Public Transit Investment and Traffic Congestion Policy[*]

Justin Beaudoin[†], Y. Hossein Farzin[‡] and C.-Y. Cynthia Lin Lawell[§]

August 9, 2018

Abstract

We develop a theoretical model to analyze what role, if any, public transit investment should play in addressing traffic congestion in urban transportation networks. In particular, we evaluate the extent to which traffic congestion should be accounted for when determining the optimal second-best level of investment in public transit infrastructure in the absence of a first-best Pigouvian congestion tax on auto travel. Our model of second-best public transit investment contributes to the literature by allowing for both demand and cost interdependencies between the auto and transit modes. In particular, owing to cost interdependencies between the auto and transit modes when transit shares the right-of-way with auto traffic, 'mixed traffic' transit investment can affect the equilibrium volume of auto travel through shifts in the auto travel cost function as well as the demand function. Our results indicate that the level of transit investment should be higher relative to that chosen when the congestion-reduction effects of transit are not accounted for, but the importance of this consideration is dependent upon the interaction of demand and cost interdependencies between the auto and transit modes, which may vary across regions. We calibrate our theoretical model with panel data from 96 urban areas across the United States over the period 1991 to 2011, and find that, due to differences in cost interdependence and cross-modal substitution, fixed guideway transit investments are expected to yield higher congestion-reduction benefits than mixed transit modes in dense regions. Our results suggest that urban mass transit may have a co-benefit of congestion reduction. As a consequence, prospective public transit projects should not be evaluated exclusively in terms of the forecasted net welfare generated by public transit users, but instead should also include interactions between auto and transit users in the cost-benefit analysis framework.

JEL Classifications: R41, R42, R48, R53, Q58

Keywords: traffic congestion, public transit investment, urban transportation, second-best policies, externality regulation

[*]We thank Erik Hurst, Doug Larson, Marcel Oesterich, Chad Sparber, and Matthew Turner; participants at the 2013 International Transportation Economics Association Conference and the 2013 Canadian Resource and Environmental Economics Study Group Conference; and participants in seminars at the University of California at Davis, California State University at Fullerton, Colgate University, University of Washington Tacoma, and Furman University for helpful comments and discussions. Beaudoin gratefully acknowledges the support of a Doctoral Dissertation Grant from the University of California Transportation Center (UCTC) and a Provost's Dissertation Year Fellowship in the Arts, Humanities and Social Sciences from the University of California at Davis. Farzin is an emeritus member and Lin Lawell is a former member of the Giannini Foundation of Agricultural Economics. All errors are our own.

[†]Corresponding author: jbea@uw.edu, Assistant Professor, School of Interdisciplinary Arts & Sciences, University of Washington Tacoma.

[‡]Professor Emeritus, Department of Agricultural and Resource Economics, University of California at Davis; Oxford Centre for the Analysis of Resource Rich Economies (OxCarre), Department of Economics, University of Oxford, UK.

[§]Associate Professor, Dyson School of Applied Economics and Management, Cornell University.

1 Introduction

Studies have shown that traffic congestion is the number one concern of individuals in rapidly growing areas in the U.S., often ranked higher than crime, school over-crowding, and housing shortages (Morrison and Lin Lawell, 2016). Congestion costs comprise the majority of the external costs of automobile travel for urban commuters in the U.S.[1] In 2011, the costs of traffic congestion (including the valuation of travel time and the increased operating and fuel costs of auto travel) were estimated to be $121 billion in the U.S. (Schrank et al., 2012).[2]

Public transit is often advocated as a means to decrease traffic congestion within urban transportation networks. Recent expenditures on public transit capital in the U.S. have exceeded $18 billion per year (American Public Transportation Association, 2012). However, despite public transit investment and other existing transportation policies, congestion and other market failures from urban transportation remain. As Winston (2000, pp. 411) notes: "Large public transit deficits, low transit load factors and severe highway congestion...suggest that the US public sector is not setting urban transportation prices and service to maximise net benefits." The U.S. Government Accountability Office has recently outlined the failure of transportation infrastructure investment programs to incorporate rigorous economic analysis and the ongoing absence of a link between investment and system performance, and there has been increasing concern about the fiscal sustainability of highway and transit operations (Libermann, 2009).[3]

In this paper, we develop a theoretical model to analyze what role, if any, public transit investment should play in traffic congestion policy. In particular, we evaluate the extent to which traffic congestion should be accounted for when determining the optimal second-best level of investment in public transit infrastructure in the absence of a first-best Pigouvian congestion tax on auto travel.

In order to determine the optimal level of investment in public transit, one should account for the effect of public transit investment on overall net social welfare, taking into account the cost of the initial investment, as well as any required operating subsidies over the life of the project.[4]

[1] Of the combined per vehicle-mile costs of congestion, accidents, and environmental externalities for urban commuters in the U.S., congestion costs represent 71.7% of the short-run average variable social cost of auto travel and 74.3% of the short-run marginal variable social cost (Small and Verhoef, 2007, pp. 98). Similarly, of the externalities associated with gasoline consumption that Lin and Prince (2009) analyze in their study of the optimal gasoline tax for the state of California, the congestion externality is the largest and should be taxed the most heavily, followed by oil security, accident externalities, local air pollution, and global climate change.

[2] Congestion can be particularly costly if individuals exhibit preferences for urgency owing to time constraints, schedule constraints, and possible penalties for being late (Bento et al., 2017).

[3] Glaeser and Ponzetto (2017) find that, owing to politics, there is over-investment in transportation projects in low-density and less educated areas, and under-investment in areas with more educated and organized urban voters.

[4] The broader issue of the funding of public transit within the U.S. urban transportation sector has been discussed by others (see e.g., Viton (1981) Winston and Shirley (1998)); in this paper we focus on the congestion-reduction

Moreover, the optimal level of public transit infrastructure and service to provide depends on the policy instruments employed to address traffic congestion.

In the first-best, a Pigouvian congestion tax would be levied on auto travel, which generates a direct price for the congestion externality and not only limits the deviation from the socially optimal level of travel and helps utilize existing capacity more efficiently, but also results in a volume of travel that provides an appropriate signal for the optimal level of capacity investment in the future. In this case, efficient public transit investment can be determined in a first-best setting, with the relevant welfare effects of public transit investment confined to the direct effects in the transit market.

The critical assumption required for the first-best framework to be appropriate is that there are no uncorrected distortions in the transportation market and its related markets. In general, however, this is not the case for urban transportation in the U.S. As has been well-documented,[5] transportation involves a number of social costs that are not currently being internalized by individual users, with the distortion receiving the most attention being the absence of marginal cost pricing related to the congestion externality associated with fixed road capacity. Congestion taxes remain underutilized in practice, due to a combination of economic factors (for example, the transaction costs of implementing the tax) and political reticence (Anas and Lindsey, 2011). Thus, since the first-best Pigouvian congestion tax is not (or cannot be) levied on auto travel, second-best public transit investment policies (rather than first-best public transit investment policies) are appropriate, and this is indeed the policy-relevant landscape in the U.S. at present.

If it is accepted *a priori* that policy instruments that would in theory achieve a first-best outcome cannot be employed due to various economic and political constraints, then it is of interest to analyze potential second-best solutions available to policymakers. The general concept of subsidizing a substitute good in the presence of an uncorrected distortion has long been established (Baumol and Bradford, 1970); in this paper, we apply this concept by developing a model of public transit investment in the absence of congestion pricing on auto travel to evaluate the effects of public transit supply on equilibrium traffic congestion. Specifically, we address the following question: Is there a theoretical justification for increasing public transit investment as a means of dealing with traffic congestion?

Our second-best model of public transit investment contributes to the literature by allowing for cost interdependencies between the auto and transit modes in addition to demand substitutability across modes, with auto travel costs potentially varying with the type and level of transit capacity

effect of public transit, which is a potentially important component of this broader evaluation process.

[5] See Small and Verhoef (2007, Table 3.3, pp. 98) and Parry et al. (2007) for recent empirical estimates of the internal and external costs of automobile travel.

provided due to the interaction between auto and transit vehicles in the roadways.

With ‘fixed guideway’ transit investment, whereby transit has its own separate right-of-way and does not directly interact with auto traffic, the auto travel cost is independent of transit capacity. In contrast, with ‘mixed traffic’ transit investment – whereby transit shares the right-of-way with auto traffic – this interaction of transit vehicles and autos implies that the auto travel cost is functionally dependent on the level of transit capacity. Owing to cost interdependencies between the auto and transit modes when transit shares the right-of-way with auto traffic, ‘mixed traffic’ transit investment can affect the equilibrium volume of auto travel through shifts in the auto travel cost function as well as the demand function.

Our results indicate that the level of transit investment should be higher relative to the level that would be chosen when the congestion-reduction effects of transit are not accounted for.[6] However, the importance of this consideration is dependent upon these demand and cost interdependencies, which may vary across regions.

We calibrate our theoretical model with panel data from 96 urban areas across the United States over the period 1991 to 2011 to numerically evaluate the relative importance of four factors that vary across regions and influence the extent to which transit investment decreases congestion: (1) the existing level of congestion, (2) the effect of transit investment on the generalized cost of transit travel, (3) the sensitivity of auto demand to changes in the generalized cost of transit travel, and (4) the magnitude of cross-modal cost interactions.

Results of our numerical model show that, due to differences in cost interdependence and cross-modal substitution, fixed guideway transit investments are expected to yield higher congestion-reduction benefits than mixed transit modes in dense regions. The ambiguous predictions of our model and the potential spatial heterogeneity of the congestion-reduction effect of public transit help to reconcile the existing inconclusive empirical evidence in the literature regarding the impact of public transit on congestion levels.

Our results suggest that urban mass transit may have a co-benefit of congestion reduction. As a consequence, prospective public transit projects should not be evaluated exclusively in terms of the forecasted net welfare generated by public transit users, but instead should also include interactions between auto and transit users in the cost-benefit analysis framework.

The balance of our paper proceeds as follows. We review the literature in Section 2. We present our

[6] If other transportation externalities, such as pollution, were taken into account, then the second-best investment level would be higher still.

theoretical model in Section 3 and describe the results and insights from our theory model in Section 4. We present the numerical results from calibrating our theory model with panel data from 96 urban areas across the United States over the period 1991 to 2011 in Section 5. Section 6 concludes.

2 Literature Review

Mohring and Harwitz (1962) and Vickrey (1969) were among the first to recognize the link between the price of auto travel and the optimal level of investment; Lindsey (2012) provides a more recent discussion. Recognizing the institutional reality that the first-best benchmark for transportation investment is generally not attainable due to the absence of a congestion tax on auto travel, a variety of 'second-best' approaches have since been examined as a result.

Most studies relating to second-best investment have focused on how distortions in the transportation market should be accounted for when evaluating road investments (see e.g., Wheaton, 1978; Friedlaender, 1981; d'Ouville and McDonald, 1990; Gillen, 1997). Recent research has incorporated endogenous investment in public transit capacity along with second-best pricing, including applications to the 'two-mode problem' that accounts for the interaction between auto and transit (see Berechman (2009) for a discussion of first-best versus second-best public transit investment, and models by Henderson, 1985; Arnott and Yan, 2000; Pels and Verhoef, 2007; Ahn, 2009; and Kraus, 2012).

Most of the previous second-best models with endogenous transit capacity have relied on an assumption that the costs of travel for auto are independent of the level of transit supplied. This assumption leads Kraus (2012) to find a global result that second-best transit capacity – accounting for the distortion in the auto market – is higher than the first-best capacity that does not account for this distortion. A key feature of our model that distinguishes it from the existing second-best investment models in the literature is that it relaxes this assumption and allows for both demand and cost interdependencies between the auto and transit modes.

Overall, the existing empirical evidence of the effect of transit investment on traffic congestion is mixed (Baum-Snow and Kahn, 2005; Winston and Langer, 2006; Winston and Maheshri, 2007; Nelson et al., 2007; Duranton and Turner, 2011; Anderson, 2014; Hamilton and Wichman, 2018).[7] Beaudoin and Lin Lawell (2018) find that the time horizon and regional heterogeneity are important when analyzing the effect of past public transit investment on the demand for automobile transportation, and help reconcile the literature's seemingly conflicting evidence regarding the empirical

[7] See Beaudoin, Farzin and Lin Lawell (2015) and Beaudoin and Lin Lawell (2017) for detailed discussions and comparisons of these studies.

relationship between transit investment and the volume of auto travel.

3 Theoretical Model

We develop a model of public transit investment in the presence of an uncorrected congestion externality related to auto travel. In our model, the transportation network of an urban area consists of two modes of travel: auto and public transit. Aggregate modal travel volumes are denoted by V_j for each mode $j \in \{\text{auto (A), transit (T)}\}$. Our model allows for both demand and cost interdependencies between the auto and transit modes.

3.1 Demand and Cost Interdependencies Between Auto and Transit

Demand interdependencies between the auto and transit modes arise when there is demand substitutability across modes. In this case, transit investment can affect the equilibrium volume of auto travel through cross-modal substitution that leads to shifts in the auto demand curve.

Demand interdependencies lead to the countervailing effects of substitution and induced demand. On the one hand, an increase in transit supply may cause some commuters to substitute transit travel for trips previously taken by automobile (the "substitution effect"), thereby decreasing auto travel. On the other hand, by reducing congestion, increasing accessibility, increasing economic activity, and/or attracting additional residents and workers to the area, transit investment may generate additional automobile trips that were previously not undertaken (the "induced demand effect"). The "equilibrium effect" accounts for both the substitution effect and the induced demand effect (Beaudoin and Lin Lawell, 2018).

Cost interdependencies between the auto and transit modes arise when auto travel costs and transit travel costs are interdependent. For example, auto travel costs may potentially vary with the type and level of transit capacity provided due to the interaction between auto and transit vehicles in the roadways.

With 'fixed guideway' transit investment, whereby transit has its own separate right-of-way and does not directly interact with auto traffic, the auto travel cost is independent of transit capacity. An example of a fixed guideway transit investment is a new light rail system.

In contrast, with 'mixed traffic' transit investment – whereby transit shares the right-of-way with auto traffic – this interaction of transit vehicles and autos implies that the auto travel cost is functionally dependent on the level of transit capacity. An example of a mixed traffic transit investment is an expanded mixed traffic bus service that interacts with autos on congested roadways. Owing to cost interdependencies between the auto and transit modes when transit shares the right-of-way with auto traffic, mixed traffic transit investment can affect the equilibrium volume of auto travel through shifts in the auto travel cost function as well as the demand function.

3.2 Travel Demand

The inverse demand functions $D_A(\cdot)$ and $D_T(\cdot)$ for auto and transit, respectively, represent the marginal willingness to pay for travel via each mode[8]. We assume that the marginal private benefit of travel is equal to the marginal social benefit of travel.[9]

3.3 Infrastructure Investment and Capital Provision

We assume a fixed level of auto capacity, denoted by $\overline{K}_A$.

Public transit investment can occur along two dimensions. First, the size of the public transit network can be expanded by increasing the route coverage. Such investment occurs along the extensive margin and is related to the accessibility of public transit service. Transit network size is denoted by K_T^S.

Second, the capacity of the public transit network can be expanded by increasing the service frequency provided across the public transit network. This investment occurs along the intensive margin and is related to the waiting and travel time associated with public transit travel. Transit capacity is denoted by K_T^C.

Together, transit investment is denoted via the vector $\mathbf{K_T} = \left(K_T^S, K_T^C\right)$, and the public transit investment cost function is given by $I_T\left(\mathbf{K_T}; \overline{K}_A\right)$.

[8] The aggregate inverse demand curves are derived by summing the individual travel decisions of a region's residents, including the decision about whether or not to undertake a trip. We assume the existence of continuous aggregate modal demand functions; while *individual* travelers may face a discrete choice between the two modes, the *aggregate* demand for the entire network – on a per unit of travel basis – is well represented as a continuous function of the marginal unit cost of travel. We do not model other margins of travel behavior (such as route choice and trip timing), given the aggregate network-level of analysis undertaken

[9] While there may be external benefits associated with the construction of infrastructure in some cases (primarily relating to economies of agglomeration due to improved accessibility, and reductions in market power brought about by reduced transaction costs), the marginal unit of travel is unlikely to confer such positive externalities, particularly in highly-developed urban regions in the U.S. (see discussion in Small and Verhoef, 2007).

3.4 Travel Cost

The generalized cost of travel encapsulates the full per-unit[10] cost of travel via mode k, combining both monetary and non-monetary aspects, and represents an individual's opportunity cost of travel.

To reflect the different effects that the size of the transport network and the capacity provided over the network can have on travelers' choices, and the differences in the value of time associated with different travel activities, the amount of travel time $\tilde{T}_j^i$ is identified separately by mode and activity, where j represents the mode and i represents the activity, with $i \in \{\text{access (A), wait (W), travel (T)}\}$. Similar notation is used for the value of time VOT_j^i, which is also assumed to differ across modes and activity types. The monetized value of time spent per activity is then calculated as $T_A^T = \tilde{T}_A^T \cdot VOT_A^T$ for auto travel time, and $T_T^i = \tilde{T}_T^i \cdot VOT_T^i$ for transit travel activity i.

3.4.1 Marginal Private Cost of Auto Travel

The marginal per-unit private cost of auto travel MPC_A is given by the sum of the monetary cost of auto travel P_A (which includes the variable out-of-pocket expenses such as fuel), the monetized value of time T_A^T, and the per-unit tax levied on auto travel τ:

$$MPC_A\left(V_A, V_T, \tau; K_T^C, \overline{K}_A\right) \equiv P_A\left(\frac{V_A}{\overline{K}_A}, \frac{K_T^C}{\overline{K}_A}, \frac{V_T}{K_T^C}\right) + T_A^T\left(\frac{V_A}{\overline{K}_A}, \frac{K_T^C}{\overline{K}_A}, \frac{V_T}{K_T^C}\right) + \tau. \tag{1}$$

With a fixed level of auto capacity, increases in travel volume beyond a threshold lead to congestion, and each marginal unit of travel thus imposes an external congestion cost on all other users in the network; this effect is manifested in longer travel times for auto, and potentially longer travel and waiting times for transit if there is physical interaction between auto and transit vehicles within the transportation network.[11] It is assumed that auto travel time is homogeneous of degree zero in auto travel volume and capacity, such that a proportionate increase in volume and capacity leaves average travel time unchanged; however, given a fixed level of auto capacity, the auto travel time function is assumed to be convex with respect to travel volume, consistent with empirical estimates of travel speed-flow relationships (see Small and Verhoef, 2007).

The cost of auto travel depends on the congestion experienced over the network, determined by the following factors: (1) the volume-to-capacity ratio $\frac{V_A}{\overline{K}_A}$, representing the relationship between auto congestion and travel speeds; (2) $\frac{K_T^C}{\overline{K}_A}$, representing the congestion effect associated with transit vehicles interacting with autos on the roadways, which causes the cost of auto travel to depend

[10] Per-unit measures are on a 'per mile of travel' basis throughout.

[11] Access and wait times are assumed to be negligible for auto travel.

on the level of transit capacity; and (3) $\frac{V_T}{K_T^C}$, representing the congestion effect associated with passengers boarding and disembarking transit vehicles and affecting auto travel speeds through the auto-transit interaction, which also causes the cost of auto travel to depend on transit supply. These potential effects of transit capacity on the cost of auto travel cost have been discussed by Sherman (1971), Viton (1981), Ahn (2009) and Basso and Silva (2014).

3.4.2 Marginal Private Cost of Transit Travel

Similarly, the marginal per-unit private cost of transit travel MPC_T is given by the following sum of the transit fare P_T and the monetized values of access, wait, and travel times:

$$MPC_T\left(V_A, V_T, K_T^S, K_T^C; \overline{K}_A\right) \equiv P_T + T_T^A\left(K_T^S\right) + T_T^W\left(\frac{K_T^C}{K_T^S}, \frac{V_A}{\overline{K}_A}, \frac{V_T}{K_T^C}\right) + T_T^T\left(\frac{V_A}{\overline{K}_A}, \frac{V_T}{K_T^C}, \frac{K_T^C}{\overline{K}_A}, \frac{K_T^S}{\overline{K}_A}\right). \quad (2)$$

The transit access time cost T_T^A is a function of the size of the public transit network K_T^S. Transit wait time T_T^W is affected by the following factors: (1) $\frac{K_T^C}{K_T^S}$, which is a measure of the average headway per route; (2) $\frac{V_A}{\overline{K}_A}$, which represents auto congestion and affects transit schedule delay costs associated with uncertain departure times, and which causes the cost of transit travel to depend on the auto mode; and (3) $\frac{V_T}{K_T^C}$, which represents the transit congestion associated with passengers boarding and disembarking the vehicle and also influences the probability of encountering a full transit vehicle and the requirement to wait for the subsequent vehicle's arrival. Transit travel time T_T^T is affected by the following factors:[12] (1) auto congestion, measured by $\frac{V_A}{\overline{K}_A}$, which also causes the cost of transit travel to depend on the auto mode; (2) the direct transit congestion effect, represented by $\frac{V_T}{K_T^C}$; (3) the transit congestion associated with transit vehicles interacting with autos on the roadways, represented by $\frac{K_T^C}{\overline{K}_A}$, which similarly causes the cost of transit travel to depend on the auto mode; and (4) the average potential trip distance that is influenced by overall transit route coverage, measured by $\frac{K_T^S}{\overline{K}_A}$, and again causes the cost of transit travel to depend on the auto mode. Appendix A shows how increased transit investment leads to a reduction in the marginal private cost of transit travel.

3.4.3 The Marginal External Cost of Travel

Each individual views the various cost components as parametric with respect to their travel decision and thus independent of their own travel volume (analogous to the incentive structure of

[12] Empirical evidence regarding the effects of auto and transit vehicle interaction was provided by the introduction of the congestion pricing scheme in central London in 2003, where a 15% reduction in auto travel was associated with a 6% increase in bus travel speeds in the area (Small and Verhoef, 2007).

an open access resource). In doing so, each individual ignores the external effects of their travel decisions on the other individuals in the network, and in congested conditions each individual's travel adversely affects the other individuals in the network.

The marginal *social* cost of travel MSC_j for each mode j accounts for both the marginal private costs incurred by the individual, as well as the external effects transmitted via the marginal increase in the average travel cost throughout the network that they generate; this is a technological externality whereby each individual's average travel cost is dependent upon the travel volumes chosen by other users across the network. Since the marginal private cost functions specified above are interdependent across modes, there are both intra- and inter-mode externalities.

The marginal *external* cost of auto travel MEC_A is then the difference between the marginal social cost and marginal private cost of travel. In congested conditions, each marginal unit of auto travel increases the average cost of auto travel through increases in operating costs and travel time, and increases the average cost of transit travel through increases in waiting and travel time:

$$MEC_A = MSC_A - MPC_A = V_A \frac{\partial MPC_A}{\partial V_A} + V_T \frac{\partial MPC_T}{\partial V_A}. \tag{3}$$

For a given auto capacity level $\overline{K}_A$, there is a threshold travel volume $\overline{V}_A$ where the congestion externality becomes relevant:

$$\frac{\partial MPC_A}{\partial V_A} \text{ and } \frac{\partial MPC_T}{\partial V_A} = 0 \text{ if } V_A \leq \overline{V}_A, \tag{4}$$

so that the marginal external cost of auto travel $MEC_A = 0$ when the travel volume V_A is less than or equal to the threshold $\overline{V}_A$, but $MEC_A > 0$ when the travel volume V_A exceeds the threshold $\overline{V}_A$.

It is assumed that observed equilibrium travel volumes exceed the threshold $\overline{V}_A$, consistent with peak travel conditions in urban areas. By definition, the marginal external cost must be convex with respect to the volume of travel (see Baumol and Oates (1988) for a proof of the convexity of a general congestion externality). This implies that the marginal external cost exceeds the average external cost, with $\frac{\partial MPC_A}{\partial V_A} > 0$ and $\frac{\partial^2 MPC_A}{\partial {V_A}^2} > 0$, implying that the magnitude of the congestion externality varies with the volume-to-capacity ratio $\frac{V_A}{\overline{K}_A}$.

While there are several other potential distortions inherent to urban transportation beyond congestion – including vehicle emissions, accidents, noise, or various market distortions resulting from government intervention – for the purposes of exposition the model assumes that the congestion-related effects embedded in the generalized cost functions are the only externalities associated with urban travel. For urban commuters in the U.S., congestion costs (considering travel time, schedule delay, and uncertainty regarding travel times) comprise the majority of the external costs associated with automobile travel (see footnote 1).

3.5 The Effects of Transit Supply

Because our model allows for demand and cost interdependencies between the auto and transit modes, the level of transit supplied affects equilibrium travel volumes through two channels: (1) by shifting the demand curves for auto and transit travel; and (2) by shifting the marginal private cost MPC_A and marginal external cost MEC_A curves through its effects on the monetary cost of auto travel $P_A(\cdot)$, the monetized values of time for auto travel $T_A^T(\cdot)$, transit access time $T_T^A(\cdot)$, transit wait time $T_T^W(\cdot)$, and transit travel time $T_T^T(\cdot)$, all of which are functions of transit investment $\mathbf{K_T}$.

While the marginal private costs and marginal external costs of auto travel are assumed to be independent of the size of the transit network, i.e. $\frac{\partial MSC_A}{\partial K_T^S} = 0$, the level of capacity supplied over the transit network affects both the marginal private costs and marginal external costs of auto travel:

$$
\begin{aligned}
\frac{\partial MSC_A}{\partial K_T^C} &= \frac{\partial MPC_A}{\partial K_T^C} + \frac{\partial MEC_A}{\partial K_T^C} && (5) \\
&= \left[\frac{\partial MPC_A}{\partial K_T^C}\right] + \left[\frac{\partial MPC_A}{\partial V_A}\frac{\partial V_A}{\partial K_T^C} + V_A \frac{\partial \frac{\partial MPC_A}{\partial V_A}}{\partial K_T^C}\right] + \left[\frac{\partial MPC_T}{\partial V_A}\frac{\partial V_A}{\partial K_T^C} + V_T \frac{\partial \frac{\partial MPC_T}{\partial V_A}}{\partial K_T^C}\right].
\end{aligned}
$$

The net effect of changes in transit capacity on MSC_A is uncertain. There is an ambiguous direct effect of transit capacity K_T^C on the marginal private cost of auto travel (see Appendix B). An increase in transit capacity also has an undetermined effect on the marginal external cost of auto travel: by reducing auto travel volume V_A through cross-modal demand substitution, transit capacity serves to decrease the magnitude of the marginal external cost, while the additional transit capacity increases the magnitude of the marginal external cost by intensifying the interaction between auto and transit vehicles. As a result, the net effect depends upon the relative magnitudes of the various components; theoretically, an increase in transit investment could shift the auto cost function upwards, downwards, or leave it unaffected.

3.6 Equilibria

On the basis of the preceding demand and cost structure, we next specify several equilibria to illustrate how the evaluation of transit investment is dependent on the regulatory policy in place.

3.6.1 Case I: The Unregulated User Equilibrium

The unregulated case has no congestion tax in place, such that $\tau = 0$. Then for fixed auto capacity $\overline{K}_A$ and for any given transit investment level $\overline{\mathbf{K}}_{\mathbf{T}}$, the short-run capacity usage results in equilibrium travel volumes V_A^0 and V_T^0 that equate the marginal benefit of travel with the marginal private cost of travel for each mode. This outcome, where individuals do not account for any external costs associated with their travel decision, is denoted as the *user equilibrium*:

$$D_j\left(V_j^0\right) = MPC_j\left(V_j^0\right) \quad \forall\, \mathbf{K_T},\ \text{and } j \in \{A, T\}. \tag{6}$$

The user equilibrium is the manifestation of the 'fundamental law of traffic congestion': with the user equilibrium, any capacity expansion that decreases travel costs will subsequently induce additional travel that eliminates any (short-run) benefits from reduced congestion as equilibrium is reached. The absence of a tax on the congestion externality yields an open access congestible resource, with the equilibrium outcome failing to maximize the net social benefits of travel.

3.6.2 Case II: The First-Best Pareto Optimal Equilibrium

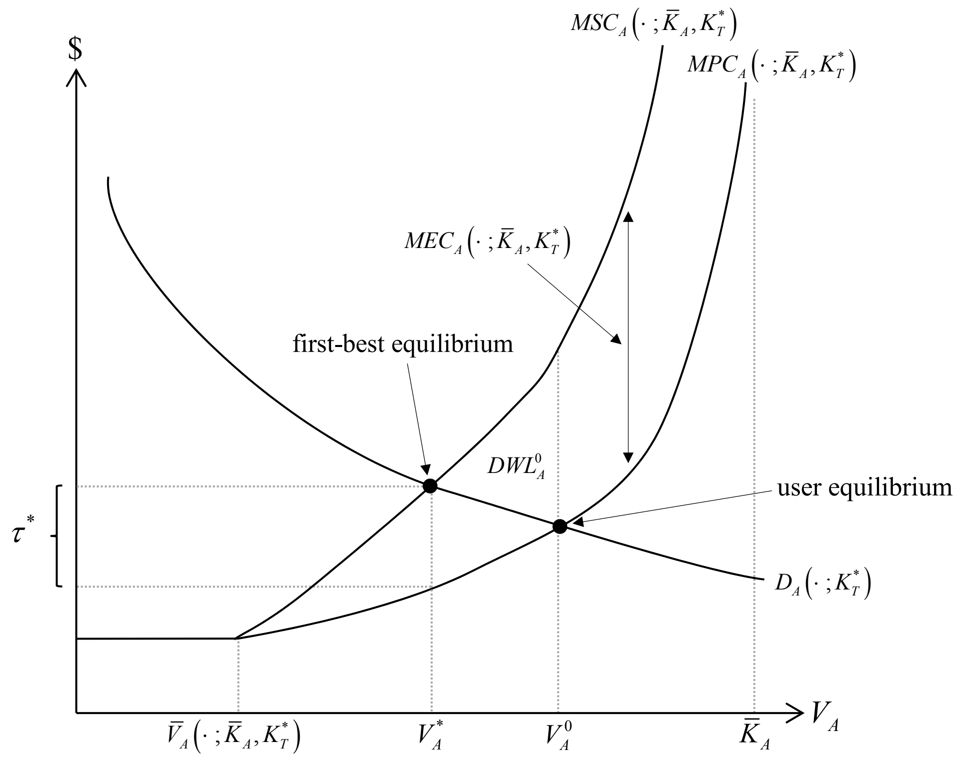
To maximize the *social* net benefits of auto and transit travel, both the volume of travel *and* the level of transit investment must be simultaneously optimized, conditional on fixed auto capacity. Here the level of public transit investment $\mathbf{K_T} = \left(K_T^S, K_T^C\right)$ is endogenous. The regulator must account for the congestion externality by levying a tax τ on each mile of auto travel. The efficient tax τ^* (i.e. the Pigouvian tax) is equal to the marginal external cost of auto travel evaluated at the efficient travel volumes and transit investment levels, while also accounting for the induced cross-modal demand and cost curve shifts brought about by the tax. If τ^* is imposed, then the user equilibrium travel volume that arises coincides with the efficient level V_A^*.

It should be noted that the Pigouvian tax is a function of the level of capital in both the auto and transit markets. With the Pigouvian tax applied, the net marginal benefit is equalized with the net marginal social cost of auto travel, and the private incentives facing each individual are aligned with the desired social incentives.

Since the marginal private costs of auto travel are convex with respect to the volume of travel as explained above, $\frac{\partial MPC_A}{\partial V_A} > 0$ and $\frac{\partial^2 MPC_A}{\partial {V_A}^2} > 0$, this implies that there is marginal external cost to auto travel when the travel volume V_A exceeds the threshold $\overline{V}_A$: $MSC_A > MPC_A\ \forall\, V_A \in \left(\overline{V}_A, \overline{K}_A\right]$. In addition, with $\frac{\partial D_A}{\partial V_A} < 0$ and $\frac{\partial^2 D_A}{\partial {V_A}^2} \geq 0$, there are unique user and first-best equilibria with auto travel volumes in the unregulated user equilibrium exceeding the first-best efficient level: $V_A^0 > V_A^*$. This yields the deadweight loss associated with the unregulated user equilibrium $DWL_A^0 > 0$ and the efficient Pigouvian tax $\tau^* > 0$.

Figure 1 illustrates the user equilibrium and the efficient equilibrium in the auto market and shows how the Pigouvian tax internalizes the congestion externality and eliminates the deadweight loss associated with the unregulated user equilibrium, DWL_A^0, by reducing auto travel from V_A^0 to V_A^*.

Figure 1: First-best equilibrium outcome vs. the user equilibrium in the auto market



3.6.3 Case III: The Second-Best Equilibrium

The general theory of the second-best applies to the urban transportation sector. If there is an uncorrected distortion in one market, then the optimality conditions in interconnected markets must be adjusted to account for the welfare implications of this distortion. Congestion taxes have had limited application in practice (apart from relatively successful implementation in Singapore, London, and Stockholm) and are essentially not employed in the U.S.; and the taxes on travel in place (predominantly fuel taxes) are not equivalent to Pigouvian taxes insofar as they are not set at the correct level, nor do they exhibit the temporal and spatial variation necessary to induce the first-best equilibrium. As a result, the second-best framework is appropriate when evaluating potential transit investments and we now consider the case where the value of τ is assumed to be set at the (fixed) suboptimal value of $\overline{\tau} \neq \tau^*$. In practice, it is usually the case that $\overline{\tau} < \tau^*$.

With this *a priori* assumption that the regulator cannot levy the optimal tax, we consider the second-best optimization problem whereby social net benefits are maximized by choosing the auto and transit volumes and the level of transit investment, given the existing auto network. In this case, the user equilibrium in (6) will be reached whereby the marginal private costs of travel are equated to the marginal benefit of travel across modes, and this equilibrium is imposed as a constraint in the optimization problem.[13] While the deadweight loss in the auto market cannot be completely eliminated in this case, the level of transit investment will influence the auto demand and travel cost functions and thus determine *which* user equilibrium is reached. This constraint also incorporates potential induced travel demand brought about by the marginal cost reductions associated with transit capacity increases, and captures the indirect effects in the auto market due to the induced modal substitution accompanying changes in the supply of transit.

The second-best constrained optimization problem is given by:

$$\begin{aligned}
\max_{\{V_A, V_T, K_T^S, K_T^C\}} \quad SNB = & \int_0^{V_A} D_A\left(v_A; MPC_T\left(V_A, V_T, K_T^S, K_T^C, \overline{K}_A\right)\right) \mathrm{d}v_A \\
& + \int_0^{V_T} D_T\left(v_T; MPC_A\left(V_A, V_T, K_T^C, \overline{\tau}, \overline{K}_A\right)\right) \mathrm{d}v_T \\
& - V_A MPC_A\left(V_A, V_T, K_T^C, \overline{\tau}, \overline{K}_A\right) - V_T MPC_T\left(V_A, V_T, K_T^S, K_T^C, \overline{K}_A\right) \\
& - I_T\left(K_T^S, K_T^C, \overline{K}_A\right) + \overline{\tau} V_A \\
\text{s.t.} \qquad K_A = & \overline{K}_A \\
\overline{\tau} < & \tau^* \\
D_A(\cdot) = & MPC_A\left(V_A, V_T, K_T^C, \overline{\tau}, \overline{K}_A\right) \qquad : \lambda_A \\
D_T(\cdot) = & MPC_T\left(V_A, V_T, K_T^S, K_T^C, \overline{K}_A\right) \qquad : \lambda_T,
\end{aligned} \tag{7}$$

where λ_A and λ_T are the Lagrange multipliers on the user equilibria constraints, representing the marginal social welfare loss of not imposing the Pigouvian tax, or equivalently the shadow price of non-optimal pricing and the marginal social benefit of an incremental movement towards the first-best equilibrium from the user equilibrium (and thus $\lambda_A \geq 0$ when $\tau < \tau^*$).

[13] This is a manifestation of Wardrop's first principle of traffic equilibrium (Wardrop, 1952), applied to the modal distribution of travel as opposed to the route distribution of travel, and an application of the "general theory of the second-best" found in early work such as Davis and Whinston (1965), Davis and Whinston (1967), and Baumol and Bradford (1970).

Assuming that there is no budget constraint relating to the level of transit investment and that transit travel is priced at its marginal social cost,[14] so that $\lambda_T = 0$, the second-best constrained optimization problem in (7) yields the first-order conditions that determine the second-best solution vector $\{V_A', V_T', K_T^{S\prime}, K_T^{C\prime}; \overline{K}_A\}$. Absent efficient pricing of auto travel, the envelope theorem does not apply and indirect effects in the auto market must be incorporated into the first-order conditions for transit investment. The resulting first-order conditions are shown in equations (8) below, with the bold terms indicating the adjustment terms in the second-best first-order conditions that do not appear in the respective first-order conditions arising from the first-best equilibrium:

$$D_A + \int_0^{V_A'} \frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial V_A}\mathrm{d}v_A - \boldsymbol{\lambda_A}\left[\boldsymbol{\frac{\partial D_A}{\partial V_A} + \frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial V_A}}\right] + \int_0^{V_T'} \frac{\partial D_T}{\partial MPC_A}\frac{\partial MPC_A}{\partial V_A}\mathrm{d}v_T$$

$$= P_A + T_A^T + \left(V_A' - \boldsymbol{\lambda_A}\right)\frac{\partial MPC_A}{\partial V_A} + V_T'\frac{\partial MPC_T}{\partial V_A} \tag{8a}$$

$$D_T + \int_0^{V_A'} \frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial V_T}\mathrm{d}v_A - \boldsymbol{\lambda_A}\left[\boldsymbol{\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial V_T}}\right] + \int_0^{V_T'} \frac{\partial D_T}{\partial MPC_A}\frac{\partial MPC_A}{\partial V_T}\mathrm{d}v_T$$

$$= P_T + \sum_{i\in\{A,W,T\}} T_T^i + \left(V_A' - \boldsymbol{\lambda_A}\right)\frac{\partial MPC_A}{\partial V_T} + V_T'\frac{\partial MPC_T}{\partial V_T} \tag{8b}$$

$$-V_T'\frac{\partial MPC_T}{\partial K_T^C} - \left(V_A' - \boldsymbol{\lambda_A}\right)\frac{\partial MPC_A}{\partial K_T^C} + \int_0^{V_A'} \frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial K_T^C}\mathrm{d}v_A - \boldsymbol{\lambda_A}\left[\boldsymbol{\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial K_T^C}}\right]$$

[14] If the Pigouvian tax on auto is in place, the optimal transit fare P_T^* satisfies the first-order conditions for Pareto optimality when:
$P_T^* = V_A^*\frac{\partial MPC_A}{\partial V_T} + V_T^*\frac{\partial MPC_T}{\partial V_T} - \int_0^{V_A^*}\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial V_T}\mathrm{d}v_A - \int_0^{V_T^*}\frac{\partial D_T}{\partial MPC_A}\frac{\partial MPC_A}{\partial V_T}\mathrm{d}v_T$. If $P_T < P_T^*$, then $\lambda_T \geq 0$, and if $P_T > P_T^*$, then $\lambda_T \leq 0$; in these cases, the under- or over-pricing of transit relative to the efficient price will require an adjustment to the conditions determining the optimal supply of transit service, analogous to the results of Wheaton (1978) for the relationship between auto pricing and investment. However, in the second-best case with inefficiently priced auto travel – auto is underpriced relative to its marginal social cost and the second-best transit fare must account for this distortion – there is a rationale for transit fare subsidies (Glaister and Lewis, 1978; Parry and Small, 2009). The transit fare would then satisfy condition (8b) when:
$P_T' = \left(V_A' - \lambda_A\right)\frac{\partial MPC_A}{\partial V_T} + V_T'\frac{\partial MPC_T}{\partial V_T} - \int_0^{V_A'}\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial V_T}\mathrm{d}v_A - \int_0^{V_T'}\frac{\partial D_T}{\partial MPC_A}\frac{\partial MPC_A}{\partial V_T}\mathrm{d}v_T + \lambda_A\left[\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial V_T}\right]$.
While future work should explore the interrelationship between transit investment and pricing in the second-best setting, this aspect is not accounted for in the present analysis in order to isolate the transit investment effect.

$$+\int_0^{V_T'} \frac{\partial D_T}{\partial MPC_A} \frac{\partial MPC_A}{\partial K_T^C} \mathrm{d}v_T = \frac{\partial I_T}{\partial K_T^C} \tag{8c}$$

$$- V_T' \frac{\partial MPC_T}{\partial K_T^S} + \int_0^{V_A'} \frac{\partial D_A}{\partial MPC_T} \frac{\partial MPC_T}{\partial K_T^S} \mathrm{d}v_A - \boldsymbol{\lambda_A} \left[\boldsymbol{\frac{\partial D_A}{\partial MPC_T} \frac{\partial MPC_T}{\partial K_T^S}} \right] = \frac{\partial I_T}{\partial K_T^S} \tag{8d}$$

$$D_A = P_A + T_A^T + \overline{\tau} \tag{8e}$$

$$D_T = P_T + \sum_{i \in \{A,W,T\}} T_T^i. \tag{8f}$$

When there is a marginal social benefit to an incremental movement towards the first-best equilibrium from the user equilibrium (i.e., $\lambda_A > 0$), both the level of transit investment and the modal travel volumes differ from the first-best case. Equations (8a)-(8b) reflect the conditions for second-best auto and transit travel usage, respectively, where the efficient travel volumes are such that the marginal benefit of travel for each mode is equated with its marginal *social* cost, incorporating cross-modal shifts in the demand and cost curves brought about by the inter- and intra-modal externalities of travel, until the modal shares equilibrate to the efficient mix across modes. The bold terms represent the welfare effects in the auto market due to the marginal unit of travel of each mode.

Equation (8c) specifies the second-best level of transit capacity, where the marginal cost of providing additional capacity is equated with the marginal benefit of reduced user costs, incorporating the cross-modal demand and cost curve shifts attributable to the level of transit capacity. The bold terms represent the welfare effects in the auto market due to the marginal unit of transit capacity. Equation (8d) determines the second-best level of transit network size, with the marginal cost of increasing the coverage of the network equated with the marginal benefit of reduced transit user costs, accounting for the welfare effects in the auto market due to the change in network size.

The preceding conditions are in the context of the travel network reaching the user equilibrium of equations (8e)-(8f). We next discuss the factors contributing to the differences between the first- and second-best equilibria, and the implications for evaluating potential transit investments.

4 Results from Theory Model

Our theory model yields several results and insights. First, if the Pigouvian tax τ^* is in place and the first-best equilibrium is achieved, then the marginal benefit of transit investment is confined to the transit sector as there is no deadweight loss associated with auto travel. In this case, efficient public transit investment can be determined in a first-best setting, with the relevant welfare effects of public transit investment confined to the direct effects in the transit market.

Second, if, on the other hand, there are uncorrected distortions in the auto market ($\tau < \tau^*$), and if there are demand and/or cost interdependencies across modes ($\frac{\partial D_A}{\partial K_T} < 0$ or $\frac{\partial MPC_A}{\partial K_T} \neq 0$), then the welfare implications of public transit investment will extend to the auto market as well.

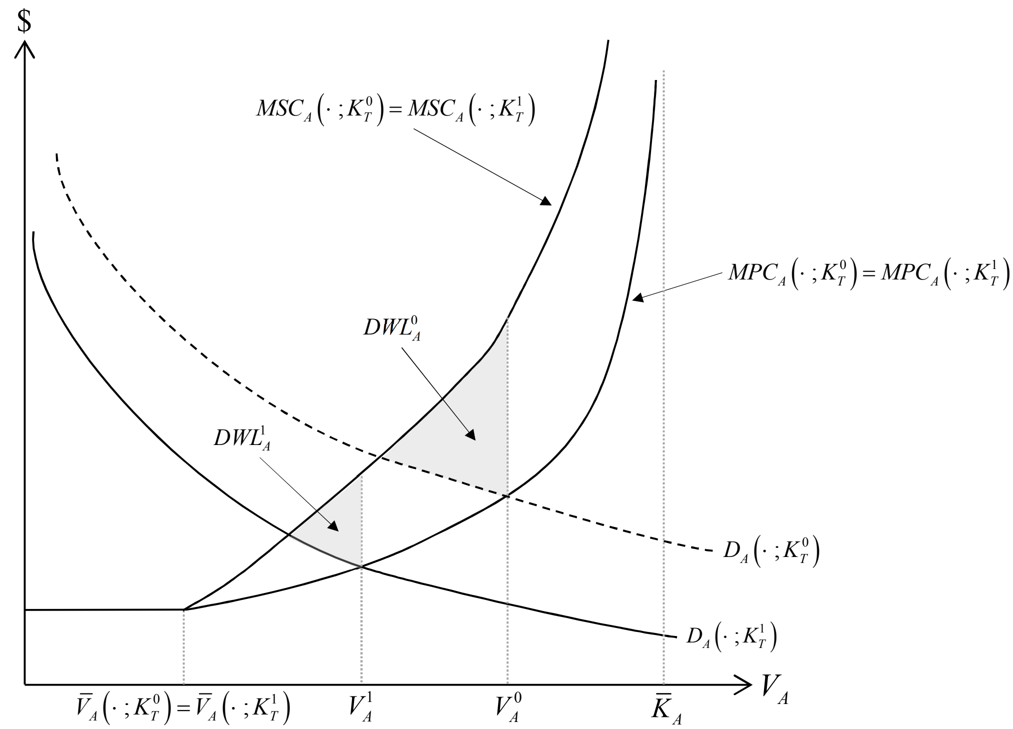
Because our model allows for demand and cost interdependencies between the auto and transit modes, transit investment can affect the user equilibrium in the auto market through two channels: (1) cross-modal substitution via shifts in the auto demand curve, and (2) shifts in the auto travel cost function, including changes to the marginal private costs MPC_A and marginal external costs MEC_A of auto travel. In this case, the change in demand and/or cost in the auto market attributable to investment in transit infrastructure may be nonzero, and these changes in demand and/or cost in the auto market should be accounted for when evaluating the net benefit of potential transit investments.[15]

If there are uncorrected distortions in the auto market ($\tau < \tau^*$), then there is a marginal social benefit to an incremental movement towards the first-best equilibrium from the user equilibrium ($\lambda_A > 0$) and the second-best investment rule will deviate from the first-best rule. From the first-order conditions in equations (8), there are four different factors affecting the magnitude of the marginal effect of transit investment on welfare in the auto market: (1) the severity of existing congestion levels, represented by λ_A; (2) the cross-elasticity of auto demand with respect to the generalized cost of transit travel, given by $\frac{\partial D_A}{\partial MPC_T}$; (3) the magnitude of the change in transit generalized costs due to transit investment, measured by $\frac{\partial MPC_T}{\partial K_T^S}$ and $\frac{\partial MPC_T}{\partial K_T^C}$; and (4) the strength of *intra*- and *inter*-mode congestion externalities, given by $\frac{\partial MPC_T}{\partial V_T}$, $\frac{\partial MPC_A}{\partial V_T}$ and $\frac{\partial MPC_T}{\partial V_A}$.

Following an investment in transit, if only the demand shift in the auto market is considered and there is no cross-modal cost interdependence, such that $\frac{\mathrm{d}MPC_A}{\mathrm{d}K_T} = 0$, then there is an unambiguous ancillary benefit in the auto market associated with a reduction in the deadweight loss (DWL) of the congestion externality, with $DWL_A^1 < DWL_A^0$; this is consistent with the theoretical results of Kraus (2012). This case is shown in Figure 2, with the magnitude of the deadweight loss reduction being proportional to the responsiveness of auto demand to changes in transit investment.

[15] For further discussion on this issue, see Small and Verhoef (2007).

Figure 2: Change in deadweight loss in auto market (no cost interdependence)



However, cost interdependencies between the auto and transit modes arise if auto travel costs and transit supply are interdependent. For example, auto travel costs may potentially vary with the type and level of transit capacity provided due to the interaction between auto and transit vehicles in the roadways.

Transit investment can be characterized as one of two types: 'fixed guideway' whereby it has its own separate right-of-way and does not directly interact with auto traffic (and the auto travel cost *function* is independent of transit capacity), and 'mixed traffic' whereby it shares the right-of-way with auto traffic (and this interaction of transit vehicles and autos implies that the auto travel cost is functionally dependent on the level of transit capacity). As a result, the functional dependence of the auto and transit cost curves may vary across these two types of transit modes.

Two scenarios where the auto travel cost functions are dependent upon the level of transit investment are shown in Figure 3. When transit investment increases from $\mathbf{K_T^0}$ to $\mathbf{K_T^1}$, transit travel increases from V_T^0 to V_T^1 and auto travel decreases from V_A^0 to V_A^1. The change in deadweight loss in the auto market is affected by the initial user equilibrium auto travel volume, V_A^0, and the associated deadweight loss, DWL_A^0, as well as the combined effects of shifts in the auto demand and travel cost curves. The net effect on the *ex post* deadweight loss in the auto market relative to the *ex ante* deadweight loss is ambiguous.

The example on the right-hand side of Figure 3 shows a case in which transit investment has a beneficial effect in the auto market, with $DWL_A^1 < DWL_A^0$, since $R < (P + Q)$. This case may be characteristic of an investment in fixed guideway transit, such as a new light rail line, where there is the potential for a sizable reduction in auto demand, and there may be a downward shift in the auto congestion cost function (due to the new light rail system substituting for mixed traffic bus service that interacts with autos).

Conversely, the example on the left-hand side of Figure 3 shows a case where transit investment has an adverse effect in the auto market, with $DWL_A^1 > DWL_A^0$, since $(F + G + H + I) > (I + J + K)$. This case may be characteristic of an increase in the supply of mixed traffic transit service; a marginal increase in bus service likely leads to a smaller reduction in auto travel, and may cause an upward shift in the congestion cost function due to the increased interaction between these buses and autos on congested roadways.

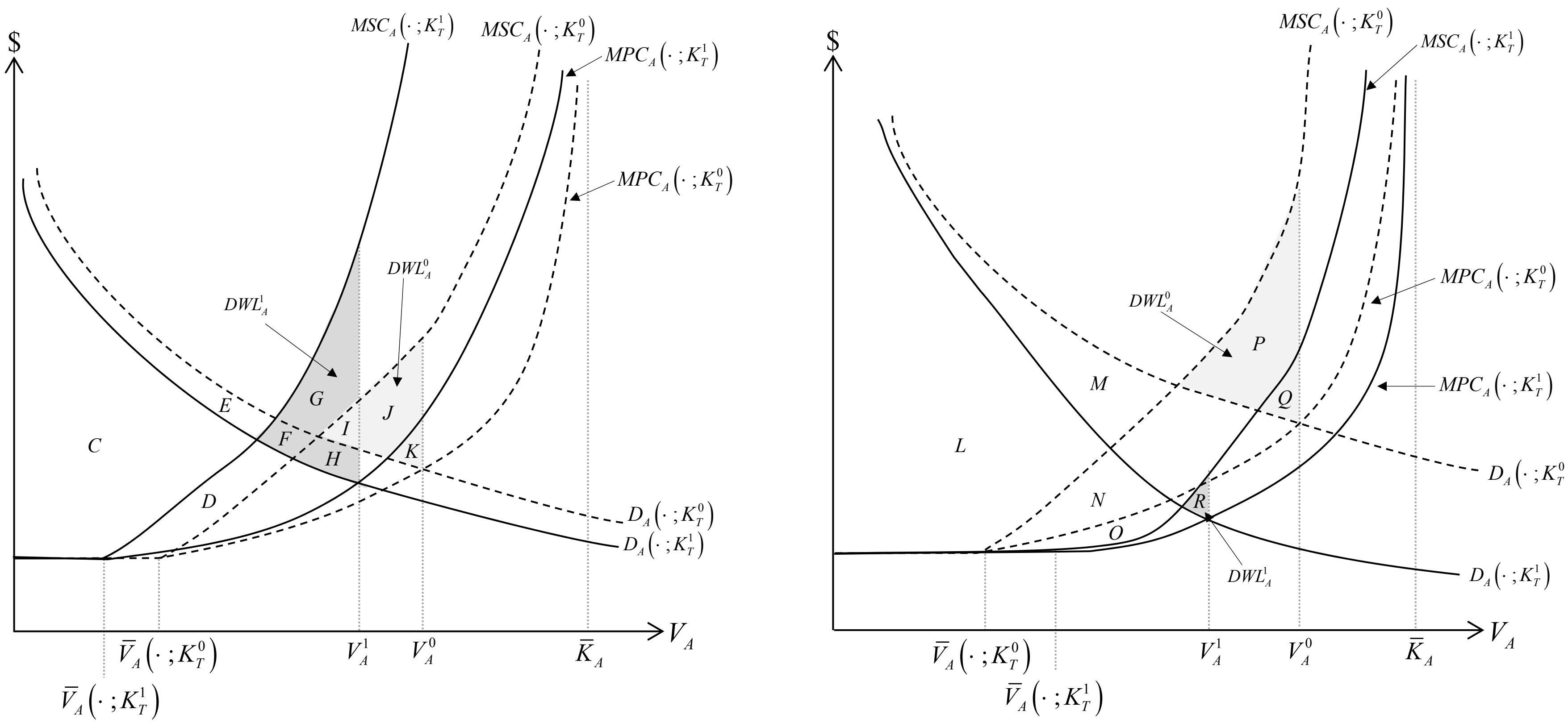
The second-best level of transit capacity is ambiguous in magnitude relative to the first-best level, due to the indeterminate sign of $\frac{\partial MPC_A}{\partial K_T^C}$ (see Appendix B for details); while the demand interdependency component in condition (8c) given by $\lambda_A \cdot \left[\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial K_T^C}\right]$ is beneficial in the auto market (provided $\frac{\partial D_A}{\partial MPC_T} < 0$), the sign and magnitude of the cost interdependency effect given by $\lambda_A \cdot \frac{\partial MPC_A}{\partial K_T^C}$ is ambiguous:

$$K_T^{C\prime} > K_T^{C*} \quad \text{iff} \quad \left[\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial K_T^C} - \frac{\partial MPC_A}{\partial K_T^C}\right] < 0. \tag{9}$$

Comparisons of first-best and second-best investment levels are complicated by the fact that not only do the investment decision rules differ, but these decision rules are evaluated at different equilibria. Based on condition (8d), however, the model implies that the second-best level of transit network size K_T^S exceeds that of the first-best case: $\lambda_A \cdot \left[\frac{\partial D_A}{\partial MPC_T}\frac{\partial MPC_T}{\partial K_T^S}\right] < 0 \Rightarrow K_T^{S\prime} > K_T^{S*}$. Here there is a co-benefit of transit investment in the auto market due to the auto demand reduction brought about by the increase in transit network size K_T^S, and it is assumed that the marginal private cost of auto travel MPC_A is independent of transit network size.

Our results therefore show that second-best transit service should be increased relative to its first-best level, provided that the net benefit of demand substitution from auto to transit outweighs any adverse effects of transit investment on the auto cost function. In particular, the level of transit investment should be higher relative to that chosen when the congestion-reduction effects of transit

Figure 3: Change in deadweight loss in auto market after transit investment: (Left) Increase in DWL (Right) Decrease in DWL



are not accounted for, but the importance of this consideration is dependent upon the interaction of demand and cost interdependencies between the auto and transit modes, which may vary across regions. Owing to the combination of cost interdependencies and differences in cross-modal substitution, fixed guideway transit investments in dense regions may yield higher congestion-reduction benefits than do mixed transit modes. The effect of the demand interdependencies depends on the relative magnitudes of the countervailing effects of substitution and induced demand.

Our model helps reconcile the mixed empirical evidence summarized in Section 2. The studies referenced have varied datasets with differing geographical scope, types of transit modes included, and time periods covered. The net effect of transit on observed congestion is the product of several factors, summarized by the extent to which the demand and cost curves shift in the auto market in response to changes in public transit investment. The parameters in the second-best 'adjustment terms' in equations (8) may be heterogeneous across different regions and different types of transit modes, in part due to demand and cost interdependencies between the auto and transit modes, and may also be affected by the structure and characteristics of the existing transportation networks. Accordingly, the ability of transit investments to reduce the deadweight loss in the auto market may also exhibit heterogeneity.

5 Numerical Analysis

We now numerically analyze the effects of transit investment on the deadweight loss of congestion in the auto market. In particular, we are interested in assessing the relative importance of the four factors discussed in Section 4 that affect the magnitude of the marginal effect of transit investment on welfare in the auto market: the existing level of congestion, the effect of transit investment on the generalized cost of transit travel, the sensitivity of auto demand to changes in the generalized cost of transit travel, and the magnitude of cross-modal cost interactions.

To calibrate our simulation model, we construct a panel dataset spanning 21 years from 1991 to 2011, covering 96 urban areas within 351 counties and 44 states across the U.S. An 'urban area' (UZA) is defined by the U.S. Census Bureau and refers to a region that is centered around a core metropolitan statistical area (MSA). We use auto data from the Texas Transportation Institute's Urban Mobility Report (Schrank, Eisele and Lomax, 2012) and transit data from the Federal Transit Administration's National Transit Database;[16] both are described in detail in Beaudoin and Lin Lawell (2018).[17]

[16] www.ntdprogram.gov/ntdprogram/data.htm.

[17] The 96 UZAs in our dataset are the 96 UZAs with both auto and transit data after merging the Texas Transportation Institute's Urban Mobility Report auto data and the Federal Transit Administration's National Transit Database transit data.

We simulate the first-best and unregulated user equilibria in Figure 1 for a representative urban area (UZA) in the U.S. We calibrate the model using plausible specifications for the functional forms of the auto travel demand and cost functions introduced in the theoretical model developed in Section 3 along with parameter values that yield a user equilibrium auto travel volume equivalent to the median value of the 96 urban areas (UZAs) in the dataset. The initial deadweight loss of the user equilibrium in the auto market is calculated as a reference point to normalize the relative reduction in deadweight loss following increased transit investment. Figure 4 below illustrates the simulation model with the assumed functional forms.

Our Base Case equilibrium provides a reference value for the unregulated and first-best equilibria, and the concomitant deadweight loss attributable to the congestion externality DWL_A. Several scenarios are then examined whereby an increase in transit supply decreases the generalized cost of transit travel and leads to some degree of modal shift from auto to transit travel, generating new user and first-best equilibria and thus a change in the deadweight loss in the auto market. Of primary interest is the percentage reduction in deadweight loss in the auto market due to a given increase in transit supply, $\frac{\%\Delta DWL_A}{\%\Delta K_T}$, which is simulated for a wide range of different parameter value combinations.

We describe our simulation model in detail in Appendix C and Table 1 summarizes the fixed parameter values used in the simulations.

We consider three scenarios representing the various pre-existing congestion levels: 'Low', 'High', and 'Severe'. To do so, we hold auto capacity K_A fixed and vary the initial user equilibrium auto travel volume V_A^0 across the three alternative levels of congestion, so that the resulting volume-to-capacity ratios correspond to the respective ranges of the volume-to-capacity ratio used by the Transportation Research Board (2010) to classify Levels of Service.

Our 'Low' congestion scenario assumption of $\left(\frac{V_A^0}{\overline{K}_A}\right) = 0.57$ is calibrated to Level of Service (LOS) 'B', which is associated with a volume-to-capacity ratio of 0.35-0.58 and "represents reasonably free-flowing conditions but with some influence by others" (Transportation Research Board, 2010). Our 'High' congestion scenario assuming $\left(\frac{V_A^0}{\overline{K}_A}\right) = 0.84$ is calibrated to LOS 'D', which occurs with a volume-to-capacity ratio of 0.75-0.90 and "represents traffic operations approaching unstable flow with high passing demand and passing capacity near zero, characterized by drivers being severely restricted in maneuverability" (Transportation Research Board, 2010). Our 'Severe' congestion scenario assumption of $\left(\frac{V_A^0}{\overline{K}_A}\right) = 1.07$ is calibrated to LOS 'F', which implies a volume-to-capacity ratio greater than 1 and "represents the worst conditions with heavily congested flow and traffic demand exceeding capacity, characterized by stop-and-go waves, poor travel time, low comfort and

Figure 4: Simulation model with assumed functional forms

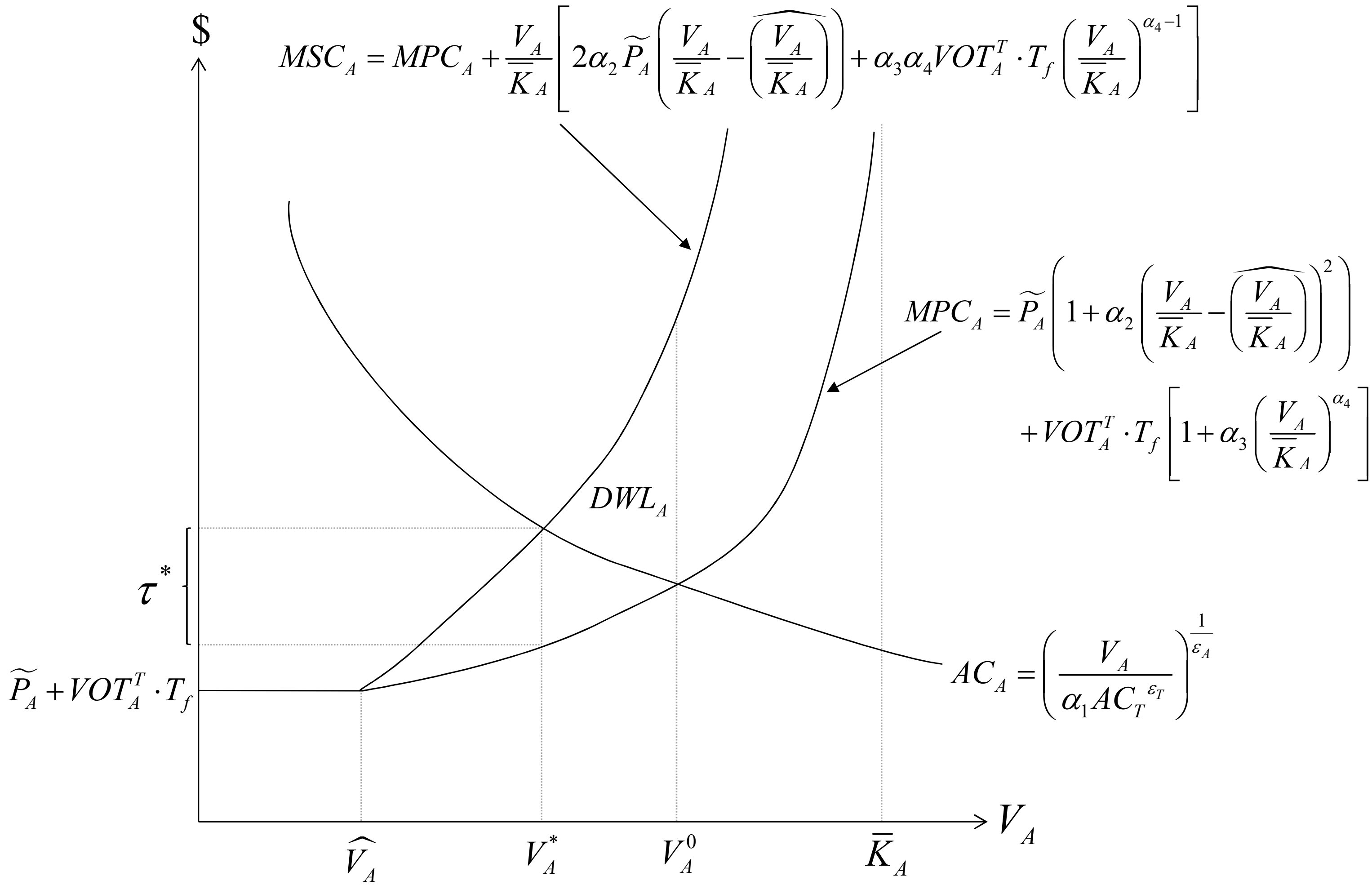


Table 1: Parameter values for simulations

		LOS 'B' **Low Congestion**	**LOS 'D'** **High Congestion**	**LOS 'F'** **Severe Congestion**
Parameter in auto travel demand function	α_1	4,536.1	6,813.7	8,939.7
Parameter in auto travel monetary cost P_A	α_2	1.7	1.7	1.7
Parameter in per-mile auto travel time $\tilde{T}_A^T$	α_3	0.2	0.2	0.2
Parameter in marginal external cost of auto travel MEC_A	α_4	10	10	10
Minimum per-unit fuel cost	$\tilde{P}_A$	0.1234	0.1234	0.1234
Fixed level of auto capacity	$\overline{K}_A$	10,000	10,000	10,000
Threshold volume-to-capacity ratio for congestion externality	$\widehat{\left(\frac{V_A}{\overline{K}_A}\right)}$	0.5	0.5	0.5
Free-flow auto travel speed with no congestion	T_f	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{60}$
Value of auto travel time	VOT_A^T	20.375	20.375	20.375
Elasticity of auto travel wrt generalized cost of auto travel	ϵ_A	-0.3	-0.3	-0.3

convenience, and increased accident exposure" (Transportation Research Board, 2010).[18] To interpret the various levels of congestion, the marginal external cost at the user equilibrium is 4% of the marginal private cost of auto travel for LOS 'B', 50% for LOS 'D', and 254% for LOS 'F'.

We evaluate the effects of two parameters: (1) the effect of transit investment on the generalized cost of transit travel, $\frac{\%\Delta MPC_T}{\%\Delta K_T}$, which translates a 10% increase in transit supply to a given percentage reduction in the (normalized) generalized cost of transit travel MPC_T, and (2) the cross-elasticity of auto demand with respect to the cost of transit travel, ϵ_T, which is a measure of the cross-modal demand substitutability.

Figure 5 shows how the deadweight loss elasticity varies as the effectiveness of transit investment in reducing the generalized cost of transit varies. Holding the cross-elasticity constant at $\epsilon_T = 0.05$, we vary $\frac{\%\Delta MPC_T}{\%\Delta K_T}$ between 0 and -1 (which implies that a 10% increase in transit supply leads to a 0-10% reduction in the average generalized cost of transit travel). The resulting deadweight loss elasticity varies from 0 to -0.2. Transit investments that significantly reduce the generalized cost of transit travel may lead to notable reductions in the deadweight loss in the auto market, even with a low cross-elasticity of demand between modes.

Figure 6 depicts the deadweight loss elasticity as the cross-price elasticity ϵ_T varies between 0.02 and 0.2, holding $\frac{\%\Delta MPC_T}{\%\Delta K_T}$ constant at -0.5. As the cross-price elasticity increases from 0 to 0.2, the deadweight loss elasticity increases from 0 to -0.4. This result highlights the importance of being able to accurately forecast the cross-price elasticity of demand when evaluating a prospective transit investment, given the heterogeneity of this parameter shown by Beaudoin and Lin Lawell (2018).

The range of elasticity values in Figures 5 and 6 mirrors the empirical estimates of Beaudoin and Lin Lawell (2018), with the elasticity of *auto travel* with respect to transit capacity ranging from -0.01 to -0.1. Due to the convexity of the congestion externality, the deadweight loss elasticity is 3.7-4.7 times the magnitude of the auto travel elasticity. The characteristics along which the empirical elasticity estimates may vary – regional population size and density, and aspects of the public transit network – are manifested in inter-regional heterogeneity of $\frac{\%\Delta MPC_T}{\%\Delta K_T}$ and ϵ_T, leading to variation across regions in the effectiveness of transit in reducing auto congestion.

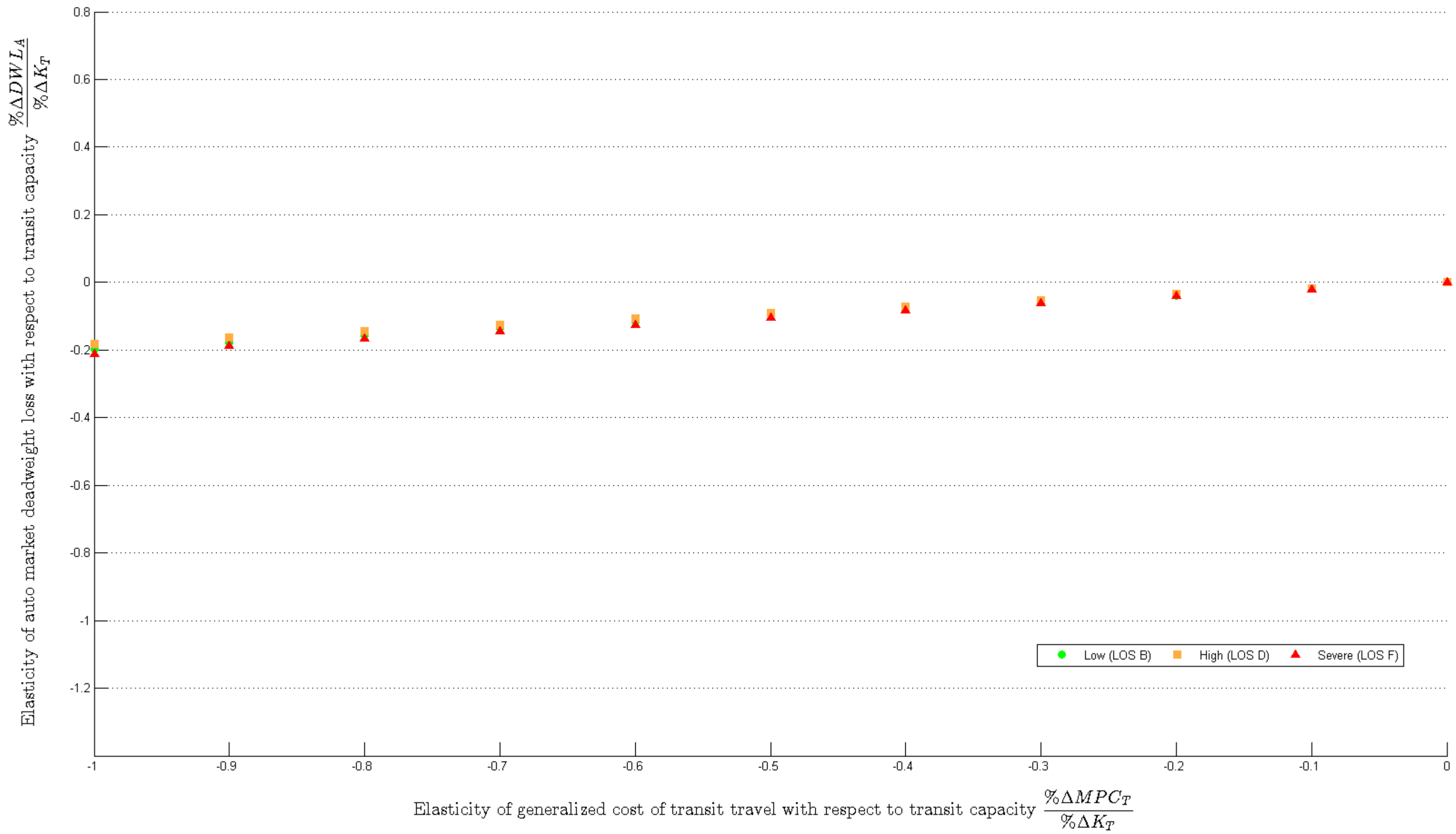
We are also interested in the effect of modal cost interdependence and the relationship between demand interdependency, which is shown in Figures 7 and 8. Figure 7 illustrates how the deadweight loss elasticity varies in relation to the degree of cost interdependence between modes when there is a low cross-price elasticity ($\epsilon_T = 0.05$). $\frac{\%\Delta MPC_A}{\%\Delta K_T}$ is varied from -0.4 to 0.4, which means

[18] It is possible for $\left(\frac{V_A^0}{\overline{K}_A}\right) > 1$, as $\overline{K}_A$ represents an engineering design capacity in relation to the level of service, and not a binding physical capacity constraint.

that a 10% increase in transit supply could lead to a *reduction* in the marginal cost of auto travel of 4% or an *increase* in the marginal cost of auto travel of 4%, and any effect in between. Due to the convexity of the congestion externality, we assume that the marginal social cost of auto travel increases/decreases at 2 times the rate of change of the marginal private cost. Figure 8 repeats Figure 7, but uses a higher cross-price elasticity ($\epsilon_T = 0.15$).

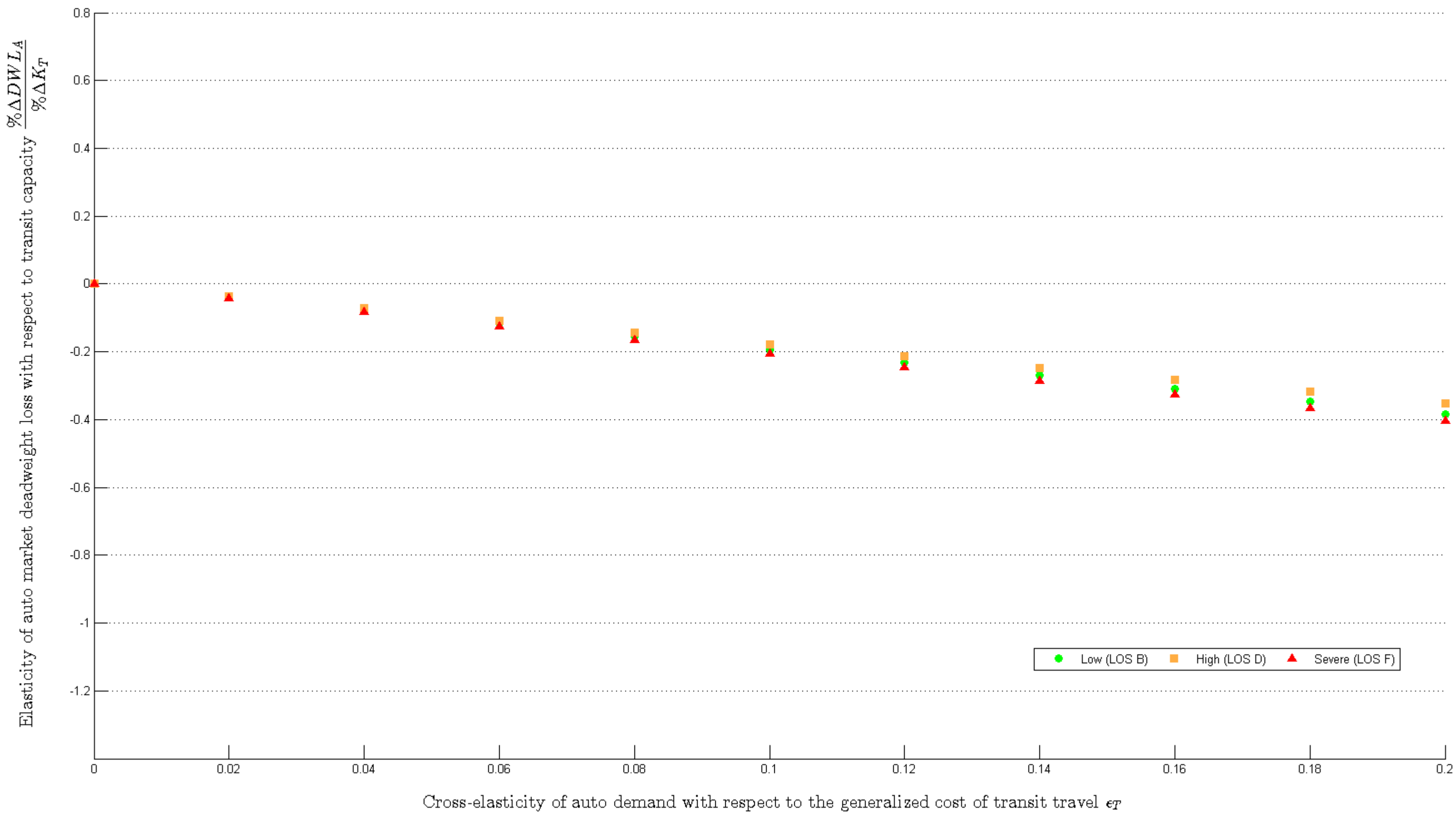
Consistent with Figure 3, the results show that transit investments that reduce the cost of auto travel (such as investments in fixed guideway transit that remove buses from the roadways) amplify the modal substitution effect and lead to significant deadweight loss reductions. Conversely, in scenarios where transit investments increase the cost of auto travel and have a low cross-elasticity of demand (such as increased supply of mixed traffic bus service), it is possible that the overall deadweight loss in the auto market actually increases.

Figure 5: Relation between effectiveness of transit investment in reducing the generalized cost of transit, and the effectiveness of transit investment in reducing auto market deadweight loss



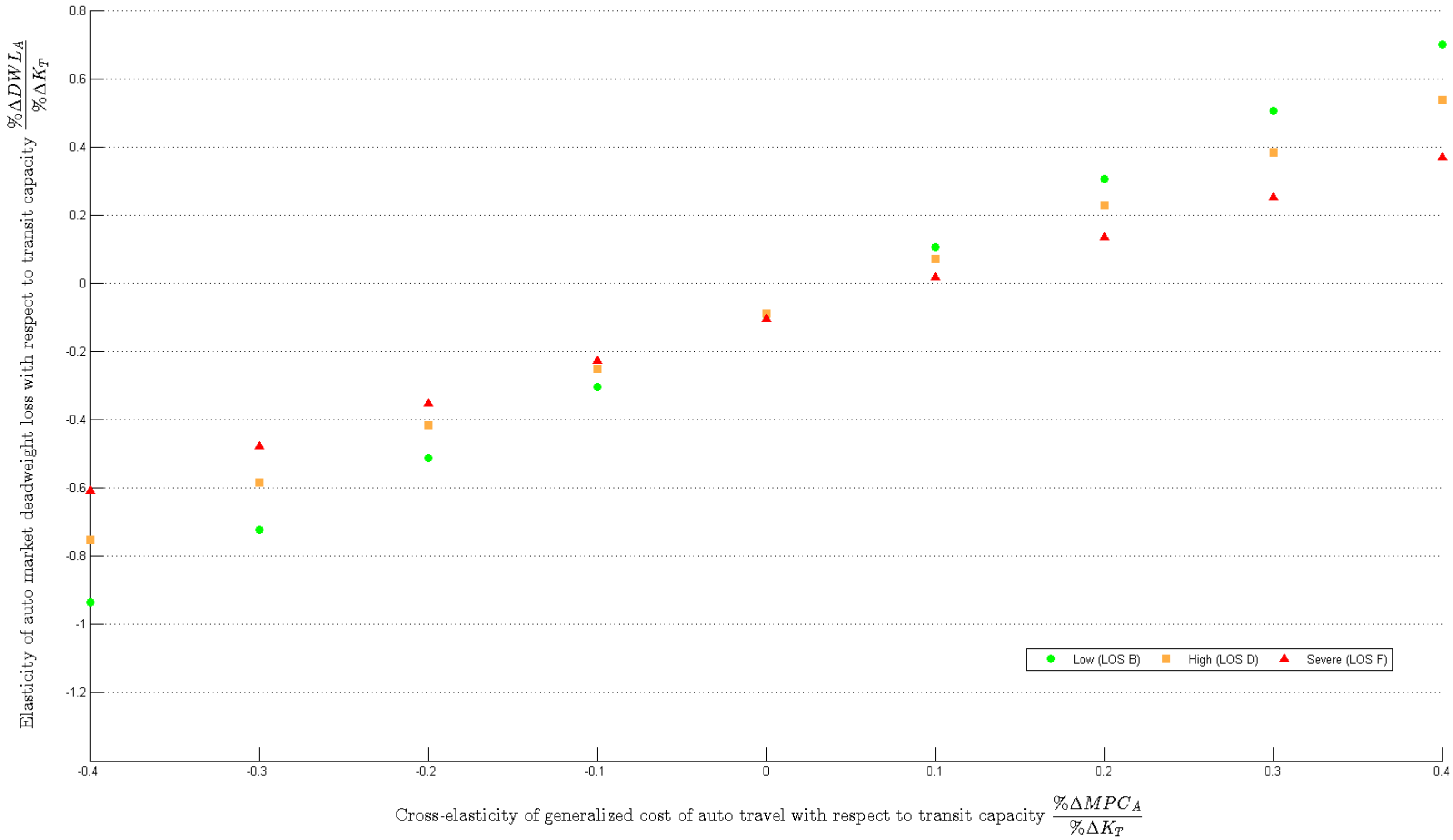
Note: Cross-elasticity of auto demand with respect to the cost of transit travel is held constant at $\epsilon_T = 0.05$.

Figure 6: Effect of cross-price elasticity on elasticity of auto market deadweight loss with respect to transit capacity



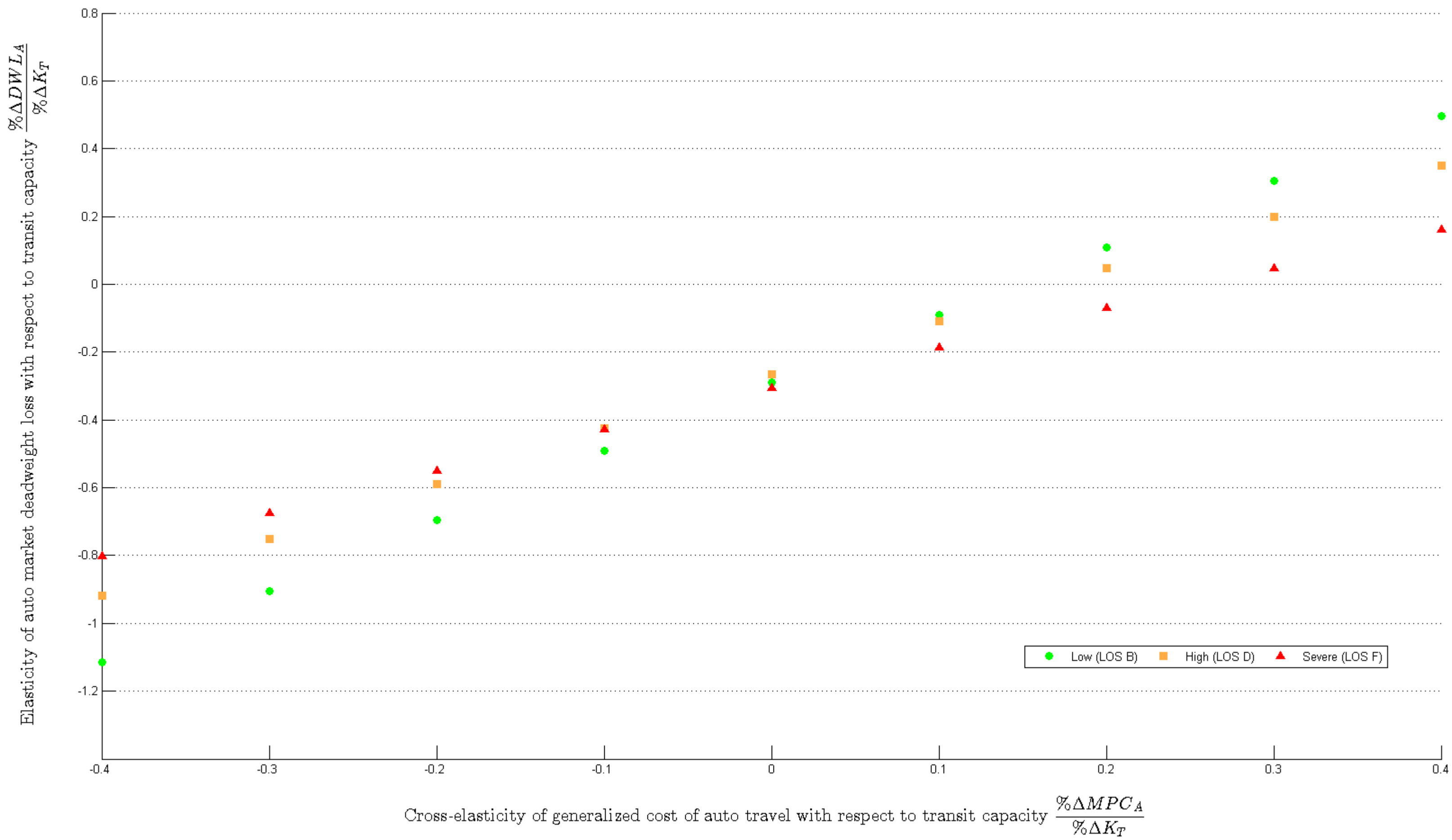
Note: Elasticity of generalized cost of transit travel with respect to transit capacity is held constant at $\frac{\%\Delta MPC_T}{\%\Delta K_T} = -0.5$.

Figure 7: Effect of cost interdependence on elasticity of auto market deadweight loss with respect to transit capacity: low cross-price elasticity



Note: Cross-elasticity of auto demand with respect to the cost of transit travel is held constant at $\epsilon_T = 0.05$.

Figure 8: Effect of cost interdependence on elasticity of auto market deadweight loss with respect to transit capacity: high cross-price elasticity



Note: Cross-elasticity of auto demand with respect to the cost of transit travel is held constant at $\epsilon_T = 0.15$.

6 Conclusion

Traffic congestion has increased significantly in the U.S. over the past several decades. The issue of congestion is attracting heightened awareness and a greater sense of urgency for policymakers as we strive for an economically and environmentally sustainable transportation sector. This paper examines the role that public transit investment can play in reducing traffic congestion.

Recent expenditures on public transit capital in the U.S. have exceeded $18 billion per year (American Public Transportation Association, 2012). It is imperative to assess whether these investment levels are efficient and being allocated appropriately, what the effects of these expenditures are on transportation activity and the environment, and what path future investment should take. More effective management of the country's transportation infrastructure can lead to a reduction in traffic congestion, a change that would directly improve the economic competitiveness of many of the country's commercial sectors and the livability of its communities. Given the aggregate costs of congestion, even modest improvements entail significant social value.

We show that if a Pigouvian tax is not levied on auto travel, there is justification for incorporating congestion-reduction benefits in the auto market brought about by transit investment when evaluating proposed changes in transit services. A general equilibrium framework that incorporates ancillary benefits in the auto market is warranted when evaluating the efficiency of public transit supply; public transit projects should not be evaluated exclusively in terms of the forecasted net welfare generated by public transit users, but instead should also include interactions between auto and transit users in the cost-benefit analysis framework.

In particular, our results show that second-best transit service should be increased relative to its first-best level, provided that the net benefit of demand substitution from auto to transit outweighs any adverse effects of transit investment on the auto cost function. In particular, the level of transit investment should be higher relative to that chosen when the congestion-reduction effects of transit are not accounted for, but the importance of this consideration is dependent upon the interaction of demand and cost interdependencies between the auto and transit modes, which may vary across regions. Owing to the combination of cost interdependencies and differences in cross-modal substitution, fixed guideway transit investments in dense regions may yield higher congestion-reduction benefits than do mixed transit modes. The effect of the demand interdependencies depends on the relative magnitudes of the countervailing effects of substitution and induced demand.

Our model helps reconcile the mixed empirical evidence summarized in Section 2. The studies referenced have varied datasets with differing geographical scope, types of transit modes included, and time periods covered. The net effect of transit on observed congestion is the product of several factors, summarized by the extent to which the demand and cost curves shift in the auto market in response to changes in public transit investment. The parameters in the second-best 'adjustment terms' in equations (8) may be heterogeneous across different regions and different types of transit modes, in part due to demand and cost interdependencies between the auto and transit modes, and may also be affected by the structure and characteristics of the existing transportation networks. Accordingly, the ability of transit investments to reduce the deadweight loss in the auto market may also exhibit heterogeneity.

Overall, our analysis indicates that the congestion-reduction effects of public transit supply warrant a higher level of public transit investment than would be provided on the basis of the isolated valuation of public transit ridership; this effect would be larger still if the additional negative externalities of auto travel were incorporated into this framework. The magnitude of this benefit is subject to considerable variability, and is dependent upon the characteristics of the existing transportation network, the technology of the proposed transit system, and the socioeconomic and geographic attributes of the region. The implication is that transit cost-benefit analyses must be carried out on a case-by-case basis and there may be limited scope for the external validity of regional studies, as past experiences in one city may not generalize to potential new transit investments in another.

While our results suggest that fixed guideway transit investments in dense regions yield higher congestion-reduction benefits than do mixed transit modes, this should not be construed as advocating for fixed guideway modes over mixed transit modes *per se*. In the analysis, we have only considered the benefits in the auto market due to transit investment, and have not considered the costs of the various transit modes. Both construction and operating costs of transit vary widely by region and type of transit.[19] Further, proponents of public transit may argue that investment in public transit today is necessary to develop transit ridership in the future and to influence land-use patterns in order to sow the roots for a more efficient public transit system in the future.

According to our theory model, if the Pigouvian tax τ^* is in place and the first-best equilibrium is achieved, then the marginal benefit of transit investment is confined to the transit sector as there is no deadweight loss associated with auto travel. While this is true in a static framework, in a dynamic model of the transportation network the effects of transit investment on the auto market should be incorporated even in the first-best, insofar as transit investment in a given time period can be expected to influence the demand – and resulting equilibria – in subsequent periods.

Since congestion is most prominent during peak periods, we abstract from the dynamics of bottleneck behavior and use a static model in this paper. Similarly, by assuming that the substitution effect outweighs the induced demand effect, we are implicitly viewing transit investment as a policy

[19] For recent estimates of the construction costs of different transit modes, see Table 3.5 in Small and Verhoef (2007).

instrument to address congestion in the short run to medium run. Following Beaudoin and Lin Lawell (2018), transit tends to reduce auto travel in the short run to the medium run; in the long run, however, it may be that transit investment induces additional auto travel that would actually lead the induced demand effect to outweigh the substitution effect, causing the auto demand curve to shift upward following investment in transit.

We hope in future work to develop a dynamic model of the *optimal* transit investment path in the presence of uncorrected auto market distortions that accounts for the transition from the time of investment until the equilibrium is reached, and that factors in the rate of induced demand and the timing of the resulting costs and benefits. Such a model can be utilized by policymakers in conducting cost-benefit analyses of potential transit investments, and can serve as a guide in formulating and evaluating long-run regional transportation plans.

This paper contributes to the literature by developing a model that allows for demand and cost interdependencies between the auto and transit modes and by accounting for heterogeneity in these interdependencies. Our results suggest that urban mass transit may have a co-benefit of congestion reduction. As a consequence, prospective public transit projects should not be evaluated exclusively in terms of the forecasted net welfare generated by public transit users, but instead should also include interactions between auto and transit users in the cost-benefit analysis framework. Whether this consideration is important will vary significantly across cities, and past experiences in one city may not generalize to potential new investments in another. While public transit investment may be able to play a complementary role, efficient pricing of auto travel remains necessary to address traffic congestion in the U.S.

References

Ahn, Kijung (2009). "Road Pricing and Bus Service Policies," *Journal of Transport Economics and Policy*, 43(1): 25-53.

American Public Transportation Association (2012). *2012 Public Transportation Fact Book*. Washington, DC.

Anas, Alex and Robin Lindsey (2011). "Reducing Urban Road Transportation Externalities: Road Pricing in Theory and in Practice," *Review of Environmental Economics and Policy*, 5(1): 66-88.

Anderson, Michael (2014). "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion," *American Economic Review*, 104(9): 2763-2796.

Arnott, Richard and An Yan (2000). "The two-mode problem: Second-best pricing and capacity," *Review of Urban and Regional Development Studies*, 12: 170-199.

Barth, Matthew and Kanok Boriboonsomsin (2009). "Traffic Congestion and Greenhouse Gases," *Access*, publication of the University of California Transportation Center, No. 35: Fall 2009.

Basso, Leonardo and Hugo Silva (2014). "Efficiency and Substitutability of Transit Subsidies and Other Urban Transport Policies," *American Economic Journal: Economic Policy*, 6(4): 1-33.

Baum-Snow, Nathaniel and Matthew Kahn (2005). "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities," *Brookings-Wharton Papers on Urban Affairs*.

Baumol, William and David Bradford (1970). "Optimal Departures from Marginal Cost Pricing," *American Economic Review*, 60(3): 265-283.

Baumol, William and Wallace Oates (1988). *The Theory of Environmental Policy, Second Edition*. Cambridge, UK: Cambridge University Press.

Beaudoin, Justin, Y. Hossein Farzin and C.-Y. Cynthia Lin Lawell (2015). "Public Transit Investment and Sustainable Transportation: A Review of Studies of Transit's Impact on Traffic Congestion and Air Quality," *Research in Transportation Economics*, 52: 15-22.

Beaudoin, Justin and C.-Y. Cynthia Lin Lawell (2017). "The effects of urban public transit investment on traffic congestion and air quality," In Hamid Yaghoubi (Ed.), *Urban Transport Systems* (Chapter 6). InTech.

Beaudoin, Justin and C.-Y. Cynthia Lin Lawell (2018). "The Effects of Public Transit Supply on the Demand for Automobile Travel," *Journal of Environmental Economics and Management*, 88: 447-467.

Bento, Antonio, Kevin Roth and Andrew R. Waxman (2017). "Avoiding Traffic Congestion Externalities? The Value of Urgency," Working paper.

Berechman, Joseph (2009). *The Evaluation of Transportation Investment Projects.* New York: Routledge.

Davis, Otto and Andrew Whinston (1965). "Welfare economics and the theory of the second best," *Review of Economic Studies*, 32(1): 1-14.

Davis, Otto and Andrew Whinston (1967). "Piecemeal policy in the theory of second-best," *Review of Economic Studies*, 34(3): 323-331.

d'Ouville, Edmond and John McDonald (1990). "Optimal Road Capacity with a Suboptimal Congestion Toll," *Journal of Urban Economics*, 28: 34-49.

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

Farzin, Y.H. (2009). "The Effect of Non-pecuniary Motivations on Labor Supply," *The Quarterly Review of Economics and Finance*, 49(4): 1236-1259.

Friedlaender, Ann (1981). "Price distortions and second best investment rules in the transportation industries," *American Economic Review, Papers and Proceedings*, 71: 389-393.

Gillen, David (1997). "Efficient use and provision of transportation infrastructure with imperfect pricing: second best rules," in *The Full Costs and Benefits of Transportation: Contributions to Theory, Method and Measurement*, by David Greene, Donald Jones and Mark Delucchi, eds. Berlin: Springer.

Glaeser, Edward L. and Giacomo A.M. Ponzetto (2017). "The Political Economy of Transportation Investment," NBER Working Paper No. 23686.

Glaister, Stephen and David Lewis (1978). "An Integrated Fares Policy for Transport in London," *Journal of Public Economics*, 9: 341-355.

Greenwood, Ian, Roger Dunn and Robert Raine (2007). "Estimating the Effects of Traffic Congestion on Fuel Consumption and Vehicle Emissions Based on Acceleration Noise," *Journal of Transportation Engineering*, 133(2): 96-104.

Hamilton, Timothy and Casey Wichman (2018). "Bicycle Infrastructure and Traffic Congestion: Evidence from DC's Capital Bikeshare," *Journal of Environmental Economics and Management*, 87: 72-93.

Henderson, J. Vernon (1985). *Economic Theory and the Cities, Second Edition.* London, UK: Academic Press.

Kraus, Marvin (2012). "Road pricing with optimal mass transit," *Journal of Urban Economics*, 72: 81-86.

Libermann, Keith (2009). *Surface Transportation: Infrastructure, Environmental Issues and Safety.* Hauppauge, N.Y: Nova Science.

Lin, C.-Y. Cynthia and Lea Prince (2009). "The optimal gas tax for California," *Energy Policy*, 37(12): 5173-5183.

Lindsey, Robin (2012). "Road pricing and investment," *Economics of Transportation*, 1: 49-63.

Litman, Todd (2013). "Understanding Transport Demands and Elasticities," Victoria Transport Policy Institute Report, www.vtpi.org/elasticities.pdf.

Mohring, Herbert and Mitchell Harwitz (1962). *Highway Benefits: An Analytical Framework*. Evanston, Illinois: Northwestern University Press.

Morrison, Geoffrey and C.-Y. Cynthia Lin Lawell (2016). "Does Employment Growth Increase Travel Time to Work?: An Empirical Analysis Using Military Troop Movements," *Regional Science and Urban Economics*, 60: 180-197.

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(2): 231-51.

Oum, Tae Hoon, William Waters and Xiaowen Fu (2008). "Transport demand elasticities," in *Handbook of Transport Modelling, 2nd ed.*, by David Hensher and Kenneth Button, eds. Oxford: Elsevier.

Parry, Ian, Margaret Walls and Winston Harrington (2007). "Automobile Externalities and Policies," *Journal of Economic Literature*, 45(2): 373-399.

Parry, Ian (2009). "Pricing Urban Congestion," *Annual Review of Resource Economics*, 1: 461-484.

Parry, Ian and Kenneth Small (2005). "Does Britain or the United States have the right gasoline tax?," *American Economic Review*, 95(4): 1276-1289.

Parry, Ian and Kenneth Small (2009). "Should Urban Transit Subsidies be Reduced?" *American Economic Review*, 99(3): 700-724.

Pels, Eric and Erik Verhoef (2007). "Infrastructure Pricing and Competition Between Modes in Urban Transport," *Environment and Planning A*, 39: 2119-2138.

Schrank, David, Bill Eisele and Tim Lomax (2012). *Texas A&M Transportation Institute's 2012 Urban Mobility Report*. College Station, TX: Texas Transportation Institute.

Sherman, Roger (1971). "Congestion interdependence and urban transit fares," *Econometrica*, 39(3): 565-576.

Skabardonis, Alexander and Richard Dowling (1996). "Improved speed-flow relationships for planning application," *Transportation Research Record*, 1572: 18-23.

Small, Kenneth and Erik Verhoef (2007). *The Economics of Urban Transportation.* New York: Routledge.

Transportation Research Board (2010). *Highway Capacity Manual 2010.* Washington, D.C.

Vickrey, William (1969). "Congestion Theory and Transport Investment," *American Economic Review*, 59: 251-261.

Viton, Philip (1981). "On Competition and Product Differentiation in Urban Transportation: The San Francisco Bay Area," *The Bell Journal of Economics*, 12(2): 362-379.

Wardrop, John (1952). "Some theoretical aspects of road traffic research," *Proceedings of the Institute of Civil Engineers*, Part II, Vol. 1: 325-378.

Wheaton, William (1978). "Price-Induced Distortions in Urban Highway Investment," *Bell Journal of Economics*, 9(2): 622-632.

Winston, Clifford and Chad Shirley (1998). *Alternate Route: Toward Efficient Urban Transportation.* Washington, D.C.: Brookings Institution.

Winston, Clifford (2000). "Government Failure in Urban Transportation," *Fiscal Studies*, 21(4): 403-425.

Winston, Clifford and Ashley Langer (2006). "The effect of government highway spending on road users' congestion costs," *Journal of Urban Economics*, 60(3): 463-483.

Winston, Clifford and Vikram Maheshri (2007). "On the Social Desirability of Urban Rail Transit Systems," *Journal of Urban Economics*, 62(2): 362-382.

Appendix A - Transit Investment and the Cost of Transit Travel

Here we show how investing in public transit by either increasing the network size or increasing capacity is expected to lower the marginal private cost of transit travel. From (2), the generalized marginal private cost of transit travel is given by:

$$MPC_T\left(V_A, V_T, K_T^S, K_T^C; \overline{K}_A\right) \equiv P_T + T_T^A\left(K_T^S\right) + T_T^W\left(\frac{K_T^C}{K_T^S}, \frac{V_A}{\overline{K}_A}, \frac{V_T}{K_T^C}\right) + T_T^T\left(\frac{V_A}{\overline{K}_A}, \frac{V_T}{K_T^C}, \frac{K_T^C}{\overline{K}_A}, \frac{K_T^S}{\overline{K}_A}\right).$$

We first consider expansion of the transit network. Assuming that the marginal benefits of reduced access and travel times following an increase in the network size outweigh the marginal disbenefit of the indirect increase in wait time due to the transit congestion effect, i.e. $\left|\frac{\partial T_T^A}{\partial K_T^S}\right| + \left|\frac{\partial T_T^T}{\partial K_T^S}\right| > \left|\frac{\partial T_T^W}{\partial K_T^S}\right|$, then $\frac{\partial MPC_T}{\partial K_T^S} < 0$ and the transit cost function shifts downward, since:

$$\frac{\partial MPC_T}{\partial K_T^S} = \frac{\partial T_T^A}{\partial K_T^S} + \frac{\partial T_T^W}{\partial K_T^S} + \frac{\partial T_T^T}{\partial K_T^S} = \frac{\partial T_T^A}{\partial K_T^S} + \frac{\partial T_T^W}{\partial \frac{K_T^C}{K_T^S}} \frac{\partial \frac{K_T^C}{K_T^S}}{\partial K_T^S} + \frac{\partial T_T^T}{\partial \frac{K_T^S}{\overline{K}_A}} \frac{\partial \frac{K_T^S}{\overline{K}_A}}{\partial K_T^S}$$

$$= \underbrace{\frac{\partial T_T^A}{\partial K_T^S}}_{(-)} + \underbrace{\left[\overbrace{\frac{\partial T_T^W}{\partial \frac{K_T^C}{K_T^S}}}^{(-)}\right]\left(\overbrace{\frac{-K_T^C}{\left(K_T^S\right)^2}}^{(-)}\right)}_{(+)} + \underbrace{\left[\overbrace{\frac{\partial T_T^T}{\partial \frac{K_T^S}{\overline{K}_A}}}^{(-)}\right]\left(\overbrace{\frac{1}{\overline{K}_A}}^{(+)}\right)}_{(-)}.$$

We next consider an increase in transit capacity. Assuming that the marginal benefit of reduced waiting time outweighs the marginal disbenefit of increased travel time due to the transit congestion effect associated with increasing transit capacity, i.e. $\left|\frac{\partial T_T^W}{\partial K_T^C}\right| > \left|\frac{\partial T_T^T}{\partial K_T^C}\right|$, then $\frac{\partial MPC_T}{\partial K_T^C} < 0$ and the transit cost function shifts downward, since:

$$\frac{\partial MPC_T}{\partial K_T^C} = \frac{\partial T_T^W}{\partial K_T^C} + \frac{\partial T_T^T}{\partial K_T^C} = \frac{\partial T_T^W}{\partial \frac{K_T^C}{K_T^S}} \frac{\partial \frac{K_T^C}{K_T^S}}{\partial K_T^C} + \frac{\partial T_T^W}{\partial \frac{V_T}{K_T^C}} \frac{\partial \frac{V_T}{K_T^C}}{\partial K_T^C} + \frac{\partial T_T^T}{\partial \frac{V_T}{K_T^C}} \frac{\partial \frac{V_T}{K_T^C}}{\partial K_T^C} + \frac{\partial T_T^T}{\partial \frac{K_T^C}{\overline{K}_A}} \frac{\partial \frac{K_T^C}{\overline{K}_A}}{\partial K_T^C}$$

$$= \underbrace{\left[\overbrace{\frac{\partial T_T^W}{\partial \frac{V_T}{K_T^C}}}^{(+)} + \overbrace{\frac{\partial T_T^T}{\partial \frac{V_T}{K_T^C}}}^{(+)}\right]\left(\overbrace{\frac{-V_T}{\left(K_T^C\right)^2}}^{(-)}\right)}_{(-)} + \underbrace{\left[\overbrace{\frac{\partial T_T^W}{\partial \frac{K_T^C}{K_T^S}}}^{(-)}\right]\left(\overbrace{\frac{1}{K_T^S}}^{(+)}\right)}_{(-)} + \underbrace{\left[\overbrace{\frac{\partial T_T^T}{\partial \frac{K_T^C}{\overline{K}_A}}}^{(+)}\right]\left(\overbrace{\frac{1}{\overline{K}_A}}^{(+)}\right)}_{(+)}.$$

Appendix B - Modal Cost Interdependency

Here we show that the cost interdependence of our theoretical model yields an uncertain effect of changes in transit supply on the cost of auto travel. From (1), the generalized marginal private cost of auto travel is given by:

$$MPC_A\left(V_A, V_T, \tau; K_T^C, \overline{K}_A\right) \equiv P_A\left(\frac{V_A}{\overline{K}_A}, \frac{K_T^C}{\overline{K}_A}, \frac{V_T}{K_T^C}\right) + T_A^T\left(\frac{V_A}{\overline{K}_A}, \frac{K_T^C}{\overline{K}_A}, \frac{V_T}{K_T^C}\right) + \tau.$$

Due to the various interaction effects across modes, the capacity of public transit service may influence the marginal private cost of auto travel as follows:

$$\begin{aligned}\frac{\partial MPC_A}{\partial K_T^C} &= \frac{\partial P_A}{\partial K_T^C} + \frac{\partial T_A^T}{\partial K_T^C} = \frac{\partial P_A}{\partial \frac{V_T}{K_T^C}}\frac{\partial \frac{V_T}{K_T^C}}{\partial K_T^C} + \frac{\partial P_A}{\partial \frac{K_T^C}{\overline{K}_A}}\frac{\partial \frac{K_T^C}{\overline{K}_A}}{\partial K_T^C} + \frac{\partial T_A^T}{\partial \frac{V_T}{K_T^C}}\frac{\partial \frac{V_T}{K_T^C}}{\partial K_T^C} + \frac{\partial T_A^T}{\partial \frac{K_T^C}{\overline{K}_A}}\frac{\partial \frac{K_T^C}{\overline{K}_A}}{\partial K_T^C} \\ &= \underbrace{\left[\overbrace{\frac{\partial P_A}{\partial \frac{V_T}{K_T^C}}}^{(+)} + \overbrace{\frac{\partial T_A^T}{\partial \frac{V_T}{K_T^C}}}^{(+)}\right]\left(\overbrace{\frac{-V_T}{\left(K_T^C\right)^2}}^{(-)}\right)}_{(-)} + \underbrace{\left[\overbrace{\frac{\partial P_A}{\partial \frac{K_T^C}{\overline{K}_A}}}^{(+)} + \overbrace{\frac{\partial T_A^T}{\partial \frac{K_T^C}{\overline{K}_A}}}^{(+)}\right]\left(\overbrace{\frac{1}{\overline{K}_A}}^{(+)}\right)}_{(+)}.\end{aligned}$$

The sign of $\frac{\partial MPC_A}{\partial K_T^C}$ is ambiguous, dependent on the relative magnitudes of (1) the benefit of reducing the cost of transit travel and inducing users to switch from auto to transit, thereby decreasing the transit congestion effect related to $\frac{V_T}{K_T^C}$, and (2) the disbenefit of transit capacity on auto congestion through the increased roadway interaction related to $\frac{K_T^C}{\overline{K}_A}$:

$$\frac{\partial MPC_A}{\partial K_T^C} = \begin{cases} > 0 \text{ if } \left|\left[\frac{\partial P_A}{\partial \frac{V_T}{K_T^C}} + \frac{\partial T_A^T}{\partial \frac{V_T}{K_T^C}}\right]\left(\frac{-V_T}{\left(K_T^C\right)^2}\right)\right| < \left|\left[\frac{\partial P_A}{\partial \frac{K_T^C}{\overline{K}_A}} + \frac{\partial T_A^T}{\partial \frac{K_T^C}{\overline{K}_A}}\right]\left(\frac{1}{\overline{K}_A}\right)\right| \\ = 0 \text{ if } \left|\left[\frac{\partial P_A}{\partial \frac{V_T}{K_T^C}} + \frac{\partial T_A^T}{\partial \frac{V_T}{K_T^C}}\right]\left(\frac{-V_T}{\left(K_T^C\right)^2}\right)\right| = \left|\left[\frac{\partial P_A}{\partial \frac{K_T^C}{\overline{K}_A}} + \frac{\partial T_A^T}{\partial \frac{K_T^C}{\overline{K}_A}}\right]\left(\frac{1}{\overline{K}_A}\right)\right| \text{ or } \frac{\partial P_A}{\partial K_T^C} = \frac{\partial T_A^T}{\partial K_T^C} = 0 \\ < 0 \text{ if } \left|\left[\frac{\partial P_A}{\partial \frac{V_T}{K_T^C}} + \frac{\partial T_A^T}{\partial \frac{V_T}{K_T^C}}\right]\left(\frac{-V_T}{\left(K_T^C\right)^2}\right)\right| > \left|\left[\frac{\partial P_A}{\partial \frac{K_T^C}{\overline{K}_A}} + \frac{\partial T_A^T}{\partial \frac{K_T^C}{\overline{K}_A}}\right]\left(\frac{1}{\overline{K}_A}\right)\right|. \end{cases}$$

As a result, the auto cost function may shift upwards or downwards (or be unaffected) following an increase in transit supply, depending on the technological characteristics of the existing and introduced transit service.

Appendix C - Simulation Model

Demand Curve for Auto Travel

The inverse demand curve for auto travel $D_A(\cdot)$ is assumed to exhibit constant elasticity with respect to the generalized costs of auto and transit travel, given by $\epsilon_A < 0$ and $\epsilon_T > 0$, respectively. Defining the quantity of auto travel as $V_A = \alpha_1 MPC_A^{\epsilon_A} MPC_T^{\epsilon_T}$ yields the desired result that $\frac{\partial V_A}{\partial MPC_A} < 0$ and $\frac{\partial V_A}{\partial MPC_T} > 0$, where the cross-elasticity of auto demand with respect to the generalized cost of transit travel incorporates shifts in the auto demand curve due to the change in the marginal private cost of transit travel following transit investment.[20] Modal prices are normalized by setting the baseline value of the per-unit cost of transit travel as $MPC_T = 1$, and the magnitude of the effect of transit supply on the generalized cost of transit travel $\frac{\partial MPC_T}{\partial K_T} < 0$ is assumed to vary across scenarios.

Litman (2013) provides a thorough review of existing studies estimating various transportation demand elasticities, summarizing several studies that have estimated ϵ_A. There is a wide variation in magnitudes depending on the context in which this elasticity is measured.[21] Oum et al. (2008) review nine studies and report that the range of ϵ_A generally lies between -0.1 and -0.5 for peak auto travel demand, and we use the midpoint value of -0.3 as our value for ϵ_A. Relatively few studies have examined ϵ_T. Past changes in transit fares or travel time have been estimated to have a minor effect on auto demand levels; Litman (2013) summarizes these past studies[22] and reports that the elasticity of auto travel demand with respect to changes in transit cost or travel time typically ranges from 0.01 to 0.09. We use the midpoint value of 0.05 as our baseline value for ϵ_T and vary this parameter across scenarios. α_1 is calibrated by initializing the baseline user equilibrium auto travel volume V_A^0 to reflect observed 2011 data.

Marginal Private Cost of Auto Travel

The marginal private cost of auto travel function in Equation (1) should have the following characteristics: (1) a threshold volume-to-capacity ratio $\widehat{\left(\frac{V_A}{K_A}\right)}$ where the congestion externality begins, and (2) the function is strictly convex beyond $\widehat{\left(\frac{V_A}{K_A}\right)}$. As we are interested in the effects of transit supply on the auto market in a second-best setting, we assume that $\tau = 0$.

The MPC_A function has two components: the monetary cost of auto travel P_A and the monetized value of travel time T_A^T. In our dataset, the median fuel cost in 2011 was \$0.1586 per vehicle-mile traveled. This value is similar to that computed by AAA in its 2011 report: the average

[20] The α values are calibration parameters throughout.

[21] See Tables 4, 9, 10, 18, and 29 in Litman (2013).

[22] See Tables 7, 31, 33, and 35 in Litman (2013).

gas cost per mile was \$0.1234, and the per-mile operating costs (including gas, maintenance and tires) were estimated to be \$0.1774.[23] Fuel efficiency is dependent upon the distribution of travel speeds and thus is a function of the level of congestion, in part due to the stop-and-start driving necessary in congested conditions. Overall, there is typically a U-shaped relationship between fuel consumption and travel speeds (see Barth and Boriboonsomsin (2009) for a related discussion of the relationship between vehicle emissions and travel speeds). Parry (2009) notes that it is typically assumed that heavily congested conditions increase fuel consumption by 30%, though there is considerable uncertainty surrounding this magnitude. According to Greenwood et al. (2007), congestion increases fuel consumption by 13-36% on average across a sample of different vehicle types. To represent this relationship, we use the following functional form:

$$P_A = \begin{cases} \tilde{P}_A & \text{if } \left(\frac{V_A}{K_A}\right) \leq \widehat{\left(\frac{V_A}{K_A}\right)} \\ \tilde{P}_A \left(1 + \alpha_2 \left(\frac{V_A}{K_A} - \widehat{\left(\frac{V_A}{K_A}\right)}\right)^2\right) & \text{if } \left(\frac{V_A}{K_A}\right) > \widehat{\left(\frac{V_A}{K_A}\right)}. \end{cases} \tag{10}$$

This assumes that fuel efficiency is maximized at a volume-to-capacity ratio less than or equal to $\widehat{\left(\frac{V_A}{K_A}\right)}$ and decreases at higher travel volumes; we assume that $\widehat{\left(\frac{V_A}{K_A}\right)} = 0.5$. The minimum value of the fuel cost per vehicle-mile is given by $\tilde{P}_A$ and we use the AAA value, since our estimated per-unit fuel cost is averaged across the largest (and generally most congested) regions. With $\tilde{P}_A = \$0.1234$, α_2 is calibrated to 1.7 to generate an increase in fuel consumption that approaches 42.5% (per-unit value of \$0.176) as the volume-to-capacity ratio increases from 0.5 to 1.

The monetized value of travel time T_A^T is the product of the value of travel time and the travel duration $\tilde{T}_A^T$. Time-averaged speed-flow functions relate the average speed over a specified period to the average vehicle inflow over that period, consistent with the static equilibrium model we are using. This relationship can be modeled as a power function of the volume-to-capacity ratio:

$$\tilde{T}_A^T = \begin{cases} T_f & \text{if } \left(\frac{V_A}{K_A}\right) \leq \widehat{\left(\frac{V_A}{K_A}\right)} \\ T_f \left[1 + \alpha_3 \left(\frac{V_A}{K_A}\right)^{\alpha_4}\right] & \text{if } \left(\frac{V_A}{K_A}\right) > \widehat{\left(\frac{V_A}{K_A}\right)} \end{cases} \tag{11}$$

where $\tilde{T}_A^T$ denotes auto travel time per mile and T_f is free-flow auto travel speed with no congestion. We assume that $T_f = \frac{1}{60}$ hours per mile. This function is used by the U.S. Department of Transportation. With $\alpha_3 = 0.2$ for freeways and $\alpha_4 = 10$, it is known as the 'updated Bureau of Public Roads (BPR) function' and was proposed by Skabardonis and Dowling (1996).

[23] See http://exchange.aaa.com/wp-content/uploads/2012/04/DrivingCosts2011.pdf.

This then yields the monetized value of travel time as:

$$T_A^T = \begin{cases} VOT_A^T \cdot T_f & \text{if } \left(\frac{V_A}{\overline{K}_A}\right) \leq \widehat{\left(\frac{V_A}{\overline{K}_A}\right)} \\ \\ VOT_A^T \cdot T_f \left[1 + \alpha_3 \left(\frac{V_A}{\overline{K}_A}\right)^{\alpha_4}\right] & \text{if } \left(\frac{V_A}{\overline{K}_A}\right) > \widehat{\left(\frac{V_A}{\overline{K}_A}\right)}. \end{cases} \tag{12}$$

Taken together, the marginal private cost of auto travel is:

$$MPC_A = \begin{cases} \tilde{P}_A + VOT_A^T \cdot T_f & \text{if } \left(\frac{V_A}{\overline{K}_A}\right) \leq \widehat{\left(\frac{V_A}{\overline{K}_A}\right)} \\ \\ \tilde{P}_A \left(1 + \alpha_2 \left(\frac{V_A}{\overline{K}_A} - \widehat{\left(\frac{V_A}{\overline{K}_A}\right)}\right)^2\right) + VOT_A^T \cdot T_f \left[1 + \alpha_3 \left(\frac{V_A}{\overline{K}_A}\right)^{\alpha_4}\right] & \text{if } \left(\frac{V_A}{\overline{K}_A}\right) > \widehat{\left(\frac{V_A}{\overline{K}_A}\right)}. \end{cases} \tag{13}$$

For simplicity, we assume there is no congestion interdependence across modes, with the auto congestion function independent of the level of transit supplied. The value of auto travel time is typically assumed to be equal to 50% of the wage rate, though the value of time in congested conditions is approximately twice as high as that in uncongested conditions, as an indication of the value that commuters place on reliable travel times (Berechman, 2009). We assume that $VOT_A^T = \$16.30$ per hour (the 2011 value of time used by Schrank et al. (2012)).[24] With an average vehicle occupancy of 1.25, this implies a per-vehicle value of time of \$20.375 per hour.

Marginal External Cost of Auto Travel

The marginal external cost of auto travel MEC_A can be derived from the marginal private cost above: $MEC_A(\cdot) = V_A \frac{\partial MPC_A}{\partial V_A} = V_A \left[\frac{\partial P_A}{\partial V_A} + \frac{\partial T_A^T}{\partial V_A}\right]$. This implies:

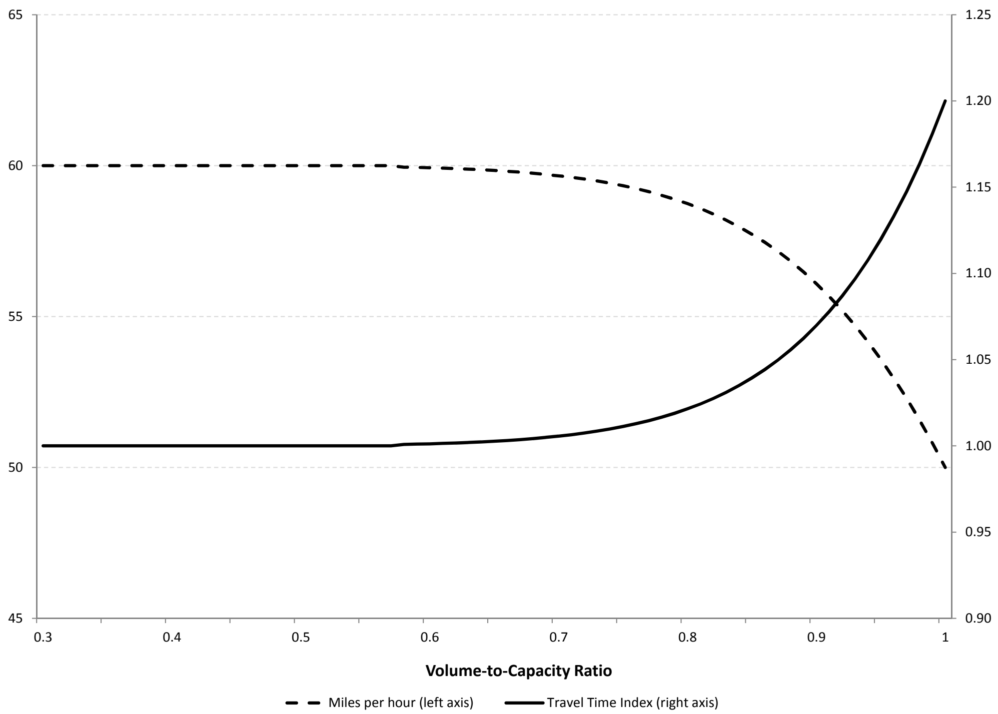
$$MEC_A = \begin{cases} 0 & \text{if } \left(\frac{V_A}{\overline{K}_A}\right) \leq \widehat{\left(\frac{V_A}{\overline{K}_A}\right)} \\ \\ \frac{V_A}{\overline{K}_A}\left[2\alpha_2 \tilde{P}_A \left(\frac{V_A}{\overline{K}_A} - \widehat{\left(\frac{V_A}{\overline{K}_A}\right)}\right) + \alpha_3 \alpha_4 VOT_A^T \cdot T_f \left(\frac{V_A}{\overline{K}_A}\right)^{\alpha_4 - 1}\right] & \text{if } \left(\frac{V_A}{\overline{K}_A}\right) > \widehat{\left(\frac{V_A}{\overline{K}_A}\right)} \end{cases} \tag{14}$$

This function is assumed to contain only the congestion externality and it is an increasing function of the volume-to-capacity ratio. The magnitude of the congestion externality thus varies across the

[24] This value abstracts from any non-pecuniary cost which may be associated with time spent traveling on a congested road. For a discussion of the effect of non-pecuniary factors on the valuation of work time, and hence on travel time, see Farzin (2009).

three scenarios outlined above: for LOS 'B' the marginal external cost of auto travel (evaluated at the user equilibrium) is 4% of the marginal private cost, while for LOS 'D' the marginal external cost is 50% of the marginal private cost, and this ratio increases to 254% for LOS 'F'. Figure 9 shows how the volume-to-capacity ratio relates to travel speeds based on the model parameters.

Figure 9: Relationship between volume-to-capacity ratio and travel speeds for simulation model



Marginal Social Cost of Auto Travel

The marginal social cost of auto travel MSC_A is derived by summing the marginal private costs and marginal external costs above, accounting for the threshold value:

$$MSC_A = \begin{cases} \tilde{P}_A + VOT_A^T \cdot T_f & \text{if } \left(\frac{V_A}{K_A}\right) \leq \widehat{\left(\frac{V_A}{K_A}\right)} \\ \tilde{P}_A \left(1 + \alpha_2 \left(\frac{V_A}{K_A} - \widehat{\left(\frac{V_A}{K_A}\right)}\right)^2\right) & \\ + VOT_A^T \cdot T_f \left[1 + \alpha_3 \left(\frac{V_A}{K_A}\right)^{\alpha_4}\right] & \text{if } \left(\frac{V_A}{K_A}\right) > \widehat{\left(\frac{V_A}{K_A}\right)} \\ + \frac{V_A}{K_A}\left[2\alpha_2 \tilde{P}_A \left(\frac{V_A}{K_A} - \widehat{\left(\frac{V_A}{K_A}\right)}\right) + \alpha_3 \alpha_4 VOT_A^T \cdot T_f \left(\frac{V_A}{K_A}\right)^{\alpha_4 - 1}\right] & \end{cases} \tag{15}$$