UC Irvine
Working Paper Series

Title
Probing a Traffic Congestion Controversy: Density and Flow Scrutinized

Permalink
https://escholarship.org/uc/item/4vr3m3fm

Author
Ohta, Hiroshi

Publication Date
1999-08-17

Probing a Traffic Congestion Controversy: Density and Flow Scrutinized

Hiroshi Ohta

Department of Economics
University of California, Irvine
Irvine, California 92697-5100, USA

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

8/17/99

Probing a Traffic Congestion Controversy: Density and Flow Scrutinized

By Hiroshi Ohta*

Abstract: Probing a die-hard traffic congestion controversy, this paper scrutinizes two key variables, density and flow, under equilibrium *versus* optimal states. We find that optimization requires flow in equilibrium to decrease under mild congestion, but flow must increase under hyper-congestion. However, while under hyper-congestion flow should be increased by decreasing density, a mild congestion requires flow to be decreased by decreasing density. Thus, inflow of vehicles should always be discouraged to either increase or decrease flow of vehicles for economic efficiency. Moreover, even when optimal policy, given traffic demand, requires flow to increase, the optimal flow itself must decrease eventually as demand increases beyond a critical level.

[R41, R42]

Communications to:

Hiroshi Ohta
Department of Economics
University of California, Irvine
Irvine, CA 92697-5100
E-mail: ota@uci.edu

After April 1, 2,000:

Hiroshi Ohta, SIPEB
Aoyama Gakuin University
4-4-25 Shibuya
Shibuya-ku, Tokyo, Japan 150-8366

Introduction

The present paper attempts to resolve a certain die-hard controversy in the traffic congestion literature. The controversy addresses in part to what constitutes the traffic demand and supply (or cost). Thus, for example, Small and Chu (1999) contend that consumers choose neither "the traffic flow given price" (quoting Evans (1992)) nor the density of vehicles as it is "a stock variable which depends on past inflows." (Footnote 1) They propose instead that the traffic demand be represented by "inflow" and the supply as "outflow". However, they do not seem to be fully content with this definition, either. For it requires such homogeneity conditions as "uniform vehicle density" along a "uniform expressway' even when demand is "changing very rapidly". Such homogeneity requirement appears to be hardly acceptable to them; and they call for a dynamic treatment of demand and supply accordingly.

This paper contends that neither homogeneity conditions nor static aspects of the basic traffic model used are responsible for, or in any way related to, the source of confusions in the congestion controversy. Indeed the Small-Chu definition of traffic demand and supply is perfectly legitimate and general (without an adjective 'static') insofar as "inflow" refers to the individuals (and their aggregate number) willing to get on a given road, and "outflow" the vehicles to be processed. The former depends on the time cost, which in turn depends on the latter. The traffic demand thus defined may be interpreted as a derived one for an intermediate input in order to complete a journey as a final demand. (This need not exclude a fun driving, however.) The individual demand depending on price (or time cost), it can be best conceived of as perfectly inelastic over prices below one's reservation price, which itself may vary among different individuals. This gets us back to both the Else-Kawashima-Evans and Walters definitions of traffic demand in the sense that both can be derived from the Small-Chu (static) definition. Individual willingness and decision making do yield aggregate outcomes in terms of either flow or density as either one is related to the other via a vehicle's speed.

Given the road capacity, the time needed to process or accommodate a given flow or density can readily be defined as the average cost imposed upon each and every passenger on the road that yields that density or flow. This average cost is thus defined as a function of either density or flow. Note in this connection that the related marginal cost concept requires a careful scrutiny. The

relevant concept here is the additional cost incurred by an additional number (density) of vehicles admitted to the road, which is to be distinguished from an additional flow of vehicles. This will be referred to as Else-Kawashima's MC to be distinguished from Walters' MC in the present paper. An intriguing feature of the former curve is that any point along its downward-sloping part can be fully stable and meaningful as is its counterpart downward-sloping AC, *contra* prevailing interpretations.(Footnote 2) Given this, what in the literature is referred to as a "demand spike" or "hypercongestion" can readily be incorporated into a simple static model of traffic flow and density.

The crucial relation to be observed in this connection is that flow F is defined as a product of speed S and density D: $F = SD$. The speed, however, is related in turn to density such that $S = f(D), f' < 0$ so that $F = Sf^{-1}(S) = g(S)$. Thus, flow is a function, but not a one-to-one function, of speed inasmuch as g' has both negative and positive signs over a given domain of S. This is the speed-flow relation that Verhoef (1999) depicts as his Figure 1 (p. 347).(Footnote 3) This strictly nonlinear relation between speed and density in turn yields a correspondingly nonlinear relation between the average traveling time and flow. It is over this nonlinear domain that the aforementioned demand spikes or hypercongestion defined over the density domain can be superimposed. The outcome will provide the two key variables of particular importance to both economists and engineers: flow and density. The two are related to each other and the other variables in a simple tractable analysis.

Our concern is more economic than engineering so that the flow dynamics *per se* is *non sequitur*. In this connection note that variables used in economics are normally homogeneous.(Footnote 4) The same homogeneity applies to density and flow in dealing with economic aspects of road congestion. Various alternative distributions including bottlenecks over a given road could be treated equivalently insofar as the time needed to pass the link remains the same. Any probabilistic distribution can be analytically transformed to any other form including a uniform distribution for certain well-defined analytical purposes such as finding equilibrium time needed for a vehicle to pass through the road. As one intends to consume its service, the road could certainly get congested. It also could yield a wave of congestion. But this does not require the road (or its service) to be treated as a heterogeneous good. A higher price of a rose will not make it a different good. A rose is a rose regardless of its price; and so is a road regardless of congestion (= time cost the passenger pays), albeit a subtle difference between the two goods does exist in a different context.(Footnote 5)

3

So, the present paper assumes a uniform vehicle density over a given homogeneous highway. But the degree of uniform density itself depends on varying numbers of passengers who may decide to get on the road. Their decision in turn depends on time needed to pass through the highway, however. The uniform density is in this sense an endogenous variable, which depends on travel time. The greater the time needed, the lesser the number of vehicles willingly getting on the road. This is the primitive relation of the present model, and referred to as the traffic demand in the literature. How it may be formed is *non sequitur*, instead it is is a primitive relation. While density is a primitive term flow is a derived term defined here by $F = SD$. For this reason density is a more relevant concept than flow is within the confines of the present model. The greater the density, however, the greater the time needed to process the number of vehicles intending to make a trip. This relationship is referred to as the traffic supply or more precisely average cost curve in the literature.

In what follows Section I sets forth the basic assumptions and related system of equations of our model. Section II derives the basic traffic equilibrium conditions under free entry over the (time cost, traffic density) quadrant. Section III in turn derives the seemingly related flow equilibrium conditions over the (time cost, traffic flow) quadrant. Section IV shows how unique density equilibrium is related or unrelated to multiple flow equilibria. Section V presents fourconsecutive diagrams to probe deeper into the subtle nature of the traffic congestion problem. In particular it shows how varying traffic demands may generate intriguing relationships between equilibrium *versus* optimal magnitudes of flow as distinguished from those of density. This section in effect summarizes our basic findings and related interpretations, thereby resolving the die-hard controversy on traffic congestion and related policy questions. Section VI concludes the paper.

I. The Basic Model and Assumptions

Consider a single path, link or loop of a given length, lanes and width: a road of a given capacity. And assume the following.

a) Whether or not one decides to get on the road (for purposes of nobody's business) depends upon the travel time t to be spent on the road; and the number of passengers willing to get on the road (during a given period of time such as a rush hour), called traffic density D, decreases linearly (for simplicity) as this travel time t increases.(Footnote 6)

4

b) The travel time t (needed to pass through the road) is inversely related to the average speed S at which a vehicle travels. (The travel time is related to speed in a manner more definitional than hypothetical, however.)

c) Speed S at which a vehicle runs decreases linearly as traffic density D increases.

Assumptions above yield the following basic system of equations.

(1) $t = a - bD$

(2) $t = \frac{\alpha}{S}$

(3) $S = c - dD$

Equation (1) above represents an inverse demand for traffic under consideration (as per Assumption a)).(Footnote 7) Traffic demand may henceforth be referred to as social marginal benefit SMB, too. Equation (2) shows how travel time is related to speed (as per Assumption b)), given the length of the road α. Thus, for example, if $\alpha = 100$ miles. Then, speed S of 100 miles/h implies $t = 1$ hour needed to pass through the link. If $S = 50$, then two hours are needed. Equation (3) being based on Assumption c) shows that speed decreases as density increases. Here the linearity of speed is just for simplicity.

Derivable from these technical/definitional relations (2) and (3) is the traffic supply or social average cost SAC:

(4) $t = \frac{\alpha}{c - dD}$

This equation shows how travel time t is related to traffic density D. This relation is referred to as the social average cost SAC curve as t represents an individual passenger's time input needed per journey for any given traffic density. The traffic equilibrium under conditions of free entry is then given by (1) and (4) at the intersection of these two curves: demand for and supply of traffic density.

Derivable from (4) is the social marginal cost SMC:

5) $\frac{d(tD)}{dD} = \frac{\alpha c}{(c - dD)^2}$

As one additional vehicle enters and congests the road, raising the travel time by t', the aggregate additional time forgone must be the sum of the entrant's own time forgone t and everybody else's time increment t', viz., $t + t'D$. A congesting action by one person deprives not only himself but everybody else on the road of his/her travel time. It goes without saying that an additional increase in 'inflow' at the toll gate would not directly affect the travel time of

5

the outflowing passengers right ahead. Crude realism aside, the model requires as if an additional density due to an additional inflow were spread evenly over the entire highway immediately.

II. The Baisc Traffic Equilibrium Conditions in Terms of Density

Note Equations (1) and (4) being equated yields competitive equilibrium in which the number of vehicles on the highway under consideration and the time needed to pass through the highway will be determined. For example, given $a = 100$ (hours), $b = 1/1000$ (hours), $\alpha = 100$ (miles), $c = 200$ (miles/h), $d = 1/500$ (miles/h), solution to equilibrium $(D*, t*)$ is $(92929, 7.07)$.(Footnote 8)

Interpretation: Suppose this highway is 100 miles long. Then automobiles will be packed like sardines, every one being approximately 2 meters apart from every one else. It will then take a passenger 7 hours to pass through, say, Tokyo's *Shuto Kosoku* (*alias*, Capital Stroke). Such equilibrium outcome is stable when people are free to enter the freeway. To see this suppose that a shorter time were needed to pass through, then a greater number of cars will be enticed to enter the freeway (as per equation (1)), but only a lesser number of cars will allow the assumed lower travel/time cost given by equation (4). This condition points to nothing but excess demand for road service: more cars than technically processible, as long as time cost remains so low. But excess demand implies a more 'inflow' (of passengers) than 'outflow', which in turn tends to cause the time cost to increase. If and when, on the other hand, time cost happens to be too high, then just an opposite process will take place.

So, traffic equilibrium is unique and stable. Socially optimal equilibrium (in the Pigouvian sense) is also obtainable by equating marginal benefit of travel represented by (1) and marginal cost thereof represented by (5). Note that there is no multiple equilibria, much less unstable equilibria here. However, this basic model can be shown to yield a seeming multi-equilibrium result to follow below.

III. The Derived Demand for (and Cost of) Traffic Flow and the Related Equilibria

We are now in position to define traffic flow F in terms of traffic density D and speed S:

(6) $F = SD$

The F in turn is related to D alone in light of this definition (6) and technical condition (3):

6

(7) $F = SD = (c - dD)D$

According to Kawashima (1990) the traffic flow is defined as "the number of vehicles to pass through the road" during a specific period of time. Clearly it depends on traffic density and the speed at which the vehicles pass through the road under consideration. Thus, our definition of the traffic flow is more specific than Kawashima's in incorporating the endogenized speed in it, but is otherwise equivalent.

Substituting (1) in (7) yields:

(8) $F = (c - d(\frac{a-t}{b}))(\frac{a-t}{b})$

This is the derived quadratic demand for 'traffic flow', derived from the linear demand for traffic 'density'. Note how one is deformed as it is related to the other. Note moreover that the derived demand is no longer linear as in (1), but instead it is quadratic.

The traffic cost SAC can also be defined in terms of traffic flow, and is derivable from (4) and (7):

(9) $F = \frac{\alpha(ct-\alpha)}{dt^2}$

Equating (8) and (9) yields the following quartic equation and solution:

(10) $(c - d(\frac{a-t}{b}))(\frac{a-t}{b}) = \frac{\alpha(ct-\alpha)}{dt^2}$,

$t = \frac{1}{2}\left(a + \sqrt{(a^2 - 4b\alpha/d)}\right), \frac{1}{2}\left(a - \sqrt{(a^2 - 4b\alpha/d)}\right), \frac{1}{2}\left(a - cb/d + \sqrt{a - cb/d)^2 + 4b\alpha/d}\right),$

or $\frac{1}{2}\left(a - cb/d - \sqrt{a - cb/d)^2 + 4b\alpha/d}\right)$

The parameters assumed in the previous section yield the following quartic equation and solution:

(10)' $-t^4 + 100t^3 - 5000t + 2500 = 0 : t = 5\sqrt{2}, -5\sqrt{2}, 50 + 35\sqrt{2},$ or $50 - 35\sqrt{2}$

IV. The Equilibrium Density Versus the Flow Equilibria

Of the four solutions obtained above note that the first one alone coincides with the unique solution obtained above by equationg (1) and (4). The second, negative solution is out of question, being irrelevant, and will be omitted from our considerations below. The remaining two solutions warrant probing, however, in comparison to the first solution. Note in this connection that the

7

first solution t value being substituted in equation (1) and (4) yields the same solution value for D. Thus,

i): $t = 7.0711 = 100 - D/1000$: $D = 92929$.
i)': $t = 7.0711 = \frac{100}{200 - D/500}$: $D = 92929$.

The other t values, being substituted in (1) and (4), yield distinctively different D values, respectively. Thus,

ii): $t = 99.497 = 100 - D/1000$: $D = 502.53$
ii)': $t = 99.497 = \frac{100}{200 - D/500}$: $D = 99497$.
iii): $t = .50253 = 100 - D/1000$: $D = 99497$.
iii)': $t = .50253 = \frac{100}{200 - D/500}$: $D = 503.45$

Note that different D values for ii)' and iii)', respectively, are obtained for obvious reasons underscoring disequilibria. Nevertheless these same t values being substituted in (8) and (9), respectively, yield unique F values as follows.

i): $t = 7.0711$ in $F = 1000(200 - 2(100 - t))(100 - t)$: $F = 1.3142 \times 10^6$
i)": $t = 7.0711$ in $F = \frac{50000(200t - 100)}{t^2}$: $F = 1.3142 \times 10^6$

ii): $t = 99.497$ in $F = 1000(200 - 2(100 - t))(100 - t)$: $F = 1.00 \times 10^5$
ii)": $t = 99.497$ in $F = \frac{50000(200t - 100)}{t^2}$: $F = 1.00 \times 10^5$

iii): $t = .50253$ in $F = 1000(200 - 2(100 - t))(100 - t)$: $F = 1.00 \times 10^5$
iii)": $t = .50253$ in $F = \frac{50000(200t - 100)}{t^2}$: $F = 1.00 \times 10^5$

We thus find a unique t value (in i)) which yields competitive equilibrium for both variables D and F. We also find certain particular t values (in ii) and iii)) which yield different Ds, thereby implying underscoring *dis*equilibrium between demand and supply, but nevertheless they yield the same Fs, as if implying equilibrium between demand and supply, which is not.

An immediate question is: How can these apparent inconsistencies be accounted for? In answer they can be readily resolved if only we prperly appreciate the two seemingly similar, but in fact entirely different key variables: density and flow. Note in this connection that the density is the number of cars running on a given highway and the flow the density multiplied by speed, i.e., $F = DS$. At a first glance, therefore, it might appear that flow is a more relevant concept than density is. In fact, opposite is the case. We consider the density is a concept more relevant to the present model of traffic demand than is the flow. This is because the density is the primitive term, but the flow is not, being defined

in terms of density and also speed as another primitive term. This implies that there is not a one-to-one correspondence between F and D. Thus, for example, it is possible that the same flow magnitude can be obtained from either combination of a high speed and a low density or a low speed and a high density. It is this peperty of F that yields a *nonmonotonic* relationship between F and t derivable from a *monotonic* relationship of either 1) or 4). No wonder that the former yields a multi-equilibrium solution, even if the latter yields a unique equilibrium.

This calls for a careful interpretation of the multi-equilibrium points *vis-a-vis* unique equilibrium point. Note initially that the latter point coresponds to one, but not the rest, of the multi-equilibrium points. The rest in fact corresponds to disequilibrium points over the (t, D) quadrant. In a related vein note secondly that excess demand, say, over D can nevertheless be transformed to equilibrium over F because a low t below equilibrium over D creates a high density demand along the demand curve 1) on one hand, and a low density supply along the supply curve 4) on the other hand. Correspondingly the former (high density) yields a low speed, and the latter a high speed, *via* equation 3), and *vice versa* when t is above equilibrium. The upshot in any case is equilibrium $F(= SD)$ for disequilibrium Ds. More important, note thirdly that the seemingly stable (or unstable) equilibria over the (t, F) quadrant are not really stable (or unstable) in light of the underscoring conditions over (t, D). To appreciate this observe Case i) which yields the only stable and unique equilibrium over (t, D), and hence also over (t, F). Note, however, that for any t below this equilibrium point a seeming excess supply of F exists, thereby seemingly implying an immanent decline in t, and toward the seemingly stable equilibrium point given by Case iii. But neither is true in light of the unique equilibrium identified by Case I over (t, D). And finally, the relevant MC defined as a monotone function as (5) can be substituted in (7) to obtain a derived MC $(= (tD)')$ over the $((tD)', F)$ quadrant which does have a backward bending portion. Thus, letting $\tau = (tD)'$ and noting $D = \frac{c \Box \sqrt{c^2 - 4dF}}{2d}$,

(11) $\frac{d(tD)}{dD} \equiv \tau = t + t'D = \frac{\alpha c}{(c - dD)^2} = \frac{4\alpha c}{(c \mp \sqrt{c^2 - 4dF})^2}$

This MC is to be distinguished from the first partial derivative of (tF) with respect to F, i.e.,

(12) $\frac{d(tF)}{dF} \equiv \tau* = tD(F) + (\partial t/\partial D)(\partial D/\partial F)F = \Box \frac{\alpha}{\sqrt{c^2 - 4dF}}$,

which does not have a backward bending portion. (See Appendix for more

details for the relationship between τ and $\tau*$.)

Note, however, that this does not make the latter relation (12) to be a more relevant or even valid representation of MC than equation (11). The reason, again, is that the relevant variable is D, and the relevant optimization calculus is to be conducted over this variable. The derived tF does not directly represent the total (aggregate) cost tD incurred by the community. If the average form of the latter changes when transformed to the former, say from a monotone to a nonmonotone, then it is no wonder that the monotone MC transfoms to a nonmonotone curve, and a backward portion of the transformed MC in F should be interpreted to correspond to a well-defined portion of the primitive MC. Related to this is the observation that a seeming excess demand (or supply) in F does not *ipso facto* correspond to that in D.

V. Diagrammatic Interpretation of Relations Between Key Variables

The relationships among the key variables are illustrated by Figures 1, 2, 3 and 4 below. The so-called back-to-back diagrams of Figures 1 and 2 illustrate how the demand and/or cost conditions depicted in the first quadrant in terms of traffic flow may be derived from and related to those in terms of density depicted in the second quadrant.

Figure 1 confines itself to cost conditions and shows how AC and MC curves in the first quadrant are related to those in the second quadrant via the purely technological relationship between flow and density depicted in the fourth quadrant (based on equation (7)). The basic message of this Figure is as follows. Any increase in density tends to slow the vehicles on the road, thereby increasing their travel time (cost) monotonically. But as long as the density remains to be low enough, any such increase in density will continue to increase flow, too, despite a decreasing speed. The upshot is that the travel time increases (decreases) with flow when density is low enough. However, as density becomes high enough to reduce the speed of vehicles large enough, the flow can no longer continue to increase, but instead will start to decline eventually. Thus, beyond a certain critical density ($= c/2d$) it becomes inversely related to flow: the higher the density, the lower the flow. The backward-bending portion of not only AC, but also MC thus reflects the relevant (high cost) portion of the cost curves on the second quadrant.

It warrants a special warning, however, that the MC depicted in the first quadrant represents the additional time costs due to an additional unit of den-

10

sity, not flow, despite its representation on the (t, F) quadrant. If the flow is treated as a primitive term, then the relevant MC needs to be defined as marginal cost of an additional flow rather than density. This is to be distinguished from the MC in terms of an additional density. The former is Walter's MC and the latter Else-Kawashima's MC, and they are entirely different animals even though both can be represented on the same (t, F) quadrant. The latter has a backward-bending portion, but the former does not. Instead the flow MC derivable from the backward-bending portion of AC can be shown to be strictly negative. This leads to an oft-alleged interpretation that the backward-bending portion of AC has little or no economic relevancy.(Footnote 9)

Underscoring such interpretation is a nothion that the flow, not density, ought to be the primitive concept. A related notion is that what passengers demand is not "being on the road," but instead "completing a trip" (Verhoef (1999)). But this ignores the fact that these two objects of demand are indeed "inseparable" (Evans (1992, p.212)). Being as if two sides of a coin, they are related to each other. One needs to get on the road in order to complete his trip. This inseparable decision depends on travel time either conjectured by or known to him, however. Thus suppose that travel time (or price to pay) were expected by many to be high enough. Then only a handful of them will get on the road while the rest won't, and both the density and flow tend to be low enough relative to the quantities technically producible. This condition of excess supply in effect will make actual travel time strictly shorter than expected. If, on the contrary, travel time happened to be (or signaled) low enough, then more passengers than the road can process would be enticed to get on the road.(Footnote 10) In any case, expected travel time will eventually be adjusted so that equilibrium density and related flow will be determined, given demand and cost parameters, when stationary state is obtained.

Figure 2 in turn shows how demand conditions may be superimposed upon the same quadrants 1 and 2 of Figure 1. Two alternative demand curves are depicted as linear lines, representing a large population (solid line) and a small one (dotted line), on the second quadrant. It is to be noted that the relevant part of the demand in the first quadrant derived from the former demand is sloping upward (solid) while the one derived from the latter is sloping downward (dotted).(Footnote 11) Note also that optimal flow is greater than equilibrium flow when demand (population) is large enough, but opposite is the case when demand is small enough. In any case, optimal density is necessarily smaller than equilibrium density. This seems to call for an extra care in compiling data

to formulate any empirically testable hypothesis. In a corresponding analytical vein we note that a high (hyper) demand and the related low flow derived in the first quadrant represent the state of "hypercongestion" (Small and Chu), "peak demand" (Verhoef) and "demand spike" (Arnott (1990)). A change in demand from a small to a large one may thus be characterized as being either "rapid," "transient" or permanent. In any case, an outcome is analytically the same within the confines of the present model and it can be interpreted as a "response of a nonlinear system" (Arnott) to the assumed change in demand.

Probing deeper into comparative statics on changes in demand parameters for more intermediate than the two extreme cases requires more microscopic observations on limited areas of cost conditions in the first quadrant. Figure 3 spotlights these areas in the first quadrant by omitting the rest and the other quadrants as well. We start with a basic (primitive) demand which is small enough and the relevant (downward-sloping) part of the derived demand is shown as MB_1, Figure 3, its intersection with the AC and MC curves yield competitive and optimal equilibrium at Ec_1 and E^*_1, respectively. Then competitive equilibrium flow at Ec_1 is strictly greater than optimal flow at E^*_1. As the basic demand increases due to, say, population increase, the derived demand shifts out to make a backward-bending curve like MB_2. We use this as a reference point as it identifies the maximum possible flow as a competitive equilibrium flow, too. (Stability conditions are satisfied despite a seeming lack thereof. The equilibrium is stable because it is based on the stable MR=MC equilibrium conditions over the (t, D) quadrant.) Note that in contrast to the MB_1, the relevant part of the MB_2 slopes downward and its intersection with the MC curve yields optimal flow which is strictly less than the maximum possible competitive equilibrium flow at Ec_2. As demand increases further, then competitive equilibrium flow starts to decline while the optimal flow continues to increase up to a peak point on the MC curve where the maximum possible flow becomes a social optimum, too. As demand increases still further, not only competitive equilibrium flow, but also optimal flow starts to decrease as the relevant upward sloping part of the MB_4 is located strictly on the left of the MB_3.

Figure 4 summarizes the findings obtained above in analytically more general terms. Note initially that the two monotonically increasing curves illustrate the impact of changes in demand parameters (a or $1/b$) upon equilibrium density D and optimal density D^*, respectively. Both of them are shown to increase monotonically with demand (population, say). Note also that equilibrium den-

sity is always greater than optimal density regardless of the demand parameters. In contrast, while both equilibrium flow F and optimal flow F^* are shown to increase with demand when demand is small enough (as in Region I), they start to decline as demand continue to increases beyond certain critical points. The critical population parameters that yield maximum F and F^* are given by $1/b = c^2/[2d(ac - 2\alpha]$ and $1/b = c^2/[2d(ac - 4\alpha]$, respectively.(Footnote 12) Note that the critical point for the equilibrium flow is located on the left to that for the optimal flow, respectively yielding the same peak flow value of $c^2/4d$. It follows that the relationship between the equilibrium flow and optimal flow is not so simple as that observable between the density counterparts. Thus, while a small enough demand (in terms of a or $1/b$) yields $F > F^*$, a large enough demand yields $F < F^*$ (as in region III.) This implies that there exists a critical demand parameter combination which yields $F = F^*$. But this does not imply that competitive equilibrium can coincide with social optimum. It never can because even in this seemingly optimal case the equilibrium density exceeds optimal density, $D > D^*$. Moreover, in this neighborhood (in Region II) the equilibrium and optimal flows move in opposite directions: when one increases the other decreases. And all these intriguing (perplexing) relations reflect the definition of flow as $F = SD$ and the inverse relationship of S to D. As D increases enough with demand speed S approaches zero, and so does equilibrium flow F and optimal flow F^* as well. A *hipercongestion* is thus not only a feasible equilibrium, but it also can be an optimal equlibrium (as in the rightmost part of Region III.).

VI. Conclusion

This paper ponders a persistent controversy on the nature of traffic congestion and related policy. We assume that the traffic demand is derived from the need for 'completing a trip', which in turn depends on travel time as price to pay for it. The smaller the traveling time, the greater the number of passengers willing to buy the service. This relation constitutes the basic traffic demand condition. The greater the number of passengers, however, the greater the traffic congestion and the greater the traveling time accordingly. This latter relation constituting the basic supply (cost) conditions can yield a unique competitive equilibrium joint with the demand conditions, *ceteris paribus.* Given this, a Pigouvian optimal tax can readily be introduced to obtain optimal congestion. Though simple and primitive these relations may appear, they do not directly reveal how many vehicles may have completed per hour or day in either compet-

itive equilibrium or optimal equilibrium conditions. This is why some writers use flow as a more primitive (relevant) variable than density.

The present paper shows how or in what way flow can become a misleading variable, if it is interpreted as a direct policy variable. However, in light of the definition of flow (as density multiplied by speed, which in turn depends on density), either variable can be managed as a policy variable in principle. The paper has provided a specific warning in doing so inasmuch as the two variables are related to each other in a *non*monotonic fashion. This is because both equilibrium flow and optimal flow keep increasing with demand only up to a certain critical point, beyond which they start to decrease. Moreover, the optimal flow peaks out only after the equilibrium flow does. But equilibrium density keeps increasing monotonically with demand and remains strictly above optimal density. It follows under these particular conditions that when demand is large enough, the optimal policy is to *increase flow by decreasing density*. But when demand is large enough, optimal flow itself could be quite small. Thus, a hypercongestion (Small and Chu) could be an optimal equilibrium, but a maximum density (that Verhoef refers to as a stationary state, p. 350) is not a feasible equilibrium. A maximum flow on the other hand can be a feasible equilibrium, either competitive or optimal.

It goes without saying, however, that neither maximizing nor even increasing flow makes a desired policy in general. Especially when demand is large enough, such a policy as to maximize flow will prove to be a fatal attraction. An arbitrary ration could certainly generate a maximum flow, but also a huge Pigouvian welfare loss.

Footnotes

1) This amounts to taking sides with Evans (1992) who nejects Walters' (1961) definition of traffic demand in terms of flow rather than density on one hand. On the other hand it also negates Else (1981), Kawashima (1990) and Evans (1992) as well who conceive traffic demand in terms of density or inflow rather than flow of vehicles.

2) Verhoef (1999, pp. 349-50) reviews opposing interpretations on which point(s) on the backward-bending AC curve may be stable or unstable, accepts it as a stationary state model, but disapproves the same part for "dynamically consistent equilibria" (pp. 356-7). Underscoring this disapproval are two related conditions: 1) dynamics being represented by a demand spike, hypercongestion, or a rapid change in demand, and 2) flow, instead of density, being treated as

14

a primitive demand. Basically the same interpretation is echoed by Small and Chu (1999, Figure 1 and related explanation.) However, when demand happens to be high enough with great many passengers willing to get on the road, the related flow demand can hardly be represented by a monotone curve like E' in Verhoef's Figures 2 and 3 inasmuch as a high density is tantamount to a low speed and a low flow accordingly. When density demand is large enough the flow demand is required to be inward-bending (Figures 2 and 3 *infra*), intersecting with an upward-bending AC (or MC) to yield a relevant competitive (or optimal) equilibrium.

3) This relation is also confirmed fairly well by its empirical counterpart given by a scattered diagram, Figure 2, that Small and Chu reproduces from Banks (1989).

4) Thus, for example, even when we refer to the diminishing MU of children, they are assumed to be all homogeneous despite the fact that they are all different, the first one being gifted with 20/20 hindsight, the second 20/20 insight, the third 20/20 hindsight, etc. Such reality may be of great interest to the parent or a genetic engineer perhaps, but normally not to an economist. Of greater interest to the economist instead is the fact that of n equally adorable children, the first-come (whichever comes first) will give the parent the highest pleasure, the second a lesser additional pleasure, and so on.

5) The peculiarity of the road in the present context is that it is a 'private good' (having rivalry in consumption,) but nevertheless is 'non-excludable' in its consumption. (Ohta (1993, pp. 5-6).) The related "tragedy of the commons" (Samuelson (1993)) stems from every passenger's move to equalize in equilibrium his time cost with everybody else's, thereby equalizing average, not marginal, time costs.

6) Here the traffic density D is equivalent to the (level of) traffic demand Q defined by T. Kawashima (1990, pp. 325-6) and also equivalent to what Erik Vanhoef (1999) refers to as "the number of users simultaneously on the road" n or his density D, too, under normalized conditions (with his $L = W = 1$).

7) A few notes are warranted on this equation. First, it represents an aggregate demand, comprising from individual passengers' demands, which are perfectly inelastic with varying reservation prices from the highest a down to zero; and the larger the a, the greater the traffic demand on individual basis. Second, the demand slope b represents the community size under consideration; and the larger the population, the smaller the b as $1/b$ represents the number of consumers having an arbitrary reservation price $\in [0, b]$. (The maximum

number of potential passengers is a/b when travel time approaches zero.) And thirdly, the demand in its inverse form represents additional benefits in terms of diminishing reservation prices of potential passengers.

8) Derivation is as follows.

$$a - bD = \frac{\alpha}{c - dD} : D = \frac{1}{2bd}ad + bc - \sqrt{(a^2d^2 - 2adbc + b^2c^2 + 4bd\alpha)} \,,$$
$$D = 92929, t = 100 - 92929/1000 = 7.0711$$

9) For example, McDonald and d'Ouville (1988) propose to interpret density as input and flow as output. It then follows that since any increase in density beyond what they call the "bottleneck density" ($=c/2d$ in our model) causes flow to diminish, the portion of the density-flow relation that yields the backward-bending AC constitutes an uneconomic region in production. Verhoef (1999) in contrast seems to admit this portion as economically relevant as it can be in stationary states. But nevertheless he considers it "dynamically infeasible". A "full-fledged" dynamic model is proposed accordingly to eradicate in effect the backward portion of both AC and MC curves.

10) This condition of excess demand is represented over the flow dimension (in the first quadrant) such that the flow producible depicted along the AC curve exceeds the flow defined along the MB curve, generating a seeming excess supply! Related to this seeming paradox is the fact that a greater density of demand generates a slower speed and a lesser flow than does a greater density supplied with a higher speed. The seeming paradox simply reflects this nonmonotonic transformation of flow from density via the definition of flow as $F = SD$, and speed as a monotonically decreasing function of density $S = S(D)$. The more passengers than processible (excess demand in terms of density) can generate a lesser flow demanded in effect than the flow technically producible (excess supply in terms of derived flow). The competitive process tends to discourage entry, thereby reducing excess demand in density on one hand and excess supply in flow on the other.

11) A few related observations are in order. First, the backward-bending MB (solid curve) on the first quadrant is equivalent to Evans' dd' in his Figure 1C (p. 213), but the present back-to-back diagram demonstrates how an exact lower (or upper) part of the primitive MB on the second quadrant corresponds to an upward-sloping (or downward-sloping) part of the derived MB on the first quadrant. Second, no demand *without* such a bending part is derivable unless the primitive demand on the second quadrant is small enough. Thus,

a monotone downward demand and a peak demand or hypercongestion (*a la* Verhoef and Small and Chu) are an impossible combination. To best appreciate why such a commonsensical relation of downward-sloping demand and upward-sloping supply is invalid, but just opposite is valid, suppose that the travel time happened to be low enough when demand is large enough. Then everybody would wish to enter the road. The density soars accordingly. But the high density implies a low speed, and a low flow accordingly. They certainly do not demand low flow or low speed, but instead only a high density when travel time (expected) is low enough. But an unintended outcome (in the aggregate) is the derived flow that eventually declines as density demanded increases. Just opposite relations apply to the supply side. The travel time decreases with density; and a lower density implies a higher speed, and hence a higher flow (until it approaches its maximum from above). The upshot is a seeming paradox: an upward-sloping demand MB and a downward-sloping supply AC yielding a stable equilibrium (at a white square). Third, what the seeming equilibria on the first quadrant other than those represented by small circles and squares may imply can readily be interpreted by applying horizontal auxiliary lines extended to the second quadrant.

12) The other parameters remaining the same as assumed in the text, the b parameter that yield these maximum flows are computed as $b \doteqdot 505$ and $b \doteqdot 510$, respectively.

References

Arnott, Richard (1990): "Signalized Intersection Queuing Theory and Central Business District Auto Congestion," *Economics Letter*, 33, 197-201.

Banks, James H. (1989): "Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations," *Transporation Research Record*, 1225, 53-60.

Else, P. K. (1981): "A Reformulation of the Theory of Optimal Congestion Taxes," *Journal of Transport Economics and Policy*, 15, 217 - 232.

Kawashima, Tatsuhiko (1990): "....", in New Frontiers in Regional Science, edited by M. Chatterji and R.E. Kuenne, Macmillan.

McDonald, John F. and Edmond L. d'Ouville (1988): Highway Traffic Flow and the 'Uneconomic' Region of Production," *Regional Science and Urban Economics*, 18, 503-509.

Ohta, Hiroshi (1993): "Introduction," (also Chapter 5) in *Does Economic Space Matter?*, ed. by Hiroshi Ohta and Jacques-Francois Thisse. Macmillan (London) and St. Martin's Press (New York), 1-18; 950110.

Samuelson, Paul A. (1993): "Tragedy of the Commons: Efficiency Rents to the Rescue of Free-Road Inefficiencies and Paradoxes," in *Does Economic Space Matter?*, ed. by Hiroshi Ohta and Jacques-Francois Thisse. Macmillan (London) and St. Martin's Press (New York), 71-80.

Small, Kenneth A. and Xuehao Chu (1999): "Hypercongestion," (mimeo).

Verhoef, Erik T. (1999): An Integrated Dynamic Model of Road Traffic Congestion Based on Simple Car-Following Theory," *Regional Science and Urban Economics*, 29, 341-369.

Wlaters, A. A. (1961): The Theory and Measurement of Private and Social Cost of Highway Congestion," *Econometrica*, 29, 676-699.

Vickery

Appendix: Else-Kawashima's MC *versus* Walters' MC

Note initially that both the traffic demand (in the form of equation (1)) and cost (social average cost SAC in the form of (4)) are defined, respectively, as a monotone function of traffic density D, viz., $t = f(D)$, $f' < 0$, and $t = g(D)$, $g' > 0$. However, both the derived traffic demand in terms of traffic flow F in the form of (8) and the corresponding SAC in (9) define traffic flow F as a *non*monotonic function of travel time t. This follows because the traffic density D is related to traffic flow F in a *non*monotonic fashion *via* equation (7). Thus, as D increases F tends to increase initially, but eventually starts to decrease due to congestion. Thus, when t is large enough, the demand for traffic density is small and so is the flow. As t decreases the quantity demanded for D increases, and so does the flow F, initially. But as t continues to decrease and keeps D to increase, the traffic flow F eventually starts to decrease. This is how both a small t and a large t are shown to yield the same small value of F on the derived demand, the first quadrant of Figure 2 (when demand is large enough); two distinct points on this demand curve are thus derived from, and related to,

18

a single t on the monotonic demand curve on the second quadrant of Figure 2. A similar relation of nonmonotonicity *vis-a-vis* monotonicity applies to the dual definition of social average cost SAC as well as marginal cost SMC in terms of either D or F.

The subtle confusion that Kawashima detected in the Walters-Else controversy stems from this dual definition. As Kawashima correctly points out it is important to properly interprete the meaning of the SAC (or SMC) dually defined in terms of either D or F. They are perfectly equivalent in the sense that either one is derivable from the other. But they are related to each other in the following fashion.

(A-1) SAC $(D) = t(D)D/D = t(D), t'(D) > 0$ (in light of (4))

(A-2) SMC $(D) \equiv \frac{d(tD(F))}{dD} \equiv \tau = t(D) + t'(D)D$, *where* $t' > 0$

Note that SAC is defined as total travel time spent by all the passengers divided by all the passengers and SMC as an additional time cost imposed upon all the passengers affected by an additional admission or 'inflow' of a passenger to the road, thereby raising traffic density by dD. Note also that both SAC and SMC are defined in terms of D. Moreover, since $t' > 0$, both SAC and SMC are monotonically increasing functions of D.

These SAC and SMC would thus normally be expressed over the (D, τ) quadrant, but they can be expressed over the (F, τ) quadrant as well by variable transformation. The transformed SAC and SMC in terms of F can be defined, respectively, as:

(A-3) SAC $= t(D(F))$

(A-4) SMC $\equiv \tau = t(D(F)) + t'(D(F))D(F)$, where D is related to F such that $F = S(D)D$.

Clearly, due to the second term on (A-4), the SMC is required to be strictly greater than SAC for all the relevant values of F. Moreover, inasmuch as the SAC is a backward bending curve, so is the SMC which on the first term on the right-hand-side of (A-4) contains the same value as the SAC defined by (A-3).

Distinguished from (A-4) is Walter's SMC that is defined as:

(A-5) SMC $\equiv \frac{d(tF)}{dF} \equiv \tau* = t(D(F)) + t'D'(F)F = t(D(F)) + t'D'(F)S(D)D$, where $S(D)D = F$.

Note that the present SMC measures additional time costs incurred by an additional flow F, not density D. It follows that the last term on the right-

hand side of (A-5) has the terms $D'(F)S(D)$ that (A-4) does not. Note in this connection that Walter's SMC slopes upward like Else-Kawashima's does only when D remains small enough so that $D'(F)$ remains positive. When it becmes large enough, then $D'(F)$ becomes negative and the slope of MC is required to become negative as the marginal curve to the downward sloping AC (in terms of F) *a la* Walters must remain below and downward sloping. Given the parameters assumed in the text, the two SMC curves under consideration can be illustrated by the following two equations and related diagrams (Appendix Figure A-1).

(A-6) $\tau = t + t'D = \frac{4\alpha c}{(c \mp \sqrt{c^2 - 4dF})^2} = \frac{80000}{(200 \mp \sqrt{40000 - .008F})^2}$

(A-7) AC (F): $F = \frac{\alpha(ct-\alpha)}{dt^2} = \frac{50000(200t-100)}{t^2}$; $t = \frac{\alpha}{2dF}(c \, \Box \, \sqrt{c^2 - 4dF})$

SMC (F): $\tau* = \frac{d(tF)}{dF} = \Box \, \frac{\alpha}{\sqrt{c^2 - 4dF}} = \Box \, \frac{100}{\sqrt{40000 - .008F}}$

Note that (A-6) yields a backward-bending MC curve on the first quadrant of Figures 1 and 2 in the text. But while (A-7) also yields a backward-bending AC curve on the same first quadrants as shown, the related SMC is monotonic. Moreover, it is either strictly positive or negative. Figure A-1 illustrates, respectively, a common AC based on (A-7), Else-Kawashima's MC or τ, and Walter's MC or τ^*(omiting a negative solution).
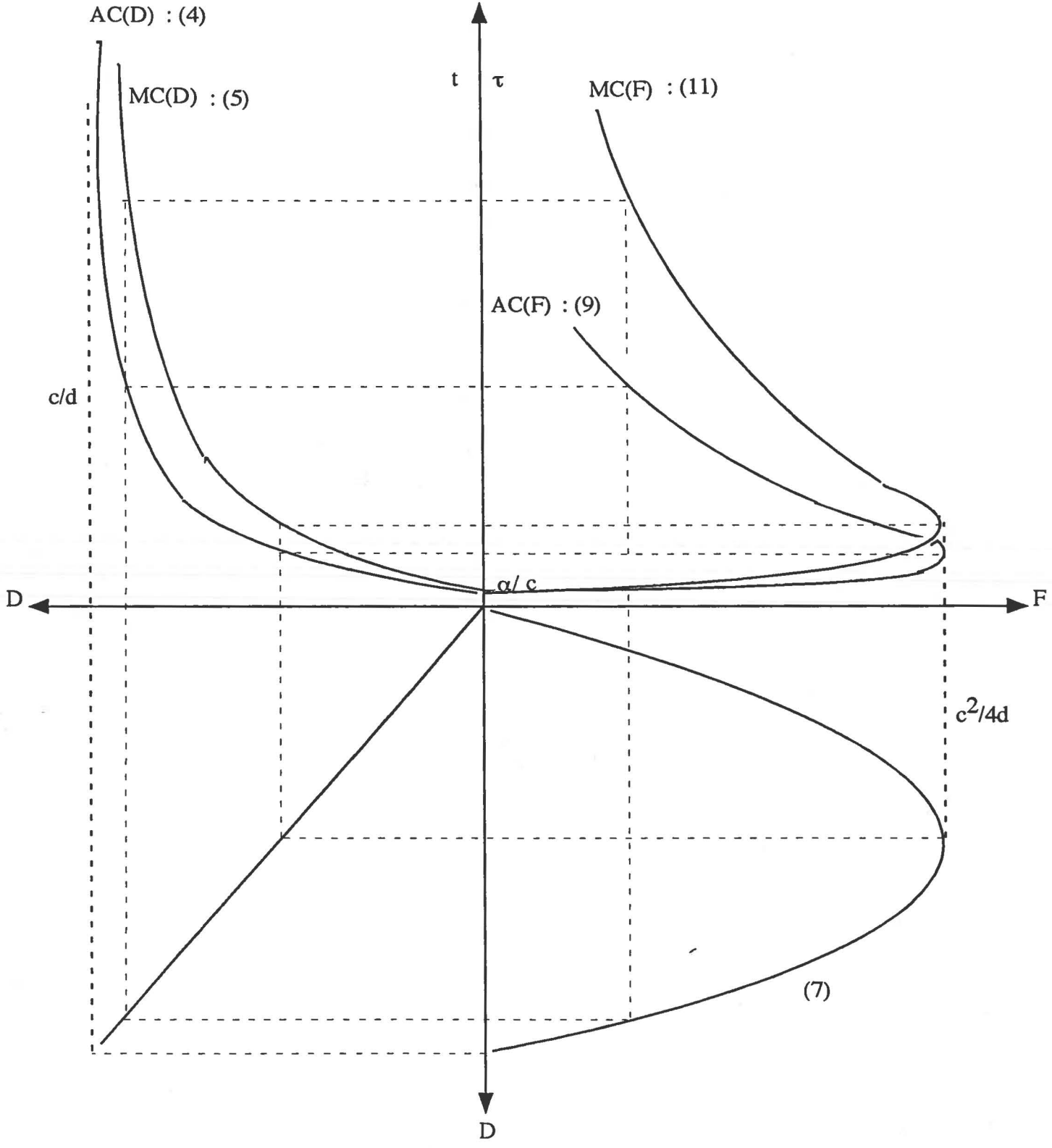
AC(D) : (4)

MC(D) : (5)

t   τ

MC(F) : (11)

AC(F) : (9)

c/d

α/ c

D   F

c²/4d

(7)

D

Figure 1 Backward-bending AC and MC Curves Derivable from the Primitive AC, MC Curves

AC(D) : (4)

MC(D) : (5)

MC(F): (11)
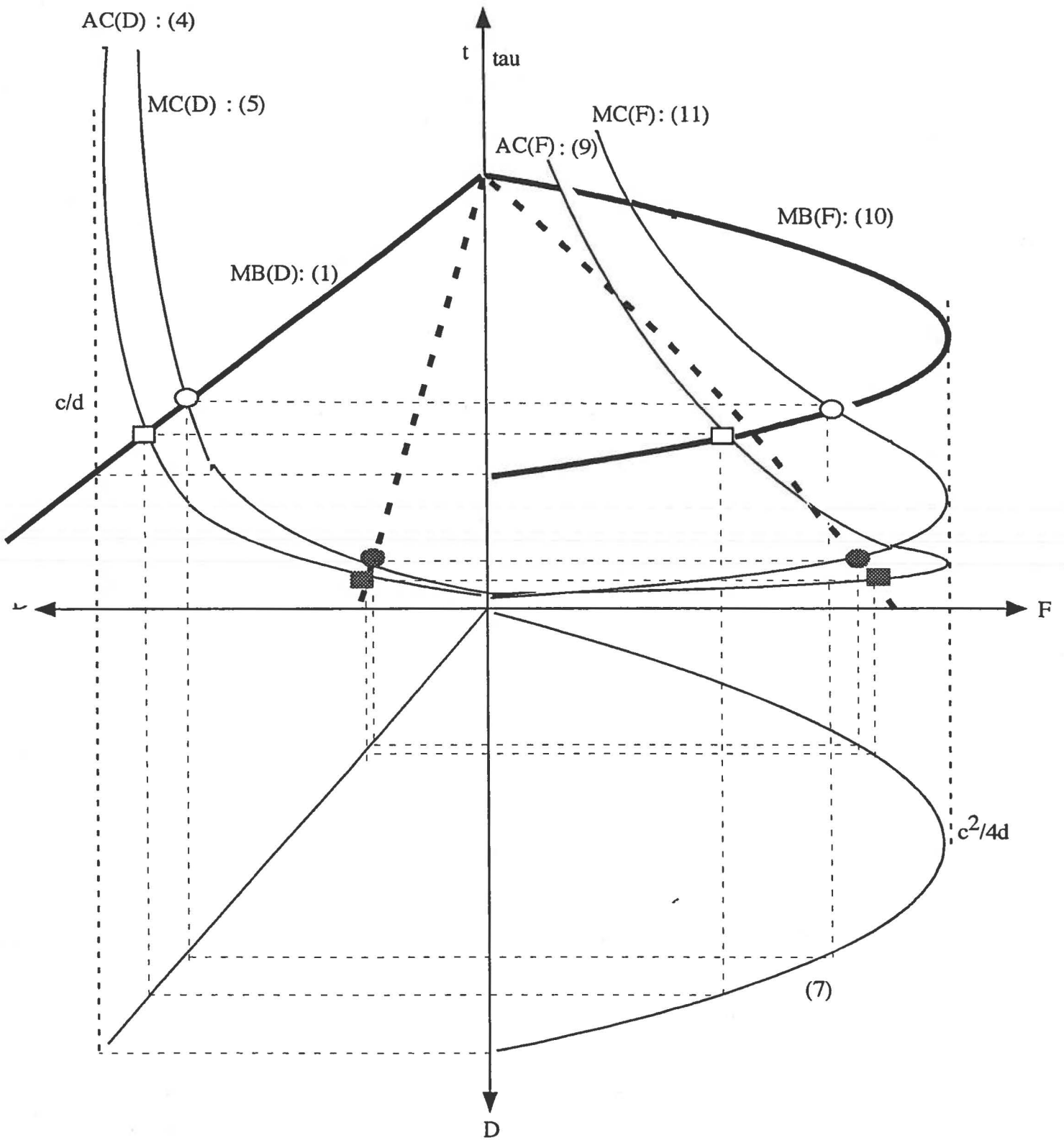
AC(F) : (9)

MB(F): (10)

MB(D): (1)

t tau

c/d

c²/4d

(7)

D

F

Figure 2 Impact of a Demand Spike upon Equilibrium vs. Optimal Density and Flow
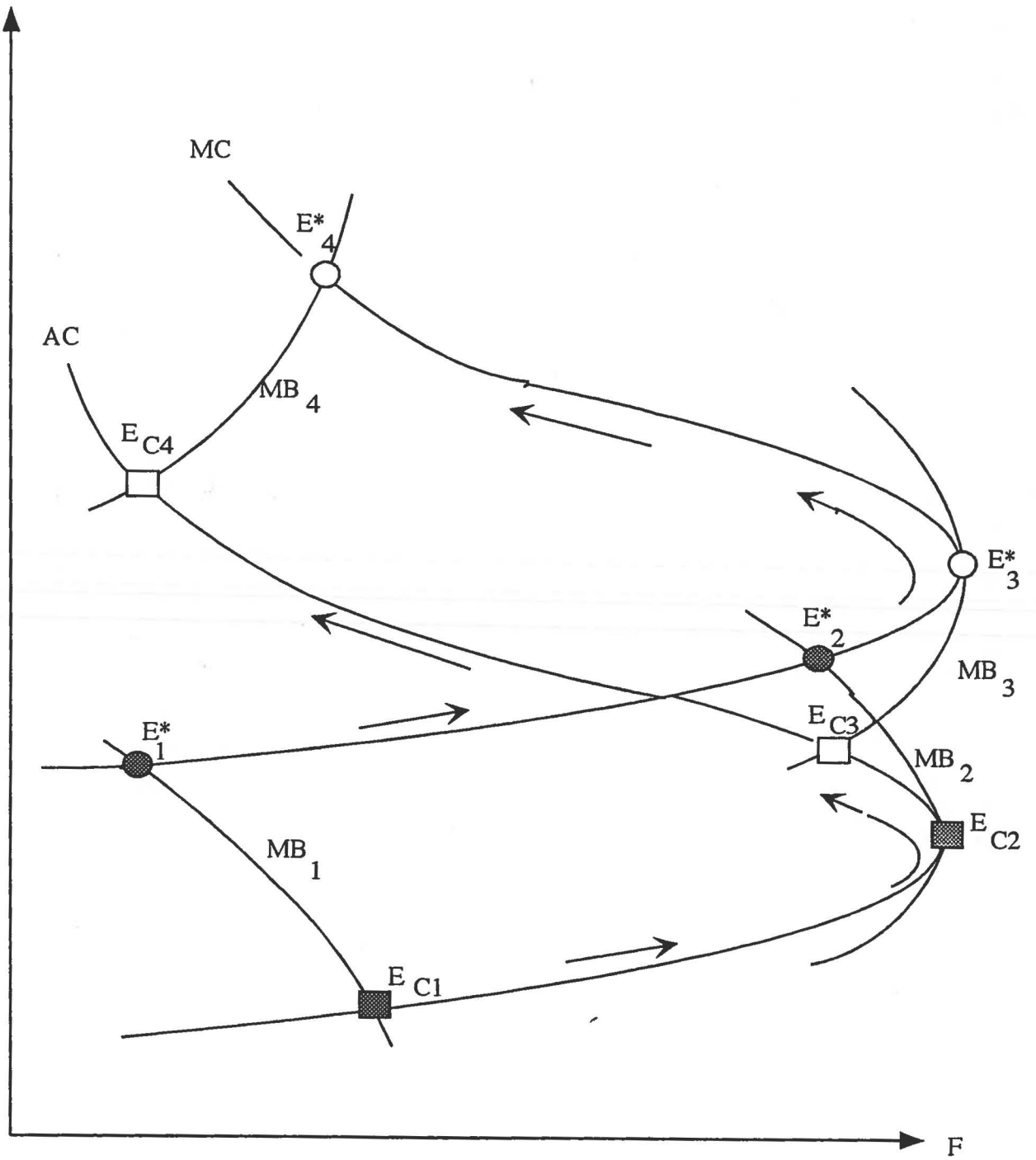
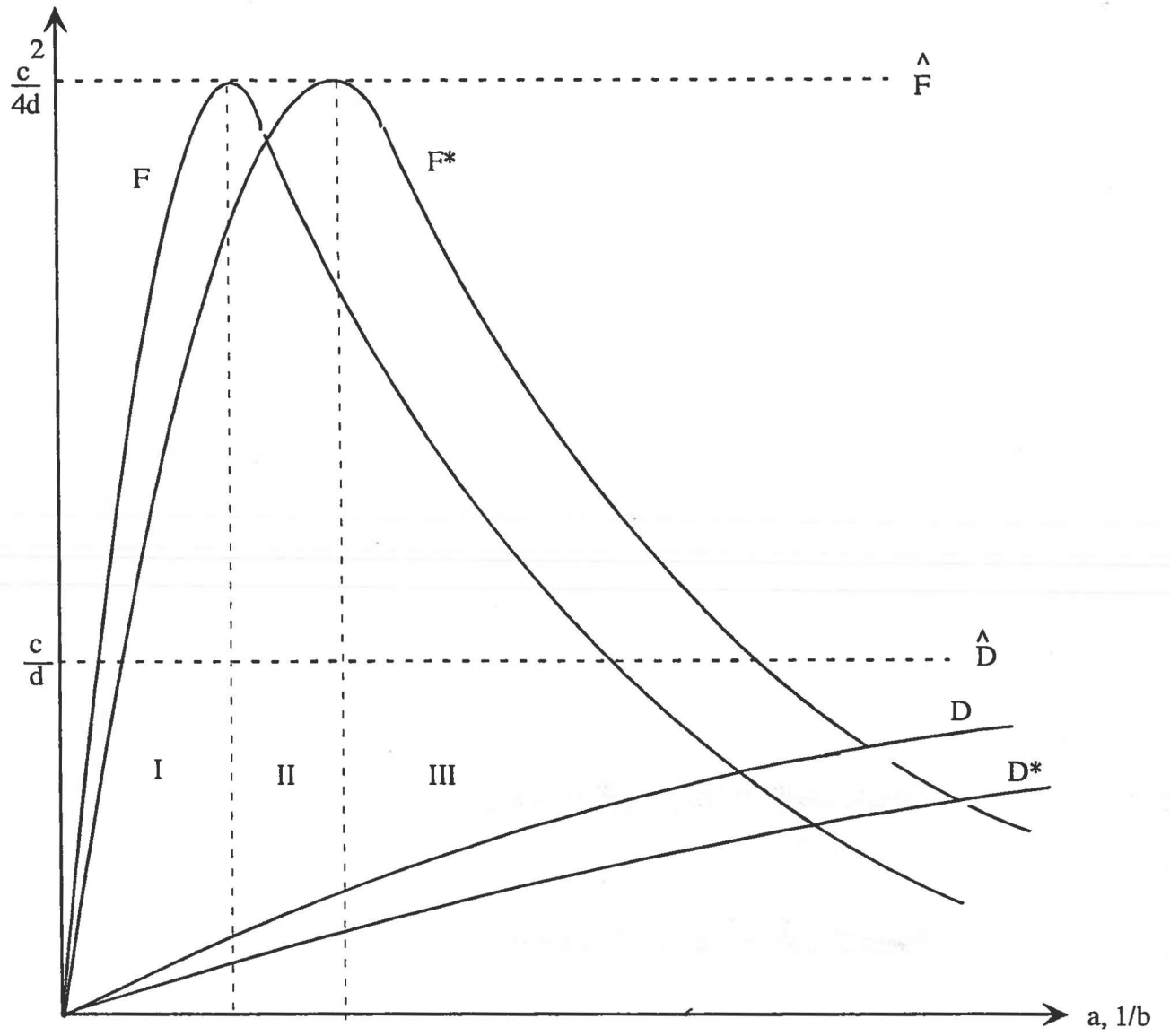Figure 3 Varying Demands Generate Equilibrium vs. Optimal Flows Along AC and MC

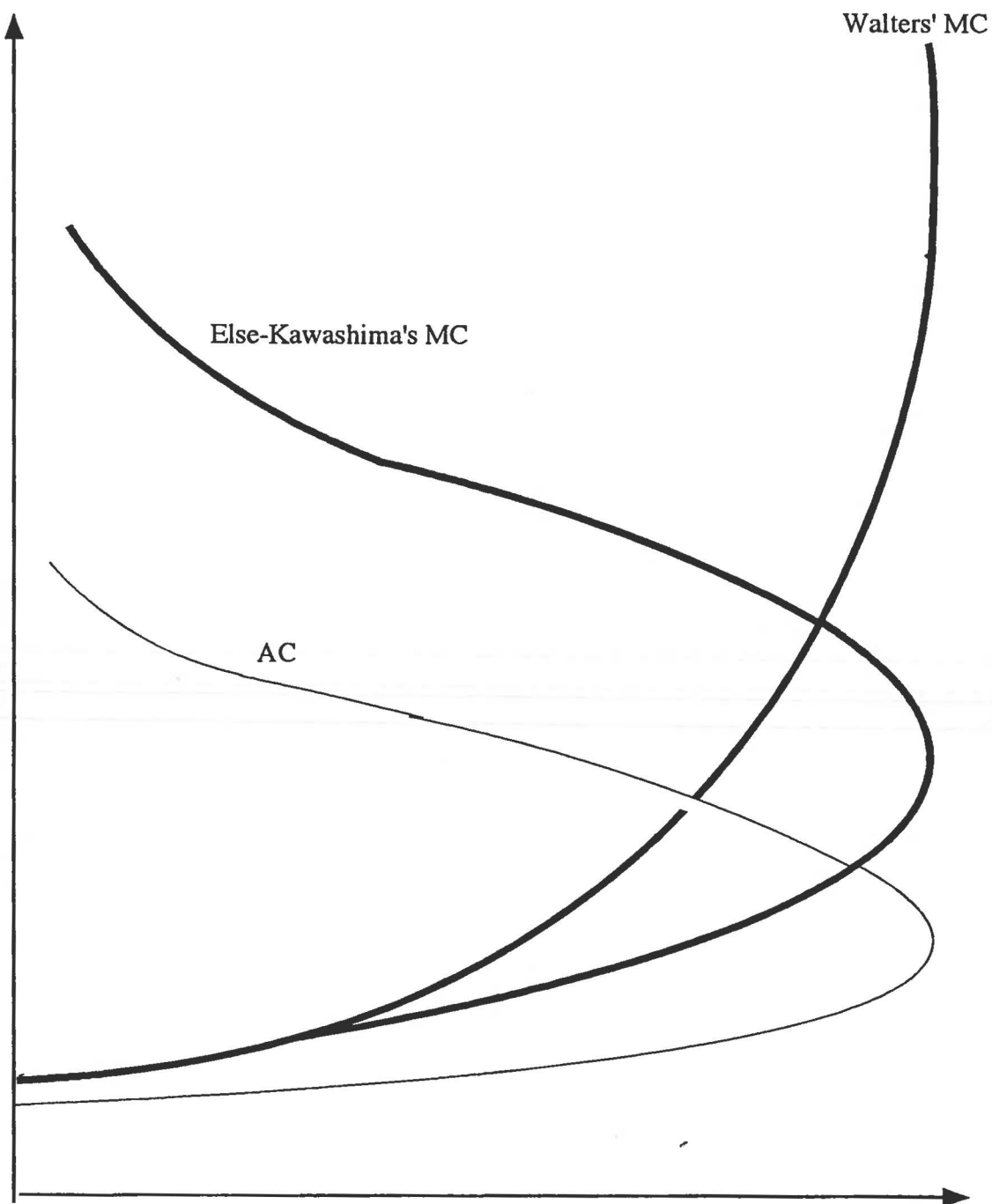Figure 4 Equilibrium vs. Optimal Density and Flow Related to Demand Parameters

Figure A-1 Else-Kawashima's vs. Walters' MC, Given AC