Transportation Research Part B 183 (2024) 102930

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Congestive mode-switching and economies of scale on a bus route[☆]

Ayush Pandey, Lewis J. Lehe [*]

Urban Traffic and Economics Lab, Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, United States of America[1]

ARTICLE INFO

Keywords:
Economics
Multiple equilibria
Feedback
Bus
Transit
Stability

ABSTRACT

This paper introduces a type of circular causation called Congestive Mode-Switching (CMS) that may arise when an increase in congestion penalizes transit relative to driving. In turn, rising congestion persuades some transit riders to drive, which exacerbates congestion further, and so on. This circular causation can beget multiple equilibria with different levels of congestion and transit ridership. The paper explores this logic with a static model of a bus route. When the bus fleet size is fixed, CMS applies because congestion raises the bus cycle time and thus lowers bus frequency, resulting in higher wait times. When the fleet size depends on bus ridership, CMS is joined by economies of scale as a second source of circular causation. We derive the system's equilibria using a static model in the vein of Walters (1961), which permits us to graphically characterize equilibria in useful ways. The comparative statics of a road improvement show how feedback alters first-order effects. A Downs-Thomson paradox is not possible, because a road improvement aids buses even more than cars. Continuous-time stability analysis shows that multiple equilibria may be stable.

1. Introduction

1.1. Background

Economic models that admit of multiple equilibria can explain many striking irregularities we see in human affairs: e.g., riots (Granovetter, 1978), economic development (Murphy et al., 1989), manufacturing belts (Krugman, 1991), and business cycles (Morris and Shin, 2000). In such models, multiple equilibria usually arise via "circular causation" (Myrdal, 1957) or "positive feedback" (Arthur, 1990): the more people select some option (e.g., opening a firm in a certain city, installing a certain operating system, etc.), the more incentive there is for others to do the same. Krugman (1993, p. 131) explains city formation this way: "Firms that have an incentive to concentrate production at a limited number of locations prefer, other things equal, to choose locations with good access to markets; but access to markets will be good precisely where a large number of firms choose to locate. This positive feedback loop drives the formation of urban centers".

Similarly, formal models with circular causation have also been shown to yield multiple equilibria in transit systems, with Kitamura et al. (1999) and Ying and Yang (2005) as early examples. David and Foucart (2014) points to such a model as an explanation for the wide variation in transit ridership among cities of similar incomes. While details vary, studies generally obtain multiple equilibria under the following premises:

☆ This research was funded by the National Science Foundation, United States grant CMMI-2052512.
* Corresponding author.
E-mail address: lehe@illinois.edu (L.J. Lehe).
1 https://lehelab.com

https://doi.org/10.1016/j.trb.2024.102930
Received 26 June 2023; Received in revised form 10 January 2024; Accepted 23 March 2024
Available online 1 April 2024
0191-2615/

(I) The quality of transit service rises with ridership.
(II) Ridership rises with the quality of transit service.

It is easy to see that, under (I)-(II), a transit system may (though not necessarily) settle at either low quality/low ridership equilibria or high quality/high ridership ones. (II) is straightforward. (I) depends somehow on *economies of scale* in the provision of transit passenger trips. Transit may exhibit economies of scale, most directly, insofar as a transit operator may put more passengers in each vehicle and thereby split the costs of operation over more trips (Jara-Díaz and Gschwender, 2009). If these savings are partly or fully passed onto customers, then the transit fare falls with ridership. But scheduled transit, famously, does not only exhibit such scale economies on the supply-side but also on the *user-side*, too. For example, if a transit operator provides frequency in proportion to ridership, then higher ridership results in lower headways and, thus, lower average passenger wait times or schedule delays—the so-called "Mohring effect" (Mohring, 1972). Alternatively, an operator may run more routes (Kocur and Hendrickson, 1982; Chang and Schonfeld, 1991) or change route structures (Fielbaum et al., 2020) so as to reduce walk times to/from stops. Similarly, "ridepooling" services such as UberPool, Didi Hitch, Via and (until recently) Lyft Shared Rides can combine a larger demand into trips with shorter detours and/or lower fares (Lehe et al., 2021; Liu et al., 2023; Fielbaum et al., 2023).

Transportation scholars have shown special interest in the dynamics of systems obeying (I)-(II): e.g., Cantarella et al. (2015), Li and Yang (2016) and Li et al. (2018) consider models in which travelers adjust their mode choice between car and bus on a day-to-day basis, while bus operators update frequencies periodically. Similarly, Iryo and Watling (2019) consider the dynamics of two-mode systems characterized by a menu of complex interactions among users, including those similar to economies of scale. This emphasis on dynamics grows partly out of concern with the "vicious" and "virtuous" cycles transit systems sometimes exhibit (Krizek and Levinson, 2007; Bar-Yosef et al., 2013).

The logic of (I)-(II) does not only multiply the potential count of equilibria; it also influences these equilibria's comparative statics—that is, how things change in response to a change in some exogenous model parameter or policy. A famous comparative static result[2] for transit systems obeying I-II is the "Downs-Thomson paradox". The "paradox" is that expanding a road that competes with a segregated transit mode (e.g., a rail or bus rapid transit line) can make congestion on the road worse. How this happens exactly is complex, but it has to do with the fact that the road expansion undermines the transit mode's economies of scale.

1.2. Congestive mode-switching

Hereafter, call the logic of (I)-(II) the **Economies of Scale (EOS) Mechanism**. This paper is about a different mechanism of circular causation that can affect mixed-traffic transit (e.g., a bus, ridepool, streetcar, etc.). It depends on these premises:

(A) A driving trip increases congestion more than a transit trip.
(B) An increase in congestion causes some people to switch from transit to driving.

Like (I) and (II), (A) and (B) can feed back on each other, and this feedback may (though not necessarily) yield multiple equilibria. When few people drive, congestion is light, which makes people take transit instead of driving. Conversely, when many people drive, congestion is heavy, which induces would-be transit riders to drive. Note this logic revolves around the amount of *driving*, rather than the amount of *transit ridership*.

Call this mechanism **Congestive Mode-Switching** (CMS). While (A) is straightforward for most transit systems, (B) demands some justification: why would increasing congestion make transit riders become drivers? Obviously, higher congestion generally makes a trip by either car or (mixed-traffic) transit worse. But for (B) to be plausible, the situation must exhibit what we will call a **Differential Penalty**: a reason that congestion penalizes transit *relative to driving*. That is, congestion must not merely render transit less attractive per se, but disadvantage it in a competition against the car. A model with CMS is thus a special case of what the traffic assignment literature calls an *asymmetric problem* (Watling, 1996). It is also, decidedly, not a general phenomenon but rather something that can only happen under particular circumstances sustaining a Differential Penalty. By contrast, in David and Foucart (2014)'s model, buses run in segregated lanes, so rising congestion favors the bus.

Only two studies seem to explicitly derive multiple equilibria from CMS, and each obtains a Differential Penalty in a different way. The first case is a brief[3] numerical example in Daganzo (1983, p. 294). On a congestible highway, a fixed flow of travelers chooses between driving and a bus. The bus frequency is fixed, so traffic speeds change only with the flow of cars. The Differential Penalty enters because Daganzo assumes the bus travels at half the car speed, which makes the absolute travel time difference between bus and car trips rise as speeds fall. The second is Lehe and Pandey (2020). In an isotropic urban zone —*isotropic* in the sense that passenger origins and destinations and traffic conditions are uniform throughout the zone, travelers choose among a ridepool service, driving and an outside option. A ridepool trip affects congestion less than driving but includes detours, which occasion the Differential Penalty: the lower the speed of traffic, the longer the detour takes, and the longer the difference in travel time between pooling and driving.

[2] See Arnott and Small (1994), Mogridge (1997) and Zhang et al. (2014, 2016) and many others studies which have explored different aspects of the paradox formally.

[3] Watling (1996) cleans up the presentation of the same example and explores it further.

1.3. Plan of paper

Neither study where CMS is explicit permits economies of scale: Daganzo (1983) assumes a fixed bus frequency, and Lehe and Pandey (2020) assumes the detour length is fixed. However, the two mechanisms are not exclusive. For example, both can be discerned working together in the history of the decline of the United States' streetcar systems (Cudahy, 1990): as automobiles took customers from streetcars, cash-starved operators cut service and maintenance; and as increasing auto traffic blocked streetcars, the travel time advantage of the auto (which can maneuver through congestion more nimbly) over the streetcar grew. While not explicit about distinguishing them, the mixed-traffic, car/bus choice model of Cantarella et al. (2015) would seem to exhibit both mechanisms: bus speed is a fraction of car speed as in Daganzo (1983), and bus frequency rises with ridership.

This paper's broad aims are (i) to draw attention to CMS as a source of novel system-level outcomes, (ii) to explicitly distinguish between the CMS and EOS mechanisms and (iii) to explore how the two mechanisms may interact. To these ends, it explores a stylized, but plausible, model of a mixed-traffic bus route. Section 2 establishes the model's setting, demands, travel times and congestion technology. Section 3 then derives equilibria when the bus fleet size is taken to be fixed, which leads to a "pure" CMS mechanism. Section 4 derives equilibria under the assumption that the bus fleet size depends on ridership, such that the CMS and EOS mechanisms both arise and interact. Section 5 concludes with ideas for future research.

Much of Sections 3 and 4 is devoted to characterizing equilibria. CMS and the EOS mechanism not only yield multiple equilibria but also qualitatively different equilibria. These differences come into relief in two ways. Firstly, as regards stability: we judge the local stability of equilibria under adjustment dynamics given as autonomous systems of differential equations; Section 3 system has one state variable (demand for driving), and Section 4 system has three (the demands for driving and bus and the bus fleet size). Secondly, as regards comparative statics: Sections 3 and 4 both consider (as in the Downs-Thomson example) the effect of a road improvement on equilibrium values of various travel statistics.

In addition to its broad intent, the paper is also distinguished from other studies in the literature in technical ways:

- The total demand for travel (the sum of demands across car and bus) is endogenous. By contrast, the car/bus choice models cited above divide an exogenous demand between the modes. Endogenizing total demand has three benefits. First, it is realistic. Second, it shows that CMS does not depend on one-to-one switching between transit and driving; traffic flow and congestion may rise even when the total flow of person-trips falls. Third, it distinguishes CMS from the EOS mechanism: when total demand is exogenous, lower bus ridership mechanically implies more car traffic (and vice versa).
- The paper obtains a Differential Penalty in an original way: as congestion rises, each bus takes longer to complete the route, which reduces the bus frequency and raises wait times for bus passengers. The Differential Penalty does not obtain by a gap between the modes' in-vehicle times that grows with congestion (as it does in Daganzo (1983)), because we assume buses move at the same speed as cars but lose a fixed amount of time per distance-unit to stops.
- The analysis hews to the classic "static traffic model" framework[4] introduced by Walters (1961). Erstwhile "supply" and "demand" curves are plotted in a space with traffic flow on the horizontal axis and a travel time index on the vertical axis. This approach aids in characterizing equilibria, and links the results to classical transportation economics.

2. Setup

2.1. Setting

The setting is an isotropic bus route. The route is a closed ring where upon completing the route buses start over. Buses and cars both use the same roadspace, so both affect and are affected by congestion. The units employed are abstract: units of time are *tu*'s, units of distance *du*'s, and units of passengers *pax*. *pcu* stands for "passenger car unit" and appears in flow metrics. Table 1 provides a glossary of notation.

2.2. Physics

Let u (tu/du) be the *unit travel time* (inverse speed) of cars. Buses move with a unit travel time of $u + \eta$ (tu/du), where $\eta > 0$ (tu/du) accounts[5] for time lost (per du) to stops or some other reason (e.g., that buses take longer to accelerate from a stop at red lights). η is exogenous (although a richer model could have it depend on passenger demands via boarding/alighting times). The *route length* is R (du), so $R(u + \eta)$ (tu) is the *cycle time* (how long it takes one bus to pass the same point on the route twice). Q (pcu/tu) is a *traffic flow* metric. Let Q_c (pcu/tu) be the flow of passenger cars, which each count as one pcu. Let B (bus) be the bus *fleet size*. Given a fleet size B, the "raw" flow of buses is $B/R(u + \eta)$ (bus/tu). κ (pcu/bus) is the weight assigned to buses in the traffic flow metric, so the weighted bus flow is

$$Q_b := \kappa B/R(u + \eta) \quad \text{(pcu/tu)}. \tag{1}$$

[4] See Verhoef (1999) for a technical analysis of the basis for these models, and De Palma and Fosgerau (2010) for a contrast between dynamic and static economic models of congestion.

[5] By contrast, in Daganzo (1983) the bus moves at a fixed fraction of the speed of cars, which implies its unit travel time is a greater-than-one multiple of the car unit travel time. While there is some evidence to support such proportionality (e.g., Levinson, 1983), other studies show a smaller-than-one relationship (McKnight et al., 2004), and others contain mixed results (Kieu et al., 2015).

Table 1
Notation.

Variable glossary			Function glossary		
Variable	Units	Meaning	Function	Units	Meaning
Q_b	pcu/tu	Bus flow	$\mu(Q)$	tu/du	Congestion technology
Q_c	pcu/tu	Car flow	$\hat{\mu}(Q_c, B)$	tu/du	Congestion technology
Q	pcu/tu	Traffic flow	$T_b(u, B)$	tu	Bus travel time
κ	pcu/bus	Weight of bus	$T_c(u)$	tu	Car travel time
u	tu/du	Car unit travel time	$W_b(t, B)$	tu	Bus wait time
η	tu/du	Extra bus unit travel time	$\rho_b(u, B)$	tu	Bus penalty
T_b	tu	Bus travel time	$X_b(T_b, \rho_b)$	pax/tu	Bus demand function
T_c	tu	Car travel time	$\chi_b(u, B)$	pax/tu	Bus demand function
ρ_b	tu	Bus penalty	$X_c(T_c, \rho_b)$	pax/tu	Bus demand function
W_b	tu	Bus wait time	$\chi_c(u, B)$	pax/tu	Car demand function
R	du	Route length	$\Theta(\cdot)$	pax/tu^2	Bus demand adjustment function
L	du	Trip length	$\Psi(\cdot)$	bus/tu	Bus fleet adjustment function
B	bus	Fleet size	$\Omega(\cdot)$	pax/tu^2	Car demand adjustment function
x_b	pax/tu	Bus demand	$\beta(x_b)$	bus	Bus provision function
x_c	pax/tu	Car demand			

Total traffic flow is the sum of the car and (weighted) bus flows:

$$Q := Q_c + Q_b = Q_c + \kappa B/R(u+\eta) \quad \text{(pcu/tu)}. \tag{2}$$

The *congestion technology function* $u = \mu(Q)$ (tu/du) relates (weighted) flows[6] to unit travel times. It has $\mu' > 0, \mu(0) > 0$. Replacing u with $\mu(Q)$ in (2) yields:

$$Q = \frac{\kappa B}{R[\mu(Q)+\eta]} + Q_c \quad \text{(pcu/tu)}. \tag{3}$$

Fixing Q_c and B, as Q rises from 0 the RHS falls monotonically from some finite value. Hence, a car flow/fleet size pair (Q_c, B) defines a unique total flow Q and unique $u = \mu(Q)$. Implicitly define the injection $u = \hat{\mu}(Q_c, B)$ by

$$\hat{\mu}(Q_c, B) := \mu\left(\frac{\kappa B}{R\left[\hat{\mu}(Q_c, B)+\eta\right]} + Q_c\right). \tag{4}$$

We now give some partial derivatives useful later. First, using (1), we obtain

$$\left(\frac{\partial Q_b}{\partial u}\right)_B = -\frac{\kappa B}{R(u+\eta)^2} < 0 \qquad \left(\frac{\partial Q_b}{\partial B}\right)_u = \frac{\kappa}{R(u+\eta)} > 0. \tag{5}$$

Let

$$\phi := \left[1 - \mu'(Q)\left(\frac{\partial Q_b}{\partial u}\right)_B\right]^{-1} \in (0,1). \tag{6}$$

ϕ is a multiplier that captures a mechanical dampening effect. It must be between 0 and 1, because $\mu' > 0$ and $\left(\partial Q_b/\partial u\right)_B < 0$. Any first-round increase in (weighted) vehicle flow raises the unit travel time (decreases the speed), which in turn lowers the flow of buses. It appears in the following partial derivatives:

$$\left(\frac{\partial Q}{\partial Q_c}\right)_B = \phi \in (0,1) \qquad \left(\frac{\partial Q}{\partial B}\right)_{Q_c} = \phi\left(\frac{\partial Q_b}{\partial B}\right)_u > 0 \tag{7}$$

$$\frac{\partial \hat{\mu}}{\partial Q_c} = \phi\mu' > 0 \qquad \frac{\partial \hat{\mu}}{\partial B} = \phi\mu'\left(\frac{\partial Q_b}{\partial B}\right)_u > 0. \tag{8}$$

Note that in giving partial derivatives of functions with defined arguments, such as $\hat{\mu}(Q_c, B)$, we do not put the subscript of the fixed argument.

2.3. Travel times

The trip length is fixed at L (du) on either mode. Hence,

$$T_c(u) = Lu \quad \text{(tu)} \tag{9}$$

[6] Thus, the route displays *flow congestion* rather than the sort of *density congestion* used in traffic flow theory, and it cannot become "hypercongested". See Zhang et al. (2018) and Gonzales (2015) for examples of bus/car models with hypercongestion.

gives the car travel time given u. A bus trip involves an in-vehicle time of $L(u+\eta)$ (tu), but also wait time at a bus stop. Assume passengers show up at stops randomly, so the average wait time experienced is one half of one headway[7]:

$$W_b(u,B) := R(u+\eta)/2B. \tag{10}$$

Therefore, the bus travel time is given by

$$T_b(u,B) := W_b(u,B) + L(u+\eta) = (R/2B + L)(u+\eta) \quad \text{(tu)}. \tag{11}$$

The *bus penalty* $\rho_b = T_b - T_c$ (tu) is the extra travel time incurred by choosing bus over car. Write it as a function:

$$\rho_b(u,B) := T_b(u,B) - T_c(u) = W_b(u,B) + L\eta \quad \text{(tu)}. \tag{12}$$

$L\eta$ is the in-vehicle time lost to stops per trip. Since L and η are fixed, the bus penalty only varies by changes in wait time. Note the partial derivatives:

$$\partial T_c/\partial u = L \qquad \partial T_b/\partial u = R/2B + L \tag{13}$$

$$\partial \rho_b/\partial u = \partial W_b/\partial u = R/2B \qquad \partial \rho_b/\partial B = \partial T_b/\partial B = \partial W_b/\partial B = -R(u+\eta)/2B^2 < 0. \tag{14}$$

The fact $\partial \rho_b/\partial u = \partial W_b/\partial u > 0$ conveys the *Differential Penalty*. As congestion rises, frequency falls (because the cycle time rises), so the wait time rises.

2.4. Passenger demands

Let x_b (pax/tu) and x_c (pax/tu) be *demands* for bus and car, respectively. But x_c winds up also being the flow of cars, too, because we assume

Assumption 1 (*One Pax Per Car*)**.** Every car trip takes place in one car: i.e., $Q_c = x_c$.

This assumption reflects a choice of units, where one pax can be thought of as the average number of people in a car trip (so long as car occupancy is exogenous).

In equilibrium, x_i is given by the *demand function* $X_i(T_i, \rho_b)$ (pax/tu) for $i = b, c$. These obey:

Assumption 2 (*Demand Assumptions*)**.** The demand functions have the properties

(a) Both demands are strictly positive for finite travel times: $X_c(T_c,\rho_b), X_b(T_b,\rho_b) > 0 \;\forall\; T_c, T_b, \rho_b < \infty$
(b) Both demands are bounded from above.
(c) Each mode's demand falls with its own travel time: $\partial X_c/\partial T_c < 0$ and $\partial X_b/\partial T_b < 0$.
(d) Passengers "switch" from bus to car when the bus penalty rises and from car to bus when the bus penalty falls: $\partial X_c/\partial \rho_b = -\partial X_b/\partial \rho_b > 0$.

We will now comment on the simplifications underlying this assumption:

- Assumption 2(a) precludes having to tediously qualify some statements and worry about limits.
- Assumption 2(b) is sensible, as there cannot be infinite demands.
- In Assumption 2(c) (and implicitly in (d)), the appearance of T_b suggests bus passengers do not distinguish between wait and in-vehicle time. But since the wait time is $W_b = Ru/2B$, the route length R can be interpreted as being magnified in such a way as to account for passengers caring more about wait than in-vehicle time.
- Assumption 2(d) suggests travelers care about the *absolute* difference in travel times. Alternatively, we might weight the two travel times differently by writing:

$$\rho_b(u,B) = \alpha_b T_b(u,B) - \alpha_c T_c(u), \tag{15}$$

where α_c and α_b reflect different sensitivity parameters (such as values of travel time savings) for car and bus, respectively. What matters to our results is that $\partial \rho_b/\partial u > 0$ (that the relative utility of bus declines with congestion). Differentiating (15) yields

$$\partial \rho_b/\partial u = L\left(\alpha_b - \alpha_c\right) + \alpha_b R/2B, \tag{16}$$

which is positive as long as

$$\alpha_c/\alpha_b < 1 + R/2BL. \tag{17}$$

Daganzo (1983), Li et al. (2012), David and Foucart (2014), Cantarella et al. (2015) and Li et al. (2018) all feature a single value of time (i.e., $\alpha_c/\alpha_b = 1$), whereby (17) holds. Likewise, if $\alpha_c/\alpha_b < 1$ (as in Model I of Tirachini et al. (2014)) then

[7] The results in our analysis do not depend on average wait time being half of the headway. The results hold as long as wait time increases with headway. See Ansari Esfeh et al. (2021) for a detailed discussion of relationship between headway and wait time.

$\partial \rho_b/\partial u > 0$. Even if $\alpha_c/\alpha_b > 1$ (if travelers are more sensitive to car in-vehicle time), then (17) *may* be satisfied, depending on the size of $R/2BL$. This is the case in Kitamura et al. (1999, Tab. 1), where the generalized cost coefficient of bus in-vehicle time is half that of car.

The model aims to demonstrate CMS and interesting consequences, so we assume absolute differences matter because it is simple and yields CMS. Research suggests that $\alpha_c/\alpha_b \geq 1$ is at least plausible. In a meta-analysis, Wardman et al. (2016, Table 8) finds statistically insignificant differences in the values of in-vehicle travel times for car and bus. Guevara (2017) notes only a minority of studies find higher values of in-vehicle time for car than for bus, and presents a clever explanation for why such measurements may arise even when people dislike spending time on the bus more than in a car.

- The bus fare is absent. Maybe the route is only one of many in a large network with uniform fare, or (as with most US agencies) only a small share of funding comes from fares. We have left fare and other such adjustments for future work because scaling the fleet size is simple and suffices to show interesting interactions between the CMS and EOS mechanisms.

3. Model with exogenous fleet size

3.1. Introduction

Section 2 gave all the pieces of a model except for how the bus fleet size B was determined. In this section, we analyze the model under the assumption

Assumption 3 (*Exogenous Fleet Size*)**.** The fleet size is fixed at $B = \bar{B}$.

This leads to CMS while excising scale economies. The assumption is helpful to exposition, since it lets CMS be analyzed in isolation, but it also worth exploring on its own. Due to administrative, financial or political constraints, an agency's fleet size may not rise and fall with ridership. This section will show that premise I (that quality rises with ridership) is not necessary for a bus system to exhibit multiple stable equilibria, nor for policy changes to be amplified by a form of positive feedback.

3.2. Equilibria

We derive the model's equilibria in the context of a static traffic model. Let

$$\chi_b(u) = X_b\left[T_b(u,\bar{B}), \rho_b(u,\bar{B})\right] \qquad \chi_c(u) = X_b\left[T_b(u,\bar{B}), \rho_b(u,\bar{B})\right] \tag{18}$$

be the demand curves for bus and car, respectively. Fig. 1(a) draws examples from a simulation described in the appendix. Note $\chi_b(u)$ falls monotonically but $\chi_c(u)$ rises and falls along different portions: i.e., a rise in congestion always decreases bus ridership but may either increase or decrease demand for driving. To see why, inspect their derivatives:

$$\chi_b' = \underbrace{\frac{\partial X_b}{\partial T_b}\frac{\partial T_b}{\partial u}}_{(<0)\,\text{cancelled trips}} + \underbrace{\frac{\partial X_b}{\partial \rho_b}\frac{\partial W_b}{\partial u}}_{(<0)\,\text{bus}\to\text{car switching}} < 0 \qquad \chi_c' = \underbrace{\frac{\partial X_c}{\partial T_c}\frac{\partial T_c}{\partial u}}_{(<0)\,\text{cancelled trips}} + \underbrace{\frac{\partial X_c}{\partial \rho_b}\frac{\partial W_b}{\partial u}}_{(>0)\,\text{bus}\to\text{car switching}} \tag{19}$$

The first terms of both derivatives are the rate travelers cancel trips. The second terms reflect switching from bus to car as u rises due to rising wait times; hence it is negative for bus demand but positive for car demand. Thus, bus ridership is always falling with u but when there is enough switching car demand may be rising: i.e., $\chi_c' > 0$. Still, total demand falls with u:

$$\chi_b' + \chi_c' = \frac{\partial X_c}{\partial T_c}\frac{\partial T_c}{\partial u} + \frac{\partial X_b}{\partial T_b}\frac{\partial T_b}{\partial u} < 0. \tag{20}$$

Next, recalling $x_c = Q_c$ per Assumption 1, let

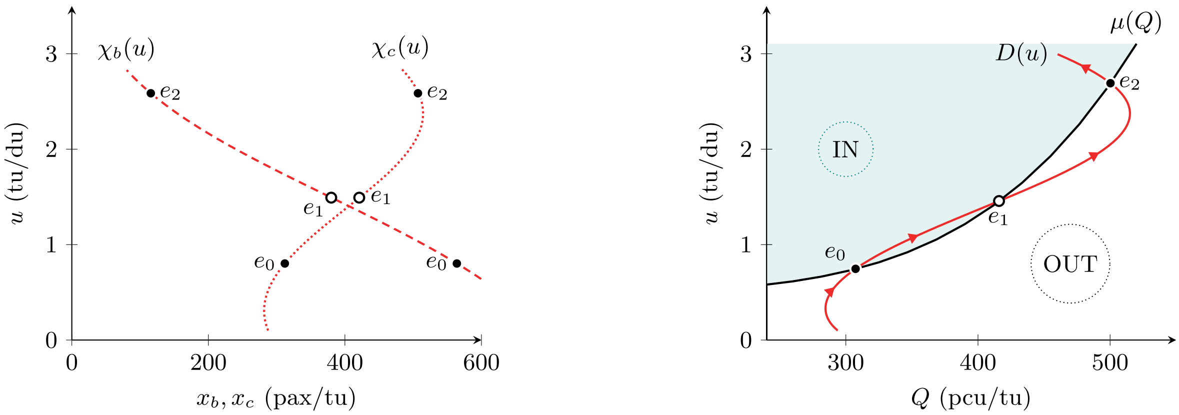
$$D(u) := \underbrace{\frac{\kappa \bar{B}}{R(u+\eta)}}_{Q_b} + \chi_c(u) \quad \text{(pcu/tu)} \tag{21}$$

give the *traffic flow demanded*. Fig. 1(b) shows an example. Like χ_c, D may rise or fall over different intervals of u. Its derivative is

$$D'(u) = \left(\partial Q_b/\partial u\right)_B + \chi_c'(u), \tag{22}$$

where the first term is negative and conveys a sort of "mechanical effect", whereby rising u reduces the flow of buses (since each takes longer to complete the route). When χ_c' is positive and sufficiently large (due to mode-switching) as to overcome the mechanical effect, then D rises with u. Factors that tend to increase χ_c' are the following:

- when drivers are less able to cancel trips (i.e., when $\partial X_c/\partial T_c$ is small);
- when bus riders more willing to switch to driving (i.e., when $\partial X_c/\partial \rho_b$ is larger);
- when congestion has a larger effect on wait times (i.e., when $\partial W_b/\partial u$ is larger).



(a) Demand curves for bus and car under Ass. 3

(b) Equilibria under Ass. 3, with "In" and "Out" regions marked

Fig. 1. Demand curves for exogenous B.

Regarding the last effect, recall that $\partial W_b/\partial u = R/2B$. Thus, congestion has a larger effect on wait times when the density of buses B/R (bus/du) is smaller (and hence the spacing between them R/B is larger).

Fig. 1(b) also shows a $u = \mu(Q)$, which plays the role of a "supply" curve. The system is in equilibrium at intersections of D and μ. An equilibrium is a point (Q^*, u^*) where $D(u^*) = Q^*$ and $\mu(Q^*) = u^*$. There may be multiple equilibria if D has a rising portion. Fig. 1(a) shows the demands associated with each equilibrium. Those with higher u, such as e_2, have higher car flow and lower bus ridership than those such as e_1 and e_2. Even though people's preferences are the same at all three equilibria, and even though there are the same number of buses at all three, the differences in bus ridership and driving are sustained at equilibria with higher u because wait times are much longer.

Observe a sort of "optical illusion" in Fig. 1(b). e_0 exhibits the highest flow of travelers (i.e., the most trips are accomplished per unit time) and e_2 the least. This must be so, since total demand $x_b + x_c$ falls with u, and u is higher at e_2 than at e_0. Despite the fact that the flow of travelers is lowest at e_2, e_2 has the highest rate of vehicle flow, because more trips are taken by car and the buses are largely empty.

To characterize equilibria for stability and comparative static analyses, divide (Q, u) space into two regions:

$$\text{In} := \{(Q,u) \quad | \quad u > \mu(Q)\}, \qquad \text{Out} := \{(Q,u) \quad | \quad u < \mu(Q)\}. \tag{23}$$

IN and OUT are labeled in Fig. 1(b), where IN is shaded. Equilibria lie on the boundary between IN and OUT, so we can distinguish them by which way D crosses μ as u rises. The arrows in Fig. 1(b) show the "direction" $D(u)$ moves in as u increases. This is equivalent to saying that:

Definition. An equilibrium with $(Q, u) = (Q^*, u^*)$ is **Outside-In** if $D'(u^*)\mu'(Q^*) < 1$ and **Inside-Out** if $D'(u^*)\mu'(Q^*) > 1$.

So in Fig. 1(b), e_0 and e_2 are Outside-In and e_2 is Inside-Out.

3.3. Local stability

We have seen there may be multiple equilibria when a rising D meets a rising μ. Similar cases in economics, such as models of fads (Becker, 1991), have raised questions about the local stability of such equilibria—that is, whether, following a small perturbation from an equilibrium, the system will return to that equilibrium. For example, Plott and Smith (1999) conducts laboratory experiments designed to create upward-sloping market demand and supply curves, to reveal which equilibria are stable and what dynamics apparently drive people's choices.

Stability analysis is important because it determines whether multiple equilibria could really be expected, and because the facts that shape an equilibrium's stability are often linked to its comparative statics. Samuelson (1941, p. 102) puts things this way: "We find ourselves confronted with this paradox: in order for the comparative statics analysis to be fruitful, we must first develop a theory of dynamics". So below, when we look at the comparative statics of road improvement, we only consider equilibria found to be stable.

The analysis here and in Section 4 proceeds under simple, continuous-time dynamics. Suppose x_c is a state variable which evolves by the rule[8]

$$\dot{x}_c = \Omega\{\chi_c(u) - x_c\}, \tag{24}$$

[8] These dynamics differ essentially from continuous-time dynamics popular in the traffic assignment literature, such as "Smith dynamics" (Smith, 1984), "network tâtonnement" (Friesz et al., 1994) and "replicator dynamics" (Iryo, 2019; Lehe and Pandey, 2024), which are all written in terms of utilities or

where $\Omega(\cdot)$ is some continuously-differentiable function such that $\Omega' > 0, \Omega(0) = 0$. This rule is a form of what game theorists (Gilboa and Matsui, 1991; Hofbauer and Sandholm, 2011) call "best-response" dynamics. The idea is that χ_c reflects the rational decisions people would make based on their innate preferences, but that at any particular time some people might not be informed, or that people need time to adjust their choices (e.g., to buy/sell a car, learn how the bus works, etc.). So if $\chi_c(u) > x_c$, then there must exist people who would like to drive but who are not doing so currently. Over time these people gradually "catch on", so car flow should rise. Conversely, if $\chi_c(u) < x_c$ then there are people currently driving who would rather not do so, and so car flow should be falling.

Next, suppose that $u = \hat{\mu}(x_c, \bar{B})$ at all times: i.e., u is not an evolving state variable but adjusts quickly to traffic conditions (or at least so much faster than demands do than it can be treated as adjusting instantaneously). Plugging $u = \mu(Q)$ into (24) yields the *autonomous system*

$$\dot{x}_c = \Omega\left\{\chi_c\left[\hat{\mu}(x_c, \bar{B})\right] - x_c\right\}. \tag{25}$$

Let $x_c = x_c^*$ be a fixed point of (25), such that $x_c^* = \chi_c\left[\hat{\mu}(x_c^*, \bar{B})\right]$ and thus $\dot{x}_c = \Omega(0) = 0$. Let

$$\delta := x_c - x_c^* \tag{26}$$

be a small perturbation away from x_c^*. Let $\omega := \Omega'(0) > 0$ be the "speed of adjustment" of x_c at a fixed point. By a Taylor Expansion,

$$\dot{\delta} = \Omega(0) + \delta\omega a + \mathcal{O}\left(\delta^2\right) \tag{27}$$

$$\text{where } a := \frac{\partial}{\partial x_c}\left[\chi_c\left\{\hat{\mu}(x_c, \bar{B})\right\} - x_c\right] = \frac{\partial \hat{\mu}}{\partial Q_c}\chi_c' - 1 \tag{28}$$

Since $\Omega(0) = 0$ and $\mathcal{O}(\delta^2)$ can be ignored when δ is small enough, for small δ we have the first order ODE $\dot{\delta} \approx \delta\omega a$. Letting $\delta(t)$ denote the value of δ at clock time t, the solution of this ODE is

$$\delta(t) = \delta(0) \cdot e^{\omega a t}\delta(t) = \delta(0) \cdot e^{\omega a t}. \tag{29}$$

Thus, $\delta \to 0$ over time if and only if $a < 0$.

Next, recalling $\partial\hat{\mu}/\partial Q_c = \phi\mu'$, write a as

$$\begin{aligned} a &= \phi\mu'\chi_c' - 1 \\ &= \phi\left[\mu'\chi_c' - 1 + \mu'\left(\frac{\partial Q_b}{\partial u}\right)_B\right] \quad \text{since } \phi = \left[1 - \mu'\left(\partial Q_b/\partial u\right)_B\right]^{-1} \\ &= \phi\left(D'\mu' - 1\right) \quad \text{by (22).} \end{aligned} \tag{30}$$

It follows that

Proposition 1. *Under the dynamics given by (24), Inside-Out ($D'\mu' > 1$) equilibria are unstable and Outside-In ($D'\mu' < 1$) equilibria are stable.*

Thus, in Fig. 1(b), e_0 and e_2 are stable but e_1 is not.

3.4. Road improvement

Consider now a generic "road improvement", specified as follows. Let ζ be a generic index of the road's efficiency. Higher ζ might follow from adding turn lanes, allocating more green time to traffic signals along the route or synchronizing signals. Hence $\partial\mu/\partial\zeta < 0$: i.e., a higher ζ means the road delivers some volume of traffic at higher speed. Writing μ with ζ as an argument, Fig. 2 illustrates an increase from ζ_0 to ζ_1, which shifts μ downward in (Q, u) space. The figure also shows the stable equilibria e_0 and e_2, which move to new places denoted by primes.

How does the road improvement change things? To answer, start by differentiating

$$\frac{\mathrm{d}u^*}{\mathrm{d}\zeta} = \frac{\mathrm{d}}{\mathrm{d}\zeta}\mu[D(u^*); \zeta] = \mu' D' \frac{\mathrm{d}u^*}{\mathrm{d}\zeta} + \frac{\partial\mu}{\partial\zeta} = \frac{\partial\mu/\partial\zeta}{1 - D'\mu'}. \tag{31}$$

Since $\partial\mu/\partial\zeta < 0$, it follows that u^* falls at stable equilibria where $D'\mu' < 1$. A road improvement relieves congestion at stable equilibria. So then do all the travel time measures T_b, T_c and W_b as well as the bus penalty ρ_b. Bus ridership rises consequently. Car demand x_c can either rise or fall—depending on the sign of χ_c' at the equilibrium. So can traffic flow Q—depending on the sign of D'.

More abstractly, stable equilibria differ by *the type of feedback* they exhibit in response to the road improvement. One may think of $\partial\mu/\partial\zeta$ as a "first-round effect" or "input" to the system: the fall in u we would observe if the road improvement did not change

generalized costs. We cannot use those easily, because this paper's specification of demand does not (by design) specify utilities/costs—only aggregate features of the demand functions X_b and X_c. There may be multiple groups of travelers (as in the simulations described in the Appendix). Or travelers might differ in their values of travel time savings, have non-linear values of travel time savings, etc.

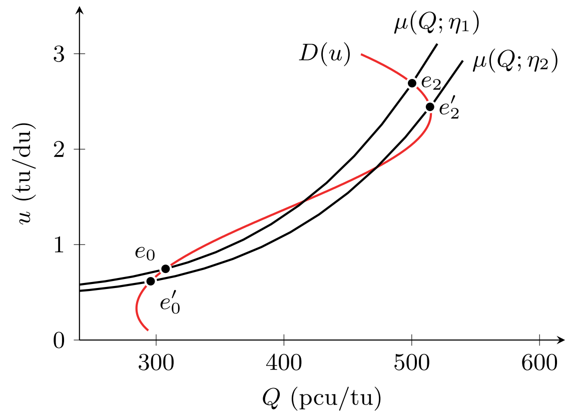


Fig. 2. Shift of congestion technology function and movement of stable equilibria.

traffic flow Q. This first-round effect is divided by $1 - D'\mu'$ to obtain a "final" effect or "output" $\mathrm{d}u^*/\mathrm{d}\zeta$. Equilibria like e_2 with $D' < 0$ exhibit *negative* feedback: $D' < 0$ implies the factor $1/(1 - D'\mu')$ is between 0 and 1, so the first-round impact $\partial\mu/\partial\zeta$ is *dampened*. By contrast, equilibria like e_0 with $D' > 0$ exhibit *positive* feedback, because $D' > 0$ and $1 > D'\mu'$ imply $1/(1 - D'\mu') > 1$. So the first-round fall in u and all downstream impacts on other variables are *amplified*.

Taking a different point of view, it may be useful to exclude the "mechanical effect" $\left(\partial Q_b/\partial u\right)_B$ from thinking about feedback: i.e., to only consider feedback from changes in demand. By this way of thinking, the change in flow from bus cycle times changing is "baked in". To do so, rewrite (31) as

$$\frac{\mathrm{d}u^*}{\mathrm{d}\zeta} = \frac{\phi\,\partial\mu/\partial\zeta}{\phi - D'\mu'\phi} = \frac{\phi\,\partial\mu/\partial\zeta}{1 - \chi_c'\,\partial\hat{\mu}/\partial Q_c}. \tag{32}$$

Now think of $\phi\,\partial\mu/\partial\zeta$ as the "first-round" effect: the decline in u we would observe if car flow were kept constant but Q_b were permitted to change via the mechanical effect. If $\chi_c' < 0$, then the denominator is larger than one, and the first-round effect is muted. There is negative feedback due to *induced demand* (Hymel et al., 2010), whereby the road improvement is attenuated by the car traffic it invites. On the other hand, if $\chi_c' > 0$, then the denominator is between 0 and 1 at stable equilibria. In this case, we have positive feedback via what might be called *reduced demand*: the road improvement reduces congestion, which (due to CMS) causes enough people to switch from car to bus that car traffic falls, which cut congestion further, and so on. An equilibrium like e_0 with declining D has rising χ_c, so it displays positive feedback and amplification from either perspective.

4. Model with endogenous fleet size

4.1. Introduction

In this section the fleet size is not fixed. Instead, to incorporate the EOS mechanism, the fleet size depends on bus ridership via the assumption

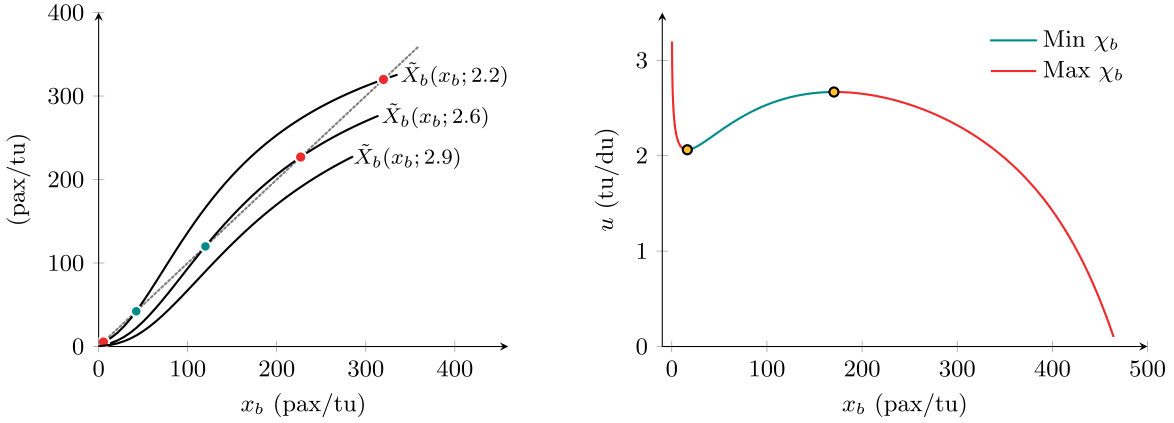
Assumption 4 (*Endogenous Fleet Size*)**.** The equilibrium fleet size is given by the function $B = \beta(x_b)$, with the following properties:

(a) $\beta'(x_b) > 0$ (b) $\beta(0) > 0$ (c) $\beta'(x_b) < R[\mu(0) + \eta]/\kappa$

(a) means the fleet size rises with ridership. Think of a budget-constrained system that pays for the marginal bus-hour with marginal fares and/or receives a subsidy proportionate to ridership: the more money comes in, the more buses they run. This back story is consistent with the historical analysis in Savage (2004), who argues that service levels, per se, have sometimes been a transit management objective. While plausible, this rule for bus provision does not seem to be found in other[9] studies. This rule yields a Differential Penalty in the sense that, for a given ridership, a rise in u raises the wait time.

Property (b) is made for analytical convenience. As for (c), $\mu(0)$ is the lowest possible u, so $\left(\partial Q_b/\partial B\right)_u \beta' < 1$ for all observed u. Hence, switching a passenger from car to bus cuts the (weighted) traffic flow, guaranteeing premise (A) from Section 1.2.

[9] Daganzo (1983) has frequency fixed. Bar-Yosef et al. (2013), Li and Yang (2016) and Li et al. (2018) have bus speeds fixed, so the choice of fleet size and frequency are equivalent. David and Foucart (2014) has bus speeds vary (their buses run in mixed traffic along at least part of the route) but makes bus frequency (rather than the fleet size) increasing with ridership. In Scenarios A and B of Cantarella et al. (2015) (where the fleet size is not fixed), the operator targets a certain bus load, which makes frequency proportionate to ridership and invariant to the cycle time.



(a) Partial equilibria obtained for discrete values of u

(b) $\tilde{\chi}_b(u)$ with three branches

Fig. 3. Partial equilibria from simulation with linear demand.

4.2. Partial equilibrium

Equilibrium under Assumption 4 involves a balance of forces on two sides: between demand and congestion and between bus ridership and the fleet size. Hence we solve for equilibrium in two stages.

To begin, rewrite the bus demand function:

$$\chi_b(u, B) = X_b\left[T_b(u, B), \rho_b(u, B)\right] \tag{33}$$

with

$$\frac{\partial \chi_b}{\partial u} = \frac{\partial X_b}{\partial u}\frac{\partial T_b}{\partial u} + \frac{\partial X_b}{\partial \rho_b}\frac{\partial W_b}{\partial u} < 0 \qquad \frac{\partial \chi_b}{\partial B} = \frac{\partial W_b}{\partial B}\frac{\partial X_b}{\partial T_b} + \frac{\partial X_b}{\partial \rho_b} > 0. \tag{34}$$

So bus demand falls with u and rises with B. Since B in turn depends on x_b, what we will call a **partial equilibrium** arises *for given* u from the interaction of the bus ridership and fleet size. A partial equilibrium ridership $x_b = \tilde{x}_b$ satisfies

$$\tilde{x}_b := \chi_b\left[u, \beta(\tilde{x}_b)\right] \tag{35}$$

Fig. 3(a) illustrates by plotting the function $f(x_b; u) := \chi_b\left[u, \beta(x_b)\right]$. $f(x_b; u)$ and the identity line meet at partial equilibria. Fig. 3(a) and subsequent figures are produced by a nested logit simulation described in Appendix A. Under our assumptions, $f(x_b; u)$ rises with x_b (because the fleet size does) and has a positive intercept (because there are always *some* buses on the route, and because demand is positive for finite travel times). Per (c) of Assumption 2, $f(x_b; u)$ is bounded above so as x_b rises f eventually meets the identity line as x_b rises. Thus, any finite u yields at least one, strictly-positive partial equilibrium ridership.

As u rises, $f(x_b; u)$ shifts downwards and the partial equilibria move. By tracing their locations as u varies, we obtain a *multi-valued* function

$$\tilde{\chi}_b(u) := \langle \tilde{x}_b \mid \tilde{x}_b = \chi_b[u, \beta(\tilde{x}_b)] \rangle \quad \text{(pax/tu)}, \tag{36}$$

which returns a *set* of partial equilibrium demands. (Such a relation is sometimes called a "demand correspondence".)

Fig. 3(b) illustrates. In the figure's example, $\tilde{\chi}_b$ has three branches colored to match corresponding equilibria from Fig. 3(a). The teal branch rises, while the red branches fall. To see why, differentiate

$$\frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} = \frac{\partial \chi_b}{\partial u} + \beta' \frac{\partial \chi_b}{\partial B}\frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} \tag{37}$$

$$= \frac{\partial \chi_b / \partial u}{1 - \beta' \, \partial \chi_b / \partial B}. \tag{38}$$

Since $\partial \chi_b / \partial u < 0$, the sign of $\mathrm{d}\tilde{x}_b/\mathrm{d}u$ is opposite to that of the denominator $1 - \beta' \, \partial \chi_b/\partial B$, where $\beta' \, \partial \chi_b / \partial B > 0$. When $\beta' \, \partial \chi_b/\partial B \in (0, 1)$, the denominator is positive: $\tilde{x}_b$ falls with u. Since $\beta' \, \partial \chi_b / \partial B = f'$, this happens at the red dots in Fig. 3 where f intersects the identity line *from above*. Hence the red dots shift *right* as u rises.

When $\beta' \, \partial \chi_b / \partial B > 1$, the denominator is negative, and $\tilde{x}_b$ *rises* with u. Visually, this happens at the teal dots in Fig. 3 where f intersects the identity line *from below*. Does this mean congestion can *boost* bus ridership? The answer is no. Such an equilibrium is a sort of (local) "minimum" sustainable level of ridership: when $\beta' \, \partial \chi_b / \partial B > 1$, then a ridership x_b^- which is slightly smaller than $\tilde{x}_b$ is insufficient to maintain a bus fleet that attracts x_b^- riders. As u rises, this (local) minimum increases, but this is not to say "higher congestion makes more people ride the bus". By converse logic, equilibria where $\beta' \, \partial \chi_b / \partial B \in (0, 1)$ are (local) maxima. Hence, we name equilibria by

Table 2
Signs of partial equilibrium derivatives for Min and Max equilibria. ? indicates an ambiguous sign.

	$\beta' \partial \chi_b / \partial B$	$\mathrm{d}\tilde{W}_b/\mathrm{d}u$	$\mathrm{d}\tilde{T}_b/\mathrm{d}u$	$\mathrm{d}\tilde{x}_b/\mathrm{d}u$	$\mathrm{d}\tilde{x}_c/\mathrm{d}u$
Max	$\in (0,1)$	+	+	−	?
Min	>1	−	?	+	−

Definition. Consider a partial equilibrium with bus ridership $\tilde{x}_b$. Equilibria with $\beta' \partial \chi_b / \partial B \in (0,1)$ are **Max**, and those with $\beta' \partial \chi_b / \partial B > 1$ are **Min**.

It will sometimes be convenient to speak of the sign of $1 - \beta' \partial \chi_b / \partial B$ not only at points but also *along branches* of χ_b. So in Fig. 3(b) the teal branch is Min and the red branches are Max. This is possible because unless there is some discontinuity in X_b, the relative slope of f and the identity line will stay the same at some equilibrium, until it perishes at a bifurcation. There are two bifurcations in Fig. 4(a), shown by golden dots.

Turning now to car demand, let

$$\chi_c(u, B) = X_c\left[T_c(u), \rho_b(u, B)\right]. \tag{39}$$

with

$$\frac{\partial \chi_c}{\partial u} = \frac{\partial X_b}{\partial u}\frac{\partial T_c}{\partial u} + \frac{\partial X_c}{\partial \rho_b}\frac{\partial W_b}{\partial u} \qquad \frac{\partial \chi_c}{\partial B} = \frac{\partial W_b}{\partial B}\frac{\partial X_c}{\partial \rho_b} < 0. \tag{40}$$

So $\partial \chi_c / \partial u$ has an ambiguous sign, but χ_c falls with B. Each $\tilde{x}_b$ implies a partial equilibrium car demand

$$\tilde{x}_c = \chi_c\left[u, \beta(\tilde{x}_b)\right], \tag{41}$$

such that

$$\frac{\mathrm{d}\tilde{x}_c}{\mathrm{d}u} = \frac{\partial \chi_c}{\partial u} + \frac{\partial \chi_c}{\partial B}\beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u}. \tag{42}$$

It is hard to get anything useful out of this expression. So let $\tilde{W}_b = W_b[u, \beta(\tilde{x}_b)]$ be a *partial equilibrium wait time* (given u). Differentiating yields

$$\frac{\mathrm{d}\tilde{W}_b}{\mathrm{d}u} = \frac{\partial W_b}{\partial u} + \frac{\partial W_b}{\partial B}\beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} \tag{43}$$

The first term is the “raw” effect caused by the fact buses take longer complete the route. The second term reflects “knock-on” effects from changes in fleet size. Using (38) and expanding leads to

$$\frac{\mathrm{d}\tilde{W}_b}{\mathrm{d}u} = \frac{1}{1 - \beta' \partial \chi_b / \partial B} \underbrace{\left\{ \frac{\partial W_b}{\partial u} + \frac{\partial W_b}{\partial B}\beta' \frac{\partial X_b}{\partial T_b} L \right\}}_{>0} \tag{44}$$

So $\tilde{W}_b$ rises with u at Max equilibria and falls at Min ones. The implied partial equilibrium bus travel time, $\tilde{T}_b = L(u + \eta) + \tilde{W}_b$, rises with u at Max equilibria and may rise or fall at Min ones.

Now rewrite $\tilde{x}_c = X_c\left[Lu, L\eta + \tilde{W}_b\right]$ and differentiate

$$\frac{\mathrm{d}\tilde{x}_c}{\mathrm{d}u} = \frac{\partial X_c}{\partial T_c} L + \frac{\partial X_c}{\partial \rho_b}\frac{\mathrm{d}\tilde{W}_b}{\mathrm{d}u} \tag{45}$$

At Min equilibria, $\tilde{W}_b$ is declining, so both terms are negative and $\tilde{x}_c$ certainly declines. At Max equilibria, the first term is negative and the second term positive because $\tilde{W}_b$ is rising, so $\tilde{x}_c$ can rise or fall. Let

$$\tilde{\chi}_c(u) = \left\langle \chi_c\left[u, \beta(\tilde{x}_b)\right] \ \ \forall \tilde{x}_b \in \chi_b(u) \right\rangle \tag{46}$$

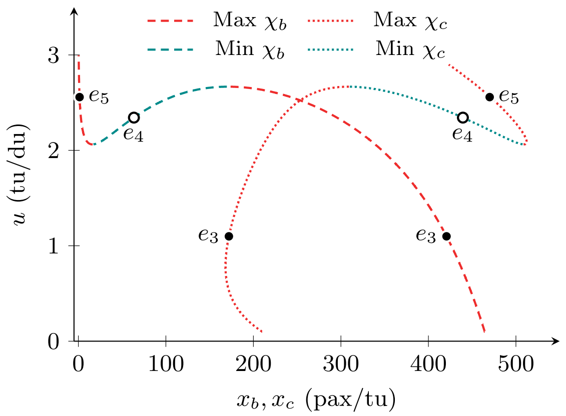
be the multi-valued car demand function returning the set of $\tilde{x}_c$’s given u. Fig. 4(a) shows $\tilde{\chi}_b$ and $\tilde{\chi}_c$ alongside each other for our simulation. In line with our derivations, $\tilde{\chi}_c$ rises and falls at different parts of the Max branches. Table 2 collects the signs of the above derivatives for Min and Max branches.

4.3. Full equilibrium

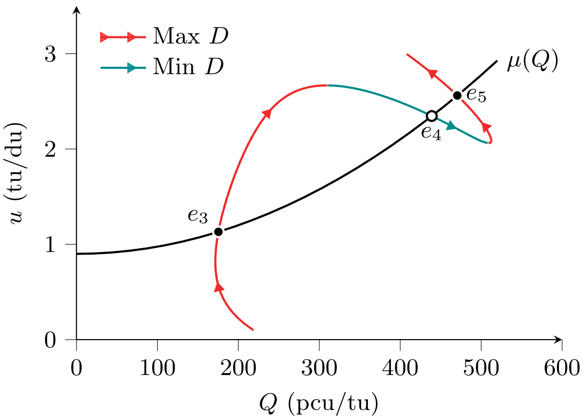
We now endogenize congestion to derive a “full equilibrium”. The “traffic flow demanded” function $D(u)$ of Section 3 becomes multi-valued

$$D(u) = \left\langle \chi_c\left[u, \beta(\tilde{x}_b)\right] + \frac{\kappa \beta(\tilde{x}_b)}{R(u+\eta)} \ \forall \ \tilde{x}_b \in \chi_b(u) \right\rangle. \tag{47}$$

Fig. 4(b) shows D from the same simulation as Fig. 4(a). D looks similar to $\tilde{\chi}_c$ because, in this simulation, traffic flow is mostly car traffic. However, if demands are such that very few people drive, then D may be dominated by changes in bus ridership (which rises with u at Min equilibria). So the shape of D is very open-ended: it can rise or fall along either type of branch.



(a) Passenger demands for bus and car under Ass. 4 with branches labeled



(b) Vehicle flows and equilibria under Ass. 4

Fig. 4. Demand curves for endogenous B.

As before, equilibria occur at intersections of D and $\mu(Q)$. But now there is more scope for differences among equilibria: equilibria may occur *along* a rising branch and *across* branches. Fig. 4(a) shows the demands associated with each equilibrium. Note how demands vary greatly across branches: e.g., the bus ridership at e_3 is hundreds of times higher than at e_5.

Hereafter, when we characterize equilibria by the sign of $1 - D'\mu'$, the D' is meant to be the *local* derivative taken at some particular point. For example, given a point on D with bus ridership $x_b = \tilde{x}_b$, the local derivative is

$$D'(u) = \frac{\partial \chi_c}{\partial u} + \frac{\partial \chi_c}{\partial B}\beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} + \left(\frac{\partial Q_b}{\partial B}\right)_u \beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} + \left(\frac{\partial Q_b}{\partial u}\right)_B . \tag{48}$$

4.4. Local stability

Adapting the dynamics of Section 3.3, suppose u is instantaneously determined by $u = \hat{\mu}(x_c, B)$ while x_c, B and x_b evolve by

$$\dot{x}_b = \Theta\Big\{\chi_b[\hat{\mu}(x_c, B), B] - x_b\Big\} \qquad \dot{x}_c = \Omega\Big\{\chi_c[\hat{\mu}(x_c, B), B] - x_c\Big\} \qquad \dot{B} = \Psi\Big\{\beta(x_b) - B\Big\}. \tag{49}$$

Θ, Ω and Ψ are evolution[10] functions. They exhibit $\Theta(0) = \Omega(0) = \Psi(0) = 0$ and $\Theta', \Omega', \Psi' > 0$. Note B evolves by the same "best response" dynamics as demands. Whenever $\beta(x_b) > B$, the revenues/subsidies coming in are sufficient for the operator to run more buses, and so it does so.

A fixed point of (49) is a triplet

$$x_b^* = \chi_b[\hat{\mu}(x_c^*, B), B] \qquad x_c^* = \chi_c[\hat{\mu}(x_c^*, B), B] \qquad B^* = \beta(x_b^*). \tag{50}$$

We judge the fixed point's local stability using the typical protocol for non-linear, continuous-time, autonomous systems. See Strogatz (2015, Ch. 5) for a clear explanation and Pandey et al. (2024) for stability analysis of a traffic system modeled with three state variables.

To begin, let $\theta = \Theta'(0), \omega = \Omega'(0)$ and $\psi = \Psi'(0)$ be, as above, the (strictly positive) speeds-of-adjustment at the fixed point. Let $\delta_{x_c} := x_c - x_c^*$, $\delta_{x_b} := x_b - x_b^*$, and $\delta_B = B - B^*$ be perturbations in the neighborhood of the fixed point. For small perturbations (i.e., when x_b, x_c, and B are near a fixed point), the trajectories of the perturbations according to (49) are approximately

$$\begin{pmatrix}\dot{\delta}_{x_c}\\ \dot{\delta}_{x_b}\\ \dot{\delta}_B\end{pmatrix} \approx \underbrace{\begin{pmatrix}-\theta & \theta a_{1,2} & \theta a_{1,3}\\ 0 & \omega a_{2,2} & \omega a_{2,3}\\ \psi\beta' & 0 & -\psi\end{pmatrix}_{(x_b^*, x_c^*, B^*)}}_{A\ (\text{Jacobian})} \begin{pmatrix}\delta_{x_c}\\ \delta_{x_b}\\ \delta_B\end{pmatrix}, \tag{51}$$

where the matrix on the RHS is the *Jacobian* of the system (49) taken at the fixed point, which we will call A. The variable entries of A (which we have not filled in above for brevity) are

$$a_{1,2} = \frac{\partial}{\partial x_c}\chi_b[\hat{\mu}(x_c, B), B] \qquad a_{1,3} = \frac{\partial}{\partial B}\chi_b[\hat{\mu}(x_c, B), B] \tag{52}$$

$$a_{2,2} = \frac{\partial}{\partial x_c}\chi_c[\hat{\mu}(x_c, B), B] - 1 \qquad a_{2,3} = \frac{\partial}{\partial B}\chi_c[\hat{\mu}(x_c, B), B] \tag{53}$$

[10] Reasonably, one might assume $\Theta = \Omega$ (i.e., that car and bus demand adjust by the same function), but this does not change results.

(51) is a system of ODE's. Its solution has the form

$$\begin{pmatrix} \delta_{x_c}(t) \\ \delta_{x_b}(t) \\ \delta_B(t) \end{pmatrix} = \boldsymbol{v}_1 e^{\lambda_1 t} + \boldsymbol{v}_2 e^{\lambda_2 t} + \boldsymbol{v}_3 e^{\lambda_3 t}, \tag{54}$$

where the $\boldsymbol{v}$'s are 3×1 *eigenvectors* of the Jacobian A and the λ's are A's *eigenvalues.* The fixed point is locally, asymptotically stable if and only if all three perturbations tend to zero in the long run (which means that all three state variables tend to the fixed point). For this to happen requires that A's eigenvalues be negative (so that the exponential terms tend to zero). The eigenvalues are the roots of A's characteristic polynomial

$$\det\left(A - \lambda I_3\right) = -\lambda^3 + m_2 \lambda^2 - m_1 \lambda + m_0 \tag{55}$$

where

$$m_0 = \theta\omega\psi\left\{a_{2,2}\left(1 - \beta' a_{1,3}\right) + a_{2,3}a_{1,2}\beta'\right\} \quad (A\text{'s determinant}) \tag{56}$$

$$m_1 = -(\theta + \psi)\omega a_{2,2} + \theta\psi\left(1 - \beta' a_{1,3}\right) \quad (\text{the sum of } A\text{'s principal minors}) \tag{57}$$

$$m_2 = \omega a_{2,2} - (\theta + \psi) \quad (A\text{'s trace}). \tag{58}$$

Next, we use the Routh–Hurwitz criteria for a third degree polynomial (see Bodson (2020) for a readable explanation) to obtain the requirements which must hold in order for the RHS of (55) to have negative roots. These criteria are

$$\text{(a) } m_0 < 0 \qquad \text{(b) } m_1 > 0 \qquad \text{(c) } m_2 < 0 \qquad \text{(d) } 0 < m_0 - m_1 m_2 \tag{59}$$

It turns out that whether these criteria are satisfied at the fixed point depends on our Outside-In/Inside-Out and Min/Max distinctions. This can be shown via two lemmas (proofs in Appendix B):

Lemma 1.

$$\frac{m_0}{\theta\omega\psi} = a_{2,2}\left(1 - \beta' a_{1,3}\right) + \beta' a_{2,3}a_{1,2} = \phi\left(1 - \beta' \frac{\partial \chi_b}{\partial B}\right)\left(D'\mu' - 1\right)$$

Lemma 2. *$a_{2,2} < 0 < 1 - \beta' a_{1,3}$ is necessary for $m_0 < 0$*

Proposition 2. *Under the dynamics given by (49), an equilibrium is stable if and only if it is either Outside-In ($D'\mu' < 1$) and Max ($\beta' \partial\chi_b/\partial B < 1$) or Inside-Out ($D'\mu' > 1$) and Min ($\beta' \partial\chi_b/\partial B < 1$).*

It may be surprising that Min equilibria may be stable. Bar-Yosef et al. (2013) also has a plot similar to Fig. 3(a), and their equivalent of Min equilibria are judged to be unstable: a slightly lower (higher) level of ridership would lead to a self-reinforcing reduction (increase) in frequency. But unlike Bar-Yosef et al. (2013), here congestion is endogenous. If an equilibrium is Min and Inside-Out, then a slightly higher ridership will not beckon more riders, because it will increase congestion too much.

Note that, just as ω's magnitude did not matter to stability in Section 3.3, here stability does not depend on the magnitudes of θ, ω and ψ. A can also be written

$$A = \begin{bmatrix} \theta & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \psi \end{bmatrix} \begin{bmatrix} -1 & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} & a_{2,3} \\ \beta' & 0 & -1 \end{bmatrix}_{x_c^*, x_b^*, B^*}. \tag{60}$$

At any stable fixed point, the right matrix has the property called *D-stability*: its eigenvalues are negative for any $\theta, \omega, \psi > 0$. Giorgi and Zuccotti (2015) review D-stability, related theorems and its history in economics.

4.5. Road improvement

Finally, reconsider the road improvement captured by an increase in the road efficiency parameter ζ. Now there are two stable equilibrium types: Max/Outside-In and Min/Inside-Out.

Max/Outside-In. Per (31), u^* falls with ζ at Outside-In equilibria. Thus, we can reverse the partial equilibrium derivatives derived earlier to sign the change in equilibrium variables. For example,

$$\frac{\mathrm{d}x_b^*}{\mathrm{d}\zeta} = \frac{\mathrm{d}u^*}{\mathrm{d}\zeta}\frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u}. \tag{61}$$

Since $\mathrm{d}\tilde{x}_b/\mathrm{d}u < 0$ at Max equilibria, and $\mathrm{d}u^*/\mathrm{d}\zeta > 0$ at Outside-In ones, the equilibrium bus ridership rises at Max/Outside-In ones. Likewise, reviewing the Max row Table 2, we see that the equilibrium bus travel time and wait time fall, too. Whether car demand x_c^* or the traffic flow Q^* rise or fall varies among Max/Outside-In equilibria. But since the travel times of both modes fall, overall demand $x_b^* + x_c^*$ must rise. The situation is basically the same as in Section 3.4.

The notable change is that Economies of Scale amplify all the derivatives. For example, compare the change in the equilibrium wait time, W_b^*, under Assumptions 3 (Exogenous Fleet Size) and 4 (Endogenous Fleet Size):

$$\text{Exogenous } B \quad \frac{\mathrm{d}W_b^*}{\mathrm{d}\zeta} = \frac{\partial\mu/\partial\zeta}{1-D'\mu'}\frac{\partial W_b}{\partial u} \tag{62}$$

$$\text{Endogenous } B\text{:} \quad \frac{\mathrm{d}W_b^*}{\mathrm{d}\zeta} = \frac{\partial\mu/\partial\zeta}{1-D'\mu'}\frac{1}{1-\beta'\,\partial\chi_b/\partial B}\left\{\frac{\partial W_b}{\partial u}+\frac{\partial W_b}{\partial B}\beta'\frac{\partial X_b}{\partial T_b}L\right\} \tag{63}$$

In the endogenous case, $\partial W_b/\partial u$ is joined by a second (positive) term inside the curly brackets. The bracketed sum is then multiplied not only by $1/(1-D'\mu')$ but also $1/(1-\beta'\,\partial\chi_b/\partial B)$. Since $1-\beta'\,\partial\chi_b/\partial B\in(0,1)$ at Max equilibria, this new term acts as a multiplier that amplifies the term in curly brackets. This multiplier captures positive feedback from economies of scale: increasing ridership increases the fleet size, which increases ridership, and so on. Ceteris paribus, the wait time falls much faster than in the Exogenous Fleet Size case. Similar amplification can be derived for other equilibrium variables.

Min/Inside-Out. Consider now the strange, but stable, Min/Inside-Out equilibria where $D'\mu'>1$ and $\beta'\,\partial\chi_b/\partial B>1$. Per (31), since $D'\mu'>1$, the unit travel time u^* actually rises. So the "raw" impact of a road improvement is *reversed*: congestion worsens. Since u^* from the Min row of Table 2 we can deduce that bus demand rises, while wait time and car traffic both fall. Next, note $1<D'\mu'$ implies $D'>0$ (a rising traffic flow demanded curve), and so traffic flow Q^* rises. Rewrite $\mathrm{d}Q^*/\mathrm{d}\zeta$ as

$$\frac{\mathrm{d}Q^*}{\mathrm{d}\zeta} = \frac{\mathrm{d}}{\mathrm{d}\zeta}\left\{x_c^*+\frac{\kappa\beta(x_b^*)}{R(u^*+\eta)}\right\} \tag{64}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\zeta}\left(x_b^*+x_c^*\right)-\frac{\mathrm{d}x_b^*}{\mathrm{d}\zeta}\underbrace{\left[1-\left(\frac{\partial Q_b}{\partial B}\right)_u\beta'\right]}_{\in(0,1)}+\left(\frac{\partial Q_b}{\partial u}\right)_B\frac{\mathrm{d}u^*}{\mathrm{d}\zeta}. \tag{65}$$

Per property (d) of Assumption 4, the term in square brackets is positive. Thus, since $\mathrm{d}u^*/\mathrm{d}\zeta>0, \mathrm{d}x_b^*/\mathrm{d}\zeta>0, \mathrm{d}Q^*/\mathrm{d}\zeta>0$, it must be that $\mathrm{d}(x_b^*+x_c^*)/\mathrm{d}\zeta>0$: the road improvement raises total demand. In turn,

$$\frac{\mathrm{d}}{\mathrm{d}\zeta}(x_b^*+x_c^*) = \frac{\partial X_b}{\partial T_b}\frac{\mathrm{d}T_b^*}{\mathrm{d}\zeta}+\frac{\partial X_c}{\partial T_c}\frac{\mathrm{d}T_c^*}{\mathrm{d}\zeta} \tag{66}$$

Since $\mathrm{d}(x_b^*+x_c^*)/\mathrm{d}\zeta>0$ and $\mathrm{d}T_c^*/\mathrm{d}\zeta>0$, it must be that $\mathrm{d}T_b^*/\mathrm{d}\zeta<0$: the road improvement lowers bus travel time. While the in-vehicle time of a bus trip rises, wait time falls more than enough to compensate.

Cosmetically, the final effects resemble the Downs-Thomson paradox, insofar as a road improvement raises traffic flow and congestion. But the situation is opposite to the Downs-Thomson paradox in other regards: the road improvement reduces driving and boosts bus ridership. Whether or not this situation could actually happen is doubtful. The increase in traffic flow and congestion are due to a bus fleet expansion so rapid as to overwhelm the drop in car traffic. But the outcome is possible under our assumptions, and it is interesting to contemplate, so we have given it some discussion for the sake of thoroughness.

Summary of effects. We have obtained many results in this section, so before concluding we will summarize some of the more interesting ones:

- At all stable equilibria, the wait time and overall bus travel time fall, while bus ridership and overall travel demand rise. There is no scope for a genuine Downs-Thomson paradox here, in the sense of the road improvement undermining the bus' economies of scale, because the bus experiences the benefits of the road improvement even more than cars do.
- At stable Outside-In equilibria there is a "Pareto improvement", insofar as bus riders and car drivers both wind up better off. The bus riders benefit more though, because they experience the same reduction in in-vehicle time ($L\mathrm{d}u^*/\mathrm{d}\zeta$) as drivers but also enjoy lower wait times. At stable Inside-Out (Min) equilibria, people who drive before and after the improvement would experience longer travel times, because so many buses enter the road that congestion rises.

5. Conclusion

5.1. Summary

This paper has worked through a model of a mixed-traffic, congestible bus route where travelers choose among driving, riding a bus and not traveling. Using a graphical approach, equilibria were derived under two assumptions: that the bus fleet size is fixed (Section 3) and that the bus fleet size rises with bus ridership (Section 4). In both cases, a rise in congestion increases the bus cycle time and thus, ceteris paribus, lowers the frequency and raises the expected wait time of bus passengers. This effect leads to a phenomenon we call Congestive Mode-Switching (CMS). CMS is said to happen when rising congestion encourages people to switch from bus to car. And since this switching creates congestion, CMS can sustain multiple equilibria. When the bus fleet size rises with ridership, CMS is joined by an "Economies of Scale Mechanism" that also underlies circular causation and makes equilibria vary dramatically from each other. The paper also derived equilibria's local stabilities under continuous-time, best-response dynamics and conducted comparative static exercises involving a road improvement. The stability analysis showed that multiple equilibria may be stable. The comparative static analysis showed how the CMS and the EOS mechanism may amplify or reverse the effects of a road improvement. A genuine Downs-Thomson Paradox never arises because traffic is mixed—unlike in the cannonical examples of the Paradox whereby transit and cars run on parallel paths. Thus, a road improvement not only lowers bus in-vehicle times but also allows buses to complete the route and start over more often, which raises bus ridership and lowers bus travel times.

5.2. Future research

Other sources of Differential Penalty. The stylized model in this paper is meant to demonstrate certain bits of logic which could have wide purchase across mixed-traffic transportation systems. A most promising way to extend the paper is to build models using similar logic. A basic question in building such models is: Why would rising congestion make people switch from transit (shared modes) to driving? Or, in our terminology: Why is there a Differential Penalty? Another seed for a Differential Penalty lies in the fact that random variation in car traffic triggers bus bunching (Daganzo, 2009; Chow et al., 2017), which increases the mean and variance of passenger wait time. If such randomness rises with the flow of cars, then there is a Differential Penalty: while both bus and car travelers experience traffic randomness while in motion, only bus travelers experience the amplified effect of bus bunching while waiting or transferring. Also, while we have assumed the trip length L is the same by both modes, in some cases the trip length of bus will be longer in cases when the bus meanders en route to the traveler's destination. This detour effect is behind CMS in the ridepooling model of Lehe and Pandey (2020).

Factors that weaken differential penalties. The model of this paper was expressly designed to demonstrate the logic of Congestive Mode-Switching and its interaction with economies of scale. However, the model is stylized and does not prove CMS is a ubiquitous phenomenon. Just as there plausibly exist factors which would tend to *amplify* a differential penalty (e.g., bus bunching) in real life, there exist other factors which would plausibly *attenuate* it. One is *crowding*. As a transit vehicle empties out, it generally becomes more attractive to ride in Haywood et al. (2017), and there exist theoretical models designed to capture such crowding effects and their impacts on travel choice (Tirachini et al., 2013; de Palma et al., 2015). A second factor is boarding/alighting delay (Lin and Wilson, 1992; Milkovits, 2008; Tirachini, 2013; Lehe and Pandey, 2024). In a model with these factors, when passengers switch to car they will (ceteris paribus) make the bus more pleasant and faster—thereby dampening the incentive to switch. Likewise the converse: as passengers switch to bus they slow it down, force others to stand, etc. Also, if people dislike time spent in the car substantially more than in the bus (enough so that they switch to bus even when its wait time is rising), the results may be reversed; we could have *anti*-congestive mode-switching.

Policy as a choice among equilibria. This paper has shown even simple traffic systems have the potential for multiple equilibria from multiple sources. While we have not considered interventions, a takeaway lesson is that the scope for policy is grander than ordinarily conceived: rather than merely translating equilibria, policy may hop between their "neighborhoods". Murphy et al. (1989), mentioned in the introduction, is about an idea from development economics called "the big push", whereby industrialization across a national economy may tip a country from an agrarian society to a richer one based on manufacturing with economies of scale. Broadly, the premise behind the "big push" is that an economy may exhibit multiple equilibria, and policy plays a role in realizing the superior ones. Similarly, Reinhold and Kearney (2008) describes how Berlin's transit operator (BVG) boosted frequency on key lines and wound up making more money—which can be thought of as using a "big push" to leap from one equilibrium to a higher one. To this end, tolls, subsidies, signal priority schemes and other measures could be incorporated into a model with CMS.

Discrete-time dynamics. The paper's system could be analyzed using discrete-time dynamics—as Bar-Yosef et al. (2013), Cantarella et al. (2015), Li and Yang (2016) and Li et al. (2018) do. Discrete-time dynamics would capture the idea people adjust choices from day-to-day. We think the results will be similar except that the bar for stability will be higher. Watling (1999) shows that the stability of a stochastic traffic equilibrium assignment under continuous-time dynamics is necessary for its stability under equivalent discrete-time dynamics (i.e., the equilibria unstable in continuous-time are unstable in discrete-time). Similarly, it may be that none of the equilibria we have judged to be unstable will be stable under plausible discrete-time dynamics, but some of the equilibria judged to be stable will be unstable given certain parameter values.

Operator adjustments. For simplicity, in this paper the operator could only adjust the fleet size, B. In reality an operator can change fares, line structures, line densities, capacities and other variables (Jara-Diaz and Gschwender, 2003; Hörcher and Tirachini, 2021). At high levels of demand the operator may also implement local and express services (Daganzo, 2010; Daganzo and Ouyang, 2019). Expanding the field of adjustments is not only worthwhile for the sake of realism but also because it leads to more powerful economies of scale. In this paper, wait time declines at a rate proportionate to $1/B$, which falls dramatically for low B but is quickly exhausted. When the operator can adapt to higher ridership in more ways the results may be more dramatic.

CRediT authorship contribution statement

Ayush Pandey: Conceptualization, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Lewis J. Lehe:** Conceptualization, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgments

The authors would like to thank Vikash Gayah and participants at the 2023 International Transportation Economics Association Annual Conference in Santander for their comments, as well as the National Science Foundation for grants CMMI-2052512 and CMMI-2052337, which provided support.

Appendix A. Simulations

Model I

The instance of the model used to create the figures in Section 3 works as follows. There are two groups of travelers. Travelers in the first group choose among driving, riding the bus and an outside option according to a nested logit model. See Train (2009, Ch.5) for an exposition of the nested logit model. The "systematic" utility of the outside option is set to zero. Those of bus and car are, respectively:

$$V_b(T_b) = 5.7 - T_b/3 \qquad V_c(T_c) = 4.5 - T_c/3. \tag{A.1}$$

Car and bus are in a nest characterized by a nest coefficient of $\sigma > 0$, where $1-\sigma$ indexes the degree to which choices between car and bus are correlated. It follows from the nested logit assumptions that the probability of someone in the first group choosing mode $i = c, b$ is

$$P_b^1(T_b, T_c) = e^{V_b(T_b)/\sigma} \frac{\left[e^{V_b(T_b)/\sigma} + e^{V_c(T_c)/\sigma}\right]^{\sigma-1}}{\left[e^{V_b(T_b)/\sigma} + e^{V_c(T_c)/\sigma}\right]^{\sigma} + 1} \qquad P_c^1(T_b, T_c) = e^{V_c(T_c)/\sigma} \frac{\left[e^{V_b(T_b)/\sigma} + e^{V_c(T_c)/\sigma}\right]^{\sigma-1}}{\left[e^{V_b(T_b)/\sigma} + e^{V_c(T_c)/\sigma}\right]^{\sigma} + 1} \tag{A.2}$$

We set $\sigma = 0.5$ for the nesting coefficient.

The second group of travelers only chooses between car and the outside option. Each has a probability of choosing to drive given by

$$P_c^2(T_c) = e^{-T_c/4}. \tag{A.3}$$

There are a total of 800 decision-makers in the first group and 200 in the second group. Thus, we can write the demand functions as

$$X_b(T_b, \rho_b) = 800 P_b^1(T_b, T_b - \rho_b) \qquad X_c(T_c, \rho_b) = 800 P_c^1(T_c + \rho_b, T_c) + 100 P_c^2(T_c). \tag{A.4}$$

For the paper's physical parameters we have $\eta = 0.1, L = 5, R = 20, \kappa = 5, \bar{B} = 5$. The function μ is a BPR function

$$\mu(Q) = 0.5 + 0.8(Q/400)^4 \tag{A.5}$$

None of the parameters were chosen for realism, but simply to produce illustrative figures.

Model II

In this simulation, things are changed to produce an illustrative plot. For the first group's systematic utility functions we have

$$V_b(T_b) = 5.7 - T_b/4 \qquad V_c(T_c) = 4.5 - T_c/4. \tag{A.6}$$

There are 600 people in the first group and 100 in the second group. The functions β and μ are

$$\beta(x_b) = 3 + x_b/20 \qquad \mu(Q) = 0.9 + 1.2(Q/400)^2. \tag{A.7}$$

Now $\eta = 0.4$.

Appendix B. Proofs

B.1. Proof of Lemma 1

We have

$$a_{1,2} = \frac{\partial \hat{\mu}}{\partial Q_c}\frac{\partial \chi_b}{\partial u} \qquad 1 - \beta' a_{1,3} = 1 - \beta' \frac{\partial \chi_b}{\partial B} - \beta' \frac{\partial \hat{\mu}}{\partial B}\frac{\partial \chi_b}{\partial u} \tag{B.1}$$

$$a_{2,2} = \frac{\partial \hat{\mu}}{\partial Q_c}\frac{\partial \chi_c}{\partial u} - 1 \qquad a_{2,3} = \frac{\partial \hat{\mu}}{\partial B}\frac{\partial \chi_c}{\partial u} + \frac{\partial \chi_c}{\partial B}. \tag{B.2}$$

Let $\alpha := 1 - \beta' \partial\chi_b/\partial B$. With this abbreviation, $a_{2,2}(1-\beta' a_{1,3}) + \beta' a_{2,3} a_{1,2}$ simplifies to

$$m_0/\psi\omega\theta = \alpha \frac{\partial \hat{\mu}}{\partial Q_c}\frac{\partial \chi_c}{\partial u} + \frac{\partial \chi_c}{\partial B}\beta' \frac{\partial \hat{\mu}}{\partial Q_c}\frac{\partial \chi_b}{\partial u} + \beta' \frac{\partial \hat{\mu}}{\partial B}\frac{\partial \chi_b}{\partial u} - \alpha \tag{B.3}$$

$$= \alpha \left\{ \frac{\partial \hat{\mu}}{\partial Q_c}\frac{\partial \chi_c}{\partial u} + \frac{\partial \chi_c}{\partial B}\beta' \frac{\partial \hat{\mu}}{\partial Q_c}\frac{1}{\alpha}\frac{\partial \chi_b}{\partial u} + \beta' \frac{\partial \hat{\mu}}{\partial B}\frac{1}{\alpha}\frac{\partial \chi_b}{\partial u} - 1 \right\} \tag{B.4}$$

$$= \alpha \left\{ \frac{\partial \hat{\mu}}{\partial Q_c}\frac{\partial \chi_c}{\partial u} + \frac{\partial \chi_c}{\partial B}\beta' \frac{\partial \hat{\mu}}{\partial Q_c}\frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} + \beta' \frac{\partial \hat{\mu}}{\partial B}\frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} - 1 \right\} \quad \left(\text{since } \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} = \frac{1}{\alpha}\frac{\partial \chi_b}{\partial u}\right) \tag{B.5}$$

$$= \alpha\left\{ \phi\mu' \frac{\partial \chi_c}{\partial u} + \phi\mu' \frac{\partial \chi_c}{\partial B}\beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} + \phi\mu' \left(\frac{\partial Q_b}{\partial B}\right)_u \beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} - 1 \right\} \quad \text{(per (8))} \tag{B.6}$$

$$= \phi\alpha\left\{ \mu' \frac{\partial \chi_c}{\partial u} + \mu' \frac{\partial \chi_c}{\partial B}\beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} + \mu' \left(\frac{\partial Q_b}{\partial B}\right)_u \beta' \frac{\mathrm{d}\tilde{x}_b}{\mathrm{d}u} + \mu' \left(\frac{\partial Q_b}{\partial u}\right)_B - 1 \right\} \quad \text{(per (38))} \tag{B.7}$$

$$= \phi\left(1 - \beta' \frac{\partial \chi_b}{\partial B}\right)\left\{D'\mu' - 1\right\} \quad \text{(per (42))} \tag{B.8}$$

B.2. Proof of Lemma 2

To start this proof, write

$$a_{1,2} = \frac{\partial}{\partial x_c} X_b\left[L(u+\eta) + W_b(u,B), L\eta + W_b(u,B)\right] \qquad a_{2,2} = \frac{\partial}{\partial x_c} X_c\left[Lu, Lu + W_b(u,B)\right] - 1 \tag{B.9}$$

$$a_{1,3} = \frac{\partial}{\partial B} X_b\left[L(u+\eta) + W_b(u,B), L\eta + W_b(u,B)\right] \qquad a_{2,3} = \frac{\partial}{\partial B} X_c\left[Lu, Lu + W_b(u,B)\right] \tag{B.10}$$

where $u = \mu(x_c, B)$. Next, note

$$\frac{\partial}{\partial B} W_b\left[\hat{\mu}(x_c,B), B\right] = \phi\mu' \left(\frac{\partial Q_b}{\partial B}\right)_u \frac{R}{2B} - \frac{R(u+\eta)}{2B^2} = -\phi \frac{R(u+\eta)}{2B^2} = \phi \frac{\partial W_b}{\partial B} < 0. \tag{B.11}$$

Thus, by expanding the above terms we obtain:

$$1 - \beta' a_{1,3} = 1 - \underbrace{\beta'\phi \frac{\partial W_b}{\partial B}\left(\frac{\partial X_b}{\partial T_b} + \frac{\partial X_b}{\partial \rho_b}\right)}_{z_{1,3}<0} - \beta' \frac{\partial \hat{\mu}}{\partial B}\frac{\partial X_b}{\partial T_b} L \qquad a_{2,3} = \frac{\partial X_c}{\partial T_c}\frac{\partial \hat{\mu}}{\partial B} L - \underbrace{\frac{\partial X_c}{\partial \rho_b}\phi \frac{\partial W_b}{\partial B}}_{z_{2,3}<0} \tag{B.12}$$

$$a_{2,2} = \phi\mu' \frac{\partial X_c}{\partial T_c} L + \underbrace{\phi\mu' \frac{\partial X_c}{\partial \rho_b}\frac{\partial W_b}{\partial u}}_{z_{2,2}>0} - 1 \qquad a_{1,2} = \phi\mu' \frac{\partial X_b}{\partial T_b} L - \underbrace{\phi\mu' \frac{\partial W_b}{\partial u}\left\{\frac{\partial X_b}{\partial T_b} + \frac{\partial X_b}{\partial \rho_b}\right\}}_{z_{1,2}<0}. \tag{B.13}$$

Both terms of $a_{1,2}$ and $a_{2,3}$ are negative, so $a_{1,2}a_{2,3} > 0$. Hence, the second term of

$$\frac{m_0}{\theta\omega\psi} = a_{2,2}\left(1 - \beta' a_{1,3}\right) + \beta' a_{2,3} a_{1,2} < 0$$

is positive. (a) thus requires either (i) $a_{2,2} < 0 < 1 - \beta' a_{1,3}$ or (ii) $1 - \beta' a_{1,3} < 0 < a_{2,2}$.

We will show that (i) is the case by contradiction. If (ii) is true, then $a_{2,2}$, while positive, cannot exceed its only positive term $z_{2,2}$. And $1 - \beta' a_{1,3}$, while negative, cannot be less than its only negative term $z_{1,3}$. Thus, $0 > a_{2,2}\left(1 - \beta' a_{1,3}\right) > z_{1,3} z_{2,2}$.

Note that $z_{2,2} z_{1,3} = -\beta' z_{2,3} z_{1,2}$. Thus,

$$\frac{m_0}{\theta\omega\psi} = a_{2,2}(1 - \beta' a_{1,3}) + \beta' a_{1,2}a_{2,3} > z_{1,3}z_{2,2} + \beta' a_{1,2}a_{2,3} = -\beta' z_{1,2}z_{2,3} + \beta' a_{1,2}a_{2,3} > 0. \tag{B.14}$$

It follows that $m_0/\theta\omega\psi < 0$ requires $a_{2,2} < 0 < 1 - \beta' a_{1,3}$. In turn, if $a_{2,2} < 0 < 1 - \beta' a_{1,3}$, then

$$m_1 = \theta\psi\left(1 - \beta' a_{1,3}\right) - (\theta + \psi)\omega a_{2,2} > 0 \tag{B.15}$$

$$m_2 = \omega a_{2,2} - (\theta + \psi) < 0, \tag{B.16}$$

so (b) and (c) are satisfied. Regarding (d), expansion reveals

$$m_0 - m_1 m_2 = \theta\psi\omega a_{1,2}a_{2,3}\beta' + (\theta + \psi)m_1 + (\theta + \psi)\omega a_{2,2}^2. \tag{B.17}$$

Since $a_{1,2}a_{2,3} > 0$, if $m_1 > 0$ then the RHS is positive, and (d) is satisfied. So if (a) is satisfied then (b)–(d) are, too.

References

Ansari Esfeh, M., Wirasinghe, S.C., Saidi, S., Kattan, L., 2021. Waiting time and headway modelling for urban transit systems – a critical review and proposed approach. Transp. Rev. 41 (2), 141–163. http://dx.doi.org/10.1080/01441647.2020.1806942.

Arnott, R., Small, K., 1994. The economics of traffic congestion. Am. Sci. 82 (5), 446–455, arXiv:29775281. URL: https://www.jstor.org/stable/29775281.

Arthur, W.B., 1990. Positive feedbacks in the economy. Sci. Am. 262 (2), 92–99, arXiv:24996687. URL: https://www.jstor.org/stable/24996687.

Bar-Yosef, A., Martens, K., Benenson, I., 2013. A model of the vicious cycle of a bus line. Transp. Res. B 54, 37–50. http://dx.doi.org/10.1016/j.trb.2013.03.010.

Becker, G.S., 1991. A note on restaurant pricing and other examples of social influences on price. J. Polit. Econ. 99 (5), 1109–1116, arXiv:2937660. URL: https://www.jstor.org/stable/2937660.

Bodson, M., 2020. Explaining the Routh–Hurwitz criterion: a tutorial presentation [focus on education]. IEEE Control Syst. Mag. 40 (1), 45–51. http://dx.doi.org/10.1109/MCS.2019.2949974.

Cantarella, G.E., Velona, P., Watling, D.P., 2015. Day-to-day dynamics & equilibrium stability in a two-mode transport system with responsive bus operator strategies. Netw. Spat. Econ. 15 (3), 485–506.

Chang, S.K., Schonfeld, P.M., 1991. Multiple period optimization of bus transit systems. Transp. Res. B 25 (6), 453–478. http://dx.doi.org/10.1016/0191-2615(91)90038-K.

Chow, A.H.F., Li, S., Zhong, R., 2017. Multi-objective optimal control formulations for bus service reliability with traffic signals. Transp. Res. B 103, 248–268. http://dx.doi.org/10.1016/j.trb.2017.02.006.

Cudahy, B.J., 1990. Cash, tokens, and transfers: a history of urban mass transit in north america. URL: https://trid.trb.org/view/356678.
Daganzo, C.F., 1983. Stochastic network equilibrium with multiple vehicle types and asymmetric, indefinite link cost Jacobians. Transp. Sci. http://dx.doi.org/10.1287/trsc.17.3.282.
Daganzo, C.F., 2009. A headway-based approach to eliminate bus bunching: systematic analysis and comparisons. Transp. Res. B 43 (10), 913–921. http://dx.doi.org/10.1016/j.trb.2009.04.002.
Daganzo, C.F., 2010. Structure of competitive transit networks. Transp. Res. B 44 (4), 434–446. http://dx.doi.org/10.1016/j.trb.2009.11.001.
Daganzo, C.F., Ouyang, Y., 2019. Public Transportation Systems. World Scientific, http://dx.doi.org/10.1142/10553.
David, Q., Foucart, R., 2014. Modal choice and optimal congestion. Reg. Sci. Urban Econ. 48, 12–20. http://dx.doi.org/10.1016/j.regsciurbeco.2014.04.005.
De Palma, A., Fosgerau, M., 2010. Dynamic and static congestion models: A review.
de Palma, A., Kilani, M., Proost, S., 2015. Discomfort in mass transit and its implication for scheduling and pricing. Transp. Res. B 71, 1–18. http://dx.doi.org/10.1016/j.trb.2014.10.001.
Fielbaum, A., Jara-Diaz, S., Gschwender, A., 2020. Beyond the mohring effect: scale economies induced by transit lines structures design. Econ. Transp. 22, 100163. http://dx.doi.org/10.1016/j.ecotra.2020.100163.
Fielbaum, A., Tirachini, A., Alonso-Mora, J., 2023. Economies and diseconomies of scale in on-demand ridepooling systems. Econ. Transp. 34, 100313. http://dx.doi.org/10.1016/j.ecotra.2023.100313.
Friesz, T.L., Bernstein, D., Mehta, N.J., Tobin, R.L., Ganjalizadeh, S., 1994. Day-to-day dynamic network disequilibria and idealized traveler information systems. Oper. Res. 42 (6), 987–1174. http://dx.doi.org/10.1287/opre.42.6.1120.
Gilboa, I., Matsui, A., 1991. Social stability and equilibrium. Econometrica 59 (3), 859–867. http://dx.doi.org/10.2307/2938230, arXiv:2938230.
Giorgi, G., Zuccotti, C., 2015. An Overview on D-stable Matrices. DEM Working Papers Series 97, University of Pavia, Department of Economics and Management.
Gonzales, E.J., 2015. Coordinated pricing for cars and transit in cities with hypercongestion. Econ. Transp. 4 (1), 64–81. http://dx.doi.org/10.1016/j.ecotra.2015.04.003.
Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociol. 83 (6), 1420–1443, arXiv:2778111. URL: https://www.jstor.org/stable/2778111.
Guevara, C.A., 2017. Mode-valued differences of in-vehicle travel time savings. Transportation 44 (5), 977–997. http://dx.doi.org/10.1007/s11116-016-9689-3.
Haywood, L., Koning, M., Monchambert, G., 2017. Crowding in public transport: who cares and why? Transp. Res. A 100, 215–227. http://dx.doi.org/10.1016/j.tra.2017.04.022.
Hofbauer, J., Sandholm, W.H., 2011. Survival of dominated strategies under evolutionary dynamics. Theor. Econ. 6 (3), 341–377. http://dx.doi.org/10.3982/TE771.
Hörcher, D., Tirachini, A., 2021. A review of public transport economics. Econ. Transp. 25 (July 2020), http://dx.doi.org/10.1016/j.ecotra.2021.100196.
Hymel, K.M., Small, K.A., Dender, K.V., 2010. Induced demand and rebound effects in road transport. Transp. Res. B 44 (10), 1220–1241. http://dx.doi.org/10.1016/j.trb.2010.02.007.
Iryo, T., 2019. Instability of departure time choice problem: A case with replicator dynamics. Transp. Res. B 126, 353–364. http://dx.doi.org/10.1016/j.trb.2018.08.005.
Iryo, T., Watling, D., 2019. Properties of equilibria in transport problems with complex interactions between users. Transp. Res. B 126, 87–114. http://dx.doi.org/10.1016/j.trb.2019.05.006.
Jara-Diaz, S.R., Gschwender, A., 2003. Towards a general microeconomic model for the operation of public transport. Transp. Rev. 23 (4), 453–469. http://dx.doi.org/10.1080/0144164032000048922.
Jara-Díaz, S.R., Gschwender, A., 2009. The effect of financial constraints on the optimal design of public transport services. Transportation 36 (1), 65–75. http://dx.doi.org/10.1007/s11116-008-9182-8.
Kieu, L.M., Bhaskar, A., Chung, E., 2015. Empirical modelling of the relationship between bus and car speeds on signalised urban networks. Transp. Plan. Technol. 38 (4), 465–482. http://dx.doi.org/10.1080/03081060.2015.1026104.
Kitamura, R., Nakayama, S., Yamamoto, T., 1999. Self-reinforcing motorization: Can travel demand management take us out of the social trap? Transp. Policy 6 (3), 135–145. http://dx.doi.org/10.1016/S0967-070X(99)00015-3.
Kocur, G., Hendrickson, C., 1982. Design of local bus service with demand equilibration. Transp. Sci. 16 (2), 149–170. http://dx.doi.org/10.1287/trsc.16.2.149.
Krizek, K.J., Levinson, D.M., 2007. Planning for Place and Plexus: Metropolitan Land Use and Transport. Routledge, London, http://dx.doi.org/10.4324/9780203935392.
Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Econ. 99 (3), 483–499. http://dx.doi.org/10.1086/261763.
Krugman, P., 1993. On the number and location of cities. Eur. Econ. Rev. 37 (2), 293–298. http://dx.doi.org/10.1016/0014-2921(93)90017-5.
Lehe, L., Gayah, V.V., Pandey, A., 2021. Increasing returns to scale in carpool matching: evidence from scoop. Findings http://dx.doi.org/10.32866/001c.25093.
Lehe, L.J., Pandey, A., 2020. Hyperdemand: A static traffic model with backward-bending demand curves. Econ. Transp. 24, http://dx.doi.org/10.1016/j.ecotra.2020.100182.
Lehe, L.J., Pandey, A., 2024. A bathtub model of transit congestion. Transp. Res. B 181, 102892. http://dx.doi.org/10.1016/j.trb.2024.102892, URL: https://www.sciencedirect.com/science/article/pii/S019126152400016X.
Levinson, H.S., 1983. Analyzing transit travel time performance. Transp. Res. Rec. (915), URL: https://trid.trb.org/view/202380.
Li, Z.-C., Lam, W.H.K., Wong, S.C., 2012. Modeling intermodal equilibrium for bimodal transportation system design problems in a linear monocentric city. Transp. Res. B 46 (1), 30–49. http://dx.doi.org/10.1016/j.trb.2011.08.002.
Li, X., Liu, W., Yang, H., 2018. Traffic dynamics in a bi-modal transportation network with information provision and adaptive transit services. Transp. Res. C 91, 77–98. http://dx.doi.org/10.1016/j.trc.2018.03.026.
Li, X., Yang, H., 2016. Dynamics of modal choice of heterogeneous travelers with responsive transit services. Transp. Res. C 68, 333–349. http://dx.doi.org/10.1016/j.trc.2016.04.014.
Lin, T.-m., Wilson, N.H.M., 1992. Dwell time relationships for light rail systems. Transp. Res. Rec. (1361), 287–295, URL: https://trid.trb.org/view/370918.
Liu, H., Devunuri, S., Lehe, L., Gayah, V.V., 2023. Scale effects in ridesplitting: A case study of the city of chicago. Transp. Res. A 173, 103690. http://dx.doi.org/10.1016/j.tra.2023.103690.
McKnight, C.E., Levinson, H.S., Ozbay, K., Kamga, C., Paaswell, R.E., 2004. Impact of traffic congestion on bus travel time in northern new jersey. Transp. Res. Rec. 1884 (1), 27–35. http://dx.doi.org/10.3141/1884-04.
Milkovits, M.N., 2008. Modeling the factors affecting bus stop dwell time. Transp. Res. Rec. 2072 (1), 125–130. http://dx.doi.org/10.3141/2072-13.
Mogridge, M.J., 1997. The self-defeating nature of urban road capacity policy: A review of theories, disputes and available evidence. Transp. Policy 4 (1), 5–23. http://dx.doi.org/10.1016/S0967-070X(96)00030-3.
Mohring, H., 1972. Optimization and scale economies in urban bus transportation. Amer. Econ. Rev. 62 (4), 591–604. http://dx.doi.org/10.2307/1806101.
Morris, S., Shin, H.S., 2000. Rethinking multiple equilibria in macroeconomic modeling. NBER Macroecon. Annu. 15, 139–161. http://dx.doi.org/10.2307/3585391, arXiv:3585391.
Murphy, K.M., Shleifer, A., Vishny, R.W., 1989. Industrialization and the big push. J. Polit. Econ. 97 (5), 1003–1026, arXiv:1831884. URL: https://www.jstor.org/stable/1831884.
Myrdal, G., 1957. Economic Theory and Under-developed Regions. G. Duckworth.
Pandey, A., Lehe, L.J., Gayah, V.V., 2024. Local stability of traffic equilibria in an isotropic network. Transp. Res. B 179, 102873. http://dx.doi.org/10.1016/j.trb.2023.102873.

Plott, C.R., Smith, J., 1999. Instability of equilibria in experimental markets: upward-sloping demands, externalities, and fad-like incentives. South. Econ. J. 65 (3), 405–426. http://dx.doi.org/10.2307/1060807, arXiv:1060807.

Reinhold, T., Kearney, A.T., 2008. More passengers and reduced costs—the optimization of the berlin public transport network. J. Public Transp. 11 (3), 57–76. http://dx.doi.org/10.5038/2375-0901.11.3.4.

Samuelson, P.A., 1941. The stability of equilibrium: comparative statics and dynamics. Econometrica 9 (2), 97–120. http://dx.doi.org/10.2307/1906872, arXiv:1906872.

Savage, I., 2004. Management objectives and the causes of mass transit deficits. Transp. Res. A 38 (3), 181–199. http://dx.doi.org/10.1016/j.tra.2003.10.003.

Smith, M.J., 1984. The stability of a dynamic model of traffic assignment—an application of a method of Lyapunov. Transp. Sci. 18 (3), 245–252, arXiv:25768135. URL: https://www.jstor.org/stable/25768135.

Strogatz, S.H., 2015. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, Second Edition. CRC Press.

Tirachini, A., 2013. Bus dwell time: The effect of different fare collection systems, bus floor level and age of passengers. Transp. A 9 (1), 28–49. http://dx.doi.org/10.1080/18128602.2010.520277.

Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp. Res. A 53, 36–52. http://dx.doi.org/10.1016/j.tra.2013.06.005.

Tirachini, A., Hensher, D.A., Rose, J.M., 2014. Multimodal pricing and optimal design of urban public transport: the interplay between traffic congestion and bus crowding. Transp. Res. B 61, 33–54. http://dx.doi.org/10.1016/j.trb.2014.01.003.

Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511805271.

Verhoef, E.T., 1999. Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing. Reg. Sci. Urban Econ. 29 (3), 341–369. http://dx.doi.org/10.1016/S0166-0462(98)00032-5.

Walters, A.A., 1961. The theory and measurement of private and social cost of highway congestion. Econometrica 29 (4), 676–699. http://dx.doi.org/10.2307/1911814, arXiv:1911814.

Wardman, M., Chintakayala, V.P.K., de Jong, G., 2016. Values of travel time in europe: review and meta-analysis. Transp. Res. A 94, 93–111. http://dx.doi.org/10.1016/j.tra.2016.08.019.

Watling, D., 1996. Asymmetric problems and stochastic process models of traffic assignment. Transp. Res. B 30 (5), 339–357. http://dx.doi.org/10.1016/0191-2615(96)00006-9.

Watling, D., 1999. Stability of the stochastic equilibrium assignment problem: A dynamical systems approach. Transp. Res. B 33 (4), 281–312. http://dx.doi.org/10.1016/S0191-2615(98)00033-2.

Ying, J.Q., Yang, H., 2005. Sensitivity analysis of stochastic user equilibrium flows in a Bi-modal network with application to optimal pricing. Transp. Res. B 39 (9), 769–795. http://dx.doi.org/10.1016/j.trb.2003.09.003.

Zhang, F., Lindsey, R., Yang, H., 2016. The downs–Thomson paradox with imperfect mode substitutes and alternative transit administration regimes. Transp. Res. B 86, 104–127. http://dx.doi.org/10.1016/j.trb.2016.01.013.

Zhang, F., Yang, H., Liu, W., 2014. The downs–Thomson Paradox with responsive transit service. Transp. Res. A 70, 244–263. http://dx.doi.org/10.1016/j.tra.2014.10.022.

Zhang, F., Zheng, N., Yang, H., Geroliminis, N., 2018. A systematic analysis of multimodal transport systems with road space distribution and responsive bus service. Transp. Res. C 96, 208–230. http://dx.doi.org/10.1016/j.trc.2018.09.009.