ as in (16). □

Proof of Proposition 3. This is part of the proof of Proposition 6. ■

Proof of Proposition 4. This is part of the proof of Proposition 7. ■

Proof of Proposition 5. We evaluate and compare the welfare expression in (23) at the equilibrium masses of travelers in the different regimes.

Let $\mathcal{B} \equiv \int_0^1 B(\theta)d\theta$. Denote by $\bar{W}_{subscript}$ the equilibrium welfare, where the *subscript* is one of the ones used in the different subsections of Section 4, that is, FB for first best, A for atomistic travelers only, AC for atomistic and corporate travelers coexisting, and C for corporate travelers only; in case of atomistic and corporate travelers coexisting, we use subscript AC when $\mu \in \left(0, \frac{1}{2}\right]$ and AC' otherwise. Then

$$\bar{W}_{FB} = \mathcal{B} - \frac{44 - 7\sqrt{7}}{108}g,$$

$$\bar{W}_A = \bar{W}_{AC} = \mathcal{B} - \frac{1}{4}g,$$

$$\bar{W}_{AC'} = \mathcal{B} - \frac{16\mu^3 - 24\mu^2 + 45\mu - 1 + \left(8\mu^2 - 10\mu - 1\right)\sqrt{4\mu^2 - 2\mu + 1}}{108\mu}g,$$

and

$$\bar{W}_C = \mathcal{B} - \frac{12-\sqrt{3}}{36}g;$$

the comparison follows through immediately. ■

Derivation of the threshold μ' such that $1-\mu = s^c_{AC} - f^c_{AC}$ under partial coverage. This is part of the proof of Proposition 6**.** □

Proof of Proposition 6. Assume μ is sufficiently large so that $1-\mu < s^c - f^c$ holds at equilibrium s^c and f^c (something that will be checked later on). The Lagrangean of the monopolist's problems in (28) is given by

$$\begin{aligned}\mathcal{G}_{AC} \equiv & \left(B\left(1-\frac{s^c+f^c}{\mu}\right) - \left(1-\frac{s^c+f^c}{\mu}\right)gs^c\right)(s^c+f^c)+\\ &+g\left(1-\frac{f^c}{\mu}\right)(s^c-f^c-f^a)f^c - \gamma(s^c+f^c-\mu).\end{aligned}$$

At the solutions to this problem, denoted by s^c_{AC}, f^c_{AC} and γ_{AC}, exploiting $f^a = 1-\mu$, Kuhn-Tucker conditions require

$$\begin{aligned}\frac{\partial \mathcal{G}_{AC}}{\partial s^c} =& B\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right) - \frac{B'\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right)\times(s^c_{AC}+f^c_{AC})}{\mu}+ \qquad \text{(A-4)}\\ &+\frac{gs^c_{AC}\left[3s^c_{AC}+4f^c_{AC}-2\mu\right]}{\mu} - \gamma_{AC} = 0,\\ \frac{\partial \mathcal{G}_{AC}}{\partial f^c} =& B\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right) - \frac{B'\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right)\times(s^c_{AC}+f^c_{AC})}{\mu}+ \qquad \text{(A-5)}\\ &+\frac{g\left[2\left(s^c_{AC}\right)^2 + f^c_{AC}\left(4(1-\mu)+3f^c_{AC}-2\right)-(1-\mu)\mu\right]}{\mu} - \gamma_{AC} = 0,\\ \frac{\partial \mathcal{G}_{AC}}{\partial \gamma} =& s^c_{AC}+f^c_{AC}-\mu \le 0, \qquad \gamma_{AC}\ge 0 \text{ and } \qquad \frac{\partial \mathcal{G}_{AC}}{\partial \gamma}\gamma_{AC} = 0. \qquad \text{(A-6)}\end{aligned}$$

Assume that $\gamma_{AC} > 0$ and $s^c_{AC}+f^c_{AC} = \mu$. Substitute $s^c_{AC} = \mu - f^c_{AC}$ in (A-4) and (A-5), equate them and solve w.r.to f^c_{AC} to obtain f^c_{AC} as in (20). Use again $s^c_{AC}+f^c_{AC} = \mu$ to obtain s^c_{AC} as in (20). Plug s^c_{AC} and f^c_{AC} thus obtained into (A-4) or (A-5) and solve w.r.to γ_{AC} to obtain $\gamma_{AC} = B(0) - B'(0) + g\frac{(4\mu-1)\sqrt{4\mu^2-2\mu+1}+8\mu^2+5\mu-1}{18\mu}$. Solve $\gamma_{AC}\ge 0$ w.r.to g to obtain $g \le g_{AC}$ as in (29). Plug s^c_{AC} and f^c_{AC} back into the inequality $1-\mu < s^c - f^c$, which becomes $\mu > \frac{1}{2}$. Hence, when the monopolist chooses s^c_{AC} and f^c_{AC}, atomistic travelers place themselves as in (16).

Assume that $\gamma_{AC} = 0$ and $s^c_{AC}+f^c_{AC}\le\mu$, so that the solution is interior. Substitute $\gamma_{AC} = 0$ in (A-4) and (A-5), equate them and solve w.r.to f^c_{AC} to obtain $f^c_{AC}(s^c_{AC})$ as in (33).

Assume now μ is sufficiently small so that $1-\mu \ge s^c - f^c$ in equilibrium (something that will be checked later on). The Lagrangean of the monopolist's problem in (25) is

$$\mathcal{L}_{AC} \equiv \left[B\left(1-\frac{c}{\mu}\right) - \left(1-\frac{c}{\mu}\right)g\frac{c+s^a+f^a}{2}\right]c - \lambda(c-\mu).$$

At the solutions, denoted by c_{AC} and λ_{AC}, Kuhn-Tucker conditions require

$$\begin{aligned}\frac{\partial \mathcal{L}_{AC}}{\partial c} =& \frac{3g\left(c_{AC}\right)^2}{2\mu} + c_{AC}\frac{g(1-2\mu) - B'\left(1-\frac{c_{AC}}{\mu}\right)}{\mu}+ \qquad \text{(A-7)}\\ &+\frac{2B\left(1-\frac{c_{AC}}{\mu}\right) - g(1-\mu)}{2} - \lambda_{AC} = 0,\\ \frac{\partial \mathcal{L}_{AC}}{\partial \lambda} =& c_{AC}-\mu \le 0, \qquad \lambda_{AC}\ge 0 \text{ and } \qquad \frac{\partial \mathcal{L}_{AC}}{\partial \lambda}\lambda_{AC} = 0. \qquad \text{(A-8)}\end{aligned}$$

Assume that $\lambda_{AC} > 0$ and $c_{AC} = \mu$. Substitute $c_{AC} = \mu$ in (A-7) and solve w.r.to λ_{AC} to obtain $\lambda_{AC} = B(0) - B'(0) + \frac{1}{2}g$. Solve $\lambda_{AC} > 0$ w.r.to g to obtain g_{AC} as in (29).

Assume that $\lambda_{AC} = 0$ and $c_{AC} < \mu$, so that the solution is interior. Substitute $\lambda_{AC} = 0$ in (A-7) to obtain c_{AC}.

We now derive the threshold μ', that is, we check under which conditions $1 - \mu < (\geq) s^c - f^c$ in equilibrium under partial coverage. First note that $1 - \mu > \frac{1}{2} > \mu \geq s^c_{AC} - f^c_{AC}$ for any s^c_{AC}, f^c_{AC}; this means that $\mu \geq \frac{1}{2}$ is a necessary condition for $1 - \mu < s^c_{AC} - f^c_{AC}$ to be fulfilled, hence for differentiation to arise. Therefore, we focus on the interval $\mu \in \left[\frac{1}{2}, 1\right)$ to calculate the threshold value μ' such that $1 - \mu = s^c_{AC} - f^c_{AC}$.

The difference $s^c_{AC} - f^c_{AC}$, computed using (33), is

$$\frac{1}{3}\left(s^c_{AC} - 2\mu + \sqrt{-4s^c_{AC} - \mu + 2s\mu + \mu^2 + 7\left(s^c_{AC}\right)^2 + 1} + 1\right).$$

We find that: $1 - \mu > s^c_{AC} - f^c_{AC}$ at $\mu = \frac{1}{2}$ for any admissible s^c_{AC} and f^c_{AC}; $1 - \mu < s^c_{AC} - f^c_{AC}$ at $\mu = 1$ for any s^c_{AC} and f^c_{AC} such that $s^c_{AC} + f^c_{AC} \neq 0$. By implicitly differentiating s^c_{AC} and f^c_{AC} w.r.to μ it is possible to establish $\frac{\partial\left(s^c_{AC} - f^c_{AC}\right)}{\partial \mu} > 0$. It follows there exists a $\mu' \in \left(\frac{1}{2}, 1\right)$ such that $1 - \mu \geq s^c_{AC} - f^c_{AC}$ for any $\mu \in \left[\frac{1}{2}, \mu'\right]$ and $1 - \mu < s^c_{AC} - f^c_{AC}$ for any $\mu \in (\mu', 1]$ ■

Proof of Proposition 7. In setting the two fares, p for the slow lane and P for the fast lane, the monopolist faces the IR constraint for the lowest type traveling and the IC constraint for the type indifferent between traveling in the slow or fast lane. When binding, these constraints now read as

$$p = B\left(1 - s - f\right) - \left(1 - s - f\right) gs, \tag{A-9}$$

$$P = p + g\left(1 - f\right)\left(s - f\right). \tag{A-10}$$

The monopolist problem can therefore be written as

$$\max_{s \geq 0, f \geq 0} \left(B\left(1 - s - f\right) - \left(1 - s - f\right) gs\right)\left(s + f\right) + \left[g\left(1 - f\right)\left(s - f\right)\right] f, \quad \text{s.t. } s + f \leq 1.$$

The Lagrangean of this problem is

$$\mathcal{L}_C \equiv \left(B\left(1 - s - f\right) - \left(1 - s - f\right) gs\right)\left(s + f\right) + \left(g\left(1 - f\right)\left(s - f\right)\right) f - \lambda(s + f - 1).$$

At the solutions, denoted by s_C, f_C and λ_C, Kuhn-Tucker conditions require

$$\frac{\partial \mathcal{L}_C}{\partial s} = B\left(1 - s_C - f_C\right) - B'\left(1 - s_C - f_C\right)\left(s_C + f_C\right) + - g s_C\left(2 - 4f_C - 3s_C\right) - \lambda_C = 0, \tag{A-11}$$

$$\frac{\partial \mathcal{L}_C}{\partial f} = B\left(1 - s_C - f_C\right) - B'\left(1 - s_C - f_C\right)\left(s_C + f_C\right) + g\left(3\left(f_C\right)^2 - 2f_C + 2\left(s_C\right)^2\right) - \lambda_C = 0, \tag{A-12}$$

$$\frac{\partial \mathcal{L}_C}{\partial \lambda} = s_C + f_C - 1 \leq 0, \qquad \lambda_C \geq 0 \text{ and } \qquad \frac{\partial \mathcal{L}_C}{\partial \lambda}\lambda_C = 0. \tag{A-13}$$

Assume that $\lambda_C = 0$ and $s_C + f_C - 1 < 0$. Substitute $\lambda_C = 0$ in (A-11) and (A-12), equate them and solve w.r.to f_C to obtain (35).

Assume instead that $\lambda_C \geq 0$ and $s_C + f_C = 1$. Substitute $s_C = 1 - f_C$ in (A-11) and (A-12), equate them and solve w.r.to f_C to obtain $f_C = \frac{1}{2} - \frac{\sqrt{3}}{6}$ as in (22). Use again $s_C = 1 - f_C$ to get $s_C = \frac{1}{2} + \frac{\sqrt{3}}{6}$ as in (22). Plug f_C and s_C thus obtained into (A-11) or (A-12) and solve w.r.to λ_C to obtain $\lambda_C = B(0) - B'(0) + g\frac{4+\sqrt{3}}{6}$. Solve $\lambda_C \geq 0$ w.r.to g to obtain $g \leq g_C$ as in (34). ■

Derivation of the sufficiency of Assumption 3 for full coverage to occur in first best and in all regimes. We show that, when Assumption 3 holds, full coverage always occurs.

First best. In the first best, full coverage occurs when $g \leq g_{FB}$, where g_{FB} is given in (11). From the concavity of $B(.)$ and Assumption 3, we may write $B'(1) \leq B'(0) \leq B(0)$. Using (11) and Assumption 1, then $g < B'(1) < g_{FB}$, which proves our claim.

Decentralized regime. Full coverage is the result of Assumptions 1 and 2 only, which ensure that all travelers enjoy a nonnegative utility when traveling and paying no fee.

Mixed regime. Note that g_{AC} in (29) is nonpositive when $\frac{B'(0)}{B(0)} \leq 1$, which proves our claim.

Centralized regime. An argument identical to the one used in the case of the mixed regime proves our claim. □

Derivation of the conditions for the existence of the Spence distortion. A sufficient condition for the mass of travelers to be larger under the centralized regime than in the first best is whenever the monopolist fully covers the market but the social planner does not, i.e., $\max\{g_{FB}, g_{AC}\} \leq g$. This occurs when $\max\left\{\frac{6}{4+\sqrt{3}}\left[B'(0) - B(0)\right]; \frac{36}{4+\sqrt{7}}B(0)\right\} \leq g < B'(1)$. This interval is not empty if and only if $B'(0) < \frac{28+\sqrt{7}+6\sqrt{3}}{36} B'(1) \cong 1.1399\, B'(1)$, thus requiring the gross benefit function $B(.)$ to be "not too" concave. □

Proof of Proposition 8. With the linear specification $B(\theta) = b_0 + b\theta$, the welfare gap $\Delta W \equiv W_{AC} - W_A$ becomes

$$\Delta W = \frac{(\mu - c_{AC})\left(-gc_{AC}^2 - 2b\mu + 2g\mu - g\mu^2 - 4\mu b_0 + 2bc_{AC} - gc_{AC} + 2gc_{AC}\mu\right)}{4\mu} \tag{A-14}$$

with

$$c_{AC} = \frac{2b - g + 2g\mu - \sqrt{g^2\mu^2 - g^2\mu - 4bg + 4b^2 + g^2 + 2bg\mu - 6g\mu b_0}}{3g}.$$

To study the sign of ΔW, we first observe that plugging $c_{AC} = 0$ into (A-14) yields

$$\Delta W(0) = \frac{\mu}{4}\left(-2b + 2g - 4b_0 - g\mu\right),$$

which is negative under Assumption 1. We then solve $\Delta W = 0$ for c_{AC} and get three solutions: μ,

$$\underline{c} = \frac{2b + g(2\mu - 1) - \sqrt{(2b-g)^2 - 4g(4b_0 - g)\mu}}{2g}$$

and

$$\overline{c} = \frac{2b + g(2\mu - 1) + \sqrt{(2b-g)^2 - 4g(4b_0 - g)\mu}}{2g}.$$

We prove our result in three steps. (i) First observe that $(2b-g)^2 - 4g(4b_0 - g)\mu < 0$ iff $\mu > \frac{(2b-g)^2}{4g(4b_0-g)}$ and $4b_0 - g > 0$. In this case, the quadratic expression in (A-14) is negative, hence $\Delta W < 0$. (ii) Second, when $\mu < \frac{(2b-g)^2}{4g(4b_0-g)}$ and $4b_0 - g > 0$, we find that $\mu < \underline{c} < \overline{c}$ and that $\frac{\partial \Delta W}{\partial c_{AC}} > 0$ at $c_{AC} = \mu$. Since $\Delta W < 0$ at $c_{AC} = 0$ and ΔW is a continuous function in $c_{AC} \in (0, \mu)$, then $\Delta W < 0$ at any $c_{AC} \in (0, \mu)$. (iii) Finally, when $4b_0 - g < 0$, we find that $\underline{c} < \mu < \overline{c}$ and $\frac{\partial \Delta W}{\partial c_{AC}} < 0$ at $c_{AC} = \mu$. This implies that $\Delta W \leq 0$ when $c_{AC} \leq \underline{c}$ and $\Delta W > 0$ when $\underline{c} < c_{AC} < \mu$.

To prove that $c_{AC} > \underline{c}$ when $\mu \to 0$ and $\max\left\{\frac{2}{3}b + 2b_0, 4b_0\right\} < g$, we proceed as follows. Note that $c_{AC} = \underline{c} = 0$ at $\mu = 0$. Moreover,

$$\lim_{\mu \to 0} \frac{\partial c_{AC}}{\partial \mu} = \frac{6b_0 + 3(2b - g)}{6(2b - g)} > 0 \tag{A-15}$$

and

$$\lim_{\mu \to 0} \frac{\partial \underline{c}}{\partial \mu} = \frac{4b_0 + 2(b - g)}{2b - g} > 0. \tag{A-16}$$

Note that $(A-15) > (A-16)$ iff $g > \frac{2}{3}b + 2b_0$, which proves our result. ■

Derivation of the marginal aggregate congestion cost in a decentralized regime. We calculate the marginal aggregate congestion cost imposed by type-0 traveler on fellow travelers in the same lane in a decentralized regime when there is no differentiation across lanes (i.e. when $\mu \in (0, \mu']$).

From (31), all travelers face the same congestion level $s_{AC} = f_{AC} = \frac{c_{AC}+1-\mu}{2}$. Aggregate congestion costs are given by $\Gamma \equiv \int_0^1 \theta g \frac{c_{AC}+1-\mu}{2} d\theta$. The marginal congestion cost is then simply given by $\frac{\partial \Gamma}{\partial c_{AC}} = \frac{g}{4}$. When choosing whether to exclude the type-0 from traveling, the planner compares this with the benefit this individual derives from traveling, i.e. $U(0) = b_0$. □

Proof of Proposition 9. When $g \leq g_{FB}$, full coverage occurs not only in the decentralized regime but also in the first best. Since the type-0 atomistic traveler gets utility $B(0) - t$ from traveling in the slow lane, Assumptions 1 and 2 implies that full coverage occurs as in the social optimum when $t = t_A$. Given t_A, substitute $\bar{s}_{FB}$ as in (12) into the IC constraint (5) to write

$$B\left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right) - \left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right) g \left(\frac{1}{2} - \frac{\sqrt{7}-2}{6}\right) - T \geq$$

$$B\left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right) - g\left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right)^2 - t_A$$

and solve it w.r. to T when holding as an equality to obtain T_A.

When $g \geq g_{FB}$, the market is fully covered under the decentralized regime, but not in the social optimum. To obtain t_A, consider that the marginal net effect on social welfare of a θ-type traveler deciding to travel in the slow lane (as opposed to not traveling) is given by the LHS of (9), while his private benefit is given by $B(\theta) - \theta g s$. The difference between this private benefit and the LHS of (9) is positive and therefore corresponds to the non-internalized component of the marginal social cost, which we set equal to t_A. Given t_A, use s_{FB} and f_{FB} in the IC constraint (5) to write

$$B(1 - f_{FB}) - (1 - f_{FB}) g f_{FB} - T = B(1 - f_{FB}) - (1 - f_{FB} - s_{FB}) g s_{FB} - t_A.$$

and solve it w.r. to T when holding as an equality to obtain T_A. ■

Proof of Proposition 10. The monopolist problem is

$$\max_{p \geq 0, P \geq 0} \ [p - t(s,f)]\, n + [P - T(s,f)]\, N \qquad \text{(A-17)}$$
$$s.t. \ \ (A-9) - (A-10)$$

where (A-9) and (A-10) are the same individual rationality and incentive compatibility constraints faced by the monopolist in the absence of taxes. Solving these two constraints w.r.to the monopolist's fares p and P and plugging them into the monopolist's problem in (A-17) allows us to rewrite it as follows

$$\max_{\substack{s \geq 0, \\ f \geq 0}} \ [B(1 - s - f) - (1 - s - f) g s - t(s,f)]\, s + \qquad \text{(A-18)}$$
$$[B(1 - s - f) - (1 - s - f) g s + g(1 - f)(s - f) - T(s,f)]\, f$$
$$\text{s.t.} \ \ s + f \leq 1.$$

Using (41), the Lagrangean of this problem is

$$\begin{aligned}\mathcal{L}_{CT} \equiv & (B(1 - s - f) - (1 - s - f) g s)\, s + \\ & + (B(1 - s - f) - (1 - s - f) g s + g(1 - f)(s - f))\, f + \\ & - (g s - z_C)\, s - (g f - z)\, f \\ & - \lambda (s + f - 1).\end{aligned}$$

At the solutions, denoted by s_{CT}, f_{CT} and γ_{CT}, Kuhn-Tucker conditions require

$$\begin{aligned}\frac{\partial \mathcal{L}_{CT}}{\partial s} = & B(1 - s_{CT} - f_{CT}) - B'(1 - s_{CT} - f_{CT})(s_{CT} + f_{CT}) + \\ & - g s_{CT}(2 - 4 f_{CT} - 3 s_{CT}) + z - \lambda_{CT} = 0;\end{aligned} \qquad \text{(A-19)}$$

$$\frac{\partial \mathcal{L}_{CT}}{\partial f} = B\left(1-s_{CT}-f_{CT}\right) - B'\left(1-s_{CT}-f_{CT}\right)\left(s_{CT}+f_{CT}\right) + \tag{A-20}$$
$$+ g\left(3(f_{CT})^2 + 2(s_{CT})^2 - 2f_{CT}\right) + z - \lambda_{CT} = 0;$$
$$\frac{\partial \mathcal{L}_{CT}}{\partial \gamma} = s_{CT} + f_{CT} - 1 \leq 0, \qquad \gamma_{CT} \geq 0 \text{ and } \qquad \frac{\partial \mathcal{L}_{CT}}{\partial \lambda}\gamma_{CT} = 0.$$

Assume that $\lambda_{CT} = 0$ and $s_{CT} + f_{CT} - 1 < 0$. Substituting $\lambda_{CT} = 0$ in (A-19) and (A-20), equalize them and solve w.r.to f to get

$$f_{CT}(s_{CT}) = \frac{1}{3}\left(2\left(1+s_{CT}\right) - \sqrt{7s_{CT}^2 - 4s_{CT} + 4}\right); \tag{A-21}$$

which is identical to (13). Notice that the level of the subsidy z does not affect $f_{CT}(s_{CT})$ in (A-21).

Assume instead that $\lambda_{CT} \geq 0$ and $s_{CT} + f_{CT} = 1$. Substituting $s_{CT} = 1 - f_{CT}$ in (A-19) and (A-20), equating them and solving w.r.to f_{CT} gives $f_{CT} = \frac{1}{2} - \frac{\sqrt{7}-2}{6}$ as in (12). Using again $s_C = 1 - f_C$ gives $s_C = \frac{1}{2} - \frac{\sqrt{7}-2}{6}$ as in (12). Plugging f_{CT} and s_{CT} thus obtained into (A-19) or (A-20) and solving w.r.to λ_{CT} it gives $\lambda_{CT} = \frac{g\left(4+\sqrt{7}\right)}{B(0)-B'(0)+z}$. Solving $\lambda_{CT} \geq 0$ w.r.to g gives

$$g \leq g_{CT} \equiv \frac{18\left[B(0) - B'(0) + z\right]}{4+\sqrt{7}}. \tag{A-22}$$

Notice that, when $z = 0$, $g_{CT} < g_{FB}$, so that $g \leq g_{CT}$ implies $g \leq g_{FB}$; that is, whenever the monopolist subject to a system of tax/subsidy as in (41) with $z = 0$ covers the market, a social planner would do it as well. This implies that, when $g \leq g_{CT}$, the subsidy that restores social optimality is equal to zero.

When instead $g > \frac{18\left[B(0)-B'(0)\right]}{4+\sqrt{7}}$, a positive subsidy is required for (A-22) to hold. The smallest subsidy is obtained by solving (A-22) w.r.to z when this holds as an equality. This gives $z = B'(0) - B(0) + g\,\frac{4+\sqrt{7}}{18}$, which is optimal as long as full coverage is socially optimal, i.e. $g \leq \frac{36\,B(0)}{4+\sqrt{7}}$.

Focus now on $g > \frac{36\,B(0)}{4+\sqrt{7}}$. In this case, the social planner partially covers the market, hence the scheme should deliver the same result in the centralized regime. The optimal subsidy is then computed as follows. Equalize the FOC w.r.to s or f in the monopolist problem in (A-18) to the FOC w.r.to s or f in the social planner problem in (8), and solve by z. This gives $z_C = B'(1 - s_{FB} - f_{FB})(s_{FB} + f_{FB}) + g\left(2f_{FB} - (s_{FB})^2 - \frac{3}{2}(f_{FB})^2\right)$. ■

The optimal tax scheme in a mixed regime. We characterize the set of taxes that restores the social optimum in a mixed regime, where atomistic and corporate travelers coexist. We show that the tax schemes markedly differ across the two groups. The tax schemes imposed to each group closely resembles that charged to that group in isolation.

The timeline of the game we solve is the following. First, the tax authority announces a tax scheme t^a, T^a, t^c and T^c, where t^a (t^c, respectively) denotes the tax imposed on atomistic (corporate) travelers traveling in the slow lane, and T^a (T^c, respectively) denotes the tax imposed on atomistic (corporate) travelers traveling in the fast lane. Second, the monopolist observes it and sets the fares p and P. Third, corporate and atomistic travelers observe the fares, and simultaneously make their travel decisions. As in the analysis contained in the main text, without any effect on our analysis, we assume that taxes on atomistic travelers are directly imposed on travelers, taxes on corporate travelers are imposed on the monopolist.

Let the taxes that replicate the social optimum be denoted as $t^a_{AC}, T^a_{AC}, t^c_{AC}$ and T^c_{AC}. Our results are summarised in the following Proposition:

Proposition. *Let*

$$x_{AC} \equiv \frac{(1-\mu)\left[(5+2s_{FB})\sqrt{7(s_{FB})^2 - 4s_P + 4} - 10(1+(s_{FB})^2)\right] + (1+17\mu)s_{FB}}{18\mu s_{FB}}. \tag{A-23}$$

For any given $\mu \in [0,1]$ and for any socially optimal allocation of travelers, the social optimum is restored by a system of per-vehicle taxes $t^a_{AC}, T^a_{AC}, t^c_{AC}$ and T^c_{AC} such that
i) taxes on atomistic travelers are

$$\begin{aligned} t^a_{AC} &= t_A, \\ T^a_{AC} &= T_A, \end{aligned} \tag{A-24}$$

where t_A and T_A are as given in Proposition 9;
ii) taxes on corporate travelers are

$$\begin{aligned} t^c_{AC} &= g\, s^c_{AC}\, x_{AC} - z_{AC}, \\ T^c_{AC} x &= g\, f^c_{AC} - z_{AC}. \end{aligned} \tag{A-25}$$

Proof. To find the first-best restoring set of taxes, we use the fact that travelers, atomistic and corporate, are exogenously allocated to their groups in fixed proportions, equal to $1-\mu$ and μ, respectively. In solving for the optimal taxes for each group, we take the allocation across lanes of the travelers in the other group as given. A system of optimal taxes is then just composed of a system of taxes on corporate travelers that restores the social optimum for this group, given the system of optimal taxes for the atomistic travelers, and viceversa.

Let the corporate travelers be allocated as to replicate, pro quota, the first best, so that $s^c = \mu s_{FB}$ and $f^c = \mu f_{FB}$. The proof of Proposition 9 applies almost verbatim to prove result *i)*.

Next, let the atomistic travelers be allocated as to replicate, pro quota, the first best, so that $s^a = (1-\mu)s_{FB}$ and $f^a = (1-\mu)f_{FB}$. First, we note that the individual rationality and incentive compatibility constraints faced by the monopolist are the same as in the absence of taxes. We then solve these two constraints w.r.to the monopolist's fares as in (26) and (27) and plug them into the monopolist's objective function to write its problem as follows

$$\begin{aligned} \max_{\substack{s^c\geq 0,\\ f^c\geq 0}} & \left\{ B\left(1-\frac{s^c+f^c}{\mu}\right) - \left[1-\frac{s^c+f^c}{\mu}\right] g\left[(1-\mu)s_P + s^c\right] - t^c \right\} s^c + \\ & \left\{ B\left(1-\frac{s^c+f^c}{\mu}\right) - \left[1-\frac{s^c+f^c}{\mu}\right] g\left[(1-\mu)s_{FB} + s^c\right] + \right. \\ & \left. \left[1-\frac{f^c}{\mu}\right] g\left[(1-\mu)(s_{FB}-f_{FB}) + s^c - f^c\right] - T^c \right\} f^c \\ \text{s.t.} \quad & s^c + f^c \leq 1-(1-\mu)(s_{FB}+f_{FB}) \end{aligned} \tag{A-26}$$

where we have also made use of $s^a = (1-\mu)s_{FB}$ and $f^a = (1-\mu)f_{FB}$.

Using $t^c = gs^c x - z$ and $T^c = gf^c - z$, the Lagrangean of this problem is

$$\begin{aligned} \mathcal{L}_T \equiv & \left\{ B\left(1-\frac{s^c+f^c}{\mu}\right) - \left[1-\frac{s^c+f^c}{\mu}\right] g\left[(1-\mu)s_{FB} + s^c\right] \right\} s^c + \\ & \left\{ B\left(1-\frac{s^c+f^c}{\mu}\right) - \left[1-\frac{s^c+f^c}{\mu}\right] g\left[(1-\mu)s_{FB} + s^c\right] + \right. \\ & \left. \left[1-\frac{f^c}{\mu}\right] g\left[(1-\mu)(s_{FB}-f_{FB}) + s^c - f^c\right] \right\} f^c + \\ & -(gs^c x - z)s^c - (gf^c - z)f^c + \\ & -\lambda\left(s^c + f^c - (1-(1-\mu)(s_P+f_P))\right). \end{aligned}$$

At the solutions, denoted by s^c_T, f^c_T and λ_T, Kuhn-Tucker conditions require

$$\begin{aligned} \frac{\partial \mathcal{L}_T}{\partial s^c} = & \left(-\frac{B'\left(1-\frac{s^c_T+f^c_T}{\mu}\right)}{\mu} + g\frac{(1-\mu)s_{FB}+s^c_T}{\mu} - \left(1-\frac{s^c_T+f^c_T}{\mu}\right) g \right) (s^c_T + f^c_T) + \\ & + B\left(1-\frac{s^c_T+f^c_T}{\mu}\right) - \left(1-\frac{s^c_T+f^c_T}{\mu}\right) g[(1-\mu)s_{FB}+s^c_T] + \\ & -2gxs^c_T + \left(1-\frac{f^c_T}{\mu}\right) gf^c_T + s - \lambda_T = 0; \end{aligned} \tag{A-27}$$

$$\frac{\partial \mathcal{L}_T}{\partial f^c} = \left(-\frac{B'\left(1-\frac{s_T^c+f_T^c}{\mu}\right)}{\mu} + g\frac{(1-\mu)s_{FB}+s_T^c}{\mu} \right)(s_T^c+f_T^c)+ \tag{A-28}$$

$$- \left(\frac{(1-\mu)(s_{FB}-f_{FB})+s_T^c-f_T^c}{\mu} + \left(1-\frac{f_T^c}{\mu}\right)+2 \right) g f_T^c +$$

$$+ B\left(1-\frac{s_T^c+f_T^c}{\mu}\right) - \left(1-\frac{s_T^c+f_T^c}{\mu}\right) g[(1-\mu)s_{FB}+s_T^c]+$$

$$+ \left(1-\frac{f_T^c}{\mu}\right) g\left[(1-\mu)(s_{FB}-f_{FB})+s_T^c-f_T^c\right] + s - \lambda_T = 0;$$

$$\frac{\partial \mathcal{L}_T}{\partial \lambda} = s_T^c + f_T^c \leq 1-(1-\mu)(s_{FB}+f_{FB}); \quad \lambda_T \geq 0 \text{ and } \frac{\partial \mathcal{L}_T}{\partial \lambda}\lambda_T = 0.$$

Assume that $\lambda_T = 0$ and $s_T + f_T < 1-(1-\mu)(s_{FB}+f_{FB})$. Substituting $\lambda_T = 0$ in (A-27) and (A-28), equalize them and solve w.r.to f_T^c to get

$$f_T^c(s_T^c;x) = \frac{1}{3}(2\mu-(1-\mu)(f_{FB}-s_{FB})+2s_T^c)+ \tag{A-29}$$

$$-\frac{1}{3}\Big[(\mu^2((f_{FB}-s_{FB})^2+f_{FB}-s_{FB}+4)-(\mu(1-4s_T^c)+4s_T^c)(f_{FB}-s_{FB})+$$

$$-6s_T^c\mu x+(1-2\mu)(f_{FB}-s_{FB})^2+7(s_T^c)^2+2s_T^c\mu)\Big]^{\frac{1}{2}}$$

that gives the profit maximizing choice of f_T^c as a function of s_T^c and the parameter x. Notice that $f_{CT}^c(s_T^c;x)$ does not depend on the level of the subsidy z.

From $s^a = (1-\mu)s_{FB}$ and $f^a = (1-\mu)f_{FB}$, it must be that $\frac{f^a}{s^a} = \frac{f_{FB}}{s_{FB}}$. Let $\tilde{f}^c$ and $\tilde{s}^c$ denote any equilibrium allocation of corporate travelers (in the fast and slow lane, respectively). Thus, any equilibrium allocation of corporate travelers that replicates the social optimum must satisfy $\frac{\tilde{f}^c}{\tilde{s}^c} = \frac{f_{FB}}{s_{FB}}$. At equilibrium socially optimal allocation, we can rewrite it as $\frac{\tilde{f}^c}{\tilde{s}^c} = \frac{f_{FB}(s_{FB})}{s_{FB}}$, where $f_{FB}(s_{FB})$ is as in (13). Solving it w.r.to $\tilde{f}^c$ gives $\tilde{f}^c(\tilde{s}^c) = \frac{f_{FB}(s_{FB})}{s_{FB}}\tilde{s}^c$. The parameter x which ensures that $f_T^c(s_T^c;x) = \tilde{f}^c(\tilde{s}^c)$ at $\tilde{s}^c = s_T^c$ is obtained by simply solving $f_T^c(s_T^c;x) = \tilde{f}^c(\tilde{s}^c)$ w.r.to x, and it is given by

$$\tilde{x} = \frac{1}{18\mu s_T^c(s_{FB})^2}\Big\{((2\mu+4)s_T^c+6\mu(1-\mu))(s_{FB})^2-10(1-\mu)\left(\frac{3}{10}\mu+s_T^c\right)(s_{FB})^3+$$

$$+(12(s_T^c)^2+(40\mu-16)s_T^c)s_{FB}-24(s_T^c)^2+$$

$$+\left(2(1-\mu)\left(s_T^c-\frac{3}{2}\mu\right)(s_{FB})^2+(8-20\mu)s_T^c s_{FB}+12(s_T^c)^2\right)\times$$

$$\times\sqrt{7(s_{FB})^2-4s_{FB}+4}\Big\}.$$

Since, at the social optimum, it must be the case that $s_T^c = \mu s_{FB}$, this may be used in the above expression to obtain x_{AC} as in (A-23).

Having restored the incentives of the monopolist to the socially optimal degree of differentiation, the subsidy may be determined along the lines of the proof for the case of the market with corporate travelers only to elicit the optimal choice of the socially optimal masses of travelers in the slow and fast lane. This is however outside the scope of this proof. □

Appendix B: numerical simulations

In this Appendix, we sketch the methodology we use for the numerical welfare analysis in Section 5.4. First, we assume

$$B(\theta) = b_0 + b\theta, \tag{B-1}$$

with $b_0 \geq 0$ and $b > 0$. As mentioned, Assumption 1 and 2 become $g < b$ and $b_0 \geq 0$, respectively. Also, Assumption 3 becomes $b \leq b_0$, which we take never to hold to avoid duplicating the analysis contained in Section 4.4. Without further loss of generality, we normalize $b = 1$. As a result, recalling that, by construction, $\mu \in [0, 1]$, the space of parameters of interest of our analysis is $\{\mu, b_0, g\} \in [0, 1] \times [0, 1) \times (0, 1)$.

We use Maple to perform our numerical analysis. We evaluate and compare welfare under a decentralized regime ($\mu = 0$) with welfare under a mixed or a centralized regime ($0 < \mu \leq 1$).

When $\mu = 0$, we know from Proposition 2 that any allocation of travelers such that $s_A = f_A = \frac{1}{2}$ is an equilibrium and that all equilibria are payoff equivalent. Hence, we allocate travelers with $\theta \in \left[0, \frac{1}{2}\right)$ to the slow lane and travelers with $\theta \in \left[\frac{1}{2}, 1\right]$ to the fast lane and plug these values in (23), which now, because of (B-1), takes the following form

$$W_A = \int_0^{\frac{1}{2}} \left(b_0 + b\theta - \frac{1}{2}\theta g\right) d\theta + \int_{\frac{1}{2}}^1 \left(b_0 + b\theta - \frac{1}{2}\theta g\right) d\theta.$$

When $0 < \mu < 1$, we create a grid of parameters combinations, letting μ vary by 0.1 and letting the other parameters vary by 0.05. For each combination of the triplet $\{\mu, b_0, g\}$, we calculate with numerical methods the equilibrium values of s^c_{AC}, f^c_{AC}, s^a_{AC} and f^a_{AC} and then use these values to calculate social welfare. Details of these calculations are given below.

i) When $\mu \in \left(0, \frac{1}{2}\right]$, we know from Propositions 3 and 6 that there is never differentiation across lanes. Full coverage occurs when $g \geq g_{AC}$, where, because of (B-1), $g_{AC} = 2(b - b_0)$.

Full coverage. When $g \geq g_{AC}$, we know from Proposition 3 that an equilibrium is any allocation of travelers such that

$$\bar{s}^c_{AC} + \bar{s}^a_{AC} = \bar{f}^c_{AC} + \bar{f}^a_{AC} = \tfrac{1}{2},$$
$$\bar{s}^c_{AC} + \bar{f}^c_{AC} = \mu,$$
$$\bar{s}^a_{AC} + \bar{f}^a_{AC} = 1 - \mu,$$

and that all these allocations are payoff equivalent. Hence, we allocate an equal mass of both atomistic and corporate travelers to each lane and let (both types of) travelers with $\theta \in \left[0, \frac{1}{2}\right)$ travel in the slow lane and those with $\theta \in \left[\frac{1}{2}, 1\right]$ in the fast lane. We substitute these values in (23), which now, because of full coverage, no differentiation and (B-1), takes the following simple form

$$W_{AC} = \int_0^1 \left(b_0 + b\theta - \frac{1}{2}\theta g\right) d\theta.$$

Partial coverage. We know from Proposition 6 that

$$s^c_{AC} + s^a_{AC} = f^c_{AC} + f^a_{AC} = \tfrac{c_{AC}+1-\mu}{2},$$
$$s^c_{AC} + f^c_{AC} = c_{AC},$$
$$s^a_{AC} + f^a_{AC} = 1 - \mu,$$

We solve numerically for c_{AC} the FOC of the monopolist' problem in (A-7) with $\lambda_{AC} = 0$, which now, because of (B-1), becomes

$$\frac{3g\,(c_{AC})^2}{2\mu} + c_{AC}\frac{g\,(1-2\mu) - b}{\mu} + \frac{2\left(b_0 + b\left(1 - \frac{c_{AC}}{\mu}\right)\right) - g\,(1-\mu)}{2} = 0.$$

Then, we allocate an equal mass of both atomistic and corporate travelers to each lane and let (both types of) travelers with $\theta \in \left[1 - (c_{AC} + (1-\mu)), \frac{1-(c_{AC}+(1-\mu))}{2}\right)$ travel in the slow lane and those with $\theta \in \left[\frac{1-(c_{AC}+(1-\mu))}{2}, 1\right]$ to the fast lane. We use these values in (36), which now, because of partial coverage, no differentiation and (B-1), takes the following form

$$W_{AC} = \mu \int_{1-\frac{c_{AC}}{\mu}}^{1} \left(b_0 + b\theta - \theta g \frac{c_{AC} + 1 - \mu}{2}\right) d\theta + \\ + (1-\mu) \int_0^1 \left(b_0 + b\theta - \theta g \frac{c_{AC} + 1 - \mu}{2}\right) d\theta.$$

ii) When $\mu \in \left(\frac{1}{2}, 1\right)$, we know from Proposition 6 that differentiation may occur depending on the value of $\mu' \in \left(\frac{1}{2}, 1\right)$ such that $1 - \mu = s^c_{AC} - f^c_{AC}$. We assume that differentiation occurs, calculate the equilibrium solutions for the monopolist (see below) and check ex-post whether $1 - \mu < s^c_{AC} - f^c_{AC}$. If this inequality does not hold, no differentiation occurs and we proceed as detailed in the case of $\mu \in \left(0, \frac{1}{2}\right]$. If it does, we proceed as described below.

Full coverage occurs when $g \geq g_{AC}$ as in (30), which, because of (B-1), becomes $g_{AC} = K(\mu)\,(b - b_0)$.

Full coverage. We know from Proposition 3 that in equilibrium $\bar{s}^a_{AC} = 0$, $\bar{f}^a_{AC} = 1 - \mu$, while $\bar{s}^c_{AC}$ and $\bar{f}^c_{AC}$ are given by

$$\bar{s}^c_{AC} = \tfrac{1}{2} + \tfrac{\sqrt{4\mu^2 - 2\mu + 1} - 2(1-\mu)}{6}, \\ \bar{f}^c_{AC} = \tfrac{1}{2} - \tfrac{\sqrt{4\mu^2 - 2\mu + 1} + 4(1-\mu)}{6}.$$

We plug these values into the social welfare function, which now, because of full coverage, differentiation and (B-1), takes the following form

$$W_{AC'} = \mu \left[\int_0^{1-\frac{\bar{f}^c_{AC}}{\mu}} \left(b_0 + b\theta - g\theta \bar{s}_{AC}\right) d\theta + \int_{1-\frac{\bar{f}^c_{AC}}{\mu}}^{1} \left(b_0 + b\theta - g\theta \bar{f}_{AC}\right) d\theta \right] + \\ + (1-\mu) \int_0^1 \left(b_0 + b\theta - g\theta \bar{f}_{AC}\right) d\theta$$

Partial coverage. We know from Proposition 6 that

$$s^a_{AC} = 0, \\ f^a_{AC} = 1 - \mu,$$

We solve numerically for s^c_{AC} and f^c_{AC} the FOCs of the monopolist' problem in (A-4) and (A-5) with $\gamma_{AC} = 0$, which now, because of (B-1), become

$$b_0 + b\left(1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right) - \frac{b\left(s^c_{AC} + f^c_{AC}\right)}{\mu} + \\ + \frac{g s^c_{AC}\left[3 s^c_{AC} + 4 f^c_{AC} - 2\mu\right]}{\mu} = 0, \\ b_0 + b\left(1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right) - \frac{b\left(s^c_{AC} + f^c_{AC}\right)}{\mu} + \\ + \frac{g\left[2\left(s^c_{AC}\right)^2 + f^c_{AC}\left(4(1-\mu) + 3 f^c_{AC} - 2\right) - (1-\mu)\mu\right]}{\mu} = 0.$$

Then, we set $s_{AC} = s^c_{AC}$ and $f_{AC} = f^c_{AC} + f^a_{AC}$. We let corporate travelers with $\theta \in \left[1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}, 1 - \frac{f^c_{AC}}{\mu}\right)$ travel in the slow lane and those with $\theta \in \left[1 - \frac{f^c_{AC}}{\mu}, 1\right]$ travel in the fast lane, together with all atomistic travelers. We plug these values into the social welfare function, which now, because of partial coverage, differentiation and (B-1), takes the following form

$$W_{AC'} = \mu \left[\int_{1-\frac{f^c_{AC} + f^c_{AC}}{\mu}}^{1-\frac{f^c_{AC}}{\mu}} \left(b_0 + b\theta - g\theta s_{AC}\right) d\theta + \int_{1-\frac{f^c_{AC}}{\mu}}^{1} \left(b_0 + b\theta - g\theta f_{AC}\right) d\theta \right] + \\ + (1-\mu) \int_0^1 \left(b_0 + b\theta - g\theta f_{AC}\right) d\theta.$$

iii) When $\mu = 1$, full coverage occurs when $g \geq g_C$ as in (34), that, because of $(B-1)$, becomes $g_C = \frac{6}{4+\sqrt{3}}(b - b_0)$.

Full coverage. We know from Proposition 4 that

$$\bar{s}_C = \tfrac{1}{2} + \tfrac{\sqrt{3}}{6} \cong 0.7887,$$
$$\bar{f}_C = \tfrac{1}{2} - \tfrac{\sqrt{3}}{6} \cong 0.2113.$$

We plug these values into the social welfare function, which now, because of full coverage, differentiation and (B-1), takes the following form

$$W_C = \left[\int_0^{1-\bar{f}_C} (b_0 + b\theta - g\theta\bar{s}_C)\, d\theta + \int_{1-\bar{f}_{AC}}^{1} \left(b_0 + b\theta - g\theta\bar{f}_C\right) d\theta \right].$$

Partial coverage. We solve numerically for s_C and f_C the FOCs of the monopolist' problem in () and (), which now, because of (B-1), become

$$\begin{aligned} & b_0 + b(1 - s_C - f_C) - b(s_C + f_C) + \\ & \qquad - g s_C (2 - 4f_C - 3s_C) = 0, \\ & b_0 + b(1 - s_C - f_C) - b(s_C + f_C) + \\ & \qquad g\left[3(f_C)^2 - 2f_C + 2(s_C)^2\right] = 0. \end{aligned}$$

We let corporate travelers with $\theta \in [1 - s_C - f_C, 1 - f_C)$ travel in the slow lane and those with $\theta \in [1 - f_C, 1]$ travel in the fast lane, together with all atomistic travelers. We plug these values into the social welfare function, which now, because of partial coverage, differentiation and (B-1), takes the following form

$$W_C = \int_{1-s_C-f_C}^{1-f_C} (b_0 + b\theta - g\theta s_C)\, d\theta + \int_{1-s_C+f_C}^{1} (b_0 + b\theta - g\theta f_C)\, d\theta.$$