Title
First-Best Downtown Transportation Systems in the Medium Run

Authors
Arnott, Richard
Rowse, John

Publication Date
2011-08-15

University of California Transportation Center
UCTC-FR-2011-16

First-Best Downtown Transportation Systems in the Medium Run

Richard Arnott and John Rowse
University of California, Riverside
September 2011

FIRST-BEST DOWNTOWN TRANSPORTATION SYSTEMS IN THE MEDIUM RUN

Richard Arnott[+]* and John Rowse[++]

August 15, 2011

Abstract:

Keywords: bus, subway, auto, congestion, transportation system, downtown

JEL Codes:

* Corresponding author
[+] Department of Economics, University of California, Riverside, Riverside CA 92506, USA. richard.arnott@ucr.edu.
[++] Department of Economics, University of Calgary, Calgary, AB xxxx, Canada. rowse@ucalgary.ca

First-best Downtown Transportation Systems in the Medium Run

This paper investigates first-best downtown transportation systems in the medium run for a broad range of demand densities. A downtown transportation system is assumed to include a subway system that operates on its own network and a congestible street system that accommodates both buses and cars. A "subway" is any mass transit mode that operates on an exclusive right of way; a "bus" is any mass transit mode for which there congestion interaction with cars[1]. The analysis is "medium run" in the sense that the subway and road networks are fixed, as are their link capacities, and "first-best" in the sense that the planner faces only technological constraints. The analysis is "downtown" only in the sense that it focuses on high levels of demand density that for most metropolitan areas occur only downtown. The analysis is static (stationary state), ignoring the intra-day dynamics of travel and congestion.

The design of optimal transportation systems is a classic problem, and many facets of it have been considered in the literature. Particularly noteworthy are Meyer, Kain, and Wohl (1965), Mohring (1972), and Kraus (1991), and a recent, state-of-the-art contribution is Tirachini and Hensher (2011). Meyer, Kain, and Wohl investigated the relationship between per passenger operating cost and demand density for different transport modes, with an interest in determining the cost-minimizing mode by demand density. Mohring (1972) made a seminal contribution in identifying economies of scale in mass transit that operate through waiting time and walking time. A doubling of demand density can be accommodated by a doubling of service frequency, which reduces average waiting time, or by a doubling of network density, which reduces average walking time. Kraus was the first to model explicitly loading/unloading and discomfort costs. Tirachini and Hensher examine the cost-minimizing design of a designated bus corridor, adding to the usual set of decision variables vehicle capacity, fare payment

[1] According to this definition, a light rail transit system is a bus system if it contributes to auto congestion, while a designated bus corridor is a subway system since it does not contribute to auto congestion.

system, and running speed, and considering bus queues at high-demand stops.

Our paper draws on our earlier work (Arnott and Rowse, 2006) to make several contributions to this line of literature. It is the first paper, to our knowledge, to consider the optimal design of a downtown transportation system with the three generic modes. The properties of an optimal urban transportation system when two of the modes are subject to Mohring-type economies of scale are of special interest. If bus and subway travel were characterized by decreasing costs over the entire range of demand intensities, it would be optimal to employ only one of the modes, depending on their technological parameters and on the level of demand intensity. But, as modeled here, there are two reasons this may not occur: the first is congestion between buses, and between buses and cars; the second is that, in the medium run considered, the subway network may be subject to capacity constraints but the bus network is not. The paper is also the first to model the congestion interaction between buses and cars using the sounder, traffic engineering approach in which speed depends on the density rather than the flow of traffic, and the first in this branch of the literature to treat curbside and garage parking explicitly. Finally, because of the nonconvexities introduced by Mohring-type economies of scale, there may be multiple local optima. Many of the optimization programs that we experimented with identified either an inferior local optimum or even a corner minimum as the global optimum. The paper illustrates the application of a simple and intuitive but effective decomposition procedure to deal with the problem.

Mohring-type economies of scale introduce essential non-convexities into the problem. When essential non-convexities are present, formal analysis with general functional forms is very difficult. For that reason, we proceed by working with an extended numerical example. Our results identify several qualitatively different optima, but not necessarily the entire set. Our numerical work uncovered a number of interesting results.

In the near future we plan to adapt the model developed in this paper to second-best policy analysis. Distortions will be added as constraints to the planner's optimization problem, along with the equilibrium constraints that a particular mode is active in the second-best optimum only if its full price is less than or equal to the full price of all inactive modes. In particular, we plan to investigate the effects of *maximum* downtown parking requirements, which have been introduced in many cities in recent years, including Boston, San Francisco, ---- . Because auto congestion is underpriced, the auto modal share is excessive. By restricting the amount of private parking available downtown, maximum parking requirements aim to induce modal switching to mass transit. The policy has the potential added benefit of exploiting economies of scale in mass transit. Is the policy really as effective as it appears at first glance, or might its effectiveness be seriously undermined by an increase in cruising for parking or by mass transit capacity constraints? The model could be applied to a host of other second-best policy issues. One of particular interest would be to re-examine Parry and Small's (200x) conclusion that, at least in the cities they consider, mass transit should be even more heavily subsidized than it is, from the somewhat different modeling perspective of this paper.

Future work should also enrich the model in the direction of realism to include heterogeneity of travelers, intra-day traffic dynamics, locational differentiation, comfort, and travel time reliability, underpriced curbside parking, the behavior of parking garage operators, whether public or private, different transportation authorities with conflicting objectives, and the political economy of downtown transportation.

Section 2 presents the model. Section 3 provides a detailed numerical analysis of the base case optimum, focusing on how the optimum downtown transportation system changes with demand density. Section 4 presents some numerical comparative static results, investigating how the characteristics of the first-best downtown transportation system change with parameter values. Section 5 discusses directions for future research, and concludes.

2. The Model

This section starts with a thumbnail sketch of the model, then displays the notation, then presents the model equations, then specifies the full constrained resource cost minimization, and concludes by listing the base-case parameter values and explaining how they were chosen.

2.1 Thumbnail Sketch

The model describes a self-contained isotropic downtown in stationary state with three generic transport modes: auto, bus, and subway. There is an exogenous demand per unit area-time for fixed-length trips that varies neither over time nor over space. All travel decisions are made by a benevolent planner, with the aim of minimizing resource costs per unit area time.

There is a road system[2] and a subway system. The road system is shared by autos and buses and suffers from congestion. Subway cars do not interact with buses or autos, and if they interact with each other do so only via headway or platform capacity constraints. The road and subway systems have already been constructed. For the road system, this means that road capacity is fixed, for the subway system, that station spacing is fixed. Apart from deciding how to allocate trips over modes, the planner has eight decision variables: for the car, the density of curbside and garage parking spaces; for the bus, headway, passengers per bus, and bus-stop spacing; and for the subway, headway, passengers per subway car, and number of subway cars per train.

- *Auto travel*

[2] The paper considers a Manhattan grid road network of one-way streets, and a Manhattan grid subway network of one-way subway lines.

All individuals live downtown[3] and have access to a car. At home an individual parks her car in her individual garage. An auto trip entails travel for a fixed distance from home to the destination, either curbside or garage parking at the destination during the visit, which is fixed in duration, and a return journey home. Traffic congestion is modeled as classic (Lighthill-Whitham-Richards) flow congestion, with travel speed depending on the endogenous density of cars in transit and of buses, as well as the density of curbside parking spaces[4]. Road capacity is fixed. There are two components of the social cost of a trip, travel time cost and parking cost. The value of travel time is taken as fixed. Garage parking is produced at constant cost per unit time. The total social cost of auto travel per unit area-time equals auto throughput[5] per unit-area time times average travel time plus parking cost per car.

The social planner decides on the density of curbside parking spaces and of garage spaces per unit area. The density of cars in transit is endogenous.

- *Bus travel*

On a bus trip, an individual walks from her home to the nearest bus stop, waits for the bus, travels on the bus to the bus stop closest to her destination[6], walks from there to her

---

[3] An alternative interpretation of the model is that all individuals live in the suburbs and take trips to a downtown area, with the model describing their travel in the downtown area.

[4] This modeling of traffic congestion was first introduced in Arnott, Rave, and Schöb (2005), Chapter 2, and follows that employed in the transportation engineering literature. In that chapter, as well as in subsequent papers (Arnott and Inci, 2006; and Arnott and Rowse, 2010), traffic congestion depends as well on the stock of cars cruising for parking. The current paper focuses on first-best planning optima in which there is no cruising for parking.

[5] In keeping with our previous work, we make a terminological distinction between throughput whose units are units of traffic (e.g. cars or bus travelers) per unit area-time, and flow whose units are units of traffic per unit time. The Fundamental Identity of Traffic Flow is that flow equals density times velocity. *Flow* equals density, measured in traffic units *per unit time on a street*, times velocity, and is hence measured in units of traffic per unit time. In stationary state, the flow of trip originations per unit area-time equals the flow of trip terminations per unit area-time, and it is this that we refer to as throughput. Throughput equals the density of traffic per unit area divided by mean trip distance times velocity.

[6] The model assumes that an individual walks to the closest bus stop and travels on a single bus from her origin to her destination, which implicitly assumes that her destination is in the same direction as the direction of the bus at that stop. Footnote xxx

destination, and does the same on the return journey. Her travel time cost includes walking time cost, waiting time cost, and in-bus travel time cost. In-bus travel time includes the time traveling in the regular flow of traffic, time embarking and disembarking passengers, and time decelerating into and accelerating out of bus stops. The speed of a bus in the regular flow of traffic is the same as that for a car, and a bus contributes to congestion an exogenous number of passenger-car equivalents. (PCE's). The values of walking time and waiting time differ from one another, and are exogenous. The value of in-bus travel time depends on the degree of crowding on the bus. Bus passenger costs per unit area-time equal the throughput of bus passengers per unit area-time times the travel time cost incurred by each passenger.

The costs of operating a bus per unit time equal the driver's wage plus other costs, which include capital, fuel, and maintenance costs. Bus operating costs per unit area-time therefore equal bus density per unit area times bus operating costs per unit time.

The social planner decides on bus stop spacing, bus headway, and the number of passengers per bus. The density of buses is endogenous.

Mohring-type economies of scale arise through both bus stop spacing and bus headway. Holding passengers per bus and bus stop spacing fixed, doubling passenger throughput is achieved by halving headway, and hence waiting time cost. Alternatively, holding passengers per bus and bus headway fixed, doubling passenger throughput is achieved by halving bus stop spacing and hence walking time cost. Thus, higher bus passenger throughput reduces average passenger waiting and walking time costs, but also increases traffic congestion.

- *Subway travel*

Subway travel is similar to bus travel, but differs in three important respects. First, the spacing between subway stations is taken as fixed, reflecting the very high costs of modifying the subway network; second, the social planner has an additional policy variable, the number of cars per subway train; third, there is no congestion interaction between subways and either buses or cars. Subways also differ from buses in technological parameters. In the base case, no constraint is put on subway headway or

---

sketches how this assumption could be relaxed this assumption, which would require treating transfers.

the number of cars per train. A doubling of subway passenger throughout can therefore always be achieved with less than a doubling of resource cost. Subsequently, consideration will be given to constraints on train length and headway, which bind at high subway passenger throughput , and result in a form of congestion interaction between subway cars.

2.2 Notation

INSERT THE NOTATION AS TABLE 1 WITH CAPITAL ROMAN LETTERS, FOLLOWED BY SMALL ROMAN LETTERS, THEN CAPITAL GREEK LETTERS, FOLLOWED BY SMALL GREEK LETTERS, WITH ALPHABETIZATION WITHIN EACH CATEGORY.

2.3 Model Equations

The planner faces the problem of minimizing the resource cost per unit area, RC, of satisfying the exogenous travel demand. We start by presenting the constraints on the minimization problem, and then present the expression for cost.

- *constraints*

The *stationary-state condition for cars* is

$$N_a = T/(\delta t(T,0,B,P). \quad (1)$$

The left-hand side is the equilibrium inflow rate into the car in-transit pool per unit area-time, N_a. The right-hand side is the steady-state outflow rate, which equals the density of cars in transit per unit area-time, T, divided the length of time each cars spends in transit, which is (return) trip distance, δ, times travel time per unit distance, t. Travel time per unit distance is a function of T, as well as the density of cars cruising for parking, C, the density of buses, B, and the density of curbside parking spaces, P: $t = t(T,C,B,P)$. In the social optimum, there is no cruising for parking so that $C = 0$.

The analogous, *stationary-state condition for buses* is

$$N_b = B/(\delta t(T,0,B,P) + \delta r^0{}_b/\Delta_b + 4n_b r^1{}_b). \quad (2)$$

The denominator is the time spent on a bus trip. The first term is bus cruising time, the second the fixed time a passenger loses on a trip from her bus decelerating into bus stops and accelerating out of them, and the third term is the time lost due to passengers embarking and disembarking. Bus cruising time is the same as in-transit car travel time, since buses and cars are assumed to travel at the same speed when in regular traffic. Bus stops are spaced Δ_b apart, so that a bus has (approximately) δ/Δ_b stops on a trip, and r^0_b is the fixed component of the time a bus loses at each stop accelerating and decelerating. n_b is the number of passengers per bus, and on a one-way journey this number of passengers embark and the same number disembark, so that the total number of embarkations and disembarkations per round trip is $4n_b$ and r^1_b is the time a passenger takes to either embark or disembark[7].

The analogous, *stationary-state condition for subways* is

$$N_s = S/(1/v_s + \delta r^0_s/\Delta_s + 4n_s r^1_s), \qquad (3)$$

where S is the density of subway trains per unit area-time. The only qualitative difference between this condition and the analogous condition for buses is that buses are slowed down by traffic congestion while subways cruise at the speed v_s.

There is also a conservation relationship between bus headway, bus stop spacing, trip distance, bus throughput per unit area time, and the number of passengers per bus, and an analogous condition for the subway. If stations are spaced a distance Δ apart on a Manhattan grid, then there are $2/\Delta$ lines traversing each unit area. If travel on each line is one-way, which we assume, then the number of trains/buses entering each unit area-time is $2/(\Delta h)$ and the number of passengers entering each unit area-time is $2nz/(\Delta h)$, where n is the number of passengers per car and z the number of cars per train. Throughput equals this quantity times the proportion of passengers who disembark per unit area. Thus, $N = 2n/(\delta\Delta h)$.

Since, by assumption, $z = 1$ for the bus, the *bus passenger conservation condition* is

[7] This specification ignores congestion in the embarkation and disembarkation process, which will be considered in section 4.3.

$$N_b = 2n_b /(\delta\Delta_b h_b). \quad (4)$$

The corresponding *subway conservation condition is*

$$N_s = 2n_s /(\delta\Delta_s h_s). \quad (5)$$

There is also the obvious *adding-up condition* that

$$N - N_a - N_b - N_s = 0. \quad (6)$$

The decision variables in the minimization problem are N_a, N_b, N_s, T, B, P, S, n_b, n_s, h_b, h_s, Δ_b, and z. Δ_s is not a decision variable since the subway network is taken as fixed. There are non-negativity restrictions on all the decision variables.

All the above constraints are equality constraints. There are also inequality constraints. The first is that the bus headway must exceed the length of time it takes a bus to discharge and load passengers at a stop[8]. The proportion of passengers that either gets on or off a bus at a particular stop approximately equals the stop spacing divided by journey distance (which gives the probability that a passenger gets off at a stop) multiplied by 2 (which gives the probability that a passenger gets either on or off at a stop): Δ/δ. It would therefore be reasonable to specify the *bus headway constraint* as

$$h_b \geq n_b\Delta_b r^1{}_b/\delta, \quad (7)$$

and the analogous *subway headway constraint* as

$$h_s \geq n_s\Delta_s r^1{}_s/\delta. \quad (8)$$

For the subway, platform length must exceed the length of a train. In examining the base case, we shall ignore these constraints.

There is a final constraint. Let N_{ac} be the throughput of car drivers who park curbside, N_{ag} be the throughput of car drivers who garage park, P be curbside parking capacity, and G garage parking capacity. Since in the model unutilized parking spaces impose a social cost but confer no benefit, $N_{ac}\lambda = P$ and $N_{ag}\lambda = G$. Furthermore, $N_a = N_{ac} + N_{ag}$. Finally, garage capacity cannot be negative. We could treat driving and curbside parking as being

---

[8] The constraint could easily be modified to allow for bus stops to accommodate more than one bus, or to

a different mode from driving and garage parking. Instead, we have chosen to treat driving as a mode, and to treat garage parking capacity only implicitly, in which case the non-negativity constraint on G reduces to

$$N_a\lambda \geq P. \qquad (9)$$

- *resource costs*

There are three direct resource costs associated with car travel: time costs, money costs, and garage parking costs. Time costs per unit area-time may be measured as ρT. Money costs per unit area-time, which include fuel costs and the variable component of depreciation and insurance, are assumed to be proportional to travel time rather than travel distance, and may therefore be measured as $g_a T$. The cost of a garage parking spot is c per unit time. The cost of garage parking per unit area-time is therefore cG, with $G \geq 0$. With G substituted out, this cost and constraint are represented as $(N_a\lambda - P)c$ and $(N_a\lambda - P)c \geq 0$. Curbside parking makes traffic congestion worse. The cost associated with this is captured in the time costs of car drivers and bus passengers. The total resource costs per unit area-time associated with auto travel are therefore

$$RC_a = (\rho + g_a)T + (N_a\lambda - P)c \quad \text{if } N_a\lambda - P > 0 \text{ (some garage parking)}$$
$$= (\rho + g_a)T \qquad \text{(no garage parking)} \qquad (10)$$

There are four direct resource costs associated with bus travel are operating costs, walking time costs, waiting time costs, and travel time costs. Operating costs are assumed to be proportional to bus density, are operating costs per bus per unit time are divided into the bus driver wage, W_b, and other operating cost, g_b, which include the amortized capital costs, maintenance, and fuel. Bus operating costs per unit area-time are therefore $(W_b + g_b)B$. Each bus passenger takes four walks, from home to the origin bus stop, from the destination bus stop to the destination, and the reverse when homebound. Since the average distance from home to a bus stop is $\Delta_b/4$, on average a bus passenger walks a distance Δ_b , which takes Δ_b/w units of time with walking speed w, and entails a cost of $\rho_w\Delta_b/w$ with a value of walking time of ρ_w. Walking time costs per unit area-time are therefore $N_b\rho_w\Delta_b/w$. Waiting time costs per unit area-time are $N_b\rho_W h_b$. Buses

are assumed to arrive at equally spaced intervals so that each passenger expects to wait a period of time $h_b/2$ for a bus[9], and therefore a period h_b for her return trip, at a constant of $\rho_W h_b$, where ρ_W is the cost of waiting time[10]. Crowding costs enter through travel time costs. In particular, the value of travel time is assumed to be proportional to the "volume-capacity" ratio, the ratio of the number of passengers per bus to bus capacity, κ_b, with the form $\rho^0_b + \rho^1_b(n_b/\kappa_b)^{\eta b}$, where ρ^0_b ,ρ^1_b, κ_b, and η_b are exogenous[11] parameters. Thus, travel time costs per unit area-time are $Bn_b[\rho^0_b + \rho^1_b(n_b/\kappa_b)^{\eta b}]$, and total bus travel costs per unit are-time are

$$RC_b = (W_b + g_b)B + N_b(\rho_w\Delta_b/w + \rho_W h_b) + Bn_b[\rho^0_b + \rho^1_b(n_b/\kappa_b)^{\eta b}]. \qquad (11)$$

Apart from possible differences in parameter values, the costs of subway travel per unit area-time differ in only two respects from those for bus travel. First, the operating costs of a subway train per unit time are taken to include a fixed component, the driver's wage, and a variable component that is linearly proportional to the number of cars in a train. Second, a subway passenger's walking costs include the costs of walking between the street and the platform, which is Γ per one-way journey. Total subway travel costs per unit area time are therefore

$$RC_s = (W_s + zg_s)S + N_s[\rho_w(\Delta_s/w + 2\Gamma) + \rho_W h_s] + Sn_s[\rho^0_s + \rho^1_s(n_s/\kappa_s)^{\eta s}]. \qquad (12)$$

---

[9] One interpretation of this is passengers do not know the bus schedule. Another is that passengers know the bus schedule and so do not actually wait for a bus but instead must travel at an inconvenient time, which generates a schedule delay cost.

[10] Footnote xx pointed out that the model assumes that an individual travels to the closest bus stop or station, and travels in the direction of the bus or train there THIS FOOTNOTE NEEDS TO BE MODIFIED FROM HERE, which is unrealistic. With complicating the model, it could alternatively be assumed that each passenger makes one transfer, which would double waiting time. A fully satisfactory treatment would optimize the route network, and solve for the resource-cost minimizing route for each traveler.

[11] In a fuller model, these, and many other, exogenous parameters could be treated as decision variables; for example, ρ^0_b reflects the comfort of the bus, absent crowding, while the form of the value-of-travel-time function reflects the bus design, such as the ratio of seating capacity to capacity.

- form of the road congestion technology

The road congestion technology combines Greenshields' Relation with the assumption that a bus' contribution to congestion can be represented in terms of passenger-car-equivalents (PCE's). Greenshields' Relation states that there is a negative linear relationship between velocity and density,V, $v = v_0(1 - V/V_j)$, where v_0 is free-flow velocity and V_j density. Since travel time per unit distance is inversely proportional to velocity, this may be rewritten as $t = t_0/(1 - V/V_j)$ or $t = t_0V_j/(V_j - V)$. Finally, it is assumed that jam density equals maximum jam density, Ω – the jam density when no road space is devoted to parking – times the proportion of road space that is devoted to traffic flow rather than parking, $1 - P/P_{max}$, where P_{max} is maximum density of curbside parking spaces per unit area-time. Thus,

$$t(T,C,B,P) = t_0\Omega(1 - P/P_{max})/[\Omega(1 - P/P_{max}) - T - \theta C - \theta_b B]. \quad (13)$$

In the planning problem, there is no cruising for parking, so $C = 0$.

2.4 *The Planning Problem*

The planner's problem is to minimize the resource costs per unit area-time subject to the constraints listed above. The complete minimization problem is given below.

COPY FROM JOHN'S LETTER TO GAMS, AND PERHAPS ADD THE HEADWAY CONSTRAINTS.

(14)

The decision variables are N_a, N_b, N_s, T, B, P, n_b, n_s, h_b, h_s, Δ_b, and z.

The time horizon of the planning problem is the medium run, with the road and subway networks and link capacities taken as fixed[12].

[12] The analysis could be extended to treat the long-run planning problem by making Δ_s a decision variable, rather than a parameter, and by optimizing the proportion of land to allocate to road space. The latter would require the specification of a more complete model in which downtown land is allocated between road space and other uses. Allocating more land to road space would require constructing taller buildings and/or allocating less workspace to each worker.

Before proceeding, it will be useful to highlight some features of the model.

1. Note that the problem can be decomposed into three sub-problems. The first is the minimization of subway resource costs, subject to a fixed throughput of subway travelers; the second is the minimization of road resource costs, which includes the resource costs associated with both bus and car travel, subject to a fixed throughput of road travelers; the third is to allocate the population between subway travel and road travel so as to minimize total resource costs, subject to an overall population constraint. This decomposition is possible because there is congestion interaction between car drivers and bus passengers, but not between road and subway users[13].

2. There are decreasing costs to subway travel if neither the subway headway or subway platform capacity constraint bind. A doubling of subway passengers can be accommodated via a doubling of the number of cars per train, holding n_s and h_s fixed, which leaves the user cost unchanged and distributes the driver's wage over a large number of train passengers. A doubling of subway passengers can also be accommodated via halving the train headway, holding n_s and z fixed, which lowers average passenger waiting time cost while leaving operating costs per passenger unchanged. For populations of subway passengers where the subway headway capacity constraint binds, but not the platform capacity constraint, decreasing costs should prevail since an increase in subway passengers can still be accommodated via an increase in train length. For even higher levels of population, where both constraints bind, an increasing number of subway travelers can be accommodated only through an increase in passengers per car and hence increased per passenger crowding costs, and above some threshold level increasing costs set in.

Intuition suggests that as the number of road travelers from zero, average road user resource costs first rise, then fall, then rise again. When the number of road travelers

---

[13] An extension of the model would be to incorporate pedestrians. This would be important if there were a distribution of trip lengths, so that some individuals would choose to walk on shorter trips. There would also be the walking involved on bus and subway trips. When trip density is high, pedestrian congestion becomes important. Pedestrians impose congestion costs not only on one another but also on road users. In such a model, pedestrians would introduce a form of congestion interdependence between subway and road travel.

is small, with reasonable parameter values, all should travel by car, and car travel suffers from congestion. At a threshold level of road travelers, it should become profitable to start operating a bus system at a non-zero scale. As the number of road travelers is increased above this level, there are two offsetting effects. On one hand, there are decreasing waiting and walking costs for bus travel, while, on the other, road congestion increases. The former may dominate for smaller number of road travelers, the latter should dominate for larger numbers.

3. The model abstracts from a number of potentially important considerations, including heterogeneity of users, pedestrian traffic, time-of-day variation, locational differentiation, and comfort.

2.5 *Parameter Values*

The base-case parameter values are given in Table 2.

DISCUSSION OF PARAMETER VALUES ON THE BASIS OF THE VALUES USED BY JOHN AND THE NOTES OF JULY 12, 2010.

3. Base-Case Optimum

This section examines how the first-best downtown transportation system changes as demand density increases for the base case set of parameters values given in Table 2. We could just present our results, but to facilitate both exposition and comprehension have decomposed the problem. Section 3.1 presents results when all travel is by auto, section 3.2 when all travel is by bus, section 3.3 when all travel is by subway, section 3.4 when all travel is by either bus or subway, and section 3.5 when all three modes are present.

3.1 All Travel is by Auto

Table 3 presents results for four different levels of demand density, N = 3708, 7416, and 14828, which correspond to 25%, 37.5% 50%, and 100% of the base-case demand intensity of 14828.

variable	N = 3707	N = 5560.5	N = 7117.44	N = 7414	N = 14828
t	0.07061	0.07061	0.0833	0.10	*
T	1047	1571	2372	2966	*
P	4403	1037	0	0	*
APC_a^+	5.437	8.441	9.0	9.0	*
ATC_a^+	7.768	7.768	9.163	11	*
ARC_a^+	13.20	16.20	18.17	20	*
MRC_a^+	22.22	22.22	36.46	∞	*

Table 3: Auto-only Optima
Notes: * indicates that this the road system has insufficient capacity to support this level of demand intensity with only auto travel.
$^+$ ARC denotes average resource cost per capita with auto only, and MRC marginal resource cost. Subscript a denotes auto only.

DISCUSSION

3.2 All Travel is by Bus

Table 4 presents the bus-only optima for the same levels of demand density as did Table 3, as well as for a level of demand intensity equal to twice that of the base case value. Here, and throughout this section, we take θ_b – the passenger-car equivalents of a bus -- to be 2.0. This is the standard value assumed (e.g., Parry and Small, 20xx) though we believe it to be unrealistically low and in the next section compare the results obtained with our preferred estimate of $\theta_b = 7.5$.

variable	N = 3707	N = 5560.5	N = 741	N = 14832	N = 29664
t	0.0515	0.0522	0.0527	0.0558	0.0615
B	83.86	122.6	154.6	305.5	556.59
Δ_b	0.2235	0.2212	0.2205	0.2215	0.2293
h_b	0.01974	0.00752	0.006033	0.003133	0.001740
n_b	17.80	18.50	18.94	20.58	23.67
AOC_b^+	2.715	2.645	2.606	2.472	2.252
AwC_b^+	6.704	6.637	6.616	6.646	6.880
AWC_b^+	0.5370	0.3760	0.3017	0.1566	0.08702

ATC_b^+	10.95	11.17	11.31	11.85	12.84
ARC_b	20.91	20.83	20.83	21.12	22.06
MRC_b	20.58	20.76	20.93	21.89	24.18

Table 4: Bus-only Optima
Notes: $^+$ AwC denotes average walking cost, AWC average waiting cost, and ATT average travel time cost. Subscript b denotes bus only

DISCUSSION

3.3 All Travel is by Subway

Table 4 presents the subway-only optima for the same levels of demand density as did Table 3. In this base case, we ignore both the subway headway constraint and the train length constraint. As a result, there are decreasing costs through the entire range of demand density.

variable	N = 3707	N = 5560.5	N = 7416	N = 14832	N = 29664
t					
S	21.19	25.96	29.37	42.40	59.97
z	1.067	1.302	1.470	2.112	2.978
h_s	0.006859	0.005601	0.004952	0.003431	0.002427
n_s	23.83	23.92	23.97	24.09	24.16
ATO_s^+	1.014	0.9486	0.9148	0.8360	0.7838
AwC_s^+	21	21	21	21	21
AWC_s^+	0.3429	0.2801	0.2476	0.1716	0.1213
ATC_s^+	3.750	3.753	3.754	3.757	3.758
ARC_s	26.22	25.98	25.92	25.76	25.66
MRC_s	25.76	25.70	25.67	25.59	25.54

Table 4: Subway-only Optima
Notes: $^+$ AwC denotes average walking cost, AWC average waiting cost, and ATT average travel time cost. Subscript s denotes subway only

DISCUSSION

Figure 1, Panel A shows ARC for each of the three modes when only that mode is employed, as a function of demand density. Panel B displays the same diagram but for MRC.

DISCUSSION

3.4 All Travel is by Road (Auto or Bus)

Figure 3 shows minimum total resource cost per unit area time , RC, as a function of the modal split between auto and bus when all travel is by auto or bus and demand density equals the base-case demand density of 14828. The density of auto travelers is measured right from 0_a and the density of bus travelers left from 0_b such that $N_a + N_b = N_r$. With the assumed parameter values and demand density, it is efficient for almost all road users to travel by bus, with only a small proportion traveling by auto.

It will be insightful to digress briefly to consider the intuition for the result. Total resource costs per unit area-time equal the sum of total auto resource costs per unit area time, given by (10), and total bus resource costs per unit area time, given by (11). Now, by the Envelope Theorem, we may suppose that the planner adjusts bus policy only by increasing bus headway. Thus, the net social benefit from the passenger transfer when the marginal car traveler parks in a parking garage[14] is

$$dRCC/dN_a = (\rho + g_a)dT/dN_a + \lambda c + (W_b + g_b)dB/dN_a - \rho_w \Delta_b/w + n_b[\rho_{0b} + \rho_{1b}(n_b/\kappa_b)^{\eta_b}]dB/dN_a , \quad (15)$$

where dT/dN_a and dB/dN_a are obtained from total differentiation of (1) and (2) with $dP/dN_a = 0$. And the net social benefit from the passenger transfer when the marginal car traveler parks curbside is

$$dRC/dN_a = (\rho + g_a)dT/dN_a + (W_b + g_b)dB/dN_a - \rho_w \Delta_b/w + n_b[\rho_{0b} + \rho_{1b}(n_b/\kappa_b)^{\eta_b}]dB/dN_a , \quad (16)$$

[14] The full derivation is given in the Appendix.

where dT/dN_a and dB/dN_a are obtained from total differentiation of (1) and (2) with $dP/dN_a = \lambda$. At an interior minimum, $dRCC/dN_a = 0$ and $d^2RCC/dN_a^2 > 0$.

When $N_a = 0$ and when, at $N_a = 0$, it is efficient for the marginal auto traveler to park curbside, as is the case in the current example, (16) reduces to

$$[dRC/dN_a]_{Na=0} = (\rho + g_a)\delta t(0,0,B,P) + \lambda c - \rho_w \Delta_b/w \qquad (17)$$

$$- (W_b + g_b + n_b [\rho_{0b} + \rho_{1b}(n_b/\kappa_b)^{\eta b}])\{-N_b \delta\theta_b t_T + n_b\}^{-1}[\delta r_{0b}/\Delta_b + 4n_b r_{1b} + \delta t(,0,B,P)(1 - N_b \delta t_T)]).$$

THIS IS THE EXPRESSION WHEN THE MARGINAL AUTO TRAVELER PARKS IN A GARAGE AND THEREFORE NEEDS TO BE MODIFIED.

Thus, when all travel is by bus, the effect on resource cost of having a single traveler switch from bus to car can be decomposed ----

Performing the above exercise for different levels of N, we may obtain resource costs and other variables of interest, as a function of demand density, with only road travel, $RC_r(N)$, from which $ARC_r(N)$ and $MRC_r(N)$ may be derived. Table 5 displays the results for the same levels of demand density as Tables 3 and 4.

DISCUSSION

3.5 All Three Modes

When we initially solved the three-mode problem numerically, we did so as the constrained optimization problem (14). Recognizing that the non-convexities caused the Mohring-type economies of scale in bus and subway transportation would likely cause numerical problems, we solved the problem with a variety of different solvers. Disconcertingly, the solvers came up with different solutions. Some identified corner minima, others identified a local minimum that was not the global minimum. Only one obtained the correct solution!

variable	N = 3708	N = 5560.5	N = 7416	N = 14832	N = 29664
t					
T					
P					
B					
Δ_b					
h_b					
n_b					
N_a					
N_b					
APC_r					
AwC_r					
AWC_r					
ATC_b					
ARC_r					
MRC_r					

Table 4: Road-only Optima
Notes: 1. APC gives total parking cost divided by total population. Subscript r denotes road only

To deal with this problem, we decomposed the problem using the components that have been developed in the previous subsections. Since the resource cost of subway travel is independent of the number of road users, and *vice versa*, reflecting the congestion independence of road and subway travel, we solved for first-best subway travel as a function of the number of subway travelers, as described in section 3.3, $RC_s(N_s)$, and then for first-best road travel as a function of the number of road users, $RC_r(N_r)$, as described in section 3.4 above. The full optimum may then be obtained simply as the solution to

$$\text{Min } RC(N) = RC_s(N_s) + RC_r(N_r) \quad \text{s.t.} \quad N_s + N_r = N. \tag{18}$$

We used this procedure to obtain an approximate solution, and then used the approximate solution as the starting point of the numerical solution of (14).

Figure 4 displays $RC_s(N_s)$ and $RC_r(N_r)$, with the constraint $N_s + N_r = N$ imposed. $RC(N; N_a, N - N_a)$ is obtained as the vertical summation of the two curves in this space. RC(N) is the minimum of this curve[15]. The global resource cost minimum in the base case example entails only auto and bus travel, and so is the same as the road-only resource cost minimum with N = 14832, which is shown in the second rightmost column in Table 4.

4. Sensitivity Analysis

This section perform sensitivity analysis. The analysis is not exhaustive but rather illustrates the logic of how the first-best downtown transportation system changes as selected parameter values are altered, and in the process makes some policy-relevant points. We consider three changes: i) an increase in θ_b from 2.0 to 7.5; ii) introducing headway and platform capacity constraints; and iii) increasing the crowding-cost parameters.

4.1 Increasing θ_b from 2.0 to 7.5

On the basis of industry practice, Parry and Small (20xx) assume that a bus generates 2.0 PCE's of congestion. Our guess is that the actual figure is considerably higher, and around 7.5 in heavily congested traffic[16]. A bus that is loading or unloading passengers or decelerating into or accelerating out of a stop blocks an entire lane of traffic for most of a city block, and a bus that is making a turn reduces intersection capacity for several seconds. Figure 5, Panel A reproduces Figure 3, indicating how resource cost as a function of the modal split between auto and bus change with the increase in θ_b, at a demand density of 14832. Figure 5, Panel B, reproduces Figure 4, indicating how

[15] A similar construction using the corresponding marginal social cost curves would identify interior minima, but identifying the global minimum would require comparing areas.

[16] Daganzo (xxxx) states that the PCE of a bus increases with the degree of congestion.

resource cost as a function of the split between road and subway travel changes with the increase in θ_b.

In Panel A, the increase in θ_b causes the resource cost curve to shift up and to tilt so that the optimal modal split between auto and bus, holding fixed road passenger density, shifts towards auto travel. Panel B is similar, showing how optimal modal split between subway and road shifts towards the subway, holding overall passenger density fixed. The optimal downtown transportation system shifts from one with only bus and auto to one with only subway and auto. Overall resource costs increase substantially, from $20.92 per trip to $24.42 per trip.

This numerical example illustrates the sensitivity of the optimal downtown transportation system to bus PCE's and points to the potential value of obtaining better estimates of the congestion function relating auto travel speed to bus density and other relevant variables.

4.2 Headway and Platform Capacity Constraints

As explained in section 2.5, the base-case demand density was chosen to represent the downtown of a metropolitan area with a population of 1 to 2 million in developed, western countries. Most cities with such moderate demand densities do not have a subway system, and, whether or not they do, headway and platform capacity constraints are unlikely to bind. But with substantially higher demand densities, they can be expected to. If they are ignored, the per capita resource cost of the optimal metropolitan transportation system asymptotically approaches a modest upper bound as population increases without any change in the road or subway system. The reason is that the subway system exhibits decreasing cost over the full range of demand densities, since increases in subway passenger density are accommodated by increases in train length and decreases in headway.

We provide only a crude treatment of headway and platform length constraints, assuming that platform length equals the length of 4 cars and that subway headway cannot fall

below 90 seconds. With these constraints, a maximum of $(4)\kappa_s(40) = 4(133)(40) = 21280$ passengers may pass through a subway station in a particular direction per hour. With the assumed $\Delta_s = 0.5$, there are two subway lines in each of the north, south, east, and west directions, so that a maximum of 170240 passengers may travel through a square mile of subway system per hour. Since trip length is 4, the maximum throughput is 47560 passengers per ml^2-hr

Figure 6 displays per passenger subway resource cost as a function of subway demand density for densities ranging from 0 to 59312 – four times the maximum level of demand density considered in the previous section. An obvious but important point is that, at high levels of demand density, subway travel exhibits increasing costs. When the headway and platform capacity constraints bind, with fixed spacing between subway stations, the only way increased passenger density can be accommodated is through an increase in passengers per car, which raises crowding costs. When subway travel exhibits increasing costs, the optimal downtown transportation system may include all three modes. Another obvious point is that a transportation system has a maximum capacity.

The current wisdom is that Mohring-type economies of scale support the heavy subsidization of mass transit travel. This may not be the case in large metropolitan areas where the subway system is so capacity constrained that crowding costs become substantial, as is currently the case in Paris, London, and Tokyo.

4.3 Raising Crowding Cost Parameters

There are two important form of crowding costs – the discomfort associated with traveling in a crowded subway car, and the congestion associated with embarking and disembarking. The base-case parameters that characterize crowding costs are conservative -- the value of time associated with traveling in a completely full subway car or train are only twice those of traveling in an empty train. And, at capacity, the time it takes to for 20% of the passengers to disembark and for 20% to embark is only 21.6

seconds for the bus and 27 seconds for the subway. The elasticity of the value of travel time with respect to the volume-capacity ratio was assumed to be equal to 2.0, and the variable time associated with loading and unloading passengers at a stop or station was assumed to be proportional to the number of passengers getting on or off and independent of volume-capacity ratio.

In this subsection, we assume that the value of time associated with traveling in a completely full subway car or train is four times that of traveling in an empty train, while retaining the assumption that the elasticity of the variable component of the value of travel time with respect to the volume-capacity ratio equals 2.0 (the parameters ρ^1_b and ρ^1_s increase from \$25.00 to \$75.00). Also, we replace the base case's assumptions that the time it takes to load or unload n passengers at a bus stop is $0.00025n_b$ (0.9 seconds per passenger) and $0.00014n_s$ for a subway car (0.5 seconds per passenger), with the assumptions that the corresponding times are $0.0005n_b + 0.0015n_b^2/\kappa_b$ (1.8 seconds per passenger on an empty bus and 7.2 seconds on a full bus) for the bus and $0.00028n_s + 0.0015n_s/\kappa_s$ (1 second per passenger on an empty subway car and 4 seconds on a full subway car) for the subway.

Table 5 augments Table 4, showing how the increase in crowding cost parameters alters the optimal road-only allocation for the various levels of demand intensity considered in Table 4.

Variable*	N = 3708	N = 5560.5	N = 7416	N = 14832	N = 29664
t					
T					
P					
B					
Δ_b					
h_b					
n_b					
N_a					

N_b					
APC_r					
AwC_r					
AWC_r					
ATC_b					
ARC_r					
MRC_r					

Table 5: The Effects of Raising Crowding Costs
Notes: * The regular number in a cell gives the current value. The number in brackets gives the percentage change ((current value – base value) X 100) in the corresponding value from the base case.

DISCUSSION

5. Directions for Future Research and Conclusion

5.1 Directions for Future Research

5.2 Conclusions

REFERENCES

Arnott, R. and E. Inci. 2006.

Arnott, R., T. Rave, and R. Schöb. 2005.

Arnott, R. and J. Rowse. 2010.

Kraus, M. 1991. Discomfort externalities and marginal cost transit fares. Journal of Urban Economics 29, 249-259.

Meyer, J., J. Kain, and M. Wohl. 1965. The Urban Transportation Problem.

Mohring, H. 1972. Optimization in urban bus transportation. American Economic Review 62, 591-604.

Parry, I., and K. Small.

Tirachini, A., and D. Hensher. 2011. Beyond frequency and capacity: the choice of running speed and fare collection system in congested bus corridors. Mimeo.