

## **Should Urban Transit Subsidies Be Reduced?**

**Ian W.H. Parry**

Resources for the Future  
1616 P Street, N.W.  
Washington, DC 20036  
(202) 328-5151  
parry@rff.org

**Kenneth A. Small**

Department of Economics  
University of California  
Irvine, CA 92797-5100  
(949) 824-5658  
ksmall@uci.edu

August 28, 2008  
revised May 15, 2009

Forthcoming: *American Economic Review*, 99(3), June 2009

JEL Codes: R48, H21.

Key words: transit subsidies; scale economies; traffic congestion; welfare effects

We are grateful to Richard Arnott, Bruno De Borger, Stef Proost, Ian Savage, George Tolley, Kurt Van Dender, Clifford Winston, two referees, and numerous seminar participants for helpful comments and suggestions. We also thank Chun Kon Kim, Elliot Klein, and Helen Wei for research assistance. Kenneth Small thanks the University of California Energy Institute for financial support. All responsibility for accuracy of the paper lies with the authors.

# Should Urban Transit Subsidies Be Reduced?

## Abstract

This paper derives empirically tractable formulas for the welfare effects of fare adjustments in passenger peak and off-peak rail and bus transit, and for optimal pricing of those services. The formulas account for congestion, pollution, and accident externalities, scale economies, and agency adjustment of transit service offerings. We apply them using parameter values for Washington (D.C.), Los Angeles, and London. The results support the efficiency of the large current fare subsidies; even starting with fares at 50 percent of operating costs, incremental fare reductions are welfare improving in almost all cases. These findings are robust to alternative assumptions and parameters.

## 1. Introduction

Passenger fares for public transportation are in most places heavily subsidized. Across the 20 largest transit systems in the United States (ranked by passenger miles), the subsidy, as measured by the difference between operating costs and passenger fare revenues, ranges from 29 to 89 percent of operating costs for rail and from 57 to 89 percent for bus (Table 1). Kenworthy and Laube (2001) document a similar pattern across city transit systems in other developed nations.

Two classic rationales for transit subsidies are often advanced (Glaister 1974; Henderson 1977; Jansson 1979). First, scale economies imply that the marginal social cost of supplying passenger miles is less than the average cost. These scale economies may arise from fixed costs, such as track and station maintenance, but more importantly they arise from the “Mohring effect,” whereby users’ waiting or access costs decline as service frequency or route density is increased (Mohring 1972). A related point is that higher passenger density allows vehicles to be operated with higher occupancy, thereby saving on the transit provider’s costs.

The second rationale is that lower transit fares discourage automobile use, thereby reducing external costs from traffic congestion, local and global air pollution, and traffic accidents. This is a second-best argument, since it assumes that these external costs cannot be internalized through appropriate road pricing.

Determining whether current fare subsidies are warranted by these two arguments is complicated by several factors. First, the strength of each argument may vary greatly by time of

day, mode, and location. Second, the appropriate subsidy depends on how transit agencies respond to increases in passenger demand at the margin—do they expand service through more vehicle miles or do they increase vehicle occupancy? Third, transit vehicles themselves contribute to externalities such as congestion and pollution, and their passengers impose external costs on each other via crowding (Kraus 1991) and via increased boarding and alighting time. Fourth, automobile externalities are partly internalized through fuel taxes. And finally, altering the subsidy for one mode will cause substitution across modes and times of day, with secondary effects on welfare due to distortions throughout the system.

Several studies have estimated optimal transit prices, focusing on one or both of the primary rationales just mentioned and usually in just one location. None of these studies encompasses all the complications just mentioned. In fact, existing estimates of optimal transit prices (given current road prices) vary enormously, from zero to more than 100 percent of operating costs, providing a confusing guide as to whether current fare subsidies should be preserved, expanded, or eliminated.<sup>1</sup> It is difficult to discern the reasons for such diverse results because the studies apply to different regions and years, they account for different factors, they make different assumptions about transit agency response, they only sometimes distinguish among times of day, and they use different models ranging from simplified analytical models to less transparent but more detailed network models. Furthermore, some consider prices that are very different from current ones, raising questions of whether assumed parameter values remain valid.

This paper provides a general framework for evaluating existing fare subsidies and potential pricing reforms. It does so by developing a single analytical model that incorporates all the factors just described, then applies it to bus and rail at peak and off-peak periods. We derive a formula for the welfare effects from incrementally adjusting subsidies from their current levels, and we also adapt it to derive optimal subsidies. The resulting formulas clarify the contributions

---

<sup>1</sup> For London, Glaister and Lewis (1978, Table 4, line 3b) estimate optimal rail and bus fares at about 50 to 60 percent of marginal operating costs. For the San Francisco Bay Area and for Pittsburgh, Viton (1983) finds optimal fares to be virtually zero. Winston and Shirley (1998) find quite the opposite for the United States as a whole, with optimal bus and rail fares covering 84 percent and 97 percent of marginal operating costs, respectively. For a prototype Belgian city, De Borger et al. (1996) estimate optimal transit fares at 50 to 114 percent of average agency costs, depending on how service frequency adjusts to passenger demand. For Brussels, Proost and Van Dender (2008) estimate optimal transit fares to be nearly zero in peak periods and about double current fares in off-peak periods. Two recent studies of Washington, D.C., by Winston and Maheshri (2007) and Nelson et al. (2007), estimate net total benefits from transit but with conflicting results. We relate our findings to some of this literature below.

of all underlying parameters and can be empirically implemented in a spreadsheet.<sup>2</sup> Following an extensive compilation of data and estimation of parameter values, we apply the formulas to three large but very different metropolitan areas: Washington, Los Angeles, and London. Our analysis includes vehicle capital costs, which can be varied fairly quickly, but not infrastructure investments. Thus, following previous pricing literature, we explore how best to use existing infrastructure without worrying about recovering sunk capital costs.<sup>3</sup>

The most striking finding is that, in almost all cases, extending fare subsidies beyond 50 percent of operating costs—often well beyond—is welfare improving at the margin across modes, periods, and cities. This finding is robust to plausible alternative assumptions about parameters and agency behavior. The main reasons why large subsidies are welfare improving are the two classic ones, but the relative importance of these two rationales varies across different cases. We find big gains from alleviating auto traffic congestion (though not pollution), especially during peak periods. We also find big gains from increased patronage due to scale economies, mainly when initial service is low as is typical during off-peak hours and, in some cases, even for peak bus service. These latter gains may be in the form of reduced wait and access costs for users, if service is increased in response to higher patronage, or in the form of operating-cost savings, if vehicle occupancy instead of service is allowed to increase.

We would caution against using an analysis as aggregated as ours to infer a finely tuned set of optimized transit prices. Nonetheless, we believe our results are useful for policymakers in that they provide a warning to think twice before attempting to cut operating deficits by raising passenger fares, as is often done under fiscal stress.

One argument against transit subsidies is the broader efficiency costs of distortionary taxes needed to finance them. However, as emphasized in the literature on environmental tax shifts, there are important counteracting effects on tax distortions elsewhere in the economy to the extent that lower transportation costs reduce overall real prices and encourage more economic activity (Bovenberg and Goulder 2002; Parry and Bento 2001). We discuss tax distortions later; based on a rough calculation, the net impact of these distortions on optimal subsidies appears to be moderate.

---

<sup>2</sup> The spreadsheet is available at: [www.aeaweb.org/aer](http://www.aeaweb.org/aer).

<sup>3</sup> As noted below, our framework could be used to assess the benefits of infrastructure improvements. These policies are beyond our scope however, partly because they are highly case-specific.

Will our results carry over to other metropolitan areas? Marginal congestion costs are likely to be lower in most other cities; however, as revealed by our sensitivity analysis, optimal fare subsidies can still be substantial due to other factors. A more definitive answer awaits detailed parameter assessments for other cases.

Our analysis does not include distributional considerations. These of course play a significant role in political decisions regarding public transit, but they often are analytically ambiguous especially considering how changes in property values tend to shift costs and benefits. And they always involve substantial value judgments. We return to this issue briefly in our conclusion.

Probably the most important qualification is that we do not explicitly model the potentially lax incentives for cost minimization by a monopoly provider. There is evidence that subsidy programs encourage high costs through excessive compensation, misuse of high-skilled labor in low-skill tasks, and inefficient use of labor and capital.<sup>4</sup> But the causes of such adverse effects are very specific to the incentive structures built into transit management and subsidy programs, and need not preclude subsidies in any form. For example, one response to this problem would be to offer user-side subsidies, computed as a fixed amount per passenger mile (by mode and period), to either public or private transit operators. Thus the question of a desirable level of subsidy can be decoupled from that of how to encourage efficient operations. The latter is treated in a general form by Laffont and Tirole (1993), and for transit specifically by Pickrell (1983), Nash (2005), Karlaftis (2007), and Borck (2007).

The rest of the paper is organized as follows. Section 2 describes the analytical model and derives key formulas. Section 3 discusses baseline data. Section 4 presents the main quantitative results, describes sensitivity analysis, and discusses broader policy applications of the analysis. Section 5 concludes and elaborates on qualifications.

## **2. Analytical Model**

A full accounting for the factors governing transit subsidies in a given city would require network models on the supply side interacted with disaggregate behavioral models, describing individual users with heterogeneous characteristics, on the demand side. Such models exist but

---

<sup>4</sup>For recent documentation, see Winston and Shirley (1998), Small and Gomez-Ibanez (1999), De Borger and Kerstens (2000), and Savage (2004).

are difficult to compare across regions and cannot easily provide transparent intuition about the underlying reasons for particular results. Furthermore, they require such extensive data and time-consuming calibration that in practice their results can rarely be replicated by researchers outside the teams that produce them.<sup>5</sup>

However, we can learn a great deal from an aggregate-level model that contains the supply and demand features most essential to measuring the factors that motivate transit subsidies. Those factors include, in particular, transit-user wait and access costs, externalities from motor vehicles, demand substitutions across modes and times of day, and transit supplier response to changes in passenger demand. Because our purpose is not to advise about detailed local policies, such as where to locate transit lines or whether fares should vary by location, we do not need the additional disaggregation provided by planning models designed for specific metropolitan areas.

We therefore develop an analytical model of urban passenger travel by auto, rail, and bus at different times of day, in which transit user costs depend on congestion, transit frequency, route density, and vehicle crowding. Travel in our model also produces pollution and accident externalities, some of which are internalized by fuel taxes. The government chooses transit characteristics and fares subject to a budget constraint, while agents optimize over travel choices taking externalities and transit characteristics as given.

We employ a representative agent framework, recognizing that this can only approximate the aggregate behavior of a diverse population. Thus, for example, considering separate income groups would add flexibility important for certain questions such as the effects of differentiated products.<sup>6</sup> However, for the questions addressed here, such disaggregation would mainly affect the pattern of aggregate demand elasticities, which we already calibrate based on extensive empirical information. Similarly, one could explicitly model demographics and land-use patterns, but their main effects would be through parameters that we already incorporate, such as automobile congestion and initial transit share of trips.

---

<sup>5</sup> For examples of such models, see Harvey (1994), Bates *et al.* (1996), and Nelson *et al.* (2007).

<sup>6</sup> For example, heterogeneity in the value of time across drivers is important in designing second-best road pricing policies, under constraints on what road segments can be priced (Small *et al.* 2006).

## 2.1. Model Assumptions

For reference, Box 1 provides a summary of most of the notation defined below.

(i) *User utility.* The representative agent has preferences defined by utility  $U$  as follows:

$$(1a) \quad U = u(X, M, \Gamma) - Z$$

$$(1b) \quad M = M(\{M^{ij}, i = P, O; j = CAR, B, R\})$$

$$(1c) \quad \Gamma = \Gamma(T, W, A, C)$$

$$(1d) \quad T = \sum_{ij} t^{ij} M^{ij}, \quad W = \sum_{ij \neq iCAR} w^{ij} M^{ij}, \quad A = \sum_{ij \neq iCAR} a^{ij} M^{ij}, \quad C = \sum_{ij \neq iCAR} c^{ij} M^{ij}$$

where all variables are in per capita terms. In (1a),  $X$  is the quantity of a numeraire or general consumption good;  $M$  is subutility from passenger miles traveled for all purposes;  $\Gamma$  is a generalized (non-money) cost of travel; and  $Z$  is disutility from pollution and traffic accident externalities.<sup>7</sup> In (1b),  $M^{ij}$  is passenger miles traveled during period  $i$  by mode  $j$  where the two time periods are  $i = P$  (peak) and  $O$  (off-peak), and the three modes are  $j = CAR$  (auto),  $B$  (bus), and  $R$  (rail). We hold trip length constant, so variations in  $M^{ij}$  arise from variations in the number of trips. In (1c),  $T$  is total in-vehicle travel time,  $W$  is time spent waiting at transit stops,  $A$  is time spent accessing transit, and  $C$  is crowding experienced on transit.<sup>8</sup> As shown in (1d), these trip attributes are aggregations over miles traveled, each multiplied by the respective unit values  $t^{ij}$ ,  $w^{ij}$ ,  $a^{ij}$ , and  $c^{ij}$ . We assume  $u(\cdot)$  is increasing and quasi-concave in  $X$  and  $M$  and decreasing and quasi-concave in  $\Gamma$ ;  $M(\cdot)$  is quasi-concave, implying that travel by different modes and times of day are imperfect substitutes; and  $\Gamma(\cdot)$  is increasing and quasi-concave.

(ii) *Travel characteristics.* Several characteristics of transit vehicles affect user and operator costs. First is vehicle *occupancy*,  $o^{ij}$ , the average number of passengers in a bus or train:

---

<sup>7</sup> We exclude possible externalities from oil dependence because they are difficult to define. Insofar as they have been quantified (for example, Leiby 2007 puts them at 32 cents per gallon of gasoline), incorporating them would make little difference to our results.

<sup>8</sup> Recent research has emphasized the importance of reliability (*i.e.*, travel-time predictability) as well as travel time in determining travel choices. This is more important in analyzing the effects of congestion than those of scheduling. Increasing service levels would improve reliability as well as cut waiting time, so including this effect would probably increase the scale economies that we measure.

$$(2a) \quad o^{ij} = M^{ij} / V^{ij}$$

where  $V^{ij}$  is total vehicle miles per capita. Second is the *load factor*,  $l^{ij}$ , defined as the fraction of a vehicle's passenger capacity  $n^{ij}$  that is occupied:

$$(2b) \quad l^{ij} = o^{ij} / n^{ij}$$

Third is the average *service frequency*,  $f^{ij}$ , along each bus or rail transit line:

$$(2c) \quad f^{ij} = V^{ij} / (h^i D^{ij})$$

Where  $h^i$  is the duration of period I in hours per year, and  $D^{ij}$  is route density, measured as total route miles per capita within the fixed service area.<sup>9</sup>

These variables determine the unit travel characteristics in (1d) as follows. First, consider per mile wait time, access time, and crowding, for transit travel

$$(3a) \quad w^{ij} = w^{ij}(f^{ij}), \quad a^{ij} = a^{ij}(D^{ij}), \quad c^{ij} = c^{ij}(l^{ij}), \quad j=B,R; \quad w^{iCAR} = a^{iCAR} = c^{iCAR} = 0$$

where the first two functions are decreasing in their arguments, and the third is increasing.<sup>10</sup>

Next, consider in-vehicle travel time. For cars, this is determined through a performance function on combined car and bus vehicle traffic:

$$(3b) \quad t^{iCAR} = t^{CAR} (V^{iCAR} + \alpha_B V^{iB})$$

where  $\alpha_B > 1$  is the “passenger car equivalent”, or the contribution of a bus to congestion relative to that for a car. For buses, in-vehicle travel time is:

$$(3c) \quad t^{iB} = t^B (V^{iCAR} + \alpha_B V^{iB}) + \theta^B o^{iB}$$

where  $t^B(\cdot)$  is another performance function giving the inverse of bus speed while the bus is moving. Equation (3c) also includes the time a bus is stationary at transit stops, expressed per passenger mile; this is vehicle occupancy  $o^{iB}$  multiplied by  $\theta^B$ , which is the average dwell time per passenger from boarding and alighting divided by trip length. In-vehicle time on trains is like that on buses except we assume there is no congestion on the rail system itself<sup>11</sup> so that travel time while moving is a constant  $t^R$ :

$$(3d) \quad t^{iR} = t^R + \theta^R o^{iR}$$

<sup>9</sup> The arbitrary definition of time-period duration  $h^i$  does not affect our results because we use (2c) only to derive an elasticity of  $f$  with respect to  $D$ .

<sup>10</sup> Equation (3a) embodies some simplifying assumptions about how such costs vary: for example, that crowding occurs entirely in vehicles as opposed to stations.

<sup>11</sup> This assumption may be dubious for London where the Tube is reputedly close to physical capacity. One could modify our model to consider an absolute capacity limit by setting  $\varepsilon_V=0$  for London rail and representing crowding costs explicitly instead of through the sub-optimization assumption (10b).



For cars and buses, we assume that  $t^j(\cdot)$  is a weakly convex function of aggregate road traffic. We also assume that buses travel more slowly than autos, so that  $t^B(V) > t^{CAR}(V)$  for any given aggregate traffic flow  $V$ .

(iii) *Pollution and accident externalities.* The nature of these externalities has been discussed extensively elsewhere (e.g., Parry and Small 2005); we simply summarize their per capita aggregate cost by

$$(4) \quad Z = \sum_{ij} z^{ij} V^{ij}$$

where  $z^{ij}$  is the combined pollution and accident external costs per vehicle mile. Some of the costs of traffic accidents (e.g., injury risk to oneself) are internal and are implicitly taken into account in the subutility function  $M(\cdot)$  for travel.

(iv) *Household optimization.* The household budget constraint is

$$(5) \quad I - TAX = X + \sum_{ij} p^{ij} M^{ij}$$

where  $I$  is (exogenous) private income,  $TAX$  is a lump-sum tax to help finance transit deficits, and the price of  $X$  is normalized to one. For bus and rail,  $p^{ij}$  is the average fare per passenger mile, whereas for auto,  $p^{iCAR} = \bar{p}^{iCAR} + \tau^{iCAR}$ , where  $\bar{p}^{iCAR}$  is pre-tax fuel costs and  $\tau^{iCAR}$  is fuel taxes, both expressed per passenger mile.<sup>12</sup>

Households choose passenger miles and the numeraire good to maximize utility (1) subject to (5), taking  $p^{ij}$ ,  $t^{ij}$ ,  $w^{ij}$ ,  $a^{ij}$ ,  $c^{ij}$ ,  $Z$ , and  $TAX$  as given. This yields first-order conditions, summarized by

$$(6a) \quad \frac{u_{M^{ij}}}{u_X} = q^{ij} \equiv p^{ij} + \rho^T t^{ij} + \rho^W w^{ij} + \rho^A a^{ij} + \rho^C c^{ij}$$

$$(6b) \quad \rho^k \equiv -u_{\Gamma_k} / u_X, \quad k=T, W, A, C$$

The quantities  $\rho^k$  are the (marginal) dollar values of in-vehicle time, waiting time, access time, and crowding, which are taken as fixed. (We allow these values to vary by time of day, although

---

<sup>12</sup> Quantities  $\bar{p}^{iCAR}$  and  $\tau^{iCAR}$  vary by time of day because congestion affects fuel economy. Other money payments (e.g., car maintenance, parking fees) are assumed constant and are implicitly treated as subtractions from the utility of car travel.

it is not indicated by the notation.) Thus  $q^{ij}$  is a generalized price, including both money and nonmoney costs per mile. Equation (6a) says that agents equate the marginal benefit from passenger miles to this generalized price for each mode and time period. From (5), (6), and (1) we obtain the demand functions and indirect utility (the latter denoted by  $\sim$ ):

$$(7) \quad M^{ij} = M^{ij}(\{q^{xy}\}, TAX), \quad X = X(\{q^{xy}\}, TAX), \quad \tilde{U} = \tilde{u}(\{q^{xy}\}, TAX) - Z$$

where  $\{q^{xy}\}$  denotes the set of  $q^{ij}$  for all  $i, j$ .

(v) *Transit agency constraints.* The agency's total operating cost,  $OC^{ij}$ , in period  $i$  for mode  $j$ , is

$$(8a) \quad OC^{ij} = F^{ij} + K^{ij} t^{ij} V^{ij}$$

$$(8b) \quad K^{ij} = k_1^{ij} + k_2^{ij} n^{ij}$$

where  $k_1^{ij}, k_2^{ij} > 0$  are parameters. In (8a),  $F^{ij}$  is a fixed cost representing, for example, the cost of operating rail stations. We assume  $F^{iB}=0$ , so there are no scale economies or diseconomies in providing bus vehicle miles—this is approximately consistent with empirical evidence (Small and Verhoef 2007, 65). Variable operating costs equal total vehicle hours of operation  $t^{ij} V^{ij}$  multiplied by variable costs per vehicle hour,  $K^{ij}$ , which primarily reflect driver labor and vehicle capital. In (8b),  $K^{ij}$  is a linear function of vehicle capacity, with scale economies in providing seat-miles to the extent that  $k_1^{ij} > 0$ . We assume  $k_1^{Pj} > k_1^{Oj}$  because peak service does not conveniently fit an eight-hour workday, so its unit labor costs are higher; and we assume  $k_2^{Pj} > k_2^{Oj}$  because larger vehicles that are purchased primarily for peak use are also available off-peak at little or no extra cost.

The agency budget constraint is

$$(9) \quad TAX + \sum_i \tau^{iCAR} V^{iCAR} = \sum_i \sum_{j \neq CAR} (OC^{ij} - p^{ij} M^{ij})$$

That is, revenues from lump-sum taxes and fuel taxes finance the transit deficit.<sup>13</sup>

(vi) *Agency adjustment of transit characteristics.* Because there is only a limited empirical basis for quantifying access and crowding costs, we eliminate the need to do so by assuming that, for given vehicle miles, the transit agency optimizes over route density and service frequency, and

---

<sup>13</sup> In practice, fuel-tax revenues are often earmarked for road and transit infrastructure projects. Accounting for this could affect our results slightly if the social benefit per dollar of infrastructure spending differs from unity.

that for a given vehicle occupancy, it optimizes over vehicle size and load factor. As shown in Appendix A,<sup>14</sup> these assumptions imply the following first-order conditions:

$$(10a) \quad \rho^W w^{ij} \eta_w^{ij} = \rho^A a^{ij} \eta_a^{ij}$$

$$(10b) \quad \rho^C c^{ij} \eta_c^{ij} o^{ij} = t^{ij} k_2^{ij} n^{ij}$$

where  $\eta_w^{ij}$ ,  $\eta_a^{ij}$ ,  $\eta_c^{ij}$  denote wait, access, and crowding cost elasticities, all defined positively: for example,  $\eta_w^{ij} = |dw^{ij}/df^{ij}| \cdot (f^{ij}/w^{ij})$ . (10a) states that route density is increased until the incremental cost of extra waiting, resulting from less frequent service, equals the incremental reduction in access cost. (10b) states that transit vehicle size is increased until the incremental reduction in crowding costs to its occupants equals the incremental cost to the agency from operating a larger vehicle. Although these assumptions represent a neutral case,<sup>15</sup> we discuss later the implications of relaxing them. From (6a) and (10a,b) we can express the generalized user price as

$$(10c) \quad q^{ij} = p^{ij} + \rho^T t^{ij} + \rho^W w^{ij} \cdot (1 + \eta_w^{ij} / \eta_a^{ij}) + t^{ij} k_2^{ij} n^{ij} / (o^{ij} \eta_c^{ij})$$

Following an increase in demand for passenger miles, we assume that a (constant) fraction  $\varepsilon_V$  of it is accommodated through increased vehicle miles. This is accomplished by increasing service frequency and route density together so as to maintain (10a). The rest of the demand increase (fraction  $1-\varepsilon_V$ ) is accommodated through higher occupancy of transit vehicles, accomplished through simultaneous increases in average vehicle size and load factor chosen to maintain (10b).

## 2.2. Welfare and Optimal Subsidy Formulas

(i) *Marginal welfare effects.* We first consider welfare effects of marginal changes in existing transit prices. The resulting formulas are relatively robust in that they depend only on marginal rather than global assumptions about demand functions and agency adjustments. We focus on peak-period rail for exposition; the formulas for other transit modes and periods are analogous.

We totally differentiate indirect utility with respect to  $-p^{PR}$ ; that is, we consider an incremental reduction in the fare while accounting for induced changes in the agency budget and

---

<sup>14</sup> Appendix A is available at: [www.aeaweb.org/aer](http://www.aeaweb.org/aer).

<sup>15</sup> That is, we consider that equations (10a,b) are equally likely to under- or over-estimate access and crowding costs. For example, the left-hand side of (10a) will overestimate access costs if the agency overinvests in service frequency relative to route density, and it will underestimate them in the opposite case.

in user and external costs. The resulting marginal welfare effect, defined (in consumption units) as  $MW^{PR} \equiv -(d\tilde{U} / dp^{PR}) / u_X$ , can be expressed as the sum of four components (see Appendix A):

$$(11) \quad MW^{PR} \equiv \underbrace{-(MC_{supply}^{PR} - p^{PR})(-M_{PR}^{PR})}_{\text{marginal cost/price gap}} + \underbrace{(MB_{scale}^{PR} - MC_{occ}^{PR})(-M_{PR}^{PR})}_{\text{net scale economy}} + \underbrace{\sum_{ij=PR, iCAR} MC_{ext}^{ij} \cdot M_{PR}^{ij}}_{\text{externality}} \\ + \underbrace{\sum_{ij=OR, PB, OB} (MC_{supply}^{ij} + MC_{ext}^{ij} + MC_{occ}^{ij} - MB_{scale}^{ij} - p^{ij}) M_{PR}^{ij}}_{\text{other transit}}$$

In (11), the quantity  $M_{PR}^{ij} \equiv dM^{ij} / dp^{PR}$  is the marginal demand shift for mode  $ij$  induced by a peak-rail price change. Our assumptions on preferences imply that  $M_{PR}^{PR} < 0$ , and that  $M_{PR}^{ij} \geq 0$  for  $ij \neq PR$ ; that is, peak-rail ridership goes up following a decrease in the fare, diverting ridership away from autos and other transit modes.

The other expressions in (11) are defined as follows:

$$(12a) \quad MC_{supply}^{ij} = (\varepsilon_V / o^{ij}) K^{ij} t^{ij}$$

$$(12b) \quad MB_{scale}^{ij} = \varepsilon_V \rho^W w^{ij} \eta_w^{ij}, \quad MC_{occ}^{ij} = (1 - \varepsilon_V) t^{ij} k_2^{ij} n^{ij} / o^{ij}$$

$$(12c) \quad MC_{ext}^{iCAR} = \left( \frac{z^{iCAR}}{u_X} + MC_{cong}^{iCAR} - \tau^{iCAR} \right) / o^{iCAR}$$

$$MC_{ext}^{ij} = \varepsilon_V \left( \frac{z^{ij}}{u_X} + MC_{cong}^{ij} \right) / o^{ij} + (1 - \varepsilon_V) MC_{dwell}^{ij}, \quad j=B, R$$

$$(12d) \quad MC_{cong}^{iCAR} = \sum_{k=CAR, B} t_{CAR}^{ik} \rho^T M^{ik} + t_{CAR}^{iB} K^{iB} V^{iB}; \quad MC_{cong}^{iB} = \alpha_B MC_{cong}^{iCAR}; \quad MC_{cong}^{iR} = 0$$

$$MC_{dwell}^{ij} = \theta^j \cdot (\rho^T o^{ij} + K^{ij})$$

In (12a),  $MC_{supply}^{ij}$  is the marginal cost to the transit agency of supplying an extra passenger mile. It equals the product of the travel time per mile, the variable operating cost per unit of time, and the response of vehicle miles to an extra passenger mile,  $\varepsilon_V / o^{ij}$ . Compared with (8a), the marginal supply cost is likely to be *below* the average operating cost per mile, to the extent that  $\varepsilon_V < 1$  and/or there are fixed costs.

In (12b),  $MB_{scale}^{ij}$  is the marginal user benefit per extra passenger mile from scale

economies and is positive to the extent that vehicle miles respond to passenger miles, *i.e.* that  $\varepsilon_V > 0$ . This term includes the reduction in wait costs from increased service frequency and the reduction in access costs from increased route density (with the latter included as a wait-cost equivalent from eq 10a).  $MC_{occ}^{ij}$  is the marginal cost of increased vehicle occupancy per extra passenger mile and is positive to the extent that  $1 - \varepsilon_V > 0$ . It incorporates the increase in agency supply costs from increased vehicle size and the increase in crowding costs from higher load factors (with the latter expressed as an agency-cost equivalent from eq 10b).

In (12c),  $MC_{ext}^{ij}$  denotes (net) external costs per passenger mile. For autos, it equals the per-vehicle-mile external cost of pollution, accidents, and congestion (the latter denoted  $MC_{cong}^{ij}$ ), net of the fuel tax, and all divided by occupancy to convert to passenger miles. For transit,  $MC_{ext}^{ij}$  includes these same costs to the extent that vehicle miles respond to passenger miles ( $\varepsilon_V > 0$ ), except there are no congestion costs for rail. In addition, it includes the marginal cost of increased dwell time,  $MC_{dwell}^{ij}$ , applicable to the extent that vehicle occupancy increases ( $1 - \varepsilon_V > 0$ ). Fuel taxes for transit are excluded from supply costs and thus do not need to be netted out here.

In (12d),  $MC_{cong}^{iCAR}$  measures the increase in travel-time cost to all road users, plus the increase in bus operating costs due to the slower-moving traffic, caused by congestion arising from an extra vehicle mile by auto.  $MC_{cong}^{iB}$  similarly measures the cost of additional congestion caused by a bus traveling one mile. Quantities  $t_{CAR}^{iCAR}$  and  $t_{CAR}^{iB}$  are the derivatives of (3b) and (3c), respectively, with respect to car-traffic volume  $V^{iCAR}$ . Finally,  $MC_{dwell}^{ij}$  is the effect on other passengers' time costs, and on agency operating costs, due to the additional boarding and alighting time when an extra passenger mile is accommodated through higher occupancy.

Revisiting (11), each term shows a component of welfare change due to new trips attracted to peak rail due to the price reduction. The “marginal cost/price gap” term shows that welfare is reduced to the extent that the fare for peak rail already falls short of the corresponding marginal supply cost. The “net scale economy” term indicates that welfare is increased to the extent that scale economies from increased peak-rail use outweigh the costs of extra vehicle occupancy. The “externality” term shows that welfare also increases insofar as pollution,

accident, and congestion externalities from auto travel are reduced, although this is partly offset if there are similar externalities from peak rail itself. Finally, the “other transit” term indicates that welfare improves to the extent that passengers are diverted from other transit modes or times of day whose fares fall short of the corresponding marginal social cost, where that marginal social cost includes incremental supply cost, occupancy cost, and own externalities, less incremental benefits from scale economies. Later on, we discuss how (11) might be used in the evaluation of other transit policies, like infrastructure investments.

(ii) *Optimized transit subsidies.* Equation (11) tells us whether or not to increase an existing subsidy. If we want to go further and find the optimal subsidy, we can do so by setting (11) to zero. Doing so, we obtain the following result for optimal fare subsidy per passenger mile,  $\hat{s}^{PR}$ :

$$(13) \quad \hat{s}^{PR} = (OC^{PR} / M^{PR}) - \hat{p}^{PR} = \underbrace{(OC^{PR} / M^{PR}) - MC_{supply}^{PR}}_{\text{average/marginal cost gap}} + \underbrace{(MB_{scale}^{PR} - MC_{occ}^{PR})}_{\text{net scale economy}} \\ + \underbrace{\left( \sum_i MC_{ext}^{iCAR} \cdot m_{PR}^{iCAR} - MC_{ext}^{PR} \right)}_{\text{externality}} + \underbrace{\left( \sum_{ij=OR, PB, OB} (MC_{supply}^{ij} + MC_{ext}^{ij} + MC_{occ}^{ij} - MB_{scale}^{ij} - p^{ij}) m_{PR}^{ij} \right)}_{\text{other transit}}$$

where  $\hat{p}^{PR}$  is value of  $p^{PR}$  that sets (11) to zero.  $m_{PR}^{ij} = -M_{PR}^{ij} / M_{PR}^{PR}$  denotes a modal diversion ratio, which tells us how many passenger-miles are diverted away from mode  $ij$  when one passenger-mile is added to peak rail. Equation (13) implies that the optimal subsidy per peak-rail passenger mile is positive to the extent that (a) marginal supply cost is below average operating cost; (b) scale economies from increasing passenger miles outweigh costs from increased occupancy; (c) externality gains from diverting auto travel exceed the marginal external costs of the increased peak-rail travel; and (d) travel is diverted from other transit for which the overall social cost per passenger mile exceeds the fare.

(iii) *Functional forms.* We assume that marginal congestion costs  $MC_{cong}^{iB}$  and  $MC_{cong}^{iCAR}$  are constant because road traffic changes only moderately in our policy simulations; we also assume that  $z^{ij} / u_x$  and  $\eta_c^{ij}$  are constant but that  $\eta_w^{ij}$  and  $\eta_a^{ij}$  vary as discussed in Section 3.

Passenger travel demands are assumed to have constant elasticities with respect to own generalized price, and to adjust to other prices according to the modal diversion ratios. For

changes in peak rail fares, this implies

$$(14a) \quad M^{PR} = \bar{M}^{PR} \left( \frac{q^{PR}}{\bar{q}^{PR}} \right)^{\eta_q^{PR}},$$

$$(14b) \quad M^{ij} = \bar{M}^{ij} - \int_{\bar{p}^{PR}}^{p^{PR}} m_{PR}^{ij} M_{PR}^{PR} dp^{PR}, \quad ij \neq PR$$

where a bar denotes an initial (currently observed) value, and  $\eta_q^{PR}$  is the elasticity of demand for peak rail with respect to its generalized price. We take  $m_{PR}^{ij}$  to be constant. To obtain  $M_{PR}^{PR}$ , we differentiate (14a) totally:

$$(14c) \quad M_{PR}^{PR} \equiv \frac{dM^{PR}}{dp^{PR}} = \eta_q^{PR} \frac{dq^{PR}}{dp^{PR}} \frac{M^{PR}}{q^{PR}}$$

Here  $dq^{PR} / dp^{PR}$  is the total effect of a one-cent-per-mile increase in the passenger fare on the generalized cost of peak-rail travel. This is greater than one cent because the reduction in peak-rail vehicle miles increases wait and access costs (assuming  $\varepsilon_V > 0$ ), which magnifies the depressing effect on ridership.<sup>16</sup>

### 3. Parameter Values

We examine three large metropolitan areas with extensive transit systems and plenty of road congestion, making them good places to illustrate the two factors discussed in the introduction as constituting the main case for transit subsidies. Yet the three are very different. London has very high transit ridership for its size and Los Angeles rather low, with Washington in between. Also, Washington exemplifies one of the most successful new post-war US rail transit systems, in contrast with London's older but more complete subway and Los Angeles' rudimentary rail transit. Our data cover the areas served by the Washington Metropolitan Area Transit Authority (WMATA), the Los Angeles County Metropolitan Transit Authority (MTA), and Transport for London (TfL) for year 2002. Appendix B provides an extensive discussion of

---

<sup>16</sup> We make this point especially because many empirical studies of transit demand elasticities have not held wait and access costs constant while observing changes in money price. Thus the elasticities they measure involve the total money-price derivative, like that defined by (14c), rather than a partial derivative that holds service characteristics constant. As discussed in Appendix B, we account for this by assuming they measure the own-fare elasticity multiplied by  $dq^i/dp^i$ .

data sources and estimation procedures for all parameters.<sup>17</sup> Here we comment on selected baseline data summarized in Table 2; alternative assumptions with possible significance for our results are discussed later.

(i) *System aggregates and agency adjustment.* The Washington and Los Angeles transit systems each carried nearly 2 billion passenger miles across all modes and times of day in 2002. This transit usage represents 4.3 percent of total passenger miles (auto plus transit) in Washington but only 1.3 percent in Los Angeles. In London, the transit system carried more than 8 billion passenger miles, or 21.7 percent of all passenger travel. For Washington, passenger miles by rail are more than three times those for bus, while the opposite applies to Los Angeles, with its extensive bus but limited rail network (including both subway and light rail). For London, the two modes are closer in size, with passenger miles for rail exceeding those for bus by 29 percent. Average transit vehicle occupancies are broadly comparable across the cities and are 26 to 76 percent greater during peak than during off-peak periods. Train occupancy is around 5 to 10 times that for bus.

We assume for our baseline scenario that transit agencies meet a 1 percent increase in passenger demand through a 0.67 percent increase in vehicle miles and a 0.33 percent increase in vehicle occupancy, or  $\varepsilon_V = 0.67$ . As explained in Appendix A, this rule would apply, under certain simplifications, if the agency optimally traded off vehicle miles and occupancy and if wait and access times were inversely proportional to service frequency and route density, respectively.<sup>18</sup>

(ii) *Operating costs, marginal supply costs, and fares.* Our cost data enable us to compute the parameters in (8). We allocate the capital costs of vehicles to the peak period, and also incorporate an assumed 25 percent difference in per-hour labor cost between peak and off-peak periods. The resulting calculations imply that average operating costs per vehicle mile are around 60 to 100 percent larger in the peak than in the off-peak period. Peak costs are greater because they include vehicle capital costs, higher unit labor costs due to irregular work hours, and in the

---

<sup>17</sup> Appendix B is available at: [www.aeaweb.org/aer](http://www.aeaweb.org/aer).

<sup>18</sup> See also Nash (1988), Jansson (1997), and Small (2004). The result is a modification of the better-known “square-root rule” (Mohring 1972), the latter applying when route density is fixed.



case of bus, additional costs incurred because it takes longer to drive a mile on congested roads. However, average operating costs per *passenger* mile are similar for peak and off-peak travel because of the different vehicle occupancies; they vary from 33 to 99 cents per passenger mile across cities, modes, and times of day. For the U.S. cities, average operating costs per passenger mile are generally higher for bus than for rail (especially for Washington, where bus occupancies are lower than in Los Angeles). London, by contrast, has especially high costs for rail and low costs for bus, possibly because rail in London operates with an old and dilapidated infrastructure while bus service in London has benefited from cost efficiencies due to competitive tendering.

Passenger fares are 20 to 25 cents per mile for Washington and London; in Los Angeles they are only 14 cents per mile for bus and 8 cents per mile for rail.<sup>19</sup> Fare subsidies, defined as  $(OC^j - p^j M^j)$ , are substantial and exceed 50 percent of average operating costs in almost all cases. Subsidies are especially large for Los Angeles rail (82 to 83 percent), with its unusually low fares, and also for Washington bus (76 to 80 percent), which has typical fares but relatively low occupancies.

(iii) *User costs.* Average wait times at transit stops are estimated from service frequency. We assume that when vehicles are less than 15 minutes apart, travelers arrive at random, so the wait-time elasticity is one; but that as the time between vehicles exceeds 15 minutes, an increasing fraction of travelers use a timetable, thereby lowering the wait time/service frequency elasticity (see Appendix B). Expressing wait times on a per mile basis and multiplying by the value of wait time  $\rho^W$  (assumed from the empirical literature to be 49 to 104 percent of the market wage, depending on location and time period), we obtain initial wait costs that vary from 7 to 64 cents per passenger mile. Wait times are much larger during the off-peak than the peak period; they are also larger for bus than for rail.

There is less empirical basis for gauging crowding and access time elasticities; we have assumed location-specific values as explained in Appendix B. When equation (10b) applies (as assumed in our baseline), our results are not very sensitive to alternative assumptions about these elasticities because they affect only the mileage shifts via generalized cost  $q$  in (14), rather than

---

<sup>19</sup> The low rail fare in Los Angeles was so pronounced that it resulted in a suit by a bus riders' group against the operating agency in 1996. However, this resulted in lowering the bus fare rather than raising the rail fares to levels comparable to those in other cities.

entering directly into equations (11)-(13).

(iv) *Marginal benefit from scale economies and marginal occupancy costs.* These are computed from (12b) using our parameters for wait costs and vehicle capital costs.

Marginal scale economies are somewhat larger for bus than for rail and for off-peak than for peak travel, because of greater wait times at transit stops. They vary between 5 and 37 cents per mile across modes, periods, and cities. Increased occupancy costs counteract some, though usually not all, scale economies at peak period; however, they are zero in the off-peak period because all vehicle capital costs (and hence crowding costs) are attributed to the peak.

(v) *Externalities.* Marginal external costs per passenger mile for autos (bottom of Table 2) are dominated by congestion. This is particularly the case for London, where marginal congestion costs are estimated at 103 and 37 cents per passenger mile in the peak and off-peak periods, respectively. Global and local pollution and external accident costs are “only” 5 to 7 cents per passenger mile, with offsetting fuel taxes of 6 to 9 cents in London and one to two cents in the U.S. The congestion costs are estimated from data on average delay, which is measured directly in the case of Washington and Los Angeles and indirectly (from observed peak and off-peak average speeds) in the case of London. This average delay is multiplied by a factor (3.7) reflecting typical estimates of the ratio of marginal to average delay cost on urban highways.

Costs from climate change amount to less than half of one cent per mile. This low value may seem surprising in light of the issue’s current prominence, yet that is what mainstream estimates of discounted damage costs imply. Even increasing our estimate several-fold would still leave pollution costs small relative to congestion costs.<sup>20</sup> For the U.S. cities, overall external costs for auto (net of fuel taxes) are 25 to 31 cents per passenger mile in the peak period and 6 to 8 cents off-peak; figures for gridlocked London are much higher, at 119 and 48 cents, respectively.<sup>21</sup>

---

<sup>20</sup> A gallon of gasoline contains 0.0024 tons of carbon (NRC 2002), so even an extremely large carbon price of, say, \$300 per ton amounts to 72 cents per gallon, or about 4 cents per auto mile for peak periods in the United States and even less for other cases. Similarly the introduction of a domestic CO<sub>2</sub> tax, or cap-and-trade system, would have virtually no effect on our results.

<sup>21</sup> The very high UK fuel tax more than offsets the contribution of pollution and uninternalized accident costs to net external cost of automobiles, causing the latter to be slightly below marginal congestion cost (as shown in the table). Our figures are measured prior to the introduction of the London congestion charge in 2003. Given its very limited

External accident and pollution costs for bus are minimal per passenger mile because of the sufficiently high vehicle occupancies and the relative safety of transit-vehicle occupants; the marginal costs of increased dwell time are also not very large. However, marginal congestion costs are more substantial and amount to 25 to 36 cents per passenger mile for London bus (assumed passenger car equivalents for bus are between 4 and 5). Marginal external costs for rail are negligible, since we assume no congestion.

(vi) *Travel responses.* Based on literature surveys of estimated transit demand elasticities (see Appendix B), we choose peak and off-peak fare elasticities of -0.24 and -0.48 for rail and -0.4 and -0.8 for bus. These are then converted to generalized-price elasticities (see Appendix B).

Modal diversion ratios are based on available evidence and our own judgment (Appendix B). We assume that 60 to 85 percent of increased passenger mileage in response to lower fares comes from diverted auto travel for U.S. cities, and 40 to 50 percent for London, where autos account for a smaller share of passenger travel (Table 2). We assume that 10 percent of extra travel on one transit mode comes from the same mode in the other period, and that the fraction from the other transit mode within the same time period is 5 percent for Los Angeles, 10 percent for Washington, and 30 percent for London.

## 4. Results

### 4.1. Baseline Results

The upper part of Table 3 shows estimates of the marginal welfare effect of a one-cent-per-mile reduction in the passenger fare, starting either at the current subsidy level or at a subsidy level equal to 50 percent of operating costs. For ease of interpretation we have divided through by initial passenger-miles, to give  $MW/M$ ; thus, results are expressed in U.S. cents per mile (at 2002 price levels) per one-cent-per-mile fare reduction.

The most striking result is that, with the exception of Washington peak-period bus, increasing the subsidy is welfare-improving across modes, periods, and cities starting at subsidy levels of 50 percent. Even starting at current subsidy levels, which are typically well above 50

---

geographical coverage, we would expect that charge to have only modest effects on marginal congestion cost and on the total auto money price  $\tau^{iCAR}$  averaged across the entire city (see Section IV). If our study were more disaggregated geographically, congestion charging might have more significant effects on optimal transit fares within central London by reducing congestion and by increasing the value of  $\tau^{iCAR}$  there.

percent, the marginal welfare effects from further lowering transit fares are positive in nine of twelve cases.

The reasons for these results can be discerned in the figures for individual components of marginal welfare at current subsidies. In all cases the marginal supply cost exceeds the fare at current prices, causing an incremental welfare *loss* from this source between 0.04 and 1.43 times the fare reduction. However, in most cases this loss is outweighed by substantial incremental welfare *gains* from the combination of net scale economies and externality benefits, the former being especially important for off-peak service and the latter for peak service. Washington peak-period bus is the exception here because of its especially high marginal supply cost. Welfare effects from interactions among transit modes play a reinforcing but generally more modest role, given that most of the extra passengers attracted to transit by subsidies were previously driving.

Although the contributions of net scale economies and externalities to marginal welfare vary considerably, one or the other is important in almost every case. In eight of the twelve cases, net scale economies are substantial—between 0.29 and 2.0 cents per one-cent fare reduction. Net scale economies are larger for bus than for rail, and larger for off-peak than for peak travel; the reasons for this, already mentioned, are amplified by the greater price-responsiveness of passenger demand (and hence of service frequency or route density) in the cases of bus and of off-peak travel. Only for peak-rail service in London are scale economies fully offset by higher occupancy costs, presumably reflecting London’s famous subway crowding and its already high service frequency and route density. As for externalities, most of the welfare gains come from reducing road congestion, except for off-peak bus service which does little to relieve road congestion.<sup>22</sup>

Table 3 also shows the optimal subsidy level for each type of service, expressed as a percentage of average operating costs. Each subsidy level is optimized separately, holding other prices at their current levels. (We did this because simultaneous optimization of all four related transit prices resulted in large changes in patronage and service levels, which, as noted earlier, render our parameter assumptions less reliable.) We find that optimal fare subsidies are more

---

<sup>22</sup> For Washington and London off-peak bus, the contribution of the bus itself to congestion is greater than the congestion removed due to diverted autos, causing  $MB_{ext}$  to be slightly negative. This occurs for these two cases both because of parameters affecting bus congestion—a high assumed value for  $\alpha_B$  (4 or 5) and a low average off-peak bus occupancy (10 or 12)—and because of parameters that minimize the number of autos diverted, namely a low modal diversion rate from autos (60 or 40 percent) and a high average auto occupancy (1.45 or 1.51).

than two-thirds of average operating costs in eleven of twelve cases; Washington peak bus is the exception, where the optimal subsidy is “only” 46 percent. Again, some combination of net scale economies and diversion of auto externalities explains a major part of the optimal subsidy. The gap between the average and marginal cost of expanding passenger miles, due mostly to the savings in agency costs when occupancy is increased, also plays a consistently important role.<sup>23</sup> Finally in Table 3, the last row shows the percent change in passenger miles when fares are adjusted to their optimum levels; it varies between a reduction of 47.9 percent for Los Angeles peak bus to an increase of 149.1 percent for London off-peak bus.

#### 4.2. Sensitivity Analysis

To explore how robust our results are, we first vary parameters with potential significance for marginal welfare effects and then consider different assumptions about agency adjustment. These parameters are selected based on our knowledge of the sizes of various parts of the equations where they appear, as well as on trial-and-error experience. We vary each parameter in both directions, by 30 to 100 percent.

Table 4 reports results from varying travel demand elasticities, congestion costs, the passenger-car equivalent for buses, and the value of wait time at transit stops. In all cases welfare effects change in the expected direction—for example, when travel is more or less price-responsive, the size of the welfare effects are magnified or shrunk but their sign is not reversed. Welfare effects change noticeably for some perturbations. For example,  $MW/M$  for off-peak service is sensitive to the value of wait time through its effect on scale economies, and  $MW/M$  for peak service is sensitive to congestion costs. In London, the case for increasing current peak bus subsidies is undermined if buses contribute as much to congestion as 10 cars, instead of 5 cars as in our baseline assumptions.

Overall, in seven of the 96 results shown, the sign of  $MW/M$  is reversed from that in the baseline scenario. Nonetheless, our basic qualitative finding—that marginal welfare effects are positive at current subsidies in the majority of cases—is unaffected.

In Table 5 we consider two alternative values for operators’ responsiveness to patronage.

---

<sup>23</sup> While it is perfectly possible that the optimal subsidy could be greater than 100 percent, our confidence in these numbers is not sufficient to say whether or not we encountered any such results. In any case it would imply negative fares, which are impractical. Zero fares, by contrast, are practical and have been tried in various settings; Perone (2002) reviews the pros and costs of these policies.

In one, an increase in demand for passenger miles is met entirely through increased vehicle supply ( $\varepsilon_V = 1.0$ ); in the other, half of the increase is met through increased vehicle miles and half through increased occupancy ( $\varepsilon_V = 0.5$ ). Marginal welfare effects are generally smaller in the first case and larger in the second than in our baseline (for which  $\varepsilon_V = 0.67$ ). This occurs because the agency costs of accommodating extra passenger miles through more vehicle miles, net of the user benefits from that extra service, are greater than the agency and user costs of accommodating extra passengers through higher occupancy.<sup>24</sup> (This observation suggests that most current service levels are inefficiently high given current patronage—although not necessarily higher than at optimal patronage.) Again, our basic favorable finding for subsidies is robust.

Finally, suppose the agency does not suboptimize over route density and service frequency, or over vehicle size and load factor, so that conditions (10a,b) no longer hold. Suppose, for example, that service frequency is excessive relative to route density. In this case marginal wait costs will underestimate marginal access costs, and correspondingly, the marginal benefit from scale economies  $MB_{scale}^{ij}$  in (12b) will be understated by a factor  $(1 - \varepsilon_f^{ij}) \cdot \{\rho^A a^{ij} \eta_a^{ij} / \rho^W w^{ij} \eta_w^{ij} - 1\}$ , where  $\varepsilon_f^{ij}$  is the fraction of marginal changes in vehicle miles that are met through increasing service frequency (as opposed to route density). If, for illustration,  $\varepsilon_f^{ij} = 0.5$  and marginal access costs exceed marginal wait costs by 50 percent, then  $MB_{scale}^{ij}$  will be understated by 25 percent. However, we have already illustrated the effect of different values for  $MB_{scale}^{ij}$  when we varied the wait cost parameter in Table 4. Similarly, relaxing suboptimization over vehicle size and crowding has essentially the same effect as varying the cost of increasing vehicle size, which has such small effects that we have not shown them in Table 5.

### 4.3. Relation to Other Studies

While space precludes a comprehensive comparison and reconciliation of our results and those in prior literature, we note here some selected results that are directly comparable to ours.

Glaister (1984) estimates net benefits from a fare decrease using a model that accounts

---

<sup>24</sup> That is, in (12a,b) the derivative of  $(MC_{supply}^{ij} - MB_{scale}^{ij})$  with respect to  $\varepsilon_V$  is greater than that of  $(-MC_{occ}^{ij})$ .

for both automobile congestion and scale economies; he does not distinguish time periods but, unlike us, he does disaggregate rail into subway (London Transport) and commuter rail (British Rail). As updated in Glaister (2001, 57), he estimates that a reduction in London’s bus fare would produce net social benefits (weighted by annual passenger miles) of approximately 0.48 times the direct revenue loss; our corresponding figure (averaged across times of day) is much higher at 0.86. Glaister reports zero net benefits for a reduction in subway fares, mainly because of subway crowding; this is lower than our estimated ratio of 0.25, probably because he accounts for capacity constraints that might preclude increasing subway service to accommodate new passengers.

Our findings on optimal subsidies for London are somewhat higher than those of Glaister and Lewis (1978), who found them to be about 50 to 60 percent of operating costs in their preferred case. However, they obtain widely differing results for different parameter assumptions, and it is difficult to pinpoint the source of differences with our results.

Our results are broadly consistent with studies of Chicago and several Australian cities, which find that although service levels are sometimes inefficiently high, fare subsidies could generally be increased with positive effects on welfare (Savage and Schupp 1997; Dodgson 1986). Indeed, for Chicago in 1994, the estimated ratio of net benefit to additional subsidy expended – similar though not identical to our  $MW/M$  – ranges from 1.2 to 1.8 across modes and time periods (Savage and Schupp 1997, Table 6), considerably higher than our estimates of  $MW/M$  for other cities. Our findings of high optimal subsidies for peak-period transit are similar to those of Proost and Van Dender (2008, 1227) for Brussels, where there is also heavy traffic congestion. Unlike them, we also find high optimal off-peak subsidies.

#### **4.4 Applying the Analysis to Broader Transit Policies**

Finally, we note how our analysis might be applied more broadly to transit system improvements. Dividing (11) through by  $-M_{PR}^{PR}$  gives an expression for the welfare benefit per extra passenger mile attracted to peak rail,  $m_w^{PR} \equiv MW^{PR} / (-M_{PR}^{PR})$ . The “externality” and “other transit” terms in this expression will depend on the modal diversion ratios, which can be interpreted to apply to other policies, besides lowering fares, that expand transit ridership. However, one must then check whether key parameter assumptions, such as modal diversion

ratios or the operator's response to ridership increases, need to be changed.

Table 6 shows the values of  $mw$  with all parameters the same as in our baseline. For example, the first two columns tell us that if Washington can attract new rail passengers by improving its escalators, adding new lines, or expanding its vehicle fleet, it would create benefits of 19.8 cents and 13.9 cents per additional peak and off-peak passenger-mile, respectively—provided that the new passengers are drawn in the same proportions from other modes as in our baseline, that marginal supply costs are the same, and so forth. Often the assumption of unchanged parameters would not be reasonable, due to differences between the specific project and the rest of the system. For example, if a rail extension helped alleviate a critical highway bottleneck, the benefits would be greater than calculated under baseline parameters, whereas if it extended service to little-congested suburbs the benefits would be lower. Once benefits are calculated, they of course need to be aggregated and discounted over the life of the project in order to assess the desirability of an investment with a known up-front cost.

Despite these qualifications, such numbers can help assess likely orders of magnitude for common cases. To illustrate, based on Pickrell (1989), and our own calculations, the annualized investment cost for the entire Washington rail system amounts to \$2.01 per newly attracted passenger mile in 2002 prices.<sup>25</sup> Our estimates of marginal benefits in Table 6 are smaller than this by more than a factor of ten. This suggests that any new rail line intended to have area-wide impacts similar to those of the existing system would need to come in at well under one-tenth the per-passenger cost of the initial system in order to have a favorable cost-benefit ratio.

## 5. Conclusion

Our analysis suggests that today's substantial operating subsidies for transit systems are warranted on efficiency grounds, at least for the three major metropolitan areas studied. The main caveat is that some of the subsidy may be lost to inefficiency or captured by labor unions, given the evidence cited earlier of increases in wages and other costs following transit subsidies. Thus, our analysis is most applicable to a transit agency with strong incentives to minimize costs.

---

<sup>25</sup> Pickrell (1989, Table S-1) estimates costs per newly attracted transit rider for the initial investment in the Washington rail system to be \$8.75 per trip in 1988 dollars, of which 2.4 percent is operating cost and perhaps another 0.7 percent is vehicle capital (based on our own estimates for Washington underlying Table 1). Adjusting the remainder to 2002 prices (using the CPI for transportation) and dividing by the current average trip length (5.92 miles) yields \$2.01 per passenger-mile for infrastructure cost.



Policies that might promote this have been discussed extensively, for example by Wachs (1989) and Winston and Shirley (1998).

Another caveat is that we have not incorporated the burden on the broader tax system from the need to finance agencies' operating deficits. To the extent that transit subsidies require raising distortionary taxes, such as income taxes, this causes efficiency losses—for example, by deterring work effort. On the other hand, lower transportation costs have a counteracting effect by lowering the general price of goods and services, thereby raising the real wage and the return to work effort. Given these two effects and some simplifying assumptions, the optimal subsidy is given by the Pigouvian subsidy divided by the marginal cost of public funds (Bovenberg and Goulder 2002). A typical estimate for the marginal cost of public funds (which here depends on the uncompensated labor supply elasticity) is around 1.15. Making this simple adjustment for fiscal considerations would therefore imply some scaling back of the optimal subsidies calculated above, but not enough to overturn the basic finding that large subsidies are still warranted.<sup>26</sup>

We have also ignored distributional considerations. Such concerns might raise the optimal subsidy for high-density bus service, which is heavily patronized by lower-income people, and lower it for rail service, which typically benefits wealthier riders and owners of land near transit stations. Quantifying these additional adjustments is contentious, as it brings in value judgments about appropriate distributional weights and it requires the analyst to trace the shifting of costs and benefits through related markets such as that for urban land. Making such adjustments also runs counter to the common view that distributional concerns are most efficiently addressed through the broader tax and benefit system.

We do not explore how optimal fares might vary across different routes—for example, a route passing through a central business district compared with one serving reverse commutes or intrasuburban trips. Analyzing this issue would require a more disaggregated model that accounted for substitution effects among different links in the road and transit network. A related qualification is that we do not account for physical capacity constraints that may make it difficult

---

<sup>26</sup> A more exact adjustment would account for at least two additional factors, which work in opposite directions. On the one hand, if transportation policies reduce congestion and the time costs of commuting, this might have a positive feedback effect on labor supply, justifying a higher level of transit subsidy (Parry and Bento 2001). On the other hand, the efficiency cost of financing transit subsidies is larger when one accounts for other margins of behavior that are distorted by income taxes, such as the choice between ordinary spending and spending that is exempt from income taxes (Feldstein 1999).

to expand vehicle-miles in certain circumstances. Nor do we consider possible benefits from agglomeration externalities, or freeing up valuable downtown space currently used for (subsidized) parking, as transit ridership expands.

It is interesting to consider how the results might change upon adoption of a comprehensive congestion-pricing program. London adopted such a program for its central area in 2003, just after the time for which our simulations apply, imposing a charge of £5 (later raised to £8) and causing road speeds to rise by around six percent (Transport for London 2004). This must have greatly reduced the uninternalized congestion externality for peak-period automobile trips to and within Central London, thereby reducing the case for subsidizing transit trips to the center. (Congestion charging simultaneously increased the case for adding transit service, which in fact was done.) It is less clear whether we would see much difference in a system-wide calculation such as done here, however, since the congestion charging zone, even with its recent extension, covers just 2.4 percent of Greater London.

Finally, we re-emphasize that transit subsidies are only a second-best response to automobile externalities. Our results should not divert attention from the much larger welfare gains to be had from pricing those externalities more directly. For example, Parry and Small (2005) put the welfare gains from an optimized uniform toll on auto mileage in the United States at around \$40 billion per year; welfare gains from differentiating the toll by region and time of day would be larger still. Such gains are on an entirely different scale from those achievable by reforming transit prices.

## References

- Bates, John, Ian Williams, Denvil Coombe, and James Leather. 1996. "The London Congestion Charging Research Programme," *Traffic Engineering and Control* 37: 334-339.
- Borck, Rainald. 2007. "The Political Economy of Urban Transit." In Andreas Kopp (ed.), *Privatisation and Regulation of Urban Transit Systems*, Proceedings of Roundtable 138, European Conference of Ministers of Transport, Paris: Organisation for Economic Co-operation and Development, forthcoming.
- Bovenberg, A. Lans, and Lawrence H. Goulder. 2002. "Environmental Taxation." In A. Auerbach and M. Feldstein, eds., *Handbook of Public Economics* (second edition). New York: North-Holland, pp. 1471-1545.
- De Borger, Bruno, and Kristiaan Kerstens. 2000. "The Performance of Bus-Transit Operators." In David A. Hensher and Kenneth J. Button, eds., *Handbook of Transport Modelling*, Amsterdam and New York: Pergamon, pp. 577-595.
- De Borger, Bruno, Inge Mayeres, Stef Proost, and Sandra Wouters. 1996. "Optimal Pricing of Urban Passenger Transport: A Simulation Exercise for Belgium." *Journal of Transport Economics and Policy* 30: 31-54.
- Dodgson, John S. 1986. "Benefits of Changes in Urban Public Transport Subsidies in the Major Australian Cities." *Economic Record*, 62: 224-35.
- Feldstein, Martin, 1999. "Tax Avoidance and the Deadweight loss of the Income Tax." *Review of Economics and Statistics* LXXXI: 674-680.
- Glaister, Stephen. 1974. "Generalised Consumer Surplus and Public Transport Pricing." *Economic Journal* 84: 849-867.
- Glaister, Stephen. 1984. "The Allocation of Urban Public Transport Subsidy." In J. LeGrand and R. Robinson (eds.), *Privatisation and the Welfare State*, London: Allen and Unwin, 177-200.
- Glaister, Stephen. 2001. "The Economic Assessment of Local Transport Subsidies in Large Cities." In Tony Grayling (ed.), *Any More Fares? Delivering Better Bus Service*, London: Institute for Public Policy Research, 55-76.
- Glaister, Stephen and David Lewis. 1978. "An Integrated Fares Policy for Transport in London." *Journal of Public Economics* 9: 341-55.
- Harvey, Greig W. 1994. "Transportation Pricing and Travel Behavior." In National Research Council, *Curbing Gridlock: Peak-Period Fees to Relieve Traffic Congestion. Volume 2: Commissioned Papers*, Transportation Research Board Special Report 242, Washington, DC: National Academy Press, 89-114.

- Henderson, J. Vernon. 1977. *Economic Theory and the Cities*. New York: Academic Press.
- Jansson, Jan Owen. 1979. "Marginal Cost Pricing of Scheduled Transport Services." *Journal of Transport Economics and Policy* 13: 268–94.
- Jansson, Jan Owen. 1997. "Theory and Practice of Transport Infrastructure and Public Transport Pricing." In G. de Rus and C. Nash (eds.), *Recent Developments in Transport Economics*, Aldershot, UK: Ashgate, 74–134.
- Karlaftis, Matthew G. 2007. "Privatisation, Regulation and Competition: A Thirty-Year Retrospective on Transit Efficiency." In Andreas Kopp (ed.), *Privatisation and Regulation of Urban Transit Systems*, Proceedings of Roundtable 138, European Conference of Ministers of Transport, Paris: Organisation for Economic Co-operation and Development, forthcoming.
- Kenworthy, Jeff, and Felix, Laube. 2001. *The Millennium Cities Database for Sustainable Transport*. International Union of Public Transport, Brussels.
- Kraus, Marvin. 1991. "Discomfort Externalities and Marginal Cost Transit Fares." *Journal of Urban Economics* 29: 249–59.
- Laffont, Jean-Jacques, and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*, Cambridge, Mass.: MIT Press.
- Leiby, Paul N., 2007. "Estimating the Energy Security Benefits of Reduced U.S. Oil Imports," Report No. ORNL/TM-2007/028. Oak Ridge National Laboratory.
- Mohring, Herbert. 1972. "Optimization and Scale Economies in Urban Bus Transportation." *American Economic Review* 62: 591–604.
- Nash, Chris. 1988. "Integration of Public Transport: an Economic Assessment." In J.S. Dodgson and N. Topham (eds.), *Bus Deregulation and Privatisation*, Aldershot, UK: Avebury, 97–118.
- Nash, Chris. 2005. "Privatization in Transport." In Kenneth J. Button and David A. Hensher (eds.), *Handbook of Transport Strategy, Policy and Institutions*, Amsterdam and Boston: Elsevier, 97-113.
- National Research Council.** 2002. *Effectiveness and Impact of Corporate Average Fuel Economy (CAFÉ) Standards*. National Academy of Sciences, Washington, DC, National Academy Press.
- Nelson, Peter, Andrew Baglino, Winston Harrington, Elena Safirova, and Abram Lipman. 2007. "Transit in Washington DC: Current Benefits and Optimal Level of Provision." *Journal of Urban Economics* 62: 231-251.
- Parry, Ian W.H., and Antonio M. Bento. 2001. "Revenue Recycling and the Welfare Effects of Road Pricing." *Scandinavian Journal of Economics* 103: 645–71.

Parry, Ian W.H., and Kenneth A. Small. 2005. "Does Britain or the United States Have the Right Gasoline Tax?" *American Economic Review* 95: 1276–89.

Perone, Jennifer S. 2002. *Advantages and Disadvantages of Fare-Free Transit Policy*. Tampa, Florida: Center for urban Transportation Research, University of South Florida, October. [http://www.dot.state.fl.us/research-center/Completed\\_PTO.htm](http://www.dot.state.fl.us/research-center/Completed_PTO.htm)

Pickrell, Don H. 1983. "Sources of Rising Operating Deficits in Urban Bus Transit." *Transportation Research Record* 915: 18-24.

Pickrell, Don H. 1989. *Urban Rail Transit Projects: Forecast versus Actual Ridership and Costs*. Cambridge, MA: US Department of Transportation, Transportation Systems Center.

Proost, Stef, and Kurt Van Dender, 2008. "Optimal Urban Transport Pricing in the Presence of Congestion, Economies of Density and Costly Public Funds." *Transportation Research A*, 42: 1220-30.

Savage, Ian. 2004 "Management Objectives and the Causes of Mass Transit Deficits." *Transportation Research A*, 38: 181-199.

Savage, Ian, and August Schupp. 1997. "Evaluating Transit Subsidies in Chicago." *Journal of Public Transportation*, 1: 93–117.

Small, Kenneth A. 2004. "Road Pricing and Public Transport." In Georgina Santos, ed., *Road Pricing: Theory and Evidence*, Elsevier, 133–58.

Small, Kenneth A., and Jose A. Gomez-Ibanez. 1999. "Urban Transportation." In Paul Cheshire and Edwin S. Mills, eds., *Handbook of Regional and Urban Economics, Volume 3: Applied Urban Economics*, Amsterdam, North-Holland, 1937–99.

Small, Kenneth A., and Erik T. Verhoef. 2007. *The Economics of Urban Transportation*. London and New York: Routledge.

Small, Kenneth A., Clifford Winston, and Jia Yan. 2006. "Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design," *Brookings-Wharton Papers on Urban Affairs* 7: 53-96.

Transport for London. 2004. Congestion Charging Impacts Monitoring: Second Annual Report, London, April. [http://www.tfl.gov.uk/tfl/ccLondon/cc\\_monitoring-2nd-report.shtml](http://www.tfl.gov.uk/tfl/ccLondon/cc_monitoring-2nd-report.shtml)

U.S. FTA. 2003. *National Transit Databases, 2003 Database*. Federal Transit Administration, Department of Transportation, Washington. [www.ntdprogram.gov/ntdprogram](http://www.ntdprogram.gov/ntdprogram).

Viton, Philip A. 1983. "Pareto-Optimal Urban Transportation Equilibria." In Theodore E. Keeler, ed., *Research in Transportation Economics, Volume 1*. Greenwich, Connecticut: JAI Press, 75–101.

Wachs, Martin. 1989. "U.S. Transit Subsidy Policy: In Need of Reform," *Science* 244: 1545–49.

Winston, Clifford, and Chad Shirley. 1998. *Alternative Route: Toward Efficient Urban Transportation*. Brookings Institution, Washington, DC.

Winston, Clifford, and Vikram Maheshri. 2007. "On the Social Desirability of Urban Rail Transit Systems." *Journal of Urban Economics*, 62: 362-382.

### Box 1. Selected Notation

Symbol	Description	Defined in or near equation #
	<i>Note: superscripts ij for time and mode are omitted</i>	
$A$	Access time, per capita annual aggregate (minutes)	(1d)
$a$	Access time divided by $M$ (minutes/passenger-mile)	(1d)
$C$	Crowding measure, per capita annual aggregate	(1d)
$c$	Crowding measure divided by $M$ (/passenger-mile)	(1d)
$D$	Route density (route miles)	(2c)
$F$	Fixed operating cost (cents/year)	(8a)
$f$	Frequency of service ( $V/D$ )	(2c)
$K$	Variable operating cost per vehicle-hour (cents/veh-hr)	(8b)
$k_1$	Part of $K$ invariant to vehicle size (cents/veh-hr)	(8b)
$k_2$	Rate of variation of $K$ with vehicle size (cents/veh-hr/pass-space)	(8b)
$l$	Load factor ( $o/n$ )	(2b)
$M$	Passenger-miles per person (annual)	(1b)
$m_{rs}^{ij}$	Modal diversion ratio: $-M_{rs}^{ij} / M_{rs}^{rs}$	(13)
$M_{rs}^{ij}$	$dM^{ij} / dp^{rs}$	(11)
$MB_{scale}$	Marginal benefit from scale economies (cents/pass-mile)	(12b)
$MC_{dwell}$	Marginal cost due to increased dwell time (cents/pass-mile)	(12d)
$MC_{cong}$	Marginal congestion cost (cents/pass-mile)	(12c)
$MC_{ext}$	Marginal external cost (cents/pass-mile)	(12c)
$MC_{occ}$	Marginal cost due to increased occupancy (cents/pass-mile)	(12a)
$MC_{supply}$	Marginal transit supply cost (cents/pass-mile)	(12a)
$MW$	Marginal welfare of fare decrease (annual) [cents/(cents/pass-mile)]	(11)
$n$	Vehicle capacity (passenger-spaces/vehicle)	(2b)
$OC$	Operating cost (cents/year)	(8a)
$o$	Vehicle occupancy ( $M/V$ )	(2a)
$p$	Fare or money price of car travel (cents/passenger-mile)	(5)
$q$	Generalized price (cents/passenger-mile)	(6a)
$s$	Subsidy (cents/passenger-mile)	(13)
$T$	In-vehicle travel time, per capita annual aggregate (minutes)	(1d)
$t$	In-vehicle travel time divided by $M$ (minutes/passenger-mile)	(1d)
$TAX$	Taxes per capita needed to finance transit deficit, annual (cents)	(5)
$V$	Vehicle-miles per person (annual)	(2a)
$W$	Waiting time, per capita annual aggregate (minutes)	(1d)
$w$	Waiting time divided by $M$ (minutes/passenger-mile)	(1d)
$z$	External pollution & accident cost (cents/vehicle-mile)	(4)
$\alpha_B$	Congestion caused by bus relative to car	(3b)
$\varepsilon_f$	Fraction of increased vehicle-miles accommodated through increased service frequency (versus increased route density)	Sect 4.2, (A7b)
$\varepsilon_n$	Fraction of increased occupancy accommodated through increased vehicle capacity (versus increased load factor)	(A7d)
$\varepsilon_V$	Fraction of increased passenger-miles accommodated through increased vehicle-miles (versus increased occupancy)	(10c)

$\eta_a$	Minus the elasticity of $a$ with respect to $D$	(10), (3a)
$\eta_c$	Elasticity of $c$ with respect to $l$	(10), (3a)
$\eta_p$	Elasticity of use with respect to money price	(B2)
$\eta_q$	Elasticity of use with respect to generalized price	(14a)
$\eta_w$	Minus the elasticity of $w$ with respect to $f$	(10), (3a)
$\theta$	Dwell time per passenger, divided by trip length (min./mile)	(3c,d)
$\rho^k$	Marginal value of quantity $k$ , $k=T,W,A,C$	(6b)
$\tau$	Fuel tax rate (cents/vehicle-mile)	(5)

*Note:* Superscripts  $ij$  for time and mode are omitted on all symbols except  $MW$ ,  $TAX$ , and all Greek letters. Monetary and time units stated here (cents, dollars, minutes, hours) are those used in our tables and text; in the formulas, all are in common units (dollars, hours).



**Table 1. Passenger Fare Subsidies for the 20 Largest U.S. Transit Authorities in 2003**

	Fare subsidy (% of operating cost)			Passenger miles		
	rail	bus	combined	total, million	% rail	%bus
MTA New York City Transit, Brooklyn, NY	29	59	41	9,451	83	17
New Jersey Transit Corporation, Newark, NJ	50	57	53	2,548	64	36
MTA Long Island Rail Road/Bus, Jamaica, NY	53	61	53	2,302	93	7
Metro-North Commuter Railroad Co., New York, NY	40	n/a	40	2,059	100	0
Washington Metrop. Area Transit Authority, Washington, DC	40	75	55	1,899	76	24
Massachusetts Bay Transportation Authority, Boston, MA	57	79	64	1,838	82	18
Los Angeles County Metrop. Transp. Authority, Los Angeles, CA	78	72	73	1,839	21	79
Chicago Transit Authority, Chicago, IL	59	64	62	1,814	58	42
Northeast Illinois Regional Commuter Railroad Corp., Chicago, IL	56	n/a	56	1,506	100	0
Southeastern Pennsylvania Transp. Authority, Philadelphia, PA	50	62	57	1,354	65	35
San Francisco Bay Area Rapid Transit District, Oakland, CA	42	n/a	42	1,148	100	0
Metropolitan Atlanta Rapid Transit Authority, Atlanta, GA	67	71	69	722	68	32
Maryland Transit Administration, Baltimore, MD	72	74	73	631	47	53
King County Dept. of Transp. - Metro Transit Division, Seattle, WA	n/a	82	82	433	0	100
Metrop. Transit Authority of Harris County, Texas, Houston, TX	n/a	82	82	417	0	100
Tri-County Metrop. Transp. District of Oregon, Portland, OR	35	89	76	407	42	58
Miami-Dade Transit, Miami, FL	85	75	77	389	28	72
Dallas Area Rapid Transit, Dallas, TX	89	87	88	385	36	64
Denver Regional Transportation District, Denver, CO	63	80	79	371	12	88
Port Authority of Allegheny County, Pittsburgh, PA	81	73	75	305	10	90
<b>Average or total (unweighted)</b>	<b>58</b>	<b>73</b>	<b>65</b>	<b>31,819</b>	<b>54</b>	<b>46</b>
<b>Average (weighted by passenger miles)</b>	<b>44</b>	<b>69</b>	<b>54</b>		<b>72</b>	<b>28</b>

Source: US FTA (2003), Tables 19 and 26.

Notes: Includes heavy rail, light rail, and commuter rail. Excludes trolley bus, cable car, demand-responsive transit, and vanpool.

Table 2. Selected Baseline Parameter Values

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
<b>TRANSIT</b>												
Annual passenger miles, millions	1,100	339	290	161	267	126	799	662	3,302	1,265	2,115	1,432
Vehicle occupancy (pass-mi/veh-mi)	151	86	14	9	117	77	20	14	138	76	17	12
Average operating cost, \$/ veh-mi	59	36	14	7	53	33	13	6	105	68	8	4
Avg operating cost, ¢/pass-mi	39	41	99	82	45	43	68	46	76	89	47	33
Marginal supply cost, ¢/pass-mi	23	25	66	55	27	26	46	31	46	54	32	22
Fare, ¢/pass-mi	20	20	20	20	8	8	14	14	25	25	20	20
Subsidy, % of average operating cost	49	52	80	76	83	82	79	69	67	72	59	40
Cost of in-vehicle travel time, ¢/pass-mi	42	31	73	47	35	26	60	35	42	31	67	43
Wait cost, ¢/pass-mi	11	26	38	87	9	21	24	55	7	22	22	64
Generalized price, ¢/pass-mi	96	116	207	266	76	96	157	196	93	105	146	207
Marginal scale economy, ¢/pass-mi	7	17	25	49	6	14	16	31	5	15	15	43
Marginal cost of occupancy, ¢/pass-mi	3	0	9	0	3	0	5	0	5	0	5	0
Marginal external cost, ¢/pass-mi	0.7	0.7	8.8	4.6	0.8	1.0	7.7	3.9	0.8	0.9	36.4	24.7
Marg. congestion cost, ¢/pass-mi	0.0	0.0	5.2	1.0	0.0	0.0	4.8	0.9	0.0	0.0	33.5	22.0
Pollution, climate & acc cost, ¢/pass-mi	0.0	0.0	0.7	1.2	0.0	0.0	0.8	1.1	0.0	0.0	0.6	0.9
Marginal dwell cost, ¢/pass-mi	0.7	0.7	2.8	2.4	0.8	1.0	2.2	2.0	0.8	0.9	2.2	1.8
Elasticity of passenger demand wrt fare	-0.24	-0.48	-0.40	-0.80	-0.24	-0.48	-0.40	-0.80	-0.24	-0.48	-0.40	-0.80
Fraction of increased transit coming from												
auto--same period	0.70	0.60	0.70	0.60	0.85	0.75	0.85	0.75	0.50	0.40	0.50	0.40
same transit mode--other period	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
other transit mode--same period	0.10	0.10	0.10	0.10	0.05	0.05	0.05	0.05	0.30	0.30	0.30	0.30
increased overall travel demand	0.10	0.20	0.10	0.20	0.00	0.10	0.00	0.10	0.10	0.20	0.10	0.20
<b>AUTO</b>												
	Peak	Off-Peak			Peak	Off-Peak			Peak	Off-Peak		
Annual passenger-miles, millions	19,583	22,055			69,519	75,226			15,450	13,951		
Occupancy	1.34	1.45			1.34	1.45			1.41	1.53		
Marginal external cost, ¢/pass-mi	25	6			31	8			119	48		
Marg. congestion cost, ¢/pass-mi	21	2			26	3			123	49		
Poll. & acc. less fuel tax, ¢/pass-mi	4	4			5	5			-4	-2		

Source: See Appendix B.

**Table 3. Baseline Welfare and Optimal Subsidy Estimates**

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
<b>Current subsidy, % of op. cost</b>	49	52	80	76	83	82	79	69	67	72	59	40
<b>Marginal welfare effects</b>												
<i>MW/M</i> at current subsidy <sup>a</sup>	<b>0.24</b>	<b>0.34</b>	<b>-0.44</b>	<b>0.64</b>	<b>0.36</b>	<b>0.14</b>	<b>-0.06</b>	<b>1.10</b>	<b>0.49</b>	<b>-0.07</b>	<b>0.52</b>	<b>1.76</b>
marginal cost/price gap	-0.04	-0.12	-0.93	-1.43	-0.61	-1.12	-0.87	-0.93	-0.20	-0.92	-0.25	-0.09
net scale economy	0.05	0.41	0.34	2.00	0.08	0.87	0.29	1.73	0.00	0.36	0.19	1.74
externality	0.20	0.07	0.17	-0.04	0.79	0.32	0.52	0.12	0.57	0.48	0.48	-0.23
other transit	0.03	-0.02	-0.02	0.11	0.10	0.08	0.00	0.17	0.13	-0.08	0.10	0.34
<i>MW/M</i> at 50% subsidy <sup>a,b</sup>	<b>0.24</b>	<b>0.35</b>	<b>-0.03</b>	<b>0.63</b>	<b>0.35</b>	<b>0.29</b>	<b>0.14</b>	<b>1.01</b>	<b>0.46</b>	<b>0.20</b>	<b>0.51</b>	<b>1.90</b>
<b>Optimum subsidy, % of op. cost<sup>c</sup></b>	<b>&gt;90</b>	<b>88</b>	<b>46</b>	<b>&gt;90</b>	<b>&gt;90</b>	<b>89</b>	<b>74</b>	<b>&gt;90</b>	<b>&gt;90</b>	<b>78</b>	<b>&gt;90</b>	<b>&gt;90</b>
Proportion of subsidy due to												
average-marginal cost gap	0.43	0.54	0.37	0.39	0.38	0.49	0.42	0.35	0.31	0.56	0.35	0.33
net scale economy	0.10	0.41	0.46	0.59	0.05	0.34	0.22	0.55	0.00	0.21	0.15	0.61
externality	0.41	0.08	0.19	-0.01	0.51	0.13	0.37	0.04	0.56	0.27	0.42	-0.11
other transit	0.07	-0.03	-0.02	0.03	0.06	0.03	0.00	0.06	0.13	-0.04	0.09	0.17
% change in passenger miles	24.9	51.0	-47.9	28.1	11.5	21.3	-9.8	32.4	21.6	10.7	42.2	149.1

Notes

<sup>a</sup> *MW/M* is the marginal welfare gain from a one cent-per-mile reduction in the passenger fare, expressed in cents per initial passenger-mile.

<sup>b</sup> Computed with a starting subsidy of 50% for the given mode and time period, holding the others at their current levels.

<sup>c</sup> The subsidy for each time period and mode is optimized separately, holding the others at their current values.

**Table 4. Results with Alternative Parameter Values: Marginal Welfare Effects at Current Subsidies**

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
<b>Baseline results</b>	<b>0.24</b>	<b>0.34</b>	<b>-0.44</b>	<b>0.64</b>	<b>0.36</b>	<b>0.14</b>	<b>-0.06</b>	<b>1.10</b>	<b>0.49</b>	<b>-0.07</b>	<b>0.52</b>	<b>1.76</b>
<b>Travel demand elasticities</b>												
Increased by 30%	0.31	0.44	-0.58	0.84	0.46	0.18	-0.08	1.43	0.64	-0.09	0.67	2.29
Reduced by 30%	0.17	0.12	-0.19	0.13	0.25	0.05	-0.03	0.23	0.34	-0.02	0.22	0.37
<b>Marginal congestion costs</b>												
Increased by 50%	0.33	0.36	-0.35	0.66	0.70	0.22	0.18	1.15	0.84	0.19	0.83	1.79
Reduced by 50%	0.15	0.32	-0.54	0.62	0.01	0.07	-0.30	1.04	0.15	-0.33	0.21	1.74
<b>Passenger-car equiv. for bus</b>												
Increased by 100%	0.25	0.34	-0.55	0.62	0.37	0.14	-0.19	1.08	0.59	-0.07	-0.13	1.01
Reduced by 50%	0.24	0.34	-0.39	0.65	0.35	0.14	0.00	1.11	0.44	-0.07	0.84	2.14
<b>Value of wait time at transit stops</b>												
Increased by 50%	0.26	0.48	-0.25	1.56	0.41	0.51	0.11	1.90	0.49	0.11	0.61	2.52
Reduced by 50%	0.22	0.20	-0.64	-0.27	0.30	-0.22	-0.24	0.29	0.50	-0.26	0.42	1.01

Note

All values are for *MW/M* in cents per passenger mile per one-cent increase in subsidy.

**Table 5. Results for Marginal Welfare Effects with Alternative Assumptions for Agency Adjustment**

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
$\varepsilon_V = 1.0$												
<i>MW/M</i> at current subsidy	0.21	0.28	-0.66	0.64	0.20	-0.12	-0.32	1.21	0.42	-0.66	0.24	2.09
marginal cost/price gap	-0.18	-0.41	-1.57	-2.51	-1.02	-1.89	-1.47	-1.78	-0.41	-1.70	-0.56	-0.52
net scale economy	0.13	0.61	0.75	2.97	0.27	1.28	0.65	2.57	0.06	0.53	0.44	2.57
externality	0.21	0.09	0.17	0.01	0.81	0.38	0.50	0.18	0.57	0.51	0.19	-0.60
other transit	0.05	-0.01	-0.01	0.17	0.13	0.12	0.00	0.24	0.19	-0.04	0.19	0.63
<i>MW/M</i> at 50% subsidy	0.20	0.29	0.07	0.74	0.36	0.44	0.20	1.16	0.42	0.10	0.31	2.15
$\varepsilon_V = 0.5$												
<i>MW/M</i> at current subsidy	0.26	0.37	-0.33	0.64	0.44	0.28	0.07	1.04	0.53	0.20	0.66	1.59
marginal cost/price gap	0.03	0.03	-0.60	-0.86	-0.39	-0.71	-0.55	-0.49	-0.09	-0.54	-0.09	0.13
net scale economy	0.01	0.31	0.12	1.49	-0.02	0.65	0.10	1.29	-0.04	0.27	0.06	1.30
externality	0.20	0.06	0.17	-0.07	0.78	0.28	0.52	0.09	0.56	0.46	0.63	-0.03
other transit	0.02	-0.03	-0.02	0.08	0.08	0.06	-0.01	0.14	0.09	-0.09	0.06	0.19
<i>MW/M</i> at 50% subsidy	0.26	0.37	-0.05	0.58	0.36	0.27	0.13	0.95	0.48	0.29	0.63	1.74

Note

All values are in cents per passenger mile per one-cent increase in subsidy.

**Table 6. Welfare Benefits from Broader Transit-Expanding Policies**

	<b>Washington</b>				<b>Los Angeles</b>				<b>London</b>			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
<i>mw</i> at current subsidy <sup>a</sup>	<b>19.8</b>	<b>13.9</b>	<b>-22.1</b>	<b>16.0</b>	<b>11.5</b>	<b>2.3</b>	<b>-2.2</b>	<b>19.8</b>	<b>51.0</b>	<b>-3.7</b>	<b>25.4</b>	<b>43.5</b>

<sup>a</sup>*mw* is the marginal welfare gain from attracting one passenger-mile to the mode and time period in question, assuming that parameters are unaffected, expressed in cents.

## Appendix A. Analytical Derivations

Equation (10): Agency optimization over route density ( $D$ ) and vehicle size ( $n$ )

Combining (1), (4), and (5), the household's indirect utility function in (7) is defined by

$$(A1) \quad \tilde{U} = \tilde{u}(\{p^{ij}, t^{ij}, w^{ij}, a^{ij}, c^{ij}\}, TAX) - Z$$

$$= \underset{X, \{M^{ij}\}, \lambda}{Max} \left\{ \begin{aligned} &u\left[X, M(\{M^{ij}\}), \Gamma\left(\sum_{ij} t^{ij} M^{ij}, \sum_{ij} w^{ij} M^{ij}, \sum_{ij} a^{ij} M^{ij}, \sum_{ij} c^{ij} M^{ij}\right)\right] - \sum_{ij} z^{ij} V^{ij} \\ &+ \lambda \cdot [I - TAX - X - \sum_{ij} p^{ij} M^{ij}] \end{aligned} \right\}$$

From the agency's point of view, (A1) can be transformed into a social utility function by substituting the various definitions and constraints of the system, namely, (2), (3), (8), and (9). In doing so, the revenues  $\sum_{ij} p^{ij} M^{ij}$  in the government's budget constraint (9) cancel those in the individual's budget constraint in the last term of (A1); prices appear only insofar as  $M^{ij}$  depends on them through the consumer's demand functions. The resulting social utility function can be optimized by setting  $\lambda = u_X$  (the first-order condition for  $X$ ) and then by setting to zero its partial derivatives with respect to  $D$ ,  $n$ ,  $V$ , and either  $M$  or  $p$ . (Henceforth we omit the  $ij$  superscripts for simplicity and understand the preceding statement to apply to each  $i$  and  $j$ .) Here it is convenient to use  $M$  as the agency's choice variable; that is, we hold  $M$  constant in taking the other three derivatives. We consider two of those in this subsection, deferring the third ( $V$ ) till later. Each is a partial derivative, holding the other three variables constant. Thus, in optimizing route density and vehicle size, we hold constant  $M$  and  $V$ , which implies also that occupancy  $o \equiv M/V$  is constant.

Route density affects user waiting and access costs, and vehicle size affects user crowding costs and agency operating costs  $OC$ . Thus each first-order condition for optimization has two terms, and each term involves only the same  $i$  and  $j$ , so we can continue to omit the  $ij$  superscripts without ambiguity:

$$(A2) \quad 0 = \frac{\partial \tilde{U}}{\partial D} = \tilde{U}_w \frac{\partial w}{\partial D} + \tilde{U}_a \frac{\partial a}{\partial D} = -\lambda M \rho^w w_f \frac{\partial f}{\partial D} - \lambda M \rho^a a_D$$

$$(A3) \quad 0 = \frac{\partial \tilde{U}}{\partial n} = \tilde{U}_c \frac{\partial c}{\partial n} - \tilde{U}_l \frac{\partial OC}{\partial n} = -\lambda M \rho^c c_l \frac{\partial l}{\partial n} - \lambda V t \frac{dK}{dn}$$

where  $w_f$ ,  $a_D$ , and  $c_l$  are derivatives of the functions defined in (3a), and we have used the definitions of  $\rho^k$  from (6b). The partial derivatives on the right-hand sides of (A2) and (A3) can be computed using definitions (3) and (8), holding  $V$ ,  $M$ , and  $o$  constant. This yields  $\partial f / \partial D = -f / D$ ,  $\partial l / \partial n = -l / n$ , and  $dK / dn = k_2$ . Inserting these and dividing each equation by  $\lambda M$  yields (10).

*Equation (11). Marginal welfare effects of reduction in peak-rail fare*

Partially differentiating (A1) and applying (6b) gives

$$(A2a) \quad \tilde{U}_{p^{ij}} \equiv -\lambda M^{ij}; \quad \tilde{U}_{t^{ij}} = -\lambda \rho^T M^{ij}; \quad \tilde{U}_{w^{ij}} = -\lambda \rho^W M^{ij}; \quad \tilde{U}_{a^{ij}} = -\lambda \rho^A M^{ij}; \quad \tilde{U}_{c^{ij}} = -\lambda \rho^C M^{ij};$$

$$(A2b) \quad \tilde{U}_{TAX} = -\lambda = -u_X; \quad \tilde{U}_Z = -1$$

Totally differentiating (A1) shows that when the agency changes peak-rail price  $p^{PR}$ , utility changes according to

$$(A3) \quad \frac{d\tilde{U}}{dp^{PR}} = \tilde{U}_{p^{PR}} + \sum_{ij} \left\{ \tilde{U}_{t^{ij}} \frac{dt^{ij}}{dp^{PR}} + \tilde{U}_{w^{ij}} \frac{dw^{ij}}{dp^{PR}} + \tilde{U}_{a^{ij}} \frac{da^{ij}}{dp^{PR}} + \tilde{U}_{c^{ij}} \frac{dc^{ij}}{dp^{PR}} \right\} \\ + \tilde{U}_{TAX} \frac{dTAX}{dp^{PR}} - \sum_{ij} z^{ij} \frac{dV^{ij}}{dp^{PR}}$$

From (A2) and (A3), we obtain

$$(A4a) \quad MW^{PR} \equiv -\frac{1}{\lambda} \frac{d\tilde{U}}{dp^{PR}} = M^{PR} + \sum_{ij} M^{ij} \left\{ \rho^T \frac{dt^{ij}}{dp^{PR}} + \rho^W \frac{dw^{ij}}{dp^{PR}} + \rho^A \frac{da^{ij}}{dp^{PR}} + \rho^C \frac{dc^{ij}}{dp^{PR}} \right\} \\ + \frac{dTAX}{dp^{PR}} + \frac{1}{\lambda} \sum_{ij} z^{ij} V_M^{ij} \frac{dM^{ij}}{dp^{PR}}$$

where  $V_M^{ij} \equiv dV^{ij} / dM^{ij}$  is a constant ( $1/o^{iCAR}$ ) for  $j=CAR$  and depends on the transit agency's operating policy for  $j=B, R$ . To keep track of its parts, we write the components of (A4a) as

$$(A4b) \quad MW^{PR} = M^{PR} + \overline{USERTIM} + \overline{WAITACC} + \overline{CROWD} + \frac{dTAX}{dp^{PR}} + \overline{POLLACC}$$

where  $\overline{WAITACC}$  includes the terms involving  $\rho^W$  and  $\rho^A$  and  $\overline{POLLACC}$ , the last term in (A4a), represents changes in pollution and accident externality costs.

We can compute  $dTAX / dp^{PR}$  by rearranging (9) with only  $TAX$  on the left-hand side, differentiating it, and using (2a) and (8) to get



$$(A5a) \quad \frac{dTAX}{dp^{PR}} = -M^{PR} - \sum_i \tau^{iA} V_M^{iCAR} M_{PR}^{iCAR} - \sum_i \sum_{j \neq CAR} P^{ij} M_{PR}^{ij} + \sum_i \sum_{j \neq CAR} K^{ij} \cdot \left( t^{ij} V_M^{ij} M_{PR}^{ij} + V^{ij} \frac{dt^{ij}}{dp^{PR}} \right)$$

$$+ \sum_i \sum_{j \neq CAR} t^{ij} V^{ij} k_2^{ij} \frac{dn^{ij}}{dp^{PR}}$$

where we hold constant  $\tau^{iA}$  and all transit prices other than  $p^{PR}$ . It is convenient to write the terms in (A5a) as changes in particular financial flows:

$$(A5b) \quad \frac{dTAX}{dp^{PR}} = -M^{PR} - FUELREV - TRANSITREV + (OPSUPPLY + OPCONG) + VEHSIZE$$

where the first term is changes in peak-rail revenue from existing passengers; the second is changes in fuel tax revenue; the third is changes in transit fare revenue due to mode and time-of-day shifts; the fourth is changes in transit operating cost related to travel time (divided into two parts: changes due to shifts among different modes and times of day with different average supply costs, and effects of congestion); and the last is changes in transit operating cost related to vehicle size. Note that new revenues reduce the lump-sum  $TAX$  that must be levied, whereas new costs increase it.

Substituting (A5b) into (A4b), we see that the terms  $M^{PR}$  cancel, and we can rearrange the other parts into a more convenient order for further calculation, as follows:

$$(A6) \quad \begin{aligned} MW^{PR} = & (WAITACC - TRANSITREV + OPSUPPLY) \\ & + (USERTIM + OPCONG + POLLACC - FUELREV) \\ & + (CROWD + VEHSIZE) \end{aligned}$$

It is useful to summarize the definitions of elasticities of bus and rail travel characteristics, recalling that all are defined so as to be positive:

$$(A7a) \quad \eta_w^{ij} = -\frac{f^{ij}}{w^{ij}} w_f^{ij}, \quad \eta_a^{ij} = -\frac{D^{ij}}{a^{ij}} a_D^{ij}, \quad \eta_c^{ij} = \frac{l^{ij}}{c^{ij}} c_l^{ij}, \quad \forall i, j \neq CAR$$

$$(A7b) \quad \varepsilon_V = \frac{M^{ij}}{V^{ij}} V_M^{ij} = o^{ij} V_M^{ij}, \quad 1 - \varepsilon_V = \frac{M^{ij}}{o^{ij}} o_M^{ij} = V^{ij} o_M^{ij}, \quad \forall i, j \neq CAR$$

We also define how service frequency and route density change with vehicle miles, and how vehicle size and load factors change with occupancy, as follows:

$$(A7c) \quad \varepsilon_f^{ij} = \frac{V^{ij}}{f^{ij}} f_V^{ij} = D^{ij} f_V^{ij}, \quad 1 - \varepsilon_f^{ij} = \frac{V^{ij}}{D^{ij}} D_V^{ij} = f^{ij} D_V^{ij}, \quad \forall i, j \neq CAR$$

$$(A7d) \quad \varepsilon_n^{ij} = \frac{o^{ij}}{n^{ij}} n_o^{ij} = l^{ij} n_o^{ij}, \quad 1 - \varepsilon_n^{ij} = \frac{o^{ij}}{l^{ij}} l_o^{ij} = n^{ij} l_o^{ij}, \quad \forall i, j \neq CAR$$

We now proceed to compute key derivatives in (A4a) and (A5a) in terms of  $M_{PR}^{ij} \equiv dM^{ij}/dp^{PR}$ . The travel time derivative can be written, using (2a), (3), and (A7b), as

$$(A8a) \quad \frac{dt^{ij}}{dp^{PR}} = (t_{CAR}^{ij} / o^{iCAR}) M_{PR}^{iCAR} + (t_B^{ij} / o^{iB}) \varepsilon_V M_{PR}^{iB} + (\theta^j / V^{ij})(1 - \varepsilon_V) M_{PR}^{ij}$$

where  $t_{CAR}^{ij} \equiv dt^j / dV_{iCAR}$  and  $t_B^{ij} \equiv dt^j / dV_{iB} = \alpha_B t_{CAR}^{ij}$ . Note that  $t_{CAR}^{iR} = t_B^{iR} = 0$  by our assumption that rail speeds are unaffected by road traffic. Similarly, the waiting, access, and crowding derivatives in (A4), which apply only for  $j \neq CAR$ , can be written using (2), (3), and (A7) as

$$(A8b) \quad \frac{dw^{ij}}{dp^{PR}} = w_f^{ij} \frac{df^{ij}}{dM^{ij}} M_{PR}^{ij} = -w^{ij} \eta_w^{ij} \varepsilon_f^{ij} \varepsilon_V M_{PR}^{ij} / M^{ij}$$

$$(A8c) \quad \frac{da^{ij}}{dp^{PR}} = a_D^{ij} \frac{dD^{ij}}{dM^{ij}} M_{PR}^{ij} = -a^{ij} \eta_a^{ij} \cdot (1 - \varepsilon_f^{ij}) \varepsilon_V M_{PR}^{ij} / M^{ij}$$

$$(A8d) \quad \frac{dc^{ij}}{dp^{PR}} = c_l^{ij} \frac{dl^{ij}}{dM^{ij}} M_{PR}^{ij} = c^{ij} \eta_c^{ij} (1 - \varepsilon_n^{ij})(1 - \varepsilon_V) M_{PR}^{ij} / M^{ij}$$

We now examine the terms in (A6) in groups. We begin by using (A8b) and (A8c) to compute *WAITACC* as given in (A4), using (10a) to simplify:

$$(A9) \quad \begin{aligned} \text{WAITACC} &= \sum_i \sum_{j \neq CAR} \left( \rho^W \frac{dw^{ij}}{dp^{PR}} + \rho^A \frac{da^{ij}}{dp^{PR}} \right) M^{ij} = - \sum_i \sum_{j \neq CAR} \rho^W w^{ij} \eta_w^{ij} \varepsilon_V M_{PR}^{ij} \\ &= - \sum_i \sum_{j \neq CAR} MB_{scale}^{ij} M_{PR}^{ij} \end{aligned}$$

where the last equality applies definition (12b). This accounts for all the terms in (11) involving  $MB_{scale}$ . As for the other terms in the first group in (A6), we note that *TRANSITREV*, the third term in (A5a), accounts for all the terms in (11) involving  $p$ . We also see that *OPSUPPLY*, as defined by the first of the two terms involving  $K^{ij}$  in (A5a), can be written using (A7b) as

$$(A10) \quad \text{OPSUPPLY} = \sum_i \sum_{j \neq CAR} (\varepsilon_V / o^{ij}) K^{ij} t^{ij} M_{PR}^{ij} = \sum_i \sum_{j \neq CAR} MC_{supply}^{ij} M_{PR}^{ij}$$

where the last equality uses definition (12a). Thus *OPSUPPLY* accounts for all the terms in (11) involving  $MC_{supply}$ .

We now turn to the second group of terms in (A6). The terms *USERTIM* and *OPCONG*, which are the terms in (A4a) and (A5a) involving  $dt^{ij}/dp^{PR}$ , can be combined and written, using (A8a), as

$$\begin{aligned}
USERTIM + OPCONG &\equiv \sum_{ij} (M^{ij} \rho^T + K^{ij} V^{ij}) \frac{dt^{ij}}{dp^{PR}} \\
&= \sum_i \sum_{j=CAR,B} (\rho^T M^{ij} + K^{ij} V^{ij}) (t_{CAR}^{ij} / o^{iCAR}) M_{PR}^{iCAR} \\
&\quad + \sum_i \sum_{j=CAR,B} \varepsilon_V \cdot (\rho^T M^{ij} + K^{ij} V^{ij}) (t_B^{ij} / o^{iB}) M_{PR}^{iB} \\
&\quad + \sum_i \sum_{j=B,R} (1 - \varepsilon_V) (\rho^T M^{ij} + K^{ij} V^{ij}) (\theta^j / V^{ij}) M_{PR}^{ij}
\end{aligned}$$

where we have adopted the notational convention that  $K^{iCAR}=0$ . Using the fact that  $t_B^{ij} = \alpha_B t_{CAR}^{ij}$ , the definition  $o^{ij}=M^{ij}/V^{ij}$ , and definitions (12d), we obtain

$$\begin{aligned}
(A11) \quad USERTIM + OPCONG &= \sum_i (MC_{cong}^{iCAR} / o^{iCAR}) M_{PR}^{iCAR} + \sum_i \varepsilon_V \cdot (MC_{cong}^{iB} / o^{iB}) M_{PR}^{iB} \\
&\quad + \sum_i \sum_{j=B,R} (1 - \varepsilon_V) MC_{dwell}^{ij} M_{PR}^{ij}
\end{aligned}$$

These terms are components of sums of  $MC_{ext}^{ij}$  as defined in (12c). Next we obtain some other components of those same sums. Using the definition of  $\varepsilon_V$  and the fact that  $\lambda=u_X$ , the change in external costs of pollution and accidents is

$$(A12) \quad POLLACC \equiv \frac{1}{\lambda} \sum_{ij} z^{ij} V_M^{ij} M_{PR}^{ij} = \sum_i \frac{z^{iCAR}}{u_X} \cdot \frac{1}{o^{iCAR}} \cdot M_{PR}^{iCAR} + \sum_i \sum_{j \neq CAR} (\varepsilon_V / o^{ij}) \cdot \frac{z^{ij}}{u_X} \cdot M_{PR}^{ij}$$

Finally, the fuel tax revenue term in (A5) is

$$(A13) \quad -FUELREV = \sum_I (-\tau^{iCAR} / o^{iCAR}) M_{PR}^{iCAR}$$

Adding equations (A11)-(A13) and applying definitions (12c) yields

$$\sum_i MC_{ext}^{iCAR} M_{PR}^{ij} + \sum_i \sum_{j=B,R} MC_{ext}^{ij} M_{PR}^{ij}$$

which accounts for all the terms in (11) involving  $MC_{ext}$ .

Finally, we consider the last group of terms in (A6), involving crowding and the costs undertaken to avoid it. Using (A7b), A7d), (A8d), and (10b), these terms add to

$$\begin{aligned}
CROWD + VEHSIZE &= \sum_i \sum_{j \neq CAR} (1 - \varepsilon_V)(1 - \varepsilon_n^{ij}) \rho^C c^{ij} \eta_c^{ij} M_{PR}^{ij} + \sum_i \sum_{j \neq CAR} t^{ij} V^{ij} k_2^{ij} n_o^{ij} o_M^{ij} M_{PR}^{ij} \\
&= \sum_i \sum_{j \neq CAR} (1 - \varepsilon_V)(1 - \varepsilon_n^{ij}) (t^{ij} k_2^{ij} n^{ij} / o^{ij}) M_{PR}^{ij} + \sum_i \sum_{j \neq CAR} (1 - \varepsilon_V) \varepsilon_n^{ij} (t^{ij} k_2^{ij} n^{ij} / o^{ij}) M_{PR}^{ij} \\
&= \sum_i \sum_{j \neq CAR} MC_{occ}^{ij} M_{PR}^{ij}
\end{aligned}$$

which accounts for the terms in (11) involving  $MC_{occ}$ . We have now accounted for all terms in (11), which completes the proof.

### *Transit agency optimization over vehicle miles of service (V).*

Now consider what would happen if the agency also optimized with respect to  $V$ . Our model does not assume the agency actually does so, but here we use this assumption to derive a benchmark case for  $\varepsilon_V$  to use in our baseline scenario. This type of benchmark is sometimes called a “quasi-first-best” response: responding to changes according to a first-best adjustment rule even though some other factor (*e.g.* a bureaucratic bias toward offering too much service) prevents the situation from being truly first-best (Small and Verhoef 2007, p. 142).

In our case even this “quasi-first-best” value of  $\varepsilon_V$  is only approximate, as we compute it under three additional simplifying assumptions:

- Elasticities of waiting and access times (defined positively) are all equal to a common value ( $\eta_w^{ij} = \eta_a^{ij} \equiv \zeta$ );
- The transit agency ignores its own vehicles’ contributions to congestion ( $t_B^{ij} = 0$ ) and to other externalities ( $z^{iB} = z^{iR} = 0$ );
- Dwell time for entering and exiting passengers is negligible ( $\theta^B = \theta^R = 0$ ).

The first bullet is an assumption common to the simpler models of Mohring effects—for example, that of Small (2004). A special case is when average waiting time is half the interval between vehicles, and average access time is proportional to the distance between parallel transit lines; then  $\zeta = 1$ .

Those assumptions enable us to derive a simple condition for maximizing (A1) with respect to the agency’s choice variables, for given travel demands  $\{M^{ij}\}$ . In what follows, we

suppress superscripts for simplicity. Maximizing with respect to  $D$  and  $n$  again yields (10). Given our first simplifying assumption, we see immediately from (10a) that average waiting cost and access cost are equated:

$$(A14) \quad \rho^W w = \rho^A a$$

This result is also in Jansson (1997). Since  $D=V/f$ , it can be written as

$$(A15) \quad \rho^W \alpha_w f^{-\zeta} = \rho^A \alpha_a \cdot (V/f)^{-\zeta}$$

where we have substituted in the constant-elasticity functions  $w = \alpha_w f^{-\zeta}$  and  $a = \alpha_a D^{-\zeta}$  describing waiting and access times, respectively. Solving (A15) for  $f$ , we see that it is proportional to the square root of  $V$ . That is,  $f$  is adjusted when  $V$  changes with elasticity  $\varepsilon_{fV} = 1/2$ . Therefore,

$$(A16) \quad \varepsilon_{fM} = \varepsilon_{fV} \varepsilon_V = 1/2 \varepsilon_V$$

We now consider maximizing with respect to  $V$ . Given our second assumption,  $V$  affects (A1) only through the terms involving waiting time  $w$ , access  $a$ , crowding  $c$ , and operator cost  $OC$ , the latter entering through budget constraint (9). The first-order condition is therefore

$$\begin{aligned} 0 &= \frac{\partial \tilde{U}}{\partial V} = \tilde{U}_w \frac{\partial w}{\partial V} + \tilde{U}_a \frac{\partial a}{\partial V} + \tilde{U}_c \frac{\partial c}{\partial V} + \tilde{U}_{TAX} \frac{\partial OC}{\partial V} \\ &= -\lambda M \rho^W w_f f_v - \lambda M \rho^A a_D D_v - \lambda M \rho^C c_l l_o \cdot (do/dV) - \lambda K t - \lambda t V k_2 n_o \cdot (do/dV) \end{aligned}$$

where the last equality uses the definitions in (6b) and (8) and the result  $\lambda = u^X$ . Dividing by  $\lambda$  and using (A7), (2a), and (10), this implies

$$\begin{aligned} 0 &= \rho^W w \zeta \varepsilon_f M / V + \rho^A a \zeta \cdot (1 - \varepsilon_f) o + \rho^C c \eta_c \cdot (1 - \varepsilon_n) o - k_1 t - k_2 t n \cdot (1 - \varepsilon_n) \\ &= \rho^W w \zeta M / V - k_1 t \end{aligned}$$

or

$$(A17) \quad \frac{wM}{V} = \frac{k_1 t}{\rho^W \zeta}$$

Under our second assumption, the right-hand side of (A17) is a constant as far as the agency is concerned. On the left-hand side,  $w = \alpha_w f^{-\zeta}$ . Therefore,

$$(A18) \quad f^{-\zeta} M V^{-1} = \text{constant}$$

Now let  $M$  change parametrically, with all the service variables  $f$ ,  $n$ , and  $V$  changing in response. Differentiating the logarithm of (A18) with respect to  $\log(M)$  yields

$$(A19) \quad -\zeta \varepsilon_{fM} + 1 - \varepsilon_V = 0$$

Substituting (A16) into (A19) and solving yields  $\varepsilon_V = 2/(2+\zeta)$ . For the common case  $\zeta=1$ , this yields  $\varepsilon_V = 2/3$ , as in Small (2004) and a special case of Nash (1988).

The intuition for this result is somewhat subtle. If  $\zeta$  is near zero, wait and access costs are relatively unaffected by vehicle miles of service, so vehicles are operated only as necessary to handle the passenger loads; thus increased passenger loads require a proportional increase in vehicle miles, *i.e.*  $\varepsilon_V = 1$ . If  $\zeta$  is large, the operator accounts for the substantial effects on user costs by running extra vehicles for passengers' convenience even when  $M$  is small; in that case, when  $M$  increases, the operator can absorb some of the increase through higher occupancy, thereby reaping more of the advantages of scale; this means choosing a smaller value of  $\varepsilon_V$ . We take  $\zeta=1$  as our base case ( $\varepsilon_V = 2/3$ ) and consider sensitivity  $\zeta \in [0, 2]$  by treating  $\varepsilon_V = 1$  and  $\varepsilon_V = 1/2$ .

## Appendix B. Assessment of Parameter Values

Here we describe our methodology for estimating parameter values along with data sources; Table 2, which is discussed in the text, summarizes our key estimates. For some parameters, breakdowns by mode or time of day are unavailable from statistical sources; in these cases we use various estimation procedures or our own judgment. U.K. monetary numbers are converted to U.S. dollars using the average 1998–2003 exchange rate of £1.0 = US\$1.6.

*System operating characteristics.* Basic data are compiled from the operating agencies and various national statistics.<sup>27</sup> For London, we allocate total passenger miles across time of day using the observed fraction of passenger trips occurring at peak period, 0.62 for rail and 0.48 for bus, and an assumed average trip length in the peak equal to 1.6 times that in the off-peak.<sup>28</sup> Passenger miles per hour are then computed assuming that the peak period covers 6 hours per workday (30 hours per week) and the off-peak covers 10 hours every day (70 hours per week). We assume peak shares are each 0.05 higher for Washington (which has a high proportion of government employment) and 0.05 lower for Los Angeles (which has a smaller discrepancy between peak and off-peak vehicles per hour).<sup>29</sup> To obtain vehicle miles offered by time period, we assume that observed total vehicle miles are allocated across the available 100 weekly hours in proportion to passenger miles per hour to the power  $\varepsilon_V=0.67$ , our baseline assumption as discussed in Appendix A.<sup>30</sup>

For Washington and Los Angeles, automobile vehicle miles by time of day are from Schrank and Lomax (2003), and occupancy is from the 2001 National Household Travel Survey on average occupancy per trip in large metropolitan areas. For London, auto passenger miles by

---

<sup>27</sup> For the United States, see the National Transit Database (FTA 2003), and for the United Kingdom, see TfL (2003, Tables 1.1, 1.2, 3.6), TfL (2004a, b), and U.K. DfT (2003, Tables 5.3, 5.16). Rail data encompass subways and light rail but not commuter rail.

<sup>28</sup> For the entire United Kingdom, commuting trips have length around twice that of trips for education, shopping, or other personal business (U.K. DfT 2003, Table 10); however, we expect a smaller discrepancy for transit trips because of the high fixed time cost of accessing transit.

<sup>29</sup> The Washington adjustments are in line with unpublished statistics we obtained from transit agency representatives; the Los Angeles transit authority has no such data on trips by time of day.

<sup>30</sup> Total vehicle miles for U.S. rail systems were obtained by multiplying vehicle-car miles by average cars per train. For peak periods the latter is calculated by the ratio of rail cars to trains; off-peak train length is assumed to be slightly lower based on common observation. For London rail, vehicle-miles (i.e. train-miles) is measured directly (in train-km) as 65.4 million train-km from LUL (2003), p. 2; the same figure appears rounded off as 65 train-km in TfL (2003), Table 1.1.

time of day are from TfL (2003, Table 3.6), including car/motorcycle and taxi. Auto vehicle-miles are from TfL (2003, Tables 1.2, 31.); to allocate them across time periods we use information about person trips by time period and overall average trip distance, along with an assumed ratio of peak to off-peak trip distance of 1.3. Auto occupancy for London is then calculated as the ratio of passenger-miles to vehicle-miles.

*Operating costs and fares.* We assume that vehicle capital costs are proportional to capacity  $n$ , whereas other operating costs are independent of  $n$ . Thus in aggregate, vehicle capital costs constitute  $k_1tV$  and other operating costs  $k_2ntV$ , using (8). Operating costs, aside from vehicle capital costs, are taken from the operating statistics of the transit agencies. For rail, we assume that 10 percent of these are the fixed cost of maintaining stations ( $F^{iR}$  in (8a)). When expressed per vehicle hour of service, we assume that the rest of these costs are 25 percent greater during peak than off-peak periods because of difficulties in scheduling labor for split shifts; hence we obtain  $k_1^{ij}$  in (8b).

As for vehicle capital costs, we estimate them ourselves by annualizing the purchase cost of a rail or bus car, assuming lifetimes of 25 and 12 years, respectively, and a real interest rate of 7 percent.<sup>31</sup> (One advantage of this procedure is that we need not rely on agency data for vehicle purchase costs, which may be distorted by various financing mechanisms such as tax-free bonds.) We allocate vehicle capital costs entirely to the peak period, on the assumption that any increase in vehicle miles in that period requires purchasing more vehicles, whereas an increase in the off-peak period does not; hence we obtain  $k_2^{Pj}n^{Pj}$  and  $k_2^{Oj} = 0$  in (8b). Vehicle capital costs are 27 to 52 percent of other peak variable operating costs. Thus our assumption that they are the portion of costs that is proportional to  $n$  leads to results consistent with several other studies of size-related costs, as reviewed by Small (2004, 156 and note 13).

Fares were obtained by dividing agency passenger fare revenue by passenger mileage (for Washington rail, peak fares were higher than off-peak in 2002, but the discrepancy was modest and we ignore it).

---

<sup>31</sup> We use U.S. nationwide figures for all vehicle prices (from APTA 2002, Table 60) except for Los Angeles rail, for which figures were available from [www.mta.net/press/pressroom/facts\\_glance](http://www.mta.net/press/pressroom/facts_glance) (where necessary, figures are



*Wait costs.* Based on evidence summarized in Small and Verhoef (2007, 53), we assume the value of in-vehicle time  $\rho^T$  in the U.S. is half the gross wage rate; in the UK we assume the value of time is only 40 percent of the gross wage rate, due to higher labor taxes there which reduce the net wage rate relative to the gross wage. In both nations we assume the value of waiting time at transit stops,  $\rho^W$ , is  $1.8\rho^T$ , the midpoint of the range suggested by Small and Verhoef. The median gross wage rate is measured at \$16.93, \$14.19, and \$18.83 for Washington, Los Angeles, and London, respectively, and then expressed per minute.<sup>32</sup> We assume that people of different wage rates have sorted themselves into different modes and time periods initially, as follows: auto and rail travelers each have wages that are 15% above the area's median gross wage rate for peak periods, and 15 percent below for off-peak (since peak travelers are more likely to be higher-wage workers); and bus travelers have wage rates that are 80 percent of these amounts.

We obtain initial wait times and the wait time elasticity as follows. Let  $H$  be average minutes between transit vehicles at a given stop, or headway (the inverse of frequency). When  $H$  is small, it is reasonable to assume that travelers arrive randomly at a stop and incur expected wait time  $H/2$ . When headways are larger, at least some travelers will use transit timetables, which, following Tisato (1998), we assume involves three time costs. The first two are planning and precautionary time required because the exact vehicle arrival time is uncertain; we assume these are 1 and 5 minutes, respectively, and each is valued at rate  $\rho^W$ . The third is the expected cost of early arrival at the destination, assuming the traveler chooses a transit vehicle arriving prior to the desired time to ensure against late arrival. This is  $\rho^E H/2$  where  $\rho^E$  is the per minute cost of early arrival, assumed conservatively to be  $0.2\rho^W$ ; that is, a minute of early arrival is equivalent to 0.2 minutes of extra planning or precautionary time.<sup>33</sup> All these costs are therefore accounted for by setting wait time per trip,  $w \cdot d$ , to  $6+0.2H/2$  for those using a timetable.

We therefore assume that when  $H < 15$ , all users arrive randomly, so the average wait time per trip is  $w \cdot d = H/2$ ; whereas when  $H > 60$ , everyone uses a schedule, so  $w \cdot d = 6 + H/10$ . In the first case,  $\eta_w^{ij} \equiv (dw/dH) \cdot (H/w) = 1$ , whereas in the latter case,  $\eta_w^{ij} = 1/[1 + (60/H)]$ . For simplicity, we

---

updated to 2002 using the CPI for Transportation Equipment). The vehicle lifetimes chosen are common in the transit cost literature, and the interest rate is that recommended for cost-benefit analysis by U.S. OMB (1992).

<sup>32</sup> Wages are from U.S. BLS (2004), TfL (2003, p. 49) and U.K. ONS (2004).

assume that the elasticity (which is what enters our calculations directly) falls gradually as a mixture of the elasticities applying to these two regimes:

$$(B1) \quad \eta_w^{ij} = \begin{cases} 1 & \text{if } 0 < H \leq 15 \\ (1 - \lambda) + \frac{\lambda}{1 + (60/H)} & \text{if } 15 < H < 60 \\ \frac{1}{1 + (60/H)} & \text{if } H \geq 60 \end{cases}$$

where  $\lambda = (H-15)/45$ . Substituting this value for  $\lambda$ , the middle part of (B1) can be written as  $\eta_w^{ij} = (4/3) - (5/3)/[1 + (60/H)]$ , from which we can see that  $\eta_w^{ij}$  is monotonically declining in  $H$  from a value of 1.0 at  $H=15$  to 0.5 at  $H=60$ .<sup>34</sup> (It then rises gradually for  $H > 60$ , but this regime is inapplicable here because in our simulations, headways never rise above 26 minutes.)

Ideally we would compute initial headways from data on car-miles, directional route-miles, and assumed duration of each time period. However some data on route-miles are unavailable, and even if they are the resulting average headways are not realistic because individual routes are heterogeneous. Therefore we use these data as guidance, but ultimately assume initial headways based on our judgment. For the US, we assume rail headways of 5 and 10 minutes (peak and off-peak), and bus headways of 12 and 25 minutes. For London, we assume smaller headways of 3 and 8 minutes for rail, 6 and 15 for bus.

*Marginal benefits from scale economies and marginal cost from occupancy.* These are easily computed from (12b), using above values for parameters  $\varepsilon_v$ ,  $\rho^W$ ,  $w^{ij}$ ,  $\eta_w^{ij}$ , and  $k_2^{ij} n^{ij}$ .

*Marginal congestion costs.* For automobiles,  $MC_{cong}^{iCAR}$  is obtained directly from equation (12d).

The travel-time functions are assumed to follow the commonly used BPR-type functional form

---

<sup>33</sup> Arnott, de Palma, and Lindsey (1993) and others use the value of  $\rho^E/\rho^T$  estimated by Small (1982) for work trips, which is 0.61. The ratio is likely to be much lower for non-work trips, which less often have a fixed schedule. So to be conservative we use half the work-trip value, or  $\rho^E = 0.305\rho^T \approx 0.2\rho^W$ .

<sup>34</sup> In the optimal subsidy and 50% subsidy calculations, where we have non-marginal price changes, we also must compute  $w d$  itself in order to obtain the full prices for use in (14). In this case we integrate (B1) to obtain  $\ln(wd) = 5.6 + (4/3)\ln H - (5/3)\ln(H+60)$ . In our calculations,  $H > 15$  occurs in only two cases (US off-peak bus), when  $H \approx 25$ . At  $H=25$  this formula implies  $w d = 12.03$ , which suggests that about seven-eighths of travelers arrive randomly (thus having wait time  $H/2 = 12.5$ ), while one-eighth of them use a schedule (wait time  $6 + H/10 = 8.5$ ); thus the penetration rate of schedules is less than  $\lambda$ , which is 0.22 at this headway.

(US Bureau of Public Roads, ) in which travel delay (dropping the  $i$  superscript for simplicity) is proportional to  $(V/C)^\gamma$  where  $V=V^{CAR}+\alpha_B V^B$  is total traffic (in passenger-car equivalents),  $C$ =capacity, and  $\gamma$  is an exponent that we take to be 3.7 based on some aggregate relationships from the U.S.<sup>35</sup> We also assume that bus speed is a fixed proportion of car speed. It is easy to show that with these assumptions, marginal delays to cars and buses, respectively, due to an auto vehicle-mile are:

$$t_{CAR}^{CAR} \equiv dt^{CAR} / dV^{CAR} = \gamma(\text{average delay/mile})/V$$

$$t_{CAR}^B \equiv dt^B / dV^{CAR} = (t^B / t^{CAR}) t_{CAR}^{CAR}$$

Average delay/mile for the U.S. cities is obtained from total person hours of delay from Schrank and Lomax (2003), allocating 85 percent of it to the peak period, and dividing by passenger miles; this yields an average peak delay (in min/pass-mile) of 0.33 minutes per passenger mile for Washington and 0.49 for Los Angeles). Our data provide direct estimates of average traffic speeds in Greater London during the peak and daytime off-peak periods; we add 10 percent to the latter to account for evenings and nights (bus speeds are 10.3 and 12.3 mph in peak and off-peak period). Average delay is then inferred assuming a free-flow speed of 30 miles per hour, with a result of 1.91 min/mile peak and 1.08 min/mile off-peak. We assume the passenger-car equivalent for buses,  $\alpha_B$  in (12d), is 4.0 for the U.S. cities and 5.0 for London, where buses are larger and cars are smaller.<sup>36</sup> Based on agency data, we find the ratio of auto to bus speed to be 2.8 in Washington, 2.7 in Los Angeles, and 1.6 in London. In London, about one-fourth of the marginal congestion cost  $MC_{cong}^{iCAR}$  turns out to be attributed to the effect of congestion on bus passengers and operators, despite that only about one-eighth of passengers using roadways do so on buses; the difference is due to the considerable adverse effects of congestion on agency costs.

*Pollution and accident externalities.* We start with nationwide average values from the assessment by Parry and Small (2005) of U.S. and U.K. automobile externalities: namely, 2.0 cents per vehicle mile for local pollution; 6 cents per gallon of gasoline for global warming; and 3.0 and 2.4 cents per vehicle mile for accidents in the United States and the United Kingdom,

<sup>35</sup> Small (1992, 70–71) found that total delay is well approximated by a power function of traffic volume, with power 4.1 in Toronto and 3.3 in Boston; we average to set  $\gamma=3.7$ .

<sup>36</sup> U.S. FHWA (1997, Table V-23) gives the passenger-car equivalent as only 2.0; however, this is only for federal urban highways where buses stop very infrequently, and it excludes mileage on city and suburban streets.

respectively. To account for greater population exposure in urban areas, we double the local pollution figure for Washington and London, and we triple it for Los Angeles, whose topography causes pollutants to disperse especially slowly. We do not adjust external accident costs because the evidence suggests that, despite higher traffic densities in urban areas, external accident risks are not necessarily higher, given the counteracting effect of slower-moving traffic (Lindberg 2001, 406–407).

Also from Parry and Small (2005), we assume fuel taxes of 40 cents per gallon for the U.S. cities<sup>37</sup> and 280 cents per gallon in London. We use their nationwide average fuel efficiencies of 20 and 30 miles per gallon for the off-peak period (on the assumption that most travel nationwide is in conditions similar to off-peak travel in these very large metropolitan areas) but reduce them by 25 percent in the peak period to adjust for the effect of congestion on fuel economy.

For bus, accidents costs per vehicle mile are taken to be the same as for auto because buses move more slowly and are driven by professionals, offsetting their much greater weight, but pollution is taken to be triple that for automobiles.<sup>38</sup> Bus global-warming costs are computed assuming fuel efficiency of 5 miles per gallon. When expressed per passenger mile, all three of these external costs are very small for bus (one cent per passenger-mile), and are taken to be zero for rail.

Our estimates of external costs omit road wear, which is negligible for autos but perhaps not for buses due to their weight and small number of axles over which it is distributed. Still, we think these are small enough to ignore. Buses probably cause marginal road damage similar to that of lighter single-unit trucks, which has been estimated at around 3 cents per vehicle-mile on urban interstate roads (U.S. FHWA 2000, Table 13), an amount that would have negligible effect on our results. The costs might be considerably higher on city streets because of their thinner pavements, but we would not expect them to dominate the congestion costs and scale economies that we find.

*Dwell times.* For bus, we adopt the midrange values for typical boarding and alighting times

---

<sup>37</sup> The federal tax was 18.4 cents per gallon; state-level taxes in California, the District of Columbia, Virginia, and Maryland were approximately 20 cents per gallon (U.S. DOC 2003, Table 730).

<sup>38</sup> These assumptions are consistent with estimates of relative external costs per vehicle mile for heavy trucks and autos in U.S. FHWA (1997, Table 13); separate estimates for bus are not available.

from TRB (2000, Exhibit 27-9), assuming two doors for alighting and boarding. We assume cash payment for the U.S. cities and prepayment (which allows rear-door boarding) for London. This yields values of 4.275 seconds for the U.S. cities and 3.375 sec for London (for comparison, Dueker et al. 2004 estimate 5.18 sec in Portland). For rail, we use the estimate by Kraus (1991, 256) from observations in Boston, which is  $1.0/N_T$  sec where  $N_T$  is the number of cars per train. In each case we divide by trip length to specify parameter  $\theta^j$ . The marginal cost of increased dwell time is then calculated from (12d), using parameters already described.

*Generalized price of travel.* The components of  $q^{ij}$  are given by (10c); besides parameter values already described, we need the time per mile of transit vehicles  $t^{ij}$  and access and crowding elasticities  $\eta_a^{ij}$  and  $\eta_c^{ij}$ . (This is in fact the only place where we need an empirical estimate of  $\eta_c^{ij}$ .)

To calculate  $t^{ij}$ , we divide total vehicle miles by vehicle hours to give average speeds, over the day, of 23 and 11 miles per hour for Washington rail and bus, and 23 and 12 miles per hour for Los Angeles rail and bus. For London, we have a direct estimate of average speeds from the agency: namely 20 miles per hour for rail, 11 for bus. For all three cities we assume the ratio of peak to off-peak speed is 1.0 for rail, while for bus it is the same as that for autos: approximately 0.86 for Washington and London, 0.79 for Los Angeles.

The access-time elasticity  $\eta_a^{ij}$  depends on route density in a manner similar to how the wait-time elasticity depends on service frequency. It is one if people live at uniformly distributed locations and walk to the nearest transit stop, and smaller if people living farther away choose a faster access mode with a fixed cost (e.g., park and ride). The less dense the transit network, the more important these other access modes, so the lower the elasticity. We assume other access modes have minor importance in London but more in Washington and more still in Los Angeles, and so choose  $\eta_a^{ij} = 0.8, 0.65$  and  $0.5$  for these cities, respectively.

There is little empirical basis for gauging  $\eta_c^{ij}$ , which is positive only for peak service; we set it to 1.5 in the baseline, though our results are not sensitive to different assumptions (because crowding costs are relatively small).

*Own-price travel demand elasticities.* Our model calls for elasticities of each mode's passenger

demand with respect to its own generalized price  $q^{ij}$ , denoted as  $\eta_q^{ij}$ . However, most empirical evidence is based on elasticities with respect to fare  $p^{ij}$ , which we denote as  $\eta_p^{ij}$ . We first review the evidence on  $\eta_p^{ij}$ , then describe how we convert to  $\eta_q^{ij}$ .

Based on Lago, Mayworm, and McEnroe (1981), Goodwin (1992), and Pratt et al. (2000), we assume that the own-fare demand elasticity, averaged over peak and off-peak time periods, is -0.5 for bus and -0.3 for rail,<sup>39</sup> and that in each case the elasticity in the off-peak period is twice that in the peak. Given that about 70 percent of passenger mileage occurs during the peak period, the values just stated imply own-fare elasticities  $\eta_p^{ij}$  of approximately -0.40 and -0.8 for peak and off-peak bus, and -0.24 and -0.48 for peak and off-peak rail, respectively. To convert these to generalized-price elasticities  $\eta_q^{ij}$ , we assume that the empirical measurement of  $\eta_p^{ij}$  incorporates the effects of  $p^{ij}$  on  $w^{ij}$  in (10c), as discussed in the derivation of (14c); that is, we assume

$$(B2) \quad \eta_p^{ij} = \frac{p^{ij}}{M^{ij}} \frac{dM^{ij}}{dq^{PR}} \frac{dq^{ij}}{dp^{ij}} = \eta_q^{ij} \frac{p^{ij}}{q^{ij}} \frac{dq^{ij}}{dp^{ij}}$$

where the ratio and the derivative on the right-hand side are both obtained from (10c). Thus we simply invert equation (B2) to obtain our estimates of  $\eta_q^{ij}$ , which we assume to be constants.

*Modal diversion ratios,  $m_{kl}^{ij}$ .* Pratt et al. (2000, 12–41 ff.) provide several estimates for U.S. cities of the proportion of incremental transit trips that are diverted to or from other modes following a change in transit price; typical numbers, averaged across time of day, are about 65 percent and 80 percent for Atlanta and Los Angeles, respectively. Nearly all of these shifts are to or from cars. We assume that Washington is like Atlanta, and that peak values  $m_{PCAR}^{Pj}$  are 0.05 higher, and off-peak values  $m_{OCAR}^{Oj}$  0.05 lower, than these average values.

Now consider the cross-elasticities between bus and rail transit. The few studies available typically find them to be about half the direct elasticities in cities with good coverage by both

---

<sup>39</sup> A recent review of mostly U.K. studies by Paulley et al. (2006) produces somewhat larger long-run elasticities, which they suggest is because elasticities have risen in magnitude and are higher in the United Kingdom than in other nations. Many of the studies relied upon by Paulley et al. are unpublished, and we do not feel the evidence is strong enough to apply these higher elasticities to our U.K. simulations.

bus and rail transit systems, such as London and Chicago (Gilbert and Jalilian 1991, Table 3b; Talvitie 1973). Assuming equal travel volume by mode, this would imply  $m_{iR}^{iB} = m_{iB}^{iR} \approx 0.5$  for  $i=P,O$ . However, we expect the substitutability between modes to decrease as one expands beyond the city to the metropolitan area, and to decrease more for cities with less and less well developed rail networks. We also expect them to have declined considerably from the 1970s or 1980s to the year 2000 because of increasing competition from the automobile. Finally, in the newer U.S. transit systems the bus lines are typically reconfigured to serve as feeders to the rail system, with competitive routes discontinued. Therefore, we assume the cross-mode diversion ratios to be just 10 percent for Washington ( $m_{iR}^{iB} = m_{iB}^{iR} = 0.1$ ) and 5 percent for Los Angeles ( $m_{iR}^{iB} = m_{iB}^{iR} = 0.05$ ).

For London, we expect less diversion to automobile and more to the other transit mode because of the smaller initial share of automobiles and travelers' more extensive transit choices. We therefore set London's diversion ratios to be like those for Washington, except 0.20 smaller for auto in the same time period, and 0.20 larger for other transit in the same time period.

Little information is available about shifts of transit riders across time periods. We assume that in each case, 10 percent of the change in transit ridership represents such shifts, and that the shifts occur entirely to the same transit mode.

Those assumptions lead to the values shown in Table 2. The fraction of extra transit trips from increased travel demand is a residual, equal to between zero and 20 percent. The review by Pratt et al. (2000) suggests that 10 percent and 26 percent of new transit trips in Los Angeles and Atlanta, respectively, represented some combination of changes in walking, trip frequency, and destination during the 1990s. Given the likely further decline in this fraction due to metropolitan decentralization, this evidence is roughly consistent with our assumed values.

## Additional References for Appendices

APTA. 2002. *Public Transportation Fact Book*. American Public Transportation Association, Washington, DC.

Arnott, Richard, André de Palma, and Robin Lindsey. 1993. "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand." *American Economic Review* 83: 161–79.

Dueker, Kenneth J., Thomas J. Kimpel, and James G. Strathman. 2004. "Determinants of Bus Dwell Time." *Journal of Public Transportation* 7: 21–40.

Gilbert, Christopher L., and Hossein Jalilian. 1991. "The Demand for Travel and for Travelcards on London Regional Transport." *Journal of Transport Economics and Policy* 25: 3–29.

Goodwin, P.B. 1992. "A Review of New Demand Elasticities with Special Reference to Short and Long Run Effects of Price Changes." *Journal of Transport Economics and Policy*: 26: 155–69.

Lago, Armando M., Patrick D. Mayworm, and J. Matthew McEnroe. 1981. "Further Evidence on Aggregate and Disaggregate Transit Fare Elasticities." *Transportation Research Record* 799: 42–47.

Lindberg, Gunnar. 2001. "Traffic Insurance and Accident Externality Charges." *Journal of Transport Economics and Policy* 35: 399–416.

LUL. 2003. *Environmental Report 2003*. London: London Underground Ltd. <http://www.tfl.gov.uk/assets/downloads/environmental-report-2003.pdf>.

Paulley, Neil, Richard Balcombe, Roger Mackett, Helena Titheridge, John Preston, Mark Wardman, Jeremy Shires, and Peter White. 2006. "The Demand for Public Transport: The Effects of Fares, Quality of Service, Income and Car Ownership." *Transport Reviews* 13: 295–306.

Pratt, Richard H., Texas Transportation Institute, Cambridge Systematics, Parsons Brinkerhoff Quade & Douglas, SG Associates, and McCollom Management Consulting. 2000. *Traveler Response to Transportation System Changes: Interim Handbook*. Transportation Research Board, Transit Cooperative Research Program Web Document 12 (March).

Schrank, David, and Tim Lomax. 2003. *The 2003 Urban Mobility Report*, Texas Transportation Institute, Texas A&M University System, College Station, Texas.

Small, Kenneth A. 1982. "The Scheduling of Consumer Activities: Work Trips." *American Economic Review* 72: 467–479.



- Small, Kenneth A. 1992. *Urban Transportation Economics*. Fundamentals of Pure and Applied Economics, Volume 51, Harwood Academic Press, Chur, Switzerland.
- Talvitie, Antti. 1973. "A Direct Demand Model for Downtown Worktrips." *Transportation* 2: 121–52.
- TfL. 2003. *London Travel Report 2003*. London: Transport for London. [www.tfl.gov.uk/tfl/reports\\_library\\_stats.shtml](http://www.tfl.gov.uk/tfl/reports_library_stats.shtml), accessed June 11, 2004.
- TfL. 2004a. *About London Buses*. [www.tfl.gov.uk/buses/cib\\_about.shtml](http://www.tfl.gov.uk/buses/cib_about.shtml), accessed July 30, 2004.
- TfL. 2004b. *Amazing Facts*. <http://tube.tfl.gov.uk/content/faq/facts.asp>, accessed July 30, 2004.
- Tisato, Peter. 1998. "Optimal Bus Subsidy and Cross Subsidy with a Logit Choice Model." *Journal of Transport Economics and Policy* 32: 331–50.
- TRB. 2000. *Highway Capacity Manual*. Transportation Research Board, National Research Council, Washington, DC.
- U.K. DfT. 2003. *Transport Statistics Great Britain*. Department for Transport, London. <http://www.dft.gov.uk/pgr/statistics/datatablespublications/public/annualbulletins/abulletinofpublictransportst5264>.
- U.K. ONS. 2004. "Hours Worked: Growth Levels Off." In: *Labour Market in 2002*, Office for National Statistics. [www.statistics.gov.uk/cci/nugget.asp?id=696](http://www.statistics.gov.uk/cci/nugget.asp?id=696), accessed June 11, 2004.
- U.S. BLS. 2004. "2002 Metropolitan Area Occupational Employment and Wage Estimates." Bureau of Labor Statistics, Washington. [http://www.bls.gov/oes/oes\\_2002.htm](http://www.bls.gov/oes/oes_2002.htm).
- U.S. DOC. 2003. *Statistical Abstract of the United States 2003*. U.S. Department of Commerce, Census Bureau, Washington.
- U.S. FHWA. 1997. *1997 Federal Highway Cost Allocation Study*. U.S. Federal Highway Administration, Department of Transportation, Washington. <http://www.fhwa.dot.gov/policy/otps/costallocation.htm>.
- U.S. FHWA. 2000. *Addendum to the 1997 Federal Highway Cost Allocation Study Final Report*. U.S. Federal Highway Administration, Department of Transportation, Washington, D.C. <http://www.fhwa.dot.gov/policy/otps/costallocation.htm>.
- U.S. FTA. 2003. *National Transit Databases, 2003 Database*. Federal Transit Administration, Department of Transportation, Washington. [www.ntdprogram.com](http://www.ntdprogram.com).

U.S. OMB. 1992. *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs*, Circular No. A-94, Revised, Section 8. U.S. Office of Management and Budget, Washington, October.