1 The Economic Theory of Urban Traffic Congestion: A Microscopic Research Agenda

Transportation scientists make a distinction between microscopic and macroscopic models of traffic flow. The distinction is similar to that between microeconomics and macroeconomics. Microscopic models aggregate up from the behavior of the individual driver, while macroscopic models are formulated in terms of traffic aggregates. Here the terms will be used somewhat more broadly, to distinguish the level of detail employed in the analysis. The theme of this chapter is that the community of urban transport economists has relied excessively on the canonical macroscopic model of urban traffic congestion, and that a redirection of research toward more microscopic modeling will result in more useful and effective policy analysis.

The basic canonical model of urban traffic congestion employed by economists, which will be discussed at some length in the next section, is formulated in terms of the price, flow, and capacity of different modes. Models in this vein are static, and when they look at network aspects of traffic congestion do so at an aggregated level (e.g., freeways versus highways versus city streets).[1] Individuals are treated as having only two margins of choice: how much to travel and what mode to take. But in fact individuals have many more margins of choice. They decide when to travel, where to travel, what activities to undertake on a particular trip, and how long to spend at a destination. Car drivers decide additionally what route to take, what speed to drive at, how aggressively and safely to drive, and how to search for parking. The canonical model also provides only a crude and implicit treatment of urban freight transport, with a truck being treated as so many car equivalents.

Accordingly, policy advice derived from the canonical model is specified in terms of the price and capacity of different modes. But most urban traffic policy is decided at a far more microlevel. Should a particular street be widened? On how many sides of the street should parking be allowed, and should time restrictions be imposed? Should an intersection be signalized, have a four-way stop, or be converted into a traffic circle? How frequently should buses run, what should be the density of their routes, and what should be done to improve schedule reliability? Should restrictions be imposed on truck size and truck delivery hours? The broad policy insights derived from the canonical, macroscopic model are certainly useful but need to be supplemented by more detailed analysis that can be applied at the scale at which actual traffic policy decisions are made.

A more technical criticism of the canonical macroscopic model is that it treats congestion as both qualitatively homogeneous and technologically determined. But there are in fact many different sorts of urban traffic congestion—not only the standard flow congestion treated in the canonical model, but also intersection congestion, queuing congestion, and various forms of parking-related congestion, as well as complicated forms of congestion interaction between cars, buses, trucks, bicycles, pedestrians, and more recently rollerbladers. And rather than being completely determined by technology, the properties of congestion are sensitive to individual decisions: for car drivers, under what circumstances to change lanes and what merging protocol to adopt; for pedestrians, whether to cross only at the intersection and only when the walk light is on; for trucks, the size of trucks to employ for urban deliveries and the routes to take; for bicyclists, whether to cycle with the traffic stream, on the shoulder of the road, or on the sidewalk; and so on.

Applied microeconomic theory has been so successful largely because of its method, which entails working with simple, conceptually consistent models based on maximizing behavior. This method elucidates basic principles and focuses on essentials, abstracting from inessential detail. Recent policy successes of this approach include the development of tradable pollution rights for sulfur dioxide, the expanded use of auctions for resource allocation with relatively small numbers of buyers and sellers, and the application of incentive contracting to public utilities (Laffont and Tirole 1993). The canonical, macroscopic model of traffic congestion, which employs this method, has been highly successful in providing a unified theory of the economics of traffic congestion, in contrast to a patchwork of specific models. With traffic congestion, however, more so than in most other microeconomic policy contexts, there are so many margins of choice and so many policy instruments that ignoring detail can lead to overlooking many important features of the problem and many potentially effective policy tools. The theme is not that the existing body of theory based on the canonical macroscopic model should be rejected but rather than it needs to be supplemented by microscopic models that bridge the gap from principles and broad policy insights to the detailed, on-the-ground policy decisions made daily by urban transport engineers and planners.

A particularly important application of this theme is to congestion pricing of urban auto travel. It has now been almost fifty years since Vickrey (1959), the father of auto congestion pricing, outlined a scheme for congestion pricing of urban auto travel in downtown Washington, D.C. Since then, urban transport economists have continually, and with virtual unanimity, been advocating such congestion pricing, but with little success, at least until very recently. This lack of success has arguably been due to the failure of urban transport economists to move beyond advocacy of the principle, to the details of policy design, including considerations of both technological feasibility and political acceptability.

The first phase of academic literature on the urban congestion pricing, including most notably Vickrey (1955, 1959, 1963), Mohring and Harwitz (1962), and Strotz (1965), developed the theory of first-best transport pricing and capacity. The second phase explored second-best pricing and capacity: the best that can be done, taking into account constraints that preclude implementation of the first-best policy, such as the difficulty of charging cars for travel on city streets. Both of these bodies of literature are reviewed in the next section. The third generation, of which the work by Santos, Newbery, and Rojey (2001) and the work done by dePalma et al. (2003) are the prime examples, is microscopic, comparing alternative ways in which urban auto congestion pricing might be applied in practice, paying attention to such details as location of cordons on actual traffic networks.

Singapore has led the way in policy practice, moving from its initial area licensing scheme implemented in 1975, to increasingly refined measures. In the 1980s and 1990s, many other jurisdictions developed congestion pricing plans—Cambridge and Hong Kong (Borins 1986), Stockholm (Ahlstrand 1998, Abbott 1990), and the Ranstaad in Holland (Verhoef, Nijkamp, and Rietveld 1997), but backed down from implementing them due to both political opposition and doubts concerning their success.[2] The success of Mayor Ken Livingstone's recent cordon pricing policy for London is likely to have a major positive impact on attitudes toward urban auto congestion pricing among both politicians and the public, and to prompt more and more cities to seriously consider the policy and commission studies comparing alternative schemes.

One interpretation of this policy history is that urban auto congestion pricing would have been implemented more rapidly if economists had sooner developed microscopic models of urban traffic congestion, since these would have provided practical guidance on how to design actual congestion pricing schemes. An alternative interpretation is that politicians and the public have slowly been persuaded of the virtues of congestion pricing, due partly to economists' advocacy of the policy in principle, partly to the demonstrated failure of solving the congestion problem by building more and wider freeways, and partly to the demonstrated success of price-based policies in other policy contexts. Whichever interpretation is correct, the stage is now set for the development of third-phase, microscopic models in order to provide sound guidance to local governments concerning the efficient design of practical, congestion-pricing schemes.

Because of their preoccupation with the canonical, macroscopic model, urban transport economists have devoted a disproportionate amount of their attention to congestion pricing. They have written a vast amount on the subject, while almost ignoring alternative and supplementary policies for alleviating traffic congestion, such as parking policy, freight delivery policy, staggered work hours for government employees, and the economic design of urban roads and of urban auto travel regulations. This chapter argues that although the prospects for the widespread adoption of congestion pricing appear more promising today than they have in years, considerably more attention needs to be paid to alternative and supplementary policies. This will require the development of a portfolio of microscopic models, each focusing on some facet of urban travel that has been overlooked through excessive focus on the canonical macroscopic model.

The chapter next sketches the development of the canonical, macroscopic model of urban transport economic theory and provides a detailed critique of it. It then discusses a selection of topics in urban transport economics that have been understudied relative to their importance because they are not easily treated in the context of the canonical, macroscopic model.

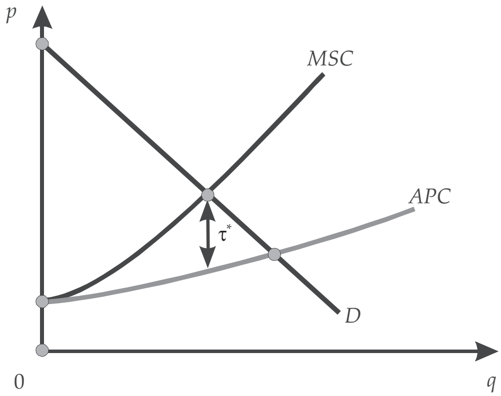

Figure 1.1
The basic model

Urban Transport Economic Theory

The development of urban transport economic theory has entailed the gradual elaboration of a canonical model. The basic model (Beckmann, McGuire, and Winsten 1956) examines travel on a single point-input, point-output road. Drivers are identical, and the only decision each makes is trip frequency. Congestion is captured by a congestion cost function that relates trip cost to travel flow and capacity.

Figure 1.1 gives a diagrammatic representation of the basic model (Walters 1961, Mohring 1976) in the short run, with capacity fixed. *D* is the demand curve, relating the flow of trips demanded, *q*, to trip price, *p*, which equals trip cost plus the toll; *APC* relates each driver's trip cost to traffic flow and is variously referred to as trip cost, average private cost, user cost, marginal private cost, and short-run average variable cost; and *MSC* is the marginal social cost of a trip. The *APC* curve is upward sloping due to congestion; as traffic flow increases, trip time, and hence trip cost, increase. User cost and marginal social cost are related in the same way as average cost and marginal cost in firm cost theory. In the absence of government intervention, the equilibrium occurs where the demand curve intersects the user cost curve, since individuals will travel up to the point where the marginal private benefit of a trip, given by the demand curve, equals the marginal private cost or user cost. The optimum occurs where the demand curve intersects the marginal social cost curve. The vertical distance between

APC and *MSC* is the marginal external congestion cost and captures the costs an individual driver imposes on other drivers by taking an extra trip through slowing them down. The minimal government intervention needed to decentralize the social optimum is the imposition of a congestion toll, τ^* in the diagram, equal to the marginal external congestion cost, evaluated at the socially optimal level of flow.

The same model may be described algebraically using either social surplus or social welfare analysis (Mayeres and Proost 1997). Here the social surplus approach will be taken. Where p is trip price and ω capacity, the demand function is $D(p)$, the user cost function $c(q, \omega)$, the social benefit function $B(q)$, and the capacity construction cost function $K(\omega)$. The direct, long-run social surplus maximization problem may then be written as

$$\max_{q,\omega} B(q) - qc(q, \omega) - K(\omega). \tag{1.1}$$

The corresponding first-order conditions are

$$q\text{: } B'(q) - (c(q, \omega) + q(\partial c(q, \omega)/\partial q)) = 0 \tag{1.2}$$

and

$$\omega\text{: } q(\partial C/\partial \omega) - K' = 0. \tag{1.3}$$

Equation 1.2 states that the optimal traffic flow is such that the marginal social benefit of a trip equals its marginal social cost, which equals user cost plus the marginal external congestion cost. Equation 1.3 states that optimal road width is such that the marginal social benefit from a marginal increase in road capacity, the reduction in travel costs holding traffic flow fixed, equals the marginal construction cost.

The above specification takes q as a decision variable. But the government does not directly control traffic flow; rather it controls it indirectly through the use of a congestion toll. The corresponding indirect social surplus maximization problem, where individuals decide on trip frequency based on trip price, the government decides on the level of the toll as well as capacity, and trip price equals user cost plus the toll, is

$$\max_{q,p,\tau,\omega} (B(q) - qp) + (q\tau - K(\omega)) \tag{1.4}$$

$$\text{s.t. (i) } q = D(p)$$

$$\text{(ii) } p = \tau + c(q, \omega),$$

which reduces to

$$\max B(D(p)) - D(p)c(D(p), \omega) - K(\omega). \tag{1.5}$$

The corresponding first-order conditions are the same as those for the direct maximization problem.

The no-toll equilibrium may be characterized as the solution to $q = D(p)$ and $p = c(q, \omega)$, or as the solution to a constrained maximization problem identical to equation 1.4 except that τ is no longer a policy variable and $\tau = 0$ is an additional constraint.

The basic model has been enriched to account for other margins of choice. Early on, the model was extended to treat route choice and modal choice. With respect to route choice, an individual chooses her route on a network, from a given origin to a given destination, so as to minimize trip price—the generalized Wardrop principle. With respect to modal choice, when modes are perfect substitutes in demand, the same principle applies, and when modes are not perfect substitutes in demand, demand for a particular mode is a function of the trip prices on all modes. A traffic network is described by a set of nodes and a set of links, with congestion assumed to occur on links but not at nodes. Congestion interaction between two modes on the same link, as between cars and buses, may be treated, but congestion interaction between different links is generally ignored. The model was also extended early on to treat user heterogeneity (Strotz 1965). Individuals from different groups have different demand functions, as well as different user cost functions, reflecting differences in values of time and vehicles driven.

It is generally assumed that individual car drivers enter the congestion cost functions symmetrically, and that treating buses, trucks, and other vehicles as so many car-equivalents captures differences between vehicle types. Under these assumptions, in the model extended to treat route choice, modal choice, and user heterogeneity, the social optimum can be decentralized by applying an anonymous toll to each car equivalent on each link in the network equal to that link's marginal external congestion cost. Furthermore, by the Envelope Theorem, when optimal congestion tolling is in place so that each traveler faces the marginal social cost of her travel, the marginal social benefit of capacity on each link can be computed in a straightforward manner as the travel cost savings on that link, without consideration of how travelers switch modes and routes in response to the incremental capacity expansion. These are very important results since they indicate that, under the assumptions of the model, little information is needed to decentralize the first-best optimum. All that is required is to measure the link marginal external congestion cost, which requires only the link congestion functions, traffic levels, and average value of time. No information is needed on the identity of travelers or on their demand functions, or how traffic flow on the network responds to incremental policy changes. It is therefore easy to understand why economists have pushed so hard for first-best congestion pricing.

These models have been applied to treat a wide range of second-best problems. Lévy-Lambert (1968), Marchand (1968), Sherman (1971), and Bertrand (1977) examined how other modes should be priced when auto congestion is unpriced (meaning that no toll is charged) or underpriced. Wheaton (1978) and Wilson (1983) considered how optimal road capacity is altered when auto congestion is unpriced or underpriced. Arnott and Yan (2000) analyzed simultaneously how second-best transit capacity, transit pricing, and road capacity should be chosen when auto congestion is underpriced. Verhoef, Emmerink, Nijkamp, and Rietveld (1996) investigated how the value of providing information on traffic conditions to car drivers is modified by not congestion-pricing car travel. Several articles (Braid 1996; Verhoef, Nijkamp, and Rietveld 1996; Liu and McDonald 1998) have been written on the proportion of the first-best efficiency gains that can be achieved when only a subset of roads can be tolled, on the assumption that individuals are identical; the conclusion reached was pessimistic. However, Small and Yan (1999) and Verhoef and Small (1999) have argued that this conclusion is modified when user heterogeneity is accounted for. With user heterogeneity, drivers will self-select over roads—those with higher values of time choosing the tolled roads, which magnifies the efficiency gains.

The body of literature we have reviewed develops what we have referred to as the canonical macroscopic model of traffic congestion. The development of this body of theory has been admirable in many respects. Through elaboration of a canonical model, the theory has moved from a very simple model to models that are increasingly descriptively realistic and incorporate more and more margins of individual and policy choice. All the model variants meet the standard criteria for good microeconomic modeling: they are thoroughly based on individual maximizing behavior and are conceptually consistent and parsimonious. Furthermore, considerable effort has gone into practical application. There is now a large literature on estimating travel demand functions and developing efficient algorithms to solve variants of the static network equilibrium problem, including the computation of second-best optimal tolls (e.g., Verhoef 2002), and there is a growing number of city-specific travel simulation models based on the above theory (e.g., Mohring 2001 for Minneapolis). These admirable qualities notwithstanding, the next section develops six major criticisms of the canonical, macroscopic model.

Before turning to them, however, mention should be made of two major lines of development in urban transport economic theory that have involved more than simply extending the canonical model: the bottleneck model and the monocentric model of transportation and land use.

The bottleneck model, first presented by Vickrey (1969) and later developed by Arnott, dePalma, and Lindsey (e.g., 1991, 1993, 1998), focuses on trip timing. Vickrey's article made two important contributions. The first was to recognize that in deciding when to travel, an individual has the choice between a convenient time when travel is congested and an inconvenient time when traffic is light. With homogeneous individuals, the equilibrium distribution of travel times will be such that utility is equalized over times when travel occurs; with heterogeneous individuals, the equilibrium condition is that no individual can improve her utility by changing her travel time. Application of this equilibrium condition to a dynamic model with Lighthill-Whitman-Richards flow congestion leads, however, to analytical intractability.[3] The second contribution of the Vickrey article was to circumvent this difficulty by assuming that congestion takes the alternative form of queuing behind a bottleneck of fixed-flow capacity. Interestingly, the reduced-form marginal external congestion cost function implied by the Vickrey bottleneck model, relating the marginal external congestion cost to the flow of travelers over the rush hour and capacity, has been shown to be consistent with the canonical macroscopic model, with the important qualification that the marginal external congestion cost function is not purely technological but incorporates individuals' trip timing decisions in such a way that a change in the time variation of the toll alters the form of the function.

The urban economic theory of transportation and land use, developed in Solow and Vickrey (1971), Solow (1972), Kanemoto (1976), and Arnott (1979), essentially embeds the basic canonical, macroscopic model of traffic congestion into the monocentric city model of the "new" urban economics. The user cost of traveling between distances

x and $x + dx$ from the central business district depends on the volume-capacity ratio at that location. Volume is measured as the number of individuals who travel on the road at x, which equals the number who live beyond distance x from the central business district, and capacity as the amount of land allocated to road use between x and $x + dx$. Anas and Kim (1996) have recently extended the model to nonmonocentric cities.

Criticisms of the Canonical Macroscopic Model

A model is a convenient and often insightful simplification of a complex reality. Use of a model, however, can distort perceptions of that reality. The broad theme of this section is that excessive reliance on the canonical macroscopic model has led urban transport economists to overlook features of urban traffic congestion that are not captured in the model and to ignore or pay too little attention to congestion alleviation policies corresponding to these features.

Many Relevant Margins of Choice Are Ignored

The essential features of most policy problems can be captured by considering only a small number of margins of choice. For example, with industrial pollution, most of the action is captured by modeling a firm's choices concerning its output level and its abatement technology, as characterized by the level of emissions of a few pollutants per unit of output. With insurance, most of the action on the consumer side is captured by viewing the consumer as choosing how much insurance to purchase as a function of its price, the size of the deductible, and the degree of co-insurance, and how much unobservable (observable margins of choice can be written into the contract) effort to expend in reducing the probability of accident (or more generally in modifying the probability distribution of accident damages), which includes expenditure on accident-reducing equipment. And with housing, most of the action can be captured by viewing the consumer as choosing location, floor area, and quality and the producer as choosing the density, durability, and quality of structures.

Traffic congestion appears to be different. The most sophisticated models of urban travel demand treat the traveler as choosing car ownership and then, conditional on car ownership, trip frequency, route, mode, and timing. But urban car drivers make many more decisions than this. A car driver continuously decides how rapidly to accelerate or decelerate, which determines his speed and the distance from the car ahead and the car behind. Periodically, he chooses whether to accept an opportunity to overtake, whether to honk his horn, whether to enter an intersection after the light has turned yellow (or in Boston, red!) or when the intersection is blocked, and whether to shift to an apparently faster lane. And if he does not have employer-provided parking, he must decide as he approaches his destination, whether to park on the street or off, and if on the street what parking search strategy to adopt, which includes how far from his destination to start cruising for parking and under what circumstances to double-park. The congestion caused by buses and trucks as well is sensitive to the behavior of their drivers.

One possible reaction to this enumeration of choices is that they are trivial. Each by itself may be trivial, but cumulatively they are very important. Think how much better traffic would flow and how less stressful driving would be if all drivers were to make socially efficient decisions or if vehicle control were automated (Ioannu and Bose 2002). Another possible reaction is that economists have little useful to say concerning these decisions, even though they are economic in the sense that drivers weigh costs against benefits in making them and that the regulation of driver behavior should be left to traffic engineers. But traffic engineers decide on traffic regulations with no explicit economic behavioral analysis, and often on the basis of insufficient data and flawed statistical analysis (Hauer 2000).[4]

If indeed there were first-best congestion pricing on every margin of choice, it would not matter that economic analysis overlooks some margins of choice. Drivers would face the right prices on every margin and would therefore make socially efficient decisions on every margin. But in practice, congestion pricing cannot be differentiated according to driver behavior.[5] Under anonymous congestion pricing, aggressive and timid drivers impose a larger marginal external congestion than socially responsible drivers but pay the same toll. As important, anonymous congestion pricing provides no incentive to drive in a socially responsible manner. A rational driver who faces the same toll independent of how he drives will drive selfishly. Thus, even with homogeneous drivers, first-best congestion pricing takes as a given selfish and inefficient driver behavior, which renders computation of the optimal toll an exercise in the theory of the second best. If driver behavior were independent of the magnitude of the toll, the first-best toll would be correctly computed. But if a larger toll causes individuals to drive faster and more aggressively, which might be the case with a congestion toll based on time in congested traffic, the optimal anonymous toll is less than the conventionally computed first-best toll.

The discussion has focused on individual margins of choice that the conventional analysis ignores. Policies associated with these margins of choice are correspondingly ignored. So too are car manufacturers' margins of choice. Automobile characteristics, such as size and acceleration, affect the congestion caused by a car. Since these characteristics are at least partially observable, congestion pricing could be based on them. Given the current state of technology, however, doing so is impractical, but then drivers have no incentive to purchase "congestion-efficient" cars or car makers to manufacture them.

The costs of traffic accidents are typically treated separately from the costs of congestion. But a substantial fraction of the time lost due to congestion results from nonrecurrent congestion, and a substantial fraction of nonrecurrent congestion is due to traffic accidents. Making cars less prone to accident would therefore reduce congestion. But neither car manufacturers in their design of cars nor consumers in their choice of car will take this into account if the congestion toll (and auto accident insurance) is independent of vehicle type.

The Congestion Function Captures Not Only Technology But Also Behavior

The canonical macroscopic model treats the congestion function as being determined by technology, but in fact the congestion cost function incorporates all the behavioral decisions related to travel as well. This point has been demonstrated formally with respect to drivers' trip-timing decisions, but applies as well to the many margins of choice related to driver behavior. Overlooking this point results in overlooking policy instruments, such as traffic regulations, that reduce congestion only through such behavioral margins and also in underestimating their benefits.[6]

A related point is that treating the value of time (how much a driver would be willing to pay to reduce travel time by one unit), which enters the congestion cost function as a datum, makes it easy to overlook policy instruments that affect the value of time. For example, traffic-calming policies that discourage aggressive driving make driving more pleasant, which reduces the value of time and hence congestion costs.

Capacity Is Too Aggregated a Policy Variable

Transport planners do not choose capacity per se. Instead, they choose road width, gradient, banking, and pavement quality, as well as speed limits and the quality of traffic signing, and so on, which together determine capacity. Since transportation planners tend to use engineering rules of thumb without reference to economic variables, their choice of how to provide a given level of capacity may differ significantly from the design that minimizes social costs. Economists have a role to play in advising transportation planners how to provide a given level of capacity in different economic environments efficiently. To do this, a set of transport economic models is needed that provides a richer treatment of traffic engineering. Two examples of excellent work along these lines are Newbery (1988) and Small, Winston, and Evans (1989). Newbery examines the economics of pavement resurfacing; Small, Winston, and Evans look at the economics of road damage, considering not only how vehicles should be charged for the road damage they cause but also how pavement durability should be chosen.

Another way economists can contribute to traffic engineering theory and practice is by developing microscopic models of traffic flow with behavioral foundations. Some work has already been done along these lines. Rotemberg (1985) and Verhoef, Rouwendal, and Rietveld (1999) provide models in which drivers decide on speed and spacing so as to maximize utility, trading off travel time against the probability of accident.[7] Further work along these lines could build on microscopic traffic flow theory (Cassidy 2002), including car-following theory, that derives aggregate traffic flow from a difference-differential equation describing the acceleration of an individual vehicle as a function of the location, speed, and acceleration of the car in front, and perhaps the car behind as well.[8]

Link Flow Congestion Is Not the Only Form of Congestion

The canonical macroscopic model treats only one form of congestion: link flow congestion, in which a driver's travel time and travel costs on a link are positively related to traffic volume or flow on the link.[9] But there are many traffic congestion phenomena inconsistent with link flow congestion. Link flow congestion ignores transient, non-steady-state flow phenomena such as shock waves and traffic jams. It ignores nodal congestion. In line-based telephone traffic, congestion at nodes (in switching circuits) is more important than congestion on links (Syski 1986). Examples of nodal congestion in the context of urban travel are intersection congestion and freeway entrance and exit congestion.[10] Link congestion is more important than nodal congestion in freeway travel but not on city streets. Other forms of congestion include pedestrian-car interaction, entry into and exit from parking, merging, and phenomena deriving from the physical length of cars such as gridlock[11] and from the interaction between different vehicle types.[12] Of these, quantitatively the most important is parking congestion, which is the topic of chapter 2.

To derive efficient urban transport systems, richer and more microscopic models of congestion will be needed.

Interaction between Urban Travel Distortions and Other Distortions in the Economy May Be Important

The conventional modeling of urban travel ignores interactions between urban travel distortions and other distortions in the economy. Two of these are especially noteworthy. The first, which has been treated in the second-best literature (e.g., Parry and Bento 2001; Calthrop, Proost, and Van Dender 2000), is the interaction between inefficiently priced traffic congestion and distortionary taxation. Of particular importance is the interaction between the labor-leisure distortion caused by income taxation and the distortion caused by traffic congestion. Suppose that work hours are standardized and noncommuting trips are taken in off-peak hours. Reducing the peak congestion toll below the marginal peak external congestion cost and raising the off-peak toll above the off-peak marginal externality cost would cause the after-tax-and-toll wage to increase, which would reduce the deadweight loss associated with the income-tax-induced labor-leisure distortion.[13]

The second, which has been discussed informally but not analyzed formally, concerns the connection between interaction externalities and urban traffic congestion. In recent years, there has been considerable research on the economics of agglomeration (Fujita and Thisse 2002 provides an excellent treatment of the subject). Most contributors to that literature consider that the nonmarket exchange of information and the informal contracting achieved through face-to-face interaction is a primary, and probably the dominant, force encouraging at least CBD (central business district) firms to cluster. If an additional worker joins the downtown labor force, he will benefit from face-to-face interaction with other workers, and he will confer benefits on the other workers through the unpriced expert services he provides them in their face-to-face interaction. Such interaction entails a positive externality. Now the equilibrium pattern of agglomeration comes about through a balancing of centripetal and centrifugal forces. The dominant centripetal force, according to current wisdom, is face-to-face interaction, while the dominant centrifugal force is transportation cost, which traffic congestion magnifies. Therefore, if the ratio of the interaction externality to the centripetal interaction force is of the same order of magnitude as the ratio of the traffic marginal external congestion to the centrifugal congestion force, then the interaction externality is of the same order of magnitude as the traffic marginal external congestion. To mitigate the deadweight loss associated with the interaction externality, interaction should be encouraged, and perhaps encouraging travel by subsidizing it is an efficient way to do this. This casual line of argument suggests that when face-to-face interaction cannot be subsidized directly, the congestion toll should be set below the marginal external congestion cost in order to encourage interaction. But this casual line of argument is too casual. If the interaction externality is distance related, it results in business land use being insufficiently concentrated, while the marginal congestion externality results in both business and residential land use being insufficiently concentrated. Little is known about the interaction externality empirically, and even less about how it interacts with the marginal congestion externality.

What implications these interactions between externalities have for urban travel policy is not at all clear. But they may be sufficiently important quantitatively to significantly alter the policies prescribed on the basis of the canonical model that ignores them.

The Demand for Travel Is Predominantly a Derived Demand

The canonical model treats individuals as deriving utility from travel per se. But individuals derive utility from activities arrayed over time and space, as well as from conventional goods and services. These activities require transportation, as well as other goods and services, as inputs. Thus, the demand for travel is primarily derived from the demand for activities. Almost everyone acknowledges the correctness of this point, yet little progress has been made in developing activity-based models of derived travel demand (but see Bhat and Koppelman 2002). The associated scheduling problems are difficult to solve, and problems involving scheduling coordination between individuals even more so. Thus, it seems that we are stuck with treating travel as a final good, which masks the possibility of alleviating traffic congestion through modifying the relationship between activities and travel. The most obvious connection between the two is land use. Concentrating land use should reduce travel; thus, minimum density controls might reduce traffic congestion and the deadweight loss associated with it. Also, mixing land uses can be effective in reducing average distances traveled in undertaking activities and in encouraging walking and bicycling rather than driving. These observations loom large in the literature on the new urbanism (Bernick and Cervero 1996) that advocates planning of sustainable cities that foster more sociable, healthier, and generally more pleasant lifestyles.

A Selection of Research Topics

The previous section argued that excessive reliance on the canonical, macroscopic model of congestion has caused urban transport economists to overlook or to pay insufficient attention to congestion alleviation policies that do not fit neatly into that model framework and that research should be redirected toward developing a portfolio of microscopic models focusing on margins of choice not captured in the canonical model. Two of the subsequent chapters present such microscopic models. Chapter 2 develops a microscopic model of saturated parking in the downtown area aimed at providing a conceptual framework for the development of a coherent downtown parking policy. Chapter 4 develops a microscopic model looking at the costs and benefits of instituting staggered work hours for employees of dominant firms, and particularly for government employees. This section presents a selection of other research topics in this vein. The intent is not to provide an exhaustive research agenda but rather to illustrate the range of important policy issues related to traffic congestion that have been overlooked by urban transport economists due to their reliance on the canonical model.

Regulation of Freight Deliveries: Time of Day and Truck Size

In downtown Boston at least, freight delivery contributes significantly to downtown traffic congestion. Large interstate (designed for freeway travel) trucks have trouble maneuvering around corners, especially on narrow streets; they completely block streets when entering and exiting from loading docks; and their double parking for deliveries where there is no loading dock severely reduces capacity.

There appear to be no modern economic studies of urban freight delivery in the context of urban travel congestion.[14] There are studies that estimate the congestion on freeways caused by various types of trucks, in terms of car equivalents. These estimates have apparently simply been applied without question to urban traffic congestion, even though on freeways, unlike on many city streets, cars can overtake trucks. By employing such a crude treatment of urban freight transport, many policies that might significantly reduce the marginal external congestion cost imposed by trucks have been overlooked.

One such policy is imposing restrictions on the times of day at which downtown freight deliveries can be made. Such a policy is on the books in Paris, though how rigorously it is enforced is unclear, and it has been discussed, though not implemented, in Boston. Evaluating such a policy would require estimating the inconvenience costs to shippers and receivers of restricted delivery hours and the noise nuisance costs to urban residents from off-peak deliveries, and then weighing these costs against the benefits from reduced traffic congestion. This would require a more sophisticated treatment of the technology of congestion interaction between cars and trucks.

Another related policy is restricting large trucks from travel on narrow streets or at times and in areas where traffic congestion is severe. Evaluating this policy would require knowledge of warehousing technology and practice. Suppose, at one extreme, that current practice is for all consumer goods transported by intercity truck to be unloaded at suburban warehouses and reloaded onto smaller trucks for delivery within the metropolitan area. In this case, restricting truck size for urban deliveries would impose relatively little additional cost. Suppose, at the other extreme, that current practice is for all goods to be transported directly from the supplier to the retailer. Regulating truck size for urban delivery would then be considerably costlier. If door-to-door deliveries were to continue, the smaller truck size would substantially increase intercity shipping costs. Otherwise, new warehousing districts would have to be constructed where goods shipped intercity in large trucks would be unloaded and loaded onto smaller trucks for urban delivery. This raises the additional question of what regulations, if any, should be applied to warehousing.

Study of the economics of urban freight transport, including the form of congestion interaction between cars and trucks, should be close to the top item on any agenda of research on urban traffic congestion. But before significant progress can be made, systematic collection of the relevant data will be necessary.

The Engineering Economics of Urban Auto Congestion

Mention has already been made of the exemplary work of Newbery (1988) and Small, Winston, and Evans (1989) related to the economics of freeway and highway design, construction, and maintenance. Their work corresponds to sound cost-benefit practice. If sound cost-benefit analysis were practiced in all aspects of urban road engineering, substantial cost savings would be achieved, especially if second-best considerations were properly accounted for (Kanemoto 1999). Though the work might be rather unglamorous and conceptually somewhat prosaic, urban transport economists can make valuable contributions by demonstrating in practical applications how cost-benefit analysis should be done and by pressing hard for the adoption of sound cost-benefit practice by state and local governments in the evaluation of all aspects of urban transportation policy. They can also contribute by applying economics to the nuts-and-bolts of road design. For example, even the most sophisticated cost-benefit procedures currently employed provide only crude treatments of uncertainty, taking no account of the literature on irreversible investment and real options (Dixit and Pindyck 1994). Rules for road resurfacing should take into account that future traffic volumes and future discount rates are generated by particular stochastic processes, that our understanding of the technology of pavement damage will improve, and that technological advances in pavement design will occur, and should accordingly be more adaptive and flexible.

Another aspect of urban transport policy to which economists can make useful contributions is the design of tendering and procurement procedures, drawing on the incentive contracting work of Laffont and Tirole (1993).

Hauer (2000) has argued that the bulk of transport engineering standards and rules is based on scant data collection and often on faulty statistical analysis. Engineering standards also tend to be applied on a one-size-fits-all basis, without reference to economic variables—the discount rate and the value of time, for example. Studies that derive more flexible standards and rules based on sound empirical analysis and economic theory should have a high return.

To do transport engineering economics well requires a good knowledge of transport engineering theory and practice. Until recently, there was little communication between traffic engineers and transportation scientists, and transport economists. The situation has been rapidly improving, however. Economics is now taught in most graduate transportation engineering programs, and the transportation science literature is drawing increasingly on economics, as is evident from extensive participation of transportation scientists in the design of congestion pricing schemes. Transportation scientists and traffic engineers quickly learned the power of prices, but they are learning more slowly the potential value of applying economics to all aspects of engineering design. On the other side, an increasing proportion of urban transport economists are keeping abreast of the relevant traffic engineering and transportation science literature published in such journals as *Transportation Science* and *Transportation Research.*[15] But urban transport economists could do more to foster communication by becoming more actively engaged in transportation engineering programs, through joint research, participation in thesis supervision, and seminar attendance and by exposing their economics students to more transportation science and traffic engineering in their courses.

Traffic Noise

Urban transport economists have given little attention to traffic noise. When they have treated it, it has usually been by incorporating noise costs as a component of generalized congestion costs. But this approach tends to treat noise as technologically determined, and therefore to overlook specific noise-reduction policies. Noise reduction would be politically popular.[16] What needs study is the cost-benefit calculus of alternative policies. It would not be difficult to design quieter engines for cars, but car makers have little incentive to do so because each driver bears only a small fraction of the engine noise cost her car generates. The same point applies to trucks and buses, which contribute disproportionately to traffic noise. Roads and tires as well can be designed to generate less noise, buses and trucks can be designed to idle more quietly, and road work and garbage collection can be made quieter. Regulation too can be effective in reducing noise; for example, horn honking, a curse of living in U.S. cities could easily be dealt with by making it a traffic offense except when done to avoid an accident, as is the case in most European cities.

Traffic Accidents

Traffic accidents are costly not only for the direct damage they cause but also for the nonrecurrent congestion they induce. Economists have paid some attention to traffic accidents (Vickrey 1968 is the seminal work), especially to the effects of insurance on the incentive to drive safely (e.g., Boyer and Dionne 1985), but have left other aspects, such as the regulation of unsafe driving, the design of roads for safety, and accident follow-up procedures, for traffic engineers to deal with, even though traffic engineers tend to choose rule-of-thumb policies with little or no reference to economics.[17] Virtually no attention has been paid by either economists or engineers to the behavioral links between traffic congestion and road accidents. But these must be important since the trade-off between reducing travel time and increasing accident risk is surely central to the many small decisions drivers make.

The Value of Time and Uncivil Driving Behavior

The standard model of traffic congestion treats an individual's value of time as exogenous. But it is not. It depends on the scheduling constraints she faces and on how pleasant or unpleasant she finds driving. A simple way to reduce the costs of traffic congestion and marginal external congestion costs is therefore to make travel more pleasant.

The market takes care of automobile comfort but not the stress level associated with driving. This stress level is strongly related to the incidence of uncivil and dangerous driving: tailgating, honking at the slightest provocation, running yellow and red lights, and making dangerous and excessive lane changes. Public policy can be effective in discouraging some forms of antisocial driving behavior, but how to deal with an angry, aggressive driver is more problematical.

Mass Transit

The bulk of the work done on the economics of urban traffic congestion has focused on cars. Relatively little has been done on the economics of mass transit, presumably because until recently, most of the innovative research in urban transport economics was done in the United States, where mass transit is relatively unimportant. Mohring (1972) explored some of the basic economic principles of mass transit, in particular economies of service frequency and service density; a substantial literature estimates mass transit cost functions (e.g., Berechman 1993) and a smaller literature examines capital-intensity bias (e.g., Frankena 1979) and the optimal scrappage of buses and subway cars. But urban transport economists have devoted little attention to the microscopics of urban mass transit.[18] Sample topics include the size, comfort, and acceleration of buses and subway cars; the length of subway trains and station platforms, and the service frequency and density of buses and subway trains; procedures to mitigate bus bunching, expedite passenger entry and exit, reduce the marginal external congestion imposed by buses on cars, and repair and maintain rolling stock; and the design of fare collection systems.

Pedestrian Traffic

The economics of pedestrian traffic is virtually unexplored. Pedestrian traffic is important in three ways. First, traffic congestion is reduced when travelers walk on shorter trips; second, pedestrian traffic can be stimulated and the costs of walking reduced by making walking more pleasant and less congested; and third, pedestrian-car congestion interaction can be important. Microscopic research topics include optimal sidewalk width, the treatment of jaywalking, the separation of bicyclists and rollerbladers from walkers on paths, and the construction of sidewalks in suburban areas.

Bicycling Traffic

Chapter 3 argues that commuting by bicycle merits more, and more serious, attention from urban transport economists than it has received. The neglect of bicycling derives from its small modal share, especially in the United States. And yet in some cities, most notably Amsterdam and Copenhagen, a substantial proportion of commuting trips are by bicycle. Why is bicycle commuting so much more popular in those cities than in apparently similar cities such as Hamburg or Brussels, or in Bordeaux or Dublin where the weather is at least as favorable? That chapter presents a number of reasons that there might be multiple equilibria in bicycle modal share As this share increases, cars and bicycles interact better in traffic, the transportation system becomes more bicycle friendly, and cultural norms change. Modeling the nonconvexities underlying these effects is a necessary step toward determining which cities would be better off with a quantum increase in bicycle commuting.

Restriction of Shopping Hours

Germany has recently liberalized its restrictions on shopping hours; previously, shops were not allowed to remain open after 6:00 P.M. on weekdays and noon on Saturdays or on Sundays or holidays. Many other jurisdictions have debated Sunday shopping. A major argument in favor of extending shopping hours is reduction in traffic congestion. Is the argument sound? It is at least conceivable that the policy would result in so many more people taking separate shopping trips rather than shopping on the way home from work that overall congestion would increase. And if the argument is sound, are liberalized shopping hours necessarily desirable? Chapter 4 discusses a policy that is similar in some respects: staggering work hours. The limited evidence available suggests that by decreasing synchronization of workdays, staggered work hours undermine productivity. Similarly, extending shopping hours might make the schedule coordination of family and other social activities more difficult.

Noncommuting Trips

Twentieth-century urban economics was criticized for being preoccupied with refining a model of the nineteenth-century city. Urban transport economics can be subject to a similar criticism—for being preoccupied with commuting traffic at a time when an increasing proportion of rush-hour travel has a noncommuting purpose. Fifteen years ago, the figure was bandied about that less than 50 percent of rush-hour trips are for commuting, and the figure for Chicago has recently fallen to 30 percent. Such figures must be treated with caution; since an ever increasing proportion of trips are chained (combining more than one activity on a single trip), the measured proportion of trips that have a commuting purpose is sensitive to how chained trips are treated. However, even if all trip chains for which commuting is a component are classified as commuting trips, the proportions of trips and of miles driven with a noncommuting purpose have been steadily increasing. This trend is easy to understand. Workers, with more leisure time and more money to spend, will naturally travel more for recreation and shopping, and an increasing proportion of nonworkers now have full- or part-time access to a car.

Noncommuting trips present difficulties for urban transport planners because they are harder to forecast and plan for. A commuting trip has a well-defined origin, destination, and desired arrival or departure time, making it relatively easy to forecast how commuters in the aggregate will respond to a change in traffic policy. Noncommuting trips, however, are derived from the scheduling of nonwork activities and have flexible origins, destinations, and desired arrival times. Unfortunately, activity-based models of travel demand are still in their infancy. Until such models have been refined and experience gained in their use, we will just have to live with the decreased accuracy in forecasting that noncommuting trips give rise to.

Transportation and Land Use

The interdependence between transportation and land use has long been recognized: land use strongly affects transportation patterns, and properties of the transportation network strongly affect land use. Urban economists have a good understanding of how the two are related in a monocentric city with only commuting travel to the central business district (e.g., Kanemoto 1980). But cities are becoming increasingly polycentric, job locations increasingly dispersed (see Anas, Arnott, and Small 1998), noncommuting trips increasingly important, and the choice of residential location less strongly tied to workplace location.[19] What the future will bring with respect to changes in land use is unclear, and how these changes in land use are likely to affect traffic volumes and patterns even more so.[20]

Land use controls on a macro scale, as practiced in Northern Europe, can certainly be very effective in reducing overall travel and congestion, but that does not make them desirable. Land use controls at the meso level, as applied by suburban jurisdictions in the United States at the metropolitan fringe, can be effective in discouraging development but also generate strong financial incentives for the controls to be relaxed.[21] Land use controls on a micro scale also affect travel demand; the pattern in the United States of separating residential from nonresidential land uses encourages car travel, while mixing land uses at the local level, as proposed by the new urbanists, discourages it. But the effects on urban travel of such policies are likely to be swamped by the effects generated by the dramatic changes in urban spatial structure noted above that are being wrought by market forces.

Conclusion

This chapter has advanced the argument that excessive reliance on a single class of models—what was termed the canonical, macroscopic model of traffic congestion—has caused urban transport economists to overlook many traffic alleviation policies that operate on margins of choice not incorporated into that model. Excessive reliance on the macroscopic model has also probably caused urban transport economists to allocate too large a proportion of their time to the study of congestion pricing, though the recent success of the London congestion-pricing scheme renders this claim less compelling.

Elaboration of the canonical, macroscopic model has generated an impressively coherent body of theory that has proved to be remarkably adaptable. At the same time, it has caused urban transport economists to pay insufficient attention to those margins of choice and those policies that do not fit neatly into the macroscopic framework. Because these margins of choice and policies appear when looking at traffic congestion in a more detailed way, they are referred to as microscopic. The argument here is not that the macroscopic model should be jettisoned but rather that it should be complemented by a portfolio of microscopic models, each of which looks at some aspect of urban traffic congestion in more detail. After detailing criticisms of the macroscopic model and enumerating many aspects of urban traffic congestion and many urban transport policies it overlooks or treats crudely, a selective microscopic research agenda was presented.

Looking at traffic congestion at the microscopic scale at which most congestion alleviation policies are actually applied will reduce the distance between urban transport economic theory and urban transport policy practice. This should lead to urban transport economists' having more influence on urban transport policy and to urban transport policy being better informed by economics and hence more efficient.

The redirection of urban transport economic theory advocated in this chapter toward more microscopic models has already begun and can be expected to continue.

A central feature of urban traffic congestion that the canonical model has treated only crudely is parking. The next chapter develops a model of saturated downtown parking and applies it to the analysis of downtown parking policy. In so doing, it illustrates the related themes developed in this chapter.

Notes

1. There is a large literature in traffic engineering and transportation science on the static, network equilibrium problem that extends the macroscopic model to networks. But economists typically examine models with only a single origin and destination, with alternative links representing alternative modes.

2. There is widespread agreement that congestion pricing is technologically feasible and could be implemented at reasonable cost. Engineers tend to advocate more sophisticated, capital-intensive schemes, while economists, recognizing that progress in the development of congestion-pricing technologies is likely to occur rapidly, thereby causing rapid technological obsolescence, argue for schemes with lower fixed costs.

3. The term *flow congestion model* is used in two ways. The first refers to any model in which the level of congestion depends on flow; the canonical macroscopic model is a flow congestion model in this sense. The second refers to models that treat traffic as a

compressible fluid and hence use the equation of continuity (conservation of mass for fluids); the seminal models in this vein were by Lighthill and Whitman (1955) and Richards (1956).

4. Because of their training, traffic engineers tend to regard the congestion function as technologically determined, whereas it incorporates all the behavioral decisions listed above.

5. Such differentiation would probably be undesirable anyway, even if it were technologically possible, since it would be unacceptably Big Brotherish.

6. The same point applies with respect to other costs associated with urban car travel: traffic noise, automobile-generated pollution, and traffic accidents.

7. Properly, the individual chooses acceleration so as to achieve the desired speed and spacing, and does so with a reaction lag.

8. While William Vickrey is best known among transport economists for his tireless crusading for congestion pricing, popularizing use of the iso-elastic congestion cost function, and development of the bottleneck model, he also pioneered in the engineering economics of congestion (not just traffic congestion but congestion on electrical networks too). He regularly attended the Transportation Research Board annual meetings, as well as international conferences in traffic engineering; he developed several models of traffic congestion other than the link flow and bottleneck models; and he also analyzed efficient subway scheduling, car seating arrangement, braking, fare collection, and platform length, and devised schemes to mitigate bus bunching.

9. The corresponding marginal external congestion cost function can also be interpreted as providing a reduced-form representation of bottleneck congestion.

10. In applications of traffic network equilibrium theory, which incorporates only link flow congestion, intersections are treated as sets of links. For example, a northbound car that makes a right-hand turn is viewed as traveling along a link joining the north-south and east-west roads. This treatment is better than nothing, but not entirely satisfactory since it ignores the interaction within the intersection of traffic traveling in different directions.

In modeling urban auto traffic congestion, Santos, Newbery and Rojey (2001) use a traffic engineering simulation model, a version of SATURN, that follows the movements of individual cars through intersections on actual urban traffic networks. They find that at the aggregate level, the congestion cost function with intersection congestion is much the same as that with bottleneck congestion.

11. Consider a Manhattan network of one-way streets with intersections. Suppose that northbound traffic backs up into the upstream intersection. This blocks the westbound traffic at that intersection, which causes traffic to back up into the corresponding upstream intersection, blocking southbound traffic at that intersection. This in turn causes traffic to back up into the corresponding upstream intersection, which blocks eastbound traffic at that intersection, which in turn backs up into the corresponding upstream intersection, which blocks the northbound traffic. When the square is completed, the grid is locked.

12. This is typically dealt with by treating vehicles other than cars as so many car-equivalents. This is clearly unsatisfactory. Consider, for example, treating a car that is cruising for parking as generating an equivalent amount of congestion as n cars in

regular traffic. On a two-way city street with no room for passing, one car cruising for parking causes all other cars on the block traveling in the same direction to slow down, whereas if the street is one way and has two lanes, the faster cars can overtake the slow car. Similarly, the car-equivalents of a bicycle have been estimated to depend on the speed and density of traffic as well as the proportion of vehicles on the road that are bicycles (Litman 1999).

13. Under this scenario, peak travel is a complement to labor. Thus, this policy follows the Atkinson-Stiglitz rule of optimal commodity taxation in the presence of a Pareto-efficient income tax: subsidize substitutes for leisure and tax complements.

14. We had intended to devote a chapter of this book to the economics of urban freight delivery but were unable to find sufficient data to generate the empirical regularities needed for persuasive modeling. One problem is that since the trucking industry is intensely competitive, trucking companies are loath to release data.

15. Much of the transportation science literature is operations research applied to transportation. This portion of the literature is easily accessible to economists.

16. The literature on noise costs seems to assume that the costs of noise depend solely on its level. However, introspection suggests that at least up to a certain volume, discontinuous increases in the noise level are more disturbing.

17. The economics of accident follow-up procedures is interesting. Accidents cause huge delays. Even a simple vehicle disablement on the shoulder of a congested freeway generates several hundred hours of vehicle delay. The direct effect of an accident is to reduce capacity. But there are three indirect effects that are at least as important. The first is "curiosity congestion"—cars in both directions slowing down to see what is going on; the second is related—secondary accidents induced by the primary accident; and the third is the congestion caused by emergency vehicles reaching the site. Should the police quickly take photos at the site, immediately tow away the vehicles involved in the accident, and conduct interviews off-site? And should not the vehicles in the accident be held responsible for the congestion costs the accident causes?

18. The major exception is Vickrey, and the bulk of his research on the subject remains unpublished.

19. There is an interesting body of literature in urban economics on "wasteful" or "excess" commuting (Hamilton and Roell 1982, Small and Song 1992, White 1988). Taking the spatial distribution of individuals and jobs as given, imagine assigning individuals to jobs so as to minimize aggregate distance traveled. The actual aggregate commuting distance traveled is about three times as large. A variety of noncompeting explanations have been advanced: Tiebout sorting (individuals consider the tax–public service package offered by different communities when deciding where to live), two-worker households, anticipation of change in future job location, moving costs, thinness of the labor market, minimum lot size controls, racial, ethnic, and income segregation, and idiosyncratic preferences for residential location.

20. Gordon and Richardson (2001) have written an interesting series of papers documenting various empirical aspects of the evolution of traffic congestion in the Los Angeles area. Of particular interest is their finding that average commuting times and distances remained steady over the period from 1970 to 1995, despite rapid population growth, relatively little new freeway construction, and no marked increase in the length of the rush hour. The most plausible explanation is that jobs have been suburbanizing at such a rate that the level of congestion remained more or less constant. Whether this is

the broad trend of the future or whether urban auto travel will become increasingly slow remains to be seen.

21. Land use forecasting is by far the weakest link of travel demand forecasting. Much of the uncertainty concerns where new development will occur, and especially where new subcenters will arise. But the uncertainty is compounded by not knowing whether land use controls at the metropolitan area are really binding.

References

Abbott, J. 1990. Key role for public transport funding boost in Stockholm's road pricing plans. *Urban Transport International* Aug.–Sept.:32–34.

Ahlstrand, I. 1998. The rise and fall of the heroic transport plan for Stockholm. *Transport Policy* 5:205–211.

Anas, A., R. Arnott, and K. Small. 1998. Urban spatial structure. *Journal of Economic Literature* 34:1426–1464.

Anas, A., and I. Kim. 1996. General equilibrium models of polycentric urban land use with endogenous traffic congestion and job agglomeration. *Journal of Urban Economics* 40:232–256.

Arnott, R. 1979. Unpriced transport congestion. *Journal of Economic Theory* 21:294–316.

Arnott, R., A. dePalma, and R. Lindsey. 1991. Does providing information to drivers reduce traffic congestion?. *Transportation Research* A 25A:309–318.

Arnott, R., A. dePalma, and R. Lindsey. 1993. A structural model of peak period congestion: A traffic bottleneck with elastic demand. *American Economic Review* 83:161–179.

Arnott, R., A. dePalma, and R. Lindsey. 1998. Recent developments in the bottleneck model. In K. Button and E. Verhoef, eds., *Road pricing, traffic congestion and the environment: Issues of efficiency and social feasibility*. Aldershot, England: Edward Elgar.

Arnott, R., and A. Yan. 2000. The two-mode problem: Second-best pricing and capacity. *Review of Urban and Regional Development Studies* 12:170–199.

Beckmann, M., C. McGuire, and C. Winsten. 1956. *Studies in the economics of transportation*. New Haven, Conn.: Yale University Press.

Berechman, J. 1993. *Public transit economics and deregulation policy*. Amsterdam: North-Holland.

Bernick, M., and R. Cervero. 1996. *Transit villages in the 21st century*. New York: McGraw-Hill.

Bertrand, T. 1977. Second-best congestion taxes in transportation systems. *Econometrica* 45:1703–1715.

Bhat, C., and F. Koppelman. 2002. Activity-based modeling of travel demand. In R. Hall, ed., *Handbook of transportation science*. Norwell, Mass.: Kluwer.

Borins, S. 1986. The political economy of road pricing: The case of Hong Kong. In *Proceedings of the Fourth World Conference on Transport Research* (vol. 2, pp. 1367–1378). Vancouver, Canada: World Conference on Transport Research.

Boyer, M., and G. Dionne. 1985. Sécurité routière: Responsabilité pour négligence et tarification. *Canadian Journal of Economics* 18:814–830.

Braid, R. 1996. Peak-load pricing of a transportation route with an unpriced substitute. *Journal of Urban Economics* 40:179–197.

Calthrop, E., S. Proost, and K. Van Dender. 2000. *Optimal road tolls in the presence of a labour tax*. Mimeo. Leuven: Centre for Economic Studies. Available on-line: ⟨http://www.soc.uoc.gr/calendar/2000EAERE/papers/PDF/F1-Calthrop.pdf⟩.

Cassidy, M. 2002. Traffic flow and capacity. In R. Hall, ed., *Handbook of transportation science* (2nd ed.). Norwell, Mass.: Kluwer.

dePalma, A., K. Kilani, R. Lindsey, P. Moilanen, E. Niskanen, S. Proost, A. Sen, and A. Vold. 2003. Welfare effects—Urban transport. Marginal Cost Pricing in Transport—Integrated Conceptual and Applied Model Analysis (MC-ICAM). Work Package 7. Unpublished report for the European Commission.

Dixit, A., and R. Pindyck. 1994. *Investment under uncertainty*. Princeton, N.J.: Princeton University Press.

Frankena, M. 1979. *Urban transportation economics: Theory and Canadian policy*. London: Butterworths.

Fujita, M., and J.-F. Thisse. 2002. *Economics of agglomeration; Cities, industrial location and regional growth*. Cambridge: Cambridge University Press.

Gordon, P., and H. Richardson. 2001. The geography of transportation and land use. In R. Holcombe and S. Stanley, eds., *Smarter growth: Market-based strategies for land use planning in the 21st century*. Westport, Conn.: Greenwood.

Hamilton, B., and A. Roell. 1982. Wasteful commuting. *Journal of Political Economy* 90:1035–1053.

Hauer, E. 2000. Safety in geometric design standards. In Krammas, R. and Brilon, W., eds. *Proceedings of the Second International Symposium of Highway Geometric Design* (pp. 11–35). Mainz, Germany: Transportation Research Board.

Ioannou, P., and A. Bose. 2002. Automated vehicle control. In R. Hall (ed.), *Handbook of transportation science* (2nd ed.). Norwell, Mass.: Kluwer.

Kanemoto, Y. 1976. Optimum, market and second-best land use patterns in a von Thunen city with congestion. *Regional Science and Urban Economics* 61:23–32.

Kanemoto, Y. 1980. *Theories of urban externalities*. Amsterdam: North-Holland.

Kanemoto, Y. 1999. Design of public procurement system. *Government Auditing Review* 6:3–20.

Laffont, J.-J., and J. Tirole. 1993. *A theory of incentives in procurement and regulation*. Cambridge, Mass.: MIT Press.

Lévy-Lambert, H. 1968. Tarification des services à qualité variable: Application aux péages de circulation. *Econometrica* 36:564–574.

Lighthill, M., and G. Whitman. 1955. On kinematic waves, II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society* A 229:317–345.

Liu, L., and J. McDonald. 1998. Efficient congestion tolls in the presence of unpriced congestion: A peak and off-peak simulation model. *Journal of Urban Economics* 44:352–366.

Marchand, M. 1968. A note on optimal tolls in an imperfect environment. *Econometrica* 36:575–581.

Mayeres, I., and S. Proost. 1997. Optimal tax and investment rules for congestion type of externalities. *Scandinavian Journal of Economics* 99:261–279.

Mohring, H. 1972. Optimization and scale economies in urban bus transportation. *American Economic Review* 62:591–604.

Mohring, H. 1976. *Transportation economics*. Cambridge, Mass.: Ballinger.

Mohring, H. 2001. Congestion. In J. Gómez-Ibáñez, B. Tye, and C. Winston, eds., *Essays in transportation economics and policy: A handbook in honor of John R. Meyer*. Washington, D.C.: Brookings Institution.

Mohring, H., and M. Harwitz. 1962. *Highway benefits: An analytical framework*. Evanston, Ill.: Northwestern University Press.

Newbery, D. 1988. Road damage externalities and road user charges. *Econometrica* 56:295–316.

Parry, I., and A. Bento. 2001. Revenue recycling and the welfare effects of road pricing. *Scandinavian Journal of Economics* 103:645–671.

Richards, P. 1956. Shockwaves on the highway. *Operations Research* 4:42–61.

Rotemberg, J. 1985. The efficiency of equilibrium traffic flows. *Journal of Public Economics* 26:191–205.

Santos, G., D. Newbery, and L. Rojey. 2001. *Traffic management with congestion charging in different English towns*. Paper presented at the Ninth World Conference on Transport Research, Seoul, Korea, July 22–27.

Sherman, R. 1971. Congestion interdependence and urban transit fares. *Econometrica* 39:565–576.

Small, K., and S. Song. 1992. "Wasteful" commuting: A resolution. *Journal of Political Economy* 100:888–898.

Small, K., C. Winston, and C. Evans. 1989. *Road work: A new highway pricing and investment policy*. Washington, D.C.: Brookings Institution.

Small, K., and J. Yan. 1999. *The value of "value pricing" of roads: Second-best pricing and product differentiation*. Working paper 99-00-02. School of Social Sciences, Irvine, California. Available on-line: ⟨http://ideas.repec.org/p/fth/calirv/99-00-02.html⟩.

Solow, R. 1972. Congestion, density and the use of land in transportation. *Swedish Journal of Economics* 74:161–173.

Solow, R., and W. Vickrey. 1971. Land use in a long, narrow city. *Journal of Economic Theory* 3:430–437.

Strotz, R. 1965. Urban transportation parables. In J. Margolis, ed., *The public economy of urban communities*. Washington, D.C.: Resources for the Future.

Syski, R. 1986. *Introduction to congestion theory in telephone systems*. Amsterdam: North-Holland.

Verhoef, E. 2002. Second-best congestion pricing in general networks: Heuristic algorithms for finding second-best optimal toll levels and toll points. *Transportation Research* 36B:707–729.

Verhoef, E., R. Emmerink, P. Nijkamp, and P. Rietveld. 1996. Information in road networks with multiple origin-destination pairs. Discussion paper TI 95-198, Tinbergen Institute, Amsterdam-Rotterdam, Netherlands.

Verhoef, E., P. Nijkamp, and P. Rietveld. 1996. Second-best congestion pricing, the case of an unpriced alternative. *Journal of Urban Economics* 40:279–302.

Verhoef, E. T., P. Nijkamp, and P. Rietveld. 1997. The social feasibility of road pricing: A case study for the Ranstaad Area. *Journal of Transportation Economics and Policy* 31:255–276.

Verhoef, E., J. Rouwendal, and P. Rietveld. 1999. Congestion caused by speed differences. *Journal of Urban Economics* 45:533–556.

Verhoef, E., and K. Small. 1999. Product differentiation on roads: Second-best congestion pricing with heterogeneity under public and private ownership. Discussion paper TI 99-066/3, Tinbergen Institute, Amsterdam-Rotterdam, Netherlands. Available on-line at ⟨http://www.uctc.net/papers/656.pdf⟩.

Vickrey, W. 1955. A proposal for revising New York's subway fare structure. *Journal of the Operations Research Society of America* 3:38–68.

Vickrey, W. 1959. *Statement on the pricing of urban street use. Hearings*. U.S. Congress Joint Committee on Metropolitan Washington Problems. 11 November 1959.

Vickrey, W. 1963. Pricing in urban and suburban transport. *American Economic Review* 53:452–465.

Vickrey, W. 1968. Automobile accidents, tort law, externalities, and insurance: An economist's critique. *Safety: Law and Contemporary Problems* 33:464–487.

Vickrey, W. 1969. Congestion theory and transport investment. *American Economic Review Proceedings* 59:251–260.

Walters, A. 1961. The theory of measurement of private and social cost of highway congestion. *Econometica* 29:676–699.

Wheaton, W. 1978. Price-induced distortions in urban highway investment. *Bell Journal of Economics* 9:622–632.

White, M. 1988. Urban commuting journeys are not wasteful. *Journal of Political Economy* 96:1097–1110.

Wilson, J. 1983. Optimal road capacity in the presence of unpriced congestion. *Journal of Urban Economics* 13:337–357.