February 21, 2018

University of Toronto

MODELING DOWNTOWN TRAFFIC CONGESTION: AN ECONOMIC PERSPECTIVE

Richard Arnott

University of California, Riverside

Modeling Downtown Traffic Congestion: A Microeconomist's Perspective

The classic literature

1. The Beckmann model
2. The Walters model
3. The Vickrey bottleneck model

The modern literature

4. Empirical regularities on freeway congestion
5. Daganzo and Geroliminis (2007)
6. Isotropic/bathtub/reservoir models

Some microfoundations of downtown traffic congestion

7. How downtown traffic congestion differs from freeway traffic congestion
8. A downtown parking model
9. Multiple downtown transportation modes
10. Queuing at intersections
11. Stochasticity

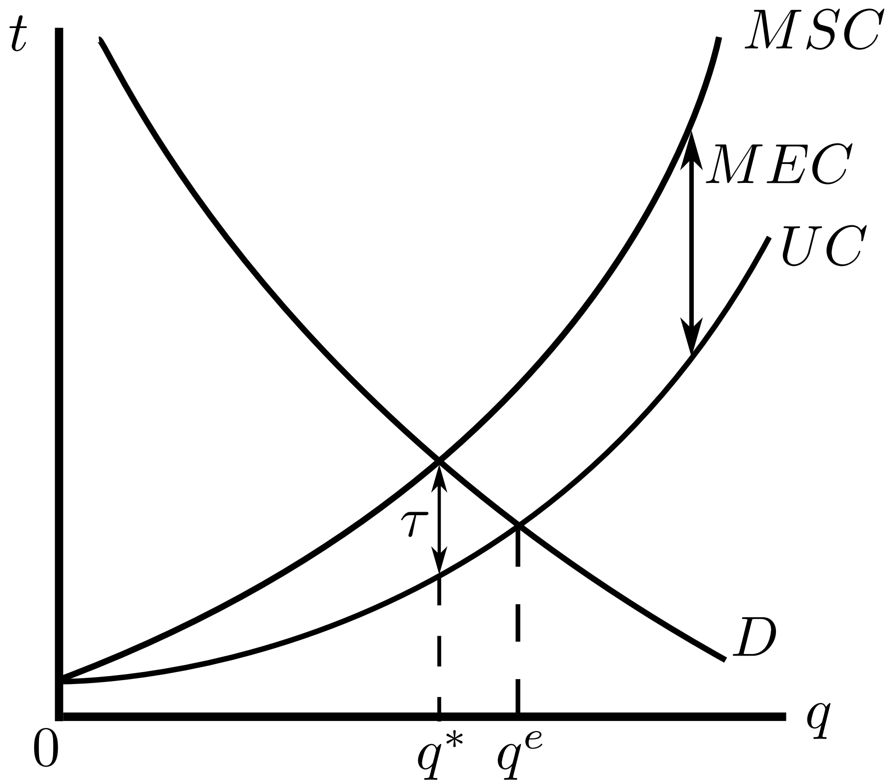
The Beckmann Model (1956)

- Transportation system viewed as a network of traffic links.
- Static, though perhaps day may be divided up into periods that are treated as independent.
- Congestion is modeled by assuming that the travel time on each link is a function of flow on that link: $t_i = t_i(q_i)$, e.g. BPR congestion function

  $$t_i = a_i + b_i \left(\frac{q_i}{s_i} \right)^{c_i}$$

  where a_i, b_i, and c_i are estimated constants, and s_i is link capacity. c_i is typically around 4.0.
- Collect trip data from travel diaries to generate a trip matrix giving the number of trips between origin j and destination k per hour or over the time period. Collect data on link travel times too.
- In "Wardrop" equilibrium, no traveler can reduce his cost by altering his route.
- Proposed transportation improvements are modeled as either the expansion of existing links or the addition of links. The benefits of a transportation improvement are measured as the value of travel time savings.

- With refinements, the Beckmann model remains the workhorse model for downtown transportation policy analysis.



- User cost, marginal social cost, marginal external cost. Textbook analysis of congestion tolling.

The Walters Model (1961) I

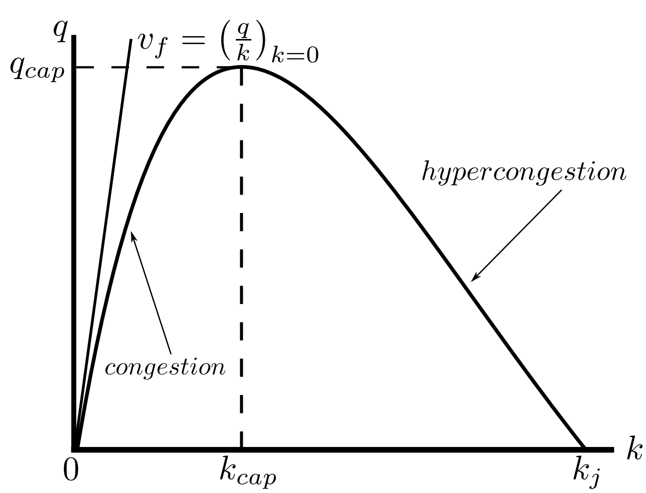
The treatment of congestion

1. Stems from a different branch of traffic flow theory in which velocity is viewed as a function of density: $v = v(k)$. E.g. $v(k) = v_f(1 - k/k_j)$ where v_f is free-flow velocity and k_j is jam density – Greenshields' Relation. Typically applied to highway or freeway traffic.

2. "Capacity" is maximum sustainable flow. The density associated with capacity flow is capacity density.

3. Makes more sense to treat travel speed as a function of density rather than of flow. Furthermore, car-following theories provide microfoundations for the steady-state relationship $v = v(k)$.

4. Combine this technological relationship with the fundamental identity of traffic flow theory that flow = density × velocity ($q = kv(k)$) to obtain the fundamental diagram:

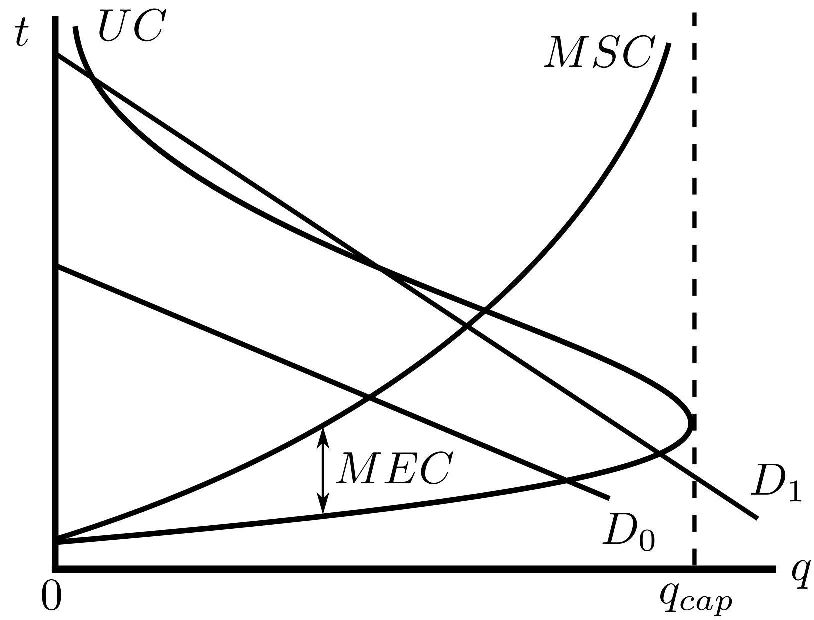


5. Distinguish between *congested* traffic flow and *hypercongested* traffic flow (traffic jam situations).

The Walters Model (1961) II

The Walters model

1. Considers a traffic link in steady state and assumes identical users. The user cost (uc) of a link trip equals the money cost, m, and time cost, with α being the value or cost of travel time. Assume that the money cost is fixed, m. Plot $\text{uc}(k) = m + \alpha L/v(k)$ against traffic flow ($q(k) = kv(k)$).



2. Standard analysis of congestion externality and congestion tolling, taking into account possibility of hypercongestion.

The Vickrey Bottleneck Model (1969) I

- Innovates in four ways.
  - Models rush-hour traffic dynamics.
  - Views congestion as deriving from a state variable.
  - Treats schedule delay costs (the costs of traveling at an inconvenient time).
  - Introduces the concept of a trip-timing equilibrium: No commuter can reduce her trip cost by altering her departure time.
- In particular, where $Q(t)$ is the length of the queue at time t, congestion is treated as a queue behind a bottleneck of fixed flow capacity, s.
- Shall treat what has come to be known as the $\alpha - \beta - \gamma$ variant of the model. Let α denote the unit value (cost) of travel time, β denote the unit cost of time early, and γ denote the unit cost of time late, and $c(t)$ denote the trip cost of a commuter who departs at time t. To simplify, assume that all commuters are identical and have the same desired arrival time, t^*. Let $\underline{t}$ denote the time of the first departure and $\bar{t}$ denote the time of the last departure.

$$c(t) = \alpha Q(t)/s + \beta(t^* - (t + Q(t)/s)) \qquad \text{early departure}$$

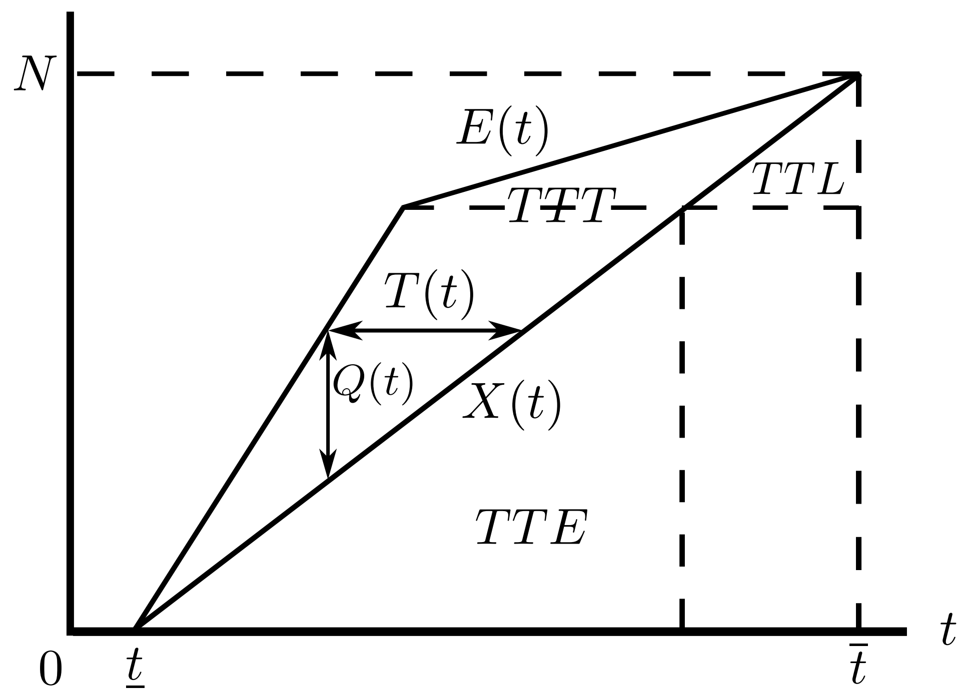
$$c(t) = \alpha Q(t)/s + \gamma(t + Q(t)/s - t^*) \qquad \text{late departure}$$

$$c(t) = \underline{c} \text{ over departure interval} \qquad \text{trip timing eqlb condition}$$

$$s(\bar{t} - \underline{t}) \qquad \text{exogenous number of commuters}$$

Vickrey's Bottleneck Model (1969) II

Cumulative arrivals and departures, queue length, and travel time.



- Since exit flow is constant over the rush hour, at the capacity of the bottleneck, hypercongestion does not occur.
- The social optimum entails the same $\underline{t}$ and $\bar{t}$ but no queuing. Travel time costs (in excess of free-flow travel time) are eliminated, and schedule delay costs are unchanged.
- The social optimum can be decentralized by a time-dependent toll that charges each driver for the congestion externality he imposes on others at the optimum.

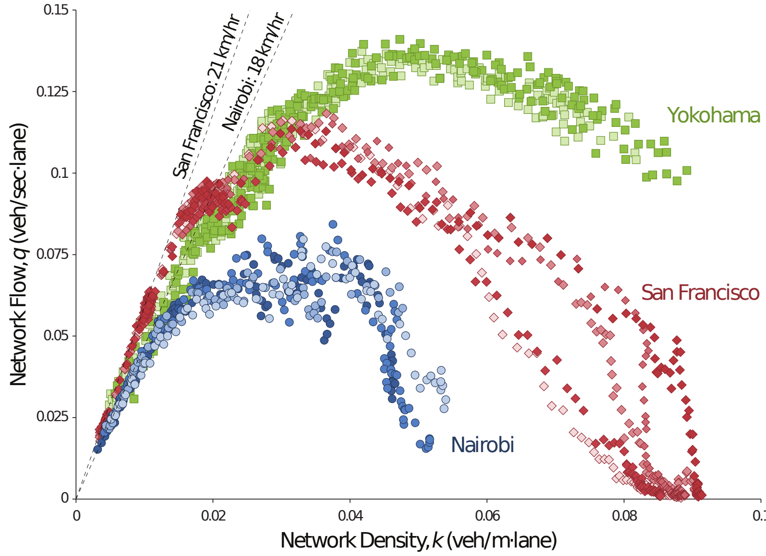
Empirical Regularities on Freeway Congestion (1980's)

- Starting in the 1980's, a huge amount of data on freeway congestion has been collected from sensors.
- At the risk of oversimplification, the following empirical regularities have been observed:
  - Up to capacity density, speed falls only slowly with density.
  - As capacity density is reached, free-flow travel becomes unstable and some incident, even a minor one, triggers a switch to a new regime with a bottleneck.
  - Queuing behind the bottlenecks may entail considerable loss of time, but flow falls only about 15%.
- Until a decade ago, virtually no data had been collected on downtown traffic congestion at the level of the neighborhood. It was therefore assumed that, at the macroscopic level, downtown traffic congestion behaves in much the same way as freeway traffic congestion.
- Theoretical work of metropolitan traffic congestion has therefore relied heavily on the bottleneck model.

Geroliminis and Daganzo (2007)

- There the literature stood for many years until Geroliminis and Daganzo (2007), which documented the existence of a stable macroscopic fundamental diagram for a neighborhood of Yokohama, Japan, using a combination of stationary and mobile sensors.

- The qualitative results have since been reproduced for many other cities.



- Geroliminis and Daganzo (2007) emphasized two empirical regularities:
  - Hypercongestion occurs systematically at the peak of the rush hour.
  - The exit rate from the traffic stream is approximately proportional to traffic density.

- The empirical work that has been done documents that, in contrast to freeway traffic, the drop in flow due to hypercongestion at the peak of the rush hour may be substantial. In downtown areas, a considerable fraction of the time loss due to traffic congestion occurs when traffic is hypercongested (Beijing and Rome).

- G&D and the empirical literature following from it have catalyzed the development of models of "downtown" rush-hour traffic flow that, in contrast to the bottleneck model, incorporate hypercongestion.

Isotropic/Bathtub/Reservoir Models I

- A fixed number of identical commuters N travel from home to work over an isotropic downtown area in the morning rush hour with LWR flow congestion ($v = v(k)$ out of steady state) replacing the bottleneck.
- Thus, essentially combines the bottleneck and the Walters model.
- Terminology is still in flux in the area.
- I shall present the social optimum problem for the "proper" model that has strict microfoundations.

$$\max_{e(t)} \int_{\underline{t}}^{\bar{t}} U(t, T(t))e(t)\ dt$$

$i)$	population constraint	$\int_{\underline{t}}^{\bar{t}} e(t)\ dt = N$
$ii)$	congestion technology	$v = v(k)$
$iii)$	accumulation	$k(t) = E(t) - X(t)$
$iv)$	FIFO	$X(t + T(t)) = E(t)$
$v)$	trip length	$\int_{t}^{t+T(t)} v(k(t))\ dt = L$
$vi)$	boundary and non-negativity conditions	$k(\underline{t}) = 0, k(\bar{t} + T(\bar{t}) = 0, e(t) \geq 0$

Differentiate the trip length condition

$$v(k(t + T(t)))(1 + \dot{T}(t)) - v(k(t)) = 0$$

Differentiate again

$$v'(k(t+T(t)))\dot{k}(t+T(t))(1+\dot{T}(t))^2+v(k(t+T(t)))\ddot{T}(t)-v'(k(t))\dot{k}(t)=0$$

Substituting in

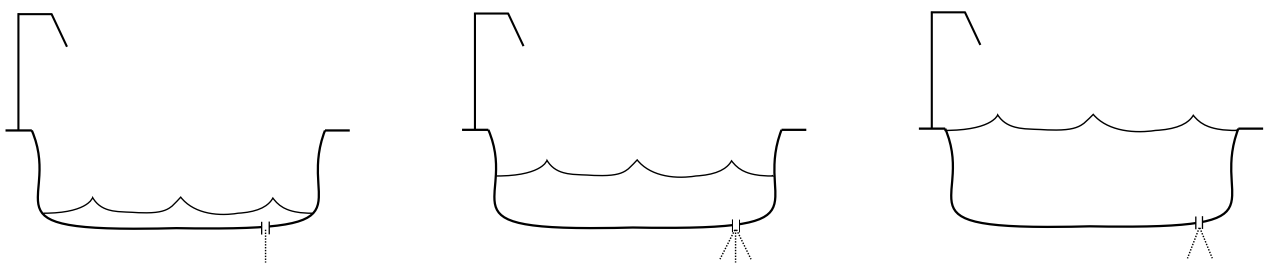
$$\dot{k}(t+T(t))=e(t+T(t))-x(t+T(t))=e(t+T(t))-e(t)$$

obtain a delay differential equation with an endogenous delay, which renders the equation system analytically intractable.

Isotropic/Bathtub/Reservoir Models II

- The central issue has been whether to work with the proper model or with an approximation to it that is analytically tractable.

- Several approximations have been proposed. One, which was first considered by Vickrey (1991), accords with the empirical regularities identified in Geroliminis and Daganzo (2007), and was explored by Arnott (2013), is to assume that $x(t) = k(t) \div (L/v(k(t)))$.



- Daganzo's students, many of whom are now leaders in transportation science, have been extending these models in many ways, including treating mass transit, parking, spatial differentiation, etc.

Microfoundations of Downtown Traffic Congestion I: How downtown traffic congestion differs from freeway traffic congestion

- We have seen that recent empirical work finds that, at the macroscopic level, the behavior of downtown traffic is significantly different from the behavior of freeway traffic.
- An active area of research is to develop microfoundations of downtown traffic congestion.
- A useful place to begin is to ask how downtown traffic congestion differs from freeway traffic congestion.
  - Link flow congestion relatively more important for freeways, and nodal congestion relatively more important on streets.
  - Nature of nodal congestion differs (interchanges and merging vs intersections).
  - There is a greater mixture of vehicle types downtown than on freeways, as well as pedestrians.
  - Parking is an essential feature of downtown traffic congestion that is not present in freeway congestion.
  - Mass transit is more important in downtown traffic.
  - Freight deliveries are important in downtown traffic but not in freeway traffic.
  - Access and egress are at discrete locations on freeways but are virtually continuous in downtown areas.

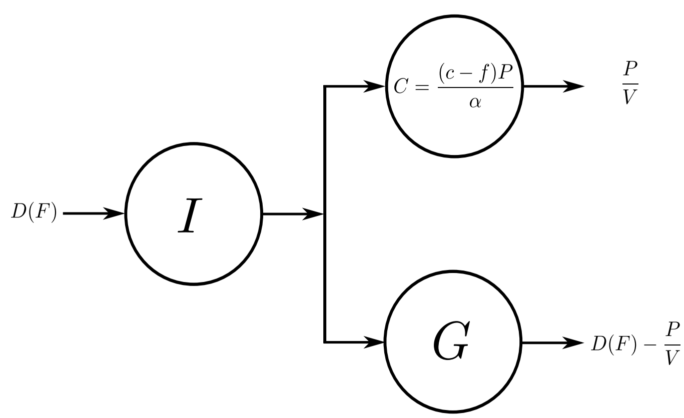
Microfoundations of Downtown Traffic Congestion II: A model of downtown parking and congestion

- The best point of entry into the literature on the economics of parking is Donald Shoup's book, The High Cost of Free Parking.

An isotropic downtown area in steady state. Each car travels a distance L from its origin to its destination, finds parking, and then visits for a length of time V. Assume that curbside parking is saturated.

Let I denote the stock of cars in transit per unit area, C be the stock of cars cruising for parking, P be the stock of curbside parking spaces, f be the downtown parking meter rate for on-street parking, c be the garage parking fee (no capacity constraints), F be the full price of a trip, $D(F)$ be the demand for trips as a function of F, t be travel time per mile, and α be the value of time.

congestion function	$t = t(I, C, P)$
full trip price with on-street P	$F_{\text{on-street}} \alpha t(I, C, P) + fV + \alpha CP/P$
full trip price with off-street P	$F_{\text{off-street}} \alpha t(I, C, P) + cV$
equalization of full prices	$F = F_{\text{on-street}} = F_{\text{on-street}}$
steady-state condition	$D(F) = I/(\lambda t(I, C, P))$



There is also a small literature that looks at cruising for parking when it is not saturated. Don Shoup has advocated that meter rates be set so that the expected curbside parking occupancy rate is 85%, and San Francisco has experimented with implementing his policy.

Microfoundations of Downtown Traffic Congestion III: Multiple Modes

- Conceptually can readily extend the analysis of congestion to treat multiple modes of transportation, e.g.

$$t_{\text{car}} = t_{\text{car}}(\text{density of buses, pedestrians, taxis, bicycles, —})$$
$$t_{\text{subway}} = t_{\text{subway}}(\text{density of passengers, density of trains})$$

- Empirical work on macroscopic models of traffic congestion is just starting to consider multiple modes. Complicated. For example, bus-car congestion interaction is complex.
- A branch of the literature considers the allocation of road capacity between modes and curbside parking.
- Economic analysis is complicated by not only the possibility of hypercongestion, but also economies of scale in mass transit and dispatch taxi service, which may lead to multiple local optima.
- A recent strand of the literature argues that much of the value of mass transit lies in reducing the number of vehicles on streets at the peak of the rush hour.

Microfoundations of Downtown Traffic Congestion IV: Intersections

- In many heavily congested downtown areas, most of the delay occurs at intersections, whether signalized or not.
- There is a well-developed body of theory in transportation science on congestion at isolated intersections (intersection queuing theory), but essentially none for networks of intersections.
- Furthermore, there has been little empirical work that looks at the form of congestion at signalized intersections.
- I have a particular research topic in mind: Ignore flow congestion and assume that all congestion delay occurs on a symmetric network of intersections. If queuing cars pass through the intersections at their capacity (flow), there would never be hypercongestion. What therefore causes hypercongestion on a symmetric network of intersections?
- Daganzo and his students have considered "spillbacks" where a queue of cars at an intersection spills back into the upstream intersection. (Queues actually don't spill back. Instead a car enters the upstream intersection and, unable to advance, blocks traffic.)
- Drawing on work by Varaiya, Daganzo's students have been considering adaptive, autonomous (centralized or decentralized) signal control as a way of dealing with this problem.
- This suggests that pricing and "traffic management" are complementary ways of managing downtown traffic congestion. Optimal mix?
- Other reasons for reduced capacity at intersections with heavy traffic congestion?

Microfoundations of Downtown Traffic Congestion V: Stochasticity

- Most traffic flow theory ignores stochasticity. When it is considered, it is typically modeled at an aggregate level – higher demand on Friday afternoons, reduced capacity due to snow or construction, etc.
- The exception to this is intersection queuing theory, where microstochasticity (stochastic arrival rate at the individual intersection) is the essence of the problem (absent stochasticity there would be no congestion).
- This raises two interesting questions that the literature on downtown traffic congestion has barely started to consider:
  - How important is microstochasticity in downtown traffic congestion?
  - If it is important, how should it be dealt with? Responsive pricing vs stochastic "traffic management".