University of Toronto
Department of Economics

Working Paper 590

Improving the fit of structural models of congestion

By Jonathan D. Hall

October 16, 2017

IMPROVING THE FIT OF STRUCTURAL MODELS OF CONGESTION*

JONATHAN D. HALL
UNIVERSITY OF TORONTO

OCTOBER 16, 2017

ABSTRACT. We need structural models of traffic congestion to answer a wide variety of questions, but the standard models fail to match the data on travel times across the day. I establish the nature and magnitude of the problem, and show its source lies in how we model agent preferences, not in the specifics of the congestion technology. The poor fit of the models suggests that we are abstracting away from features with a first-order impact on model predictions, which limits our ability to use these models to evaluate counterfactuals quantitatively and—when agents are heterogeneous—qualitatively as well. I explore several ways of improving the fit of these models, concluding with recommendations for tractable and intuitive ways of doing so.

Keywords: Structural model, Congestion, Model fit, Calibration, Dynamic, Bottleneck Model, Traffic

JEL Codes: R4, H23, H4

*I am grateful for feedback from William Strange, Gilles Duranton, Robin Lindsey, Robert McMillan, and audiences at the International Transport Economics Association. *Email:* jonathan.hall@utoronto.ca.

1. Introduction

Arnott et al. (1993) persuasively argue it is important to use structural models of congestion as the standard static model "contains ambiguities and is poorly specified" (p. 161). These structural models are inherently dynamic and are used to address a variety of important questions, including optimal highway capacity (Arnott et al. 1993), the value of travel time information (Arnott et al. 1999), optimal parking prices (Fosgerau and de Palma 2013), how tolls affect urban spatial structure (Gubins and Verhoef 2014; Takayama and Kuwahara 2017), which road segments should be tolled (de Palma et al. 2004), non-price mechanisms for addressing congestion (Nie 2015), and the distributional consequences of tolling (van den Berg and Verhoef 2011; Hall 2017b).[1]

This paper shows that while standard structural models of congestion generate a travel time profile that has the right general shape, they fit observed travel times poorly. In particular, under standard assumptions for parameter values, the predictions for either peak travel times or the length of rush hour are off by at least an order of magnitude. The poor fit suggests that these models are abstracting away from features that have a first-order impact on model predictions, which limits our ability to use them to evaluate counterfactuals.

I highlight three problems with fitting these models to the data. In decreasing order of severity, they are; first, travel times climb and fall far slower than the models predict; second, travel times fall too slowly after the peak relative to how quickly they climb before the peak; and, third, travel times are essentially flat at the peak.

The source of the poor fit lies almost entirely with how we model agent preferences over different arrival times, and has little to do with the model of congestion itself. Thus these problems affect the bottleneck model of Vickrey (1969) and Arnott et al. (1993), as well as other structural models such as those of Henderson (1974) and Chu (1995), the cell transmission model of Daganzo (1994), and the hydrodynamic traffic flow model of Lighthill and Whitham (1955) and Richards (1956).[2] This is because agent preferences determine the slope of the travel time

[1] The time-of-use decision at the heart of structural models of traffic congestion is also used in models of transportation by train (Kraus and Yoshida 2002; De Borger and Fosgerau 2012; de Palma et al. 2017) and airplane (Silva et al. 2014).

[2] These models essentially only differ in how they model congestion. In the bottleneck model drivers only slow down those who come after them, while in Henderson (1974) and Chu (1995) drivers only slow down those traveling at the same time as them. In the cell transmission and

profile, while the congestion technology then determines how many agents are traveling at each point in time.

Of course, as Friedman (1953) argues, we do not expect models to be perfect representations of reality, but useful abstractions that help us understand reality better. The problems I identify do not affect all uses of structural congestion models. However, I show they limit our ability to use these models to quantify magnitudes of some outcome (e.g., the social welfare gains from some policy) and derive qualitative results with heterogeneous agents.

I explore multiple solutions to each of the three problems, including allowing for uncertainty in arrival times and generalizing the schedule delay cost function beyond being piecewise linear.

I conclude with the following recommendations for improving the ability of structural models of congestion to fit the data. If your goal is to quantify an outcome with homogeneous agents, then it is simplest to use significantly lower parameter values for agents' schedule delay costs than are typically assumed. If your goal is a quantitative or qualitative result with heterogeneous agents, then add the assumptions that (1) agents have a continuum of desired arrival times, which allows some agents to be inframarginal, (2) agents differ in the flexibility of their schedules, so some agents actually are inframarginal, and (3) use parameter values that imply the cost of being late is less than the cost of being early. These solutions do not fix the problem of travel times being essentially flat at the peak, however, its magnitude is small after fixing the other two. If it is to be fixed, it is most tractably fixed by allowing for agents to be indifferent between arrivals in a small window around their desired arrival time.

The remainder of the paper is structured as follows. After introducing the model (Section 2), I document the three problems with the fit of the model to the data (Section 3), explore several solutions to these problems (Section 4), and show these problems affect our theoretical predictions (Section 5). I conclude in Section 6.

Lighthill and Whitham (1955)–Richards (1956) models drivers slow down both those who come after and those traveling at the same time. These last two models are less tractable, and only recently have been combined with models of agent preferences (e.g. Han et al. 2011; Ukkusuri et al. 2012; Friesz et al. 2013).

2. Model

I start with the model of agent preferences of Vickrey (1969), which is used throughout the literature, while making minimal assumptions on the congestion technology.[3]

2.1. Congestion technology. There is a single road connecting where people live to where they work. Travel time for an agent arriving at t is

$$T(t) = T^f + T^v(t),$$

where T^f is fixed travel time—the amount of time it takes to travel the road absent any congestion—and $T^v(t)$ is variable travel time. I assume $T^v(t)$ is continuous everywhere and differentiable almost everywhere.

For most of the results in this paper, this is all of the structure we need on the congestion technology. When exploring whether adding uncertainty helps fix the empirical problems with structural congestion models, I modify the model to add uncertainty; and I use the bottleneck model when discussing examples of how the problems identified in this paper affect quantitative and qualitative results.

2.2. Agent preferences. Agents choose when to arrive at work to minimize the cost of traveling.[4] Agents dislike two aspects of traveling: travel time and schedule delay—that is, arriving earlier or later than desired. These costs combine to form the trip cost; the trip cost of arriving at time t for an agent of type i is

$$p_i(t) = \alpha_i T(t) + D_i(t^* - t) \tag{1}$$

---

[3] There are other ways of modeling agent preferences that are similar to that above. Most prominent is the utility-theoretic models of Vickrey (1973) and Fosgerau and Engelson (2011), where agent preferences are defined in terms of the per-minute utility of time spent traveling, at home, and at work. If the utility rate of traveling and time at home are constant, and the utlity rate at work is piecewise-constant with a discontinuity at t^*, then this is isomorphic to what is above. In this case α is the difference between the per-minute utility at home and the per-minute utility while driving, β is the difference between the per-minute utility of being at home and the per-minute utility of being at work before t^*, γ is the difference between the per-minute utility of being at home and the per-minute utility of being at work after t^*.

[4] Nearly all papers using structural congestion models focus on the morning commute. de Palma and Lindsey (2002) point out that a key difference between the morning and evening commutes may be that in the evening agents have a preferred *departure* time rather than a preferred *arrival* time. The first and third problems discussed below are also issues with the evening commute, and the improvements that deal with those problems are likewise relevant to the evening commute.

where α is the cost per unit time traveling (i.e., the agent's value of time) and D_i is group i's schedule delay cost function. Schedule delay costs are piecewise linear,

$$D_i(x) = x \begin{cases} \beta_i & x \leq 0 \\ -\gamma_i & x > 0 \end{cases}$$

where β is the cost per unit time early to work, and γ is the cost per unit time late to work. Each of these parameters represents how much an agent is willing to pay in money to reduce travel time or schedule delay by one unit of time.

Each individual agent has zero mass, and there is mass N_i of agents of type i.

When exploring different ways of improving the fit of structural congestion models, I consider various extensions to this classic formulation of preferences.

2.3. **Definition of equilibrium.** The relevant equilibrium concept is that of a perfect-information, pure-strategy Nash equilibrium, in which no agent can reduce his trip cost by changing his arrival time.

3. Identifying the problems

To better identify the problems, it is helpful to focus on three ratios: β/α, γ/α and γ/β. I first show that these ratios have simple economic interpretations, which helps us assess reasonable values for them, and discuss the commonly assumed values for them. I then show how these ratios map into the empirical travel time profile, and use data from three cities to compare the theoretically predicted travel times given commonly assumed values for these ratios to the empirical travel time profile.

The ratios β/α and γ/α are an agent's willingness to pay in travel time to reduce schedule delay (early and late respectively) by one unit of time, and provide a measure of the inflexibility of his schedule. As in Hall (2017b), define β_i/α_i as type i's *inflexibility*. If a shift worker is late he generally faces penalties and when he is early he passes the time talking with co-workers. Since there is not much difference for the shift worker between spending time traveling or being at work early, his β/α is close to one (the largest possible β/α). Similarly, due to the penalty when late, γ/α is large. In contrast, an academic can start working whenever she gets to the office and so has a very low marginal disutility from being early or late and so her β/α is closer to zero.

It seems reasonable to assume that β/α should typically be somewhat large, and that there should exist a large number of people who have a β/α very near to one.

Table 1. Assumed parameter values

Paper	β/α	γ/β
Arnott et al. (1993)	0.61	3.9
de Palma and Lindsey (2002)	$[0.29, 0.94]$	4
de Palma et al. (2004)	0.6	4.17
Fosgerau and Karlström (2010)	0.5	4
Liu and Nie (2011)	$\{0.61, 0.78\}$	$\{3.04, 3.90\}$
van den Berg and Verhoef (2011)	$[0.33, 0.99]$	3.9
Tian et al. (2013)	$[0.19, 0.98]$	5.25
Gubins and Verhoef (2014)	0.5	4
Xiao et al. (2014)	0.61	3.9
Takayama and Kuwahara (2017)	$\{0.3, 0.4, 0.45, 0.5\}$	4

Typical assumptions, as summarized in Table 1, are for mean inflexibility to be greater than one-half, and maximum values near one.

The ratio γ/β represents the relative cost of being late to early. To the best of my knowledge, it is always assumed that arriving late is worse than arriving early, so that $\gamma/\beta > 1$, with typical values near four, as Table 1 shows.

Not only are the standard assumptions on β/α and γ/β intuitively reasonable, they also have empirical evidence supporting them. Small (1982) estimates a discrete choice model of when to arrive at work, and finds results supporting these assumptions. Indeed, most of the papers listed in Table 1 cite Small (1982) in support of their assumptions.

These preference parameters determine the shape of the travel time profile, $T(t)$. The first-order condition for choosing arrival time, $p_i'(t) = 0$, implies

$$T'(t) = \frac{\beta_i}{\alpha_i} \qquad \text{if } t < t^*, \tag{2}$$

$$T'(t) = -\frac{\gamma_i}{\alpha_i} = -\frac{\gamma_i}{\beta_i}\frac{\beta_i}{\alpha_i} \qquad \text{if } t > t^*, \text{ and} \tag{3}$$

$$-\frac{\gamma_i}{\alpha_i} \leq T'(t) \leq \frac{\beta_i}{\alpha_i} \qquad \text{if } t = t^*. \tag{4}$$

These impose the standard requirement that an agent's marginal rate of substitution be tangent to the budget line, unless at a corner solution.

For someone to be arriving at any given time t, one of (2)–(4) must be satisfied, and since someone is arriving at all times during rush hour, one of these equations must be satisfied at all times during rush hour. Thus the slope of the travel time profile is directly governed by agents' preferences, and by looking at empirical travel time profiles we can learn about agents' preferences.

I measure the travel time profile in three cities, Los Angeles, Chicago, and Boston; using data from a different source for each one.[5] Figure 1 compares the model predicted travel times using the parameters typically assumed in the literature to travel times on California State Route 91 from the center of Corona to the junction of SR-91 and I-605.[6]

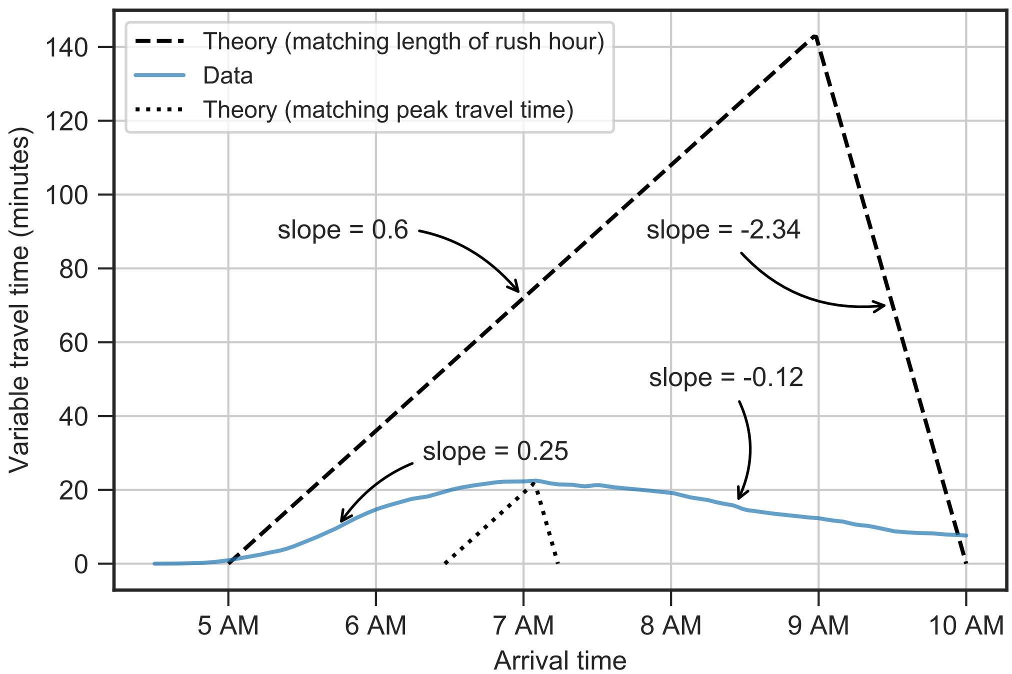

FIGURE 1. Travel times predicted by theory compared to average travel times for CA SR-91W for postmiles 16–42 for 2004. Data from Caltrans PeMS.

Notes: For the theoretical predictions I assume $\beta/\alpha = 0.6$ and $\gamma/\beta = 3.9$. The higher theoretically predicted travel times come from matching the empirical length of rush hour and the lower theoretically predicted travel times come from matching the empirical peak travel time. Slopes for the empirical travel time profile are the steepest slopes over an hour interval.

[5] I use data from California Department of Transportation (2014), Illinois Department of Transportation (2017), and data collected from Google Maps.

[6] I choose this specific segment because it roughly represents the median commute for those living in Corona who use SR-91, as calculated using data from Sullivan (1999). I calculate travel times for all business days in 2004 using loop detector data from the California Department of Transportation's Performance Measurement System ("PeMS", 2014).

As Figure 1 shows, using the standard assumptions about parameter values, the model fails to match the data. If I match peak travel times, I predict rush hour is only 46 minutes long, but if I match the length of rush hour, I predict variable travel times seven times longer than those seen in reality.[7]

Figure 1 shows three problems with matching the model to the data. The most significant problem is that the slope of the travel time profile is much lower than our intuition for the preference parameters suggest it should be. The maximum slope (over an hour interval) before the peak is 0.25, while the expected average slope is 0.6 rather than the expected 0.6, and the maximum slope after the peak is -0.12 rather than -2.34. In addition, the slope after the peak is more shallow than the slope before the peak: the ratio of the (maximum) slope after the peak to the (maximum) slope before the peak is 0.48 rather than near 4. Finally, the slope is relatively flat at the peak.

This mismatch between theory and data is not unique to California State Route 91. Figure 2 shows the pattern holds for I-55 in Chicago from I-80 to I-94 (67.8 kilometers, a long commute) and a more typical commute of I-290 from Wolf Road to I-94 (22.5 kilometers).[8] So far all of this data is just for highways, however, it holds if we include surface streets as well. Figure 3 also plots travel times from Hopkinton to Boston Commons (48.5 kilometers, a long commute) and Wellesley to Boston Commons (24.3 km, a more typical commute) using data from Google Maps.[9] In every case the slope before the peak is less than half that expected, the slope after is less than a seventh that expected, the ratio of the slope after the peak to the slope before the peak is always less than 2/5ths that expected, and the slope of the travel time profile is flat at the peak.

4. Possible solutions

Having identified the three problems with structural models of congestion, I now explore possible solutions to them. Table 2 summarizes the results of this section.

[7]In everyday speech "rush hour" refers to only the period of time when travel times are at their very worst; here I am using "rush hour" to denote the entire period of time when travel times are higher than in free flow conditions. It is this entire evolution of travel times across the day that structural congestion models are designed to explain.

[8]Data is for all weekdays, and is from Illinois Department of Transportation (2017). The mean commute in the Chicago metropolitan area is 31.2 minutes (U.S. Census Bureau 2016).

[9]Data covers all weekdays in July and August of 2017. The mean commute in the Boston metropolitan area is 30.2 minutes (U.S. Census Bureau 2016).

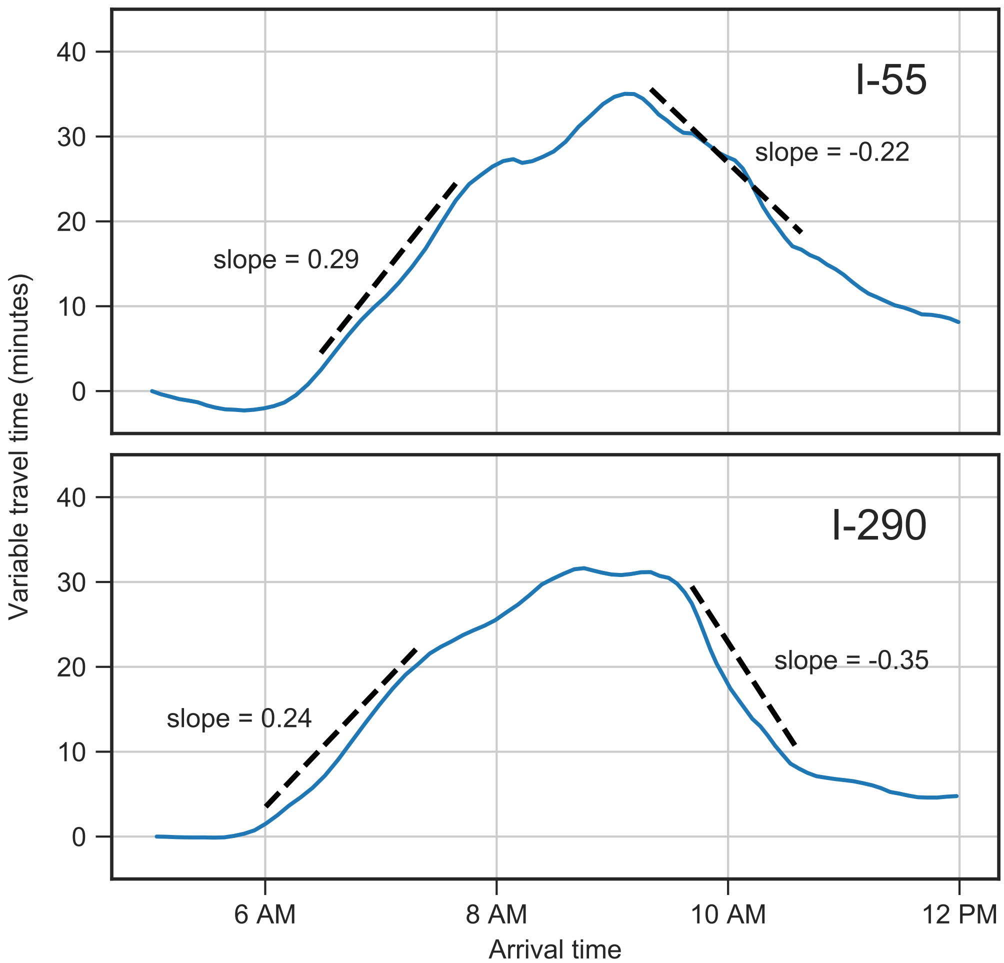

FIGURE 2. Average weekday travel time by arrival time for I-55N (from I-80 to I-94) and I-290E (from Wolf Road to I-94) in Chicago for 2004–2006. Data from Illinois Department of Transportation (2017).

4.1. **Slope of travel time profile too low.** The most significant problem is that travel times rise and fall far slower than expected.

4.1.1. *Possible solution: There are no inflexible drivers.* The simplest solution would be to conclude that the model is correct and our intuition about reasonable parameter values is wrong. Choosing values for β/α and γ/α to best match the data

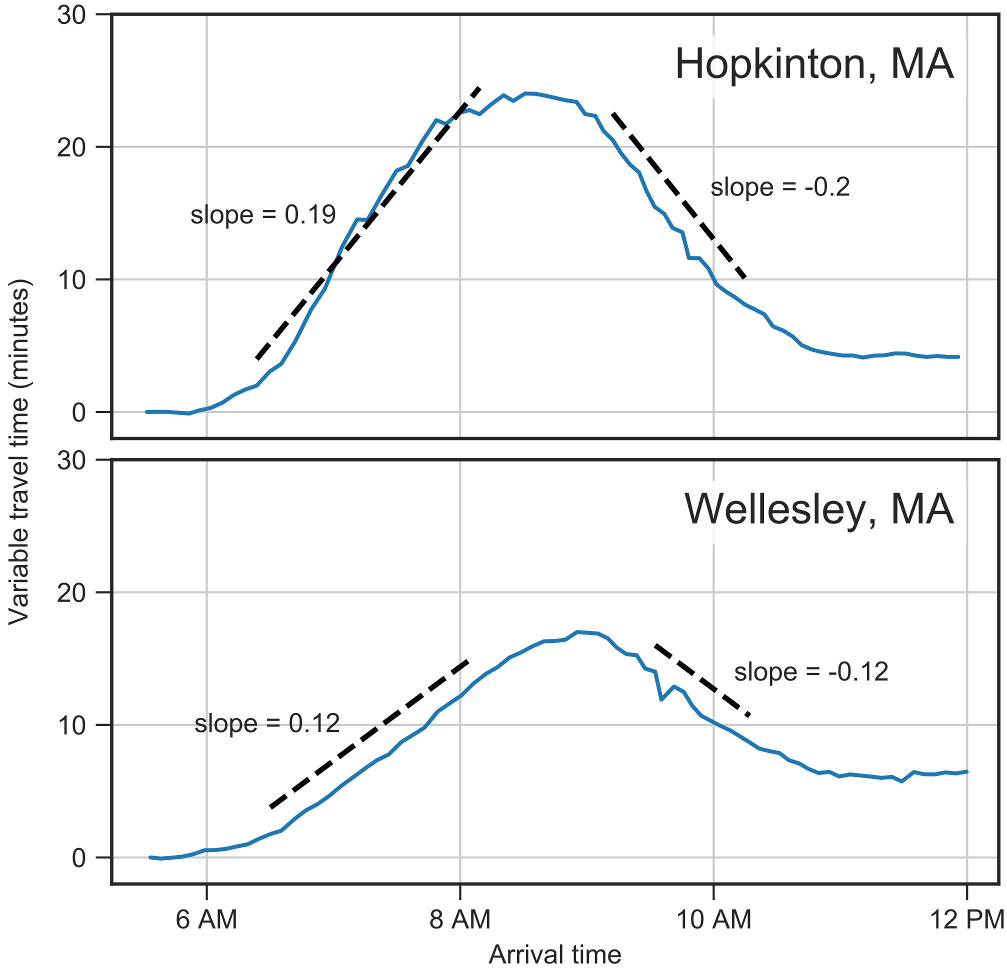

FIGURE 3. Average weekday travel time by arrival time for trips from Hopkinton to Boston Commons and from Wellesley to Boston Commons for July and August of 2017. Data collected from Google Maps.

significantly improves the fit: reducing the sum of squared errors in our prediction by 99.99 percent.[10] However, this yields estimates of $\beta/\alpha = .22$ and $-\gamma/\alpha = -.09$.

[10]Specifically, I choose the start of rush hour, the peak of rush hour, β/α, and γ/α to minimize the sum of squared errors between predicted and actual travel times between 5–10 a.m. The comparison of sum of squared errors is based on the theoretical predictions matching the length of rush hour. If I compare to the results from using the traditional parameter values and matching peak travel time, the estimated parameters reduce the sum of squared errors by 99.66 percent.

Table 2. Summary of possible solutions

Modeling change	Problem: Slopes too small	Problem: Slope after peak too small relative to slope before peak	Problem: Flat top
Change parameter values	Yes, but implausible that there are no inflexible drivers.	Yes, as long as fraction late doesn't change in counterfactual.	*n/a*
Long-run vs. short-run preferences	No	Yes	n/a
Continuum of desired arrival times	Yes	Yes, but requires all agents to arrive on-time	Yes
Uncertainty over travel times	No	Maybe, but needs heterogeneous desired arrival times.	Yes
Heterogeneous trip lengths	No	*n/a*	*n/a*
Generalize $D(t - t^*)$	Yes, but need heterogeneous desired arrival times.	*n/a*	Yes, but need heterogeneous desired arrival times.
Heterogeneous γ/β	*n/a*	Yes	*n/a*

Notes: Entries labeled "*n/a*" are those which were not applicable to the given problem.

However, these estimates imply there do not exist any agents who have even moderately inflexible schedules. This is not reasonable; 58 percent of workers in the 2009 National Household Travel Survey report they cannot choose what time they start work and thus have a β/α near one (and a large γ/α as well). While I can reduce the percentage of drivers who are inflexible by only considering trips on the interstate in the morning and assuming *all* non-work trips are flexible, I am still left with 29 percent of agents being very inflexible.[11] Inflexible agents exist, they are just not affecting equilibrium travel times in the way the model with our standard assumptions predicts.

4.1.2. *Long-run vs. short-run preferences.* A similar solution is to note that long-run and short-run preferences can differ, as in Peer et al. (2015). In the long run agents can adjust on more margins of behavior, and so are more flexible. Consider, for example, how easy it can be to arrange your schedule to arrive an hour later than usual versus the cost of arriving an hour late unexpectedly. If our intuition for reasonable parameter values is based on thinking about the short run, then we will be looking for values of inflexibility that are too large.

However, as discussed above, many workers lack the ability to choose when they arrive, even in the long run, and so making the distinction between long- and short-run preferences does not affect their inflexibility.

4.1.3. *Possible solution: Add value of reliability.* A third possible solution is to consider the value of reliability by adding in uncertainty over travel times. As Figure 4 shows, the amount of uncertainty in travel times varies over rush hour. Adding

[11] I define the morning as trips that end between 4:30–noon.

this into structural congestion models, as in Arnott et al. (1999) and Fosgerau and Karlström (2010), would provide an extra incentive for agents to leave early or late—agents could reduce the uncertainty in their travel time.

However, Fosgerau and Karlström (2010) estimate that the value of reliability accounts for about 15 percent of total time costs while Small et al. (2005) estimates it accounts for about a third of total time costs. The slope of the travel time profile is climbing roughly half as fast as theory suggests, so to close this gap the value of reliability would need to account for 50 percent of total time costs. Thus while considering uncertainty helps narrow the gap between the parameters that seem reasonable and model predictions, it does not close it.

However, even this would not be enough. A model with uncertainty predicts that travel times at the start and end of rush hour should climb at the same rate they do in a model without uncertainty. Consider a slightly transformed model that includes uncertainty. In contrast to above, we now focus on *departure* times rather than arrival times. Assume there is a distribution of possible states of the world, $F(\phi)$, where a larger ϕ means worse traffic, $\partial T(t_d, \phi)/\partial\phi > 0$; and that leaving later means arriving weakly later, $\partial T(t_d, \phi)/\partial t_d \geq -1$. Define $\hat{\phi}(t_d)$ as the largest ϕ so that an agent departing at t_d arrives on-time.

Re-writing the cost function in terms of departure times gives

$$p(t_d, \phi) = \alpha T(t_d, \phi) + D(t^* - t_d - T(t_d, \phi)). \tag{5}$$

Agents choose their departure time to minimize their expected trip cost, which yields the following first order condition:

$$\begin{aligned}\frac{dE(p(t_d, \phi))}{dt_d} = \alpha \int_0^\infty \frac{\partial T(t_d, \phi)}{\partial t_d} dF(\phi) \\ - \beta \int_0^{\hat{\phi}(t_d)} \left(1 + \frac{\partial T(t_d, \phi)}{\partial t_d}\right) dF(\phi) \\ + \gamma \int_{\hat{\phi}(t_d)}^\infty \left(1 + \frac{\partial T(t_d, \phi)}{\partial t_d}\right) dF(\phi) = 0.\end{aligned} \tag{6}$$

By considering agents who are always early or always late, I obtain explicit formulas for the slope of the expected travel time profile from (6). If an agent is always early, then

$$\frac{dE(T(t_d, \phi))}{dt_d} = \frac{\beta_i}{\alpha_i - \beta_i}, \tag{7}$$

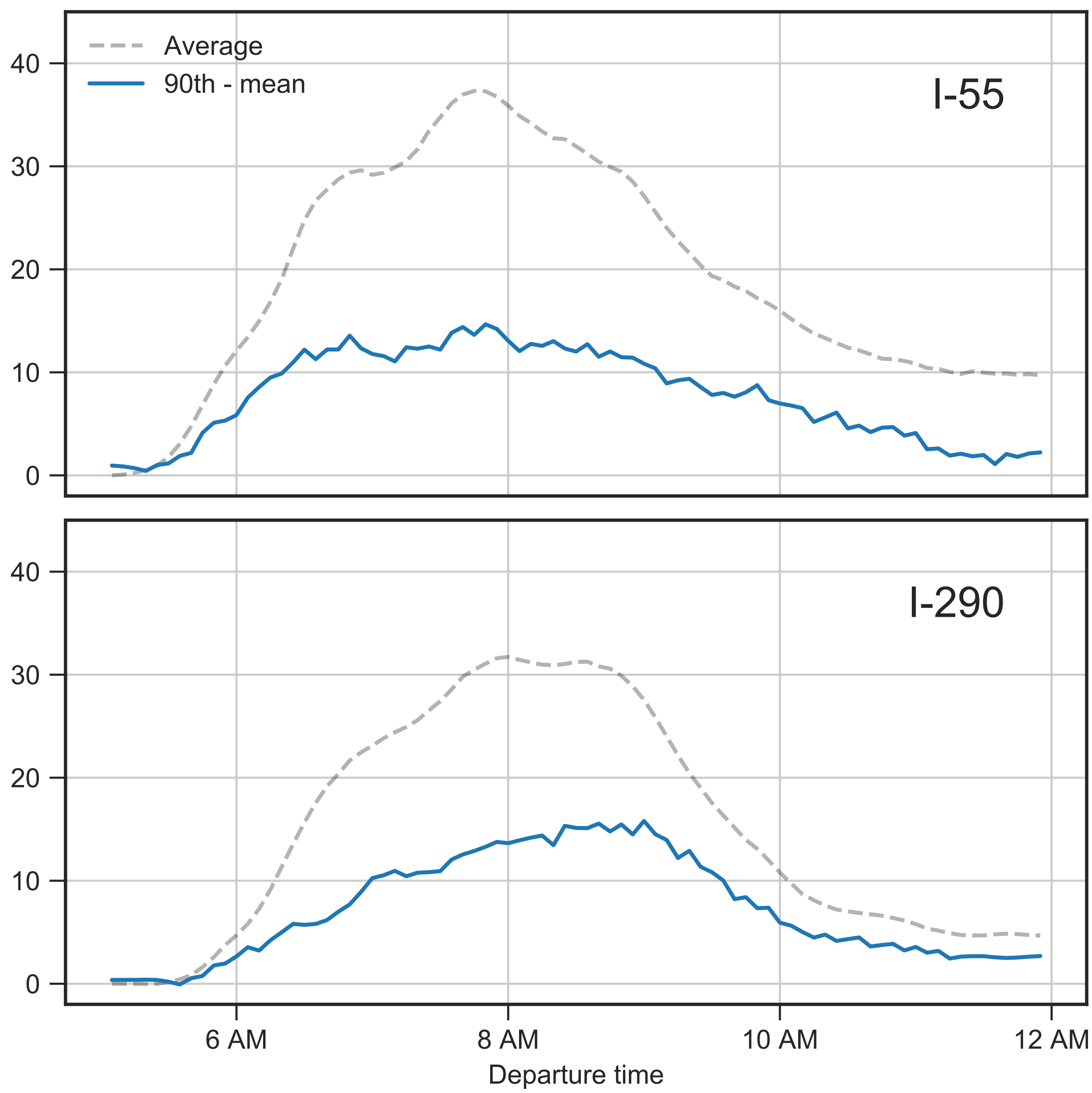

Figure 4. Average weekday variable travel time and 90th percentile minus the mean travel time by departure time for I-55N and I-290E in Chicago for 2004–2006. Data from Illinois Department of Transportation (2017).

while if an agent is always late

$$\frac{dE(T(t_d, \phi))}{dt_d} = -\frac{\gamma_i}{\alpha_i + \gamma_i}. \tag{8}$$

The slopes for always early and always late agents are identical to what we obtain in a world without uncertainty.[12] Even considering uncertainty, the slope at the start and end of rush should be much steeper than we observe in the data. Thus considering the value of uncertainty by adding uncertainty over travel times does not resolve the problem.

It is worth noting there are many possible sources of travel time uncertainty. These include demand fluctuations, as well as capacity shocks due to construction, weather, accidents (as in Fosgerau and Lindsey (2013)), or endogenous capacity breakdown (as in Hall and Savage (2017)). The analysis above covers all of these cases.

4.1.4. *Possible solution: Heterogeneous trip lengths and origin-destination pairs.* A fourth possible solution is to consider different trip lengths and origin-destination (O-D) pairs. However, the analysis above holds for any given O-D pair, and while it is possible, and even likely, that for a given O-D pair there is not someone traveling at every point in time, it is still the case that if an agent is arriving early or late, then the slope of the travel time profile at the point in time he is traveling must be tangent to his indifference curve. In the data we *never* see slopes like those we would expect given reasonable parameters, and thus considering different trip lengths and O-D pairs does not solve the problem.

4.1.5. *Possible solution: Only measuring marginal inflexibility.* A fifth possible solution is to note that we are just measuring marginal inflexibility at the equilibrium time of arrival. If we generalize the schedule delay cost function D to have more curvature, as in Lindsey (2004) and Fosgerau and Engelson (2011), then the first order conditions imply

$$T'(t) = \frac{D_i'(t - t^*)}{\alpha_i} \qquad \text{if } t \neq t^*.$$

In this case it is possible for agents to really dislike being very early or very late, but to have low marginal inflexibility because they are actually arriving very close to their desired arrival time.

This modeling change alone is not enough to resolve the puzzle. With otherwise homogeneous agents, we are still stuck concluding agents are very flexible and do not mind arriving 3 hours early.

[12] These do not match (2) and (3) because those are in terms of arrival time. For the derivation of the slope of the travel time profile in terms of departure times see Arnott et al. (1993).

One possibility is to also allow agents to have heterogeneous levels of inflexibility. With heterogeneous inflexibility, the inflexible arrive near the peak and on the margin are quite flexible, while the flexible arrive off-peak and are, likewise, on the margin quite flexible. However, maintaining the assumption of homogeneous desired arrival times and adding heterogeneity in inflexibility seems unable to fully resolve the problem. With homogeneous desired arrival times, we are still left concluding that the median agent does not mind arriving an hour early or late.

We can resolve the problem by combining the generalization of the schedule delay cost function with allowing agents to have heterogeneous desired arrival times. In this case it is possible for there to be no one who is arriving very early or very late, and so the marginal inflexibility can be low. Given how early some people arrive at work, it seems most plausible that we would need to also allow for heterogeneous inflexibility, so that there are flexible agents who are arriving extremely early or late.

4.1.6. *Possible solution: Inframarginal agents.* The sixth possible solution to the problem of travel times climbing and falling too slowly is to allow for the possibility that some drivers are inframarginal and arriving exactly on-time. This works because, as (4) shows, when agents are at a corner solution (i.e., arriving exactly on-time), their marginal rate of substitution need not be tangent to the budget line (i.e., travel time profile). As a result, we are only observing the inflexibility of those drivers who choose to arrive early or late. As the least inflexible drivers will choose to be early or late, the puzzle turns into the not particularly puzzling "the least inflexible drivers are not very inflexible."

In order to have inframarginal drivers we need to have a continuum of desired arrival times. This is necessary because for a positive measure of agents to arrive exactly on time, there must be a positive measure of desired arrival times. This modeling assumption, with otherwise homogeneous agents, appeared in the initial papers using the bottleneck model (Vickrey 1969; Hendrickson and Kocur 1981), but was subsequently dropped as it did not affect equilibrium outcomes.[13]

However, when agents are heterogeneous allowing for a continuum of desired arrival times does affect equilibrium outcomes. As Newell (1987) shows, adding this assumption means equilibrium travel times and tolls only depend on the preferences of some drivers, which allows us to rationalize our priors about agent inflexibility with the empirically observed travel time profile.

[13]The only exceptions are Newell (1987), de Palma and Lindsey (2002), and Hall (2017b).

This is also analytically tractable. With otherwise homogeneous agents it is possible to have fairly flexible specifications of desired arrival times and still find simple solutions (see Hendrickson and Kocur (1981)). Even with heterogeneous agents it is possible to find closed-form solutions for the equilibrium both when the road is free and tolled, as while as for pricing a portion of the lanes, given the assumption that desired arrival times are uniformly distributed (see Hall (2017b)).

Furthermore, the assumption of a continuum of desired arrival times is more reasonable than it initially sounds. While it is unlikely there is anyone who wants to arrive at work at 7:34:21.3, what matters for the model is when agents want to arrive at the end of the bottleneck or highway, not when they want to arrive at work. Because the distribution of distances between the end of the bottleneck and work is continuous, the distribution of desired arrival times at the end of the bottleneck is also continuous.

4.1.7. *Possible solution: All drivers are inframarginal.* An extreme special case of the previous solution would be to assume that all, or nearly all, drivers are inframarginal with respect to the choice of when to arrive. This requires allowing for heterogeneous desired arrival times. Were all drivers inframarginal, then by the logic in Section 4.1.6 the travel time profile only gives bounds on driver preferences. This solves all three problems, as in equilibrium the travel time profile could be relatively flat, leading everyone (or almost everyone) to conclude it is not worthwhile to leave early or late.[14]

That said, this seems implausible for medium and large cities. Hall (2017a) reports that 57 percent of those on California SR-91 actually leave early or late to avoid traffic, which means a majority of drivers are marginal. It also seems unlikely that many drivers have a desired arrival time of 6 a.m., and so the drivers doing so are likely marginal.

For small cities, this may very well hold. In these cities travel times are higher at the peak of rush hour, but perhaps not by enough to induce anyone to leave early or late.

[14]In the bottleneck model this requires that the density of desired arrivals at any point during rush hour exactly equals road capacity, and causes travel times to be indeterminant (that is, there is not a unique equilibrium). The model of Henderson (1974) and Chu (1995) has a unique equilibrium when using this solution. In this model, congestion only depends on the number of drivers arriving at a given time. Travel times rise and fall based on the demand for a given arrival time, without necessarily requiring any agents to arrive early or late.

4.1.8. *Comparing solutions.* Both generalizing the schedule delay cost function and having a continuum of desired arrival times solve this first problem. I prefer solving the problem by having a continuum of desired arrival times, and thus having inframarginal agents for three reasons. First, to solve the problem by generalizing the schedule delay cost function also requires heterogeneous arrival times, so adding a continuum of desired arrival times requires the fewest changes to the model. Second, it is tractable. Third, it matches our intuition that very inflexible agents exist and that many people strictly prefer their chosen arrival time over all others.

That said, allowing β/α and γ/α to be unreasonably small does allow models with homogeneous agents to give reasonable predictions for travel times. There are likely situations where doing so is appropriate.

4.2. **Slope after peak too small relative to before peak.** The second problem is that the slope of the travel time profile after the peak is too small relative to the slope before the peak. By (2) and (3), the ratio of the slope before and after is γ/β, and measures the cost of being late relative to being early. This ratio being small in absolute value suggests agents do not mind arriving late.

4.2.1. *Possible solution: Many agents do not mind arriving late.* One interpretation of this result is that the marginal driver who is late incurs lower schedule delay costs than the marginal driver who is early. Furthermore, recalling the distinction between long and short-run preferences, being late does not necessarily mean literally arriving late to an appointment, but can mean you would prefer to go to the doctor at 9 a.m. but instead schedule the appointment for 11 a.m. to avoid traffic. You arrive exactly on-time to your 11 a.m. appointment, but still have schedule delay costs.

In addition, everyday experience with the frequency of late arrivals implies many people prefer arriving late to arriving early, at least for many non-work trips.

There are three ways to add agents who do not mind being late to the model. The first is to allow for heterogeneity in γ/β so that some agents find being late very costly, while others will not mind being late. Those who find it costly will arrive early, while those who do not mind being late will arrive late. For the bottleneck model, Arnott et al. (1994) shows how to solve for equilibrium analytically when this is the only source of heterogeneity in agent preferences and

Liu et al. (2015) shows how to solve for equilibrium numerically when there are additional dimensions of heterogeneity.

A second is to explicitly model the difference between the long and short run.

However, a simpler approach is to simply use a low value for γ/β. While this amounts to assuming there are no agents who hate arriving late, this only matters if in the counterfactual policy there are agents who switch from arriving early to arriving late, or vice-versa. If the policy does not change the fraction of agents who are late (conditional on other preference parameter values), then this assumption simplifies analysis without compromising results.

4.2.2. *Possible solution: Add value of reliability.* Adding the value of reliability by allowing for uncertainty in travel times might help with this problem. If there are no agents who are always late, then the slope after the peak can be less steep than in the deterministic model.

The intuition for this is that when an agent who is sometimes early, and sometimes late, makes a small delay in their departure time, they decrease their expected schedule delay early (a good thing) while increasing their expected schedule delay late (a bad thing). Thus for a given change in departure time, total schedule delay costs change by less than they would in a model without uncertainty. In order to keep agents on their first order condition, this means expected travel times likewise change by less than they would in a model without uncertainty.

This result for the bottleneck model is implicit in Proposition 1 of Arnott et al. (1999). To derive it with more general assumptions on the congestion technology we add the assumption that $T(t_1,\phi_1) > T(t_2,\phi_1) \Leftrightarrow T(t_1,\phi_2) > T(t_2,\phi_2)$. This implies that if we exogenously increase the travel time at t_d for one ϕ, we have increased it for all ϕ, and so for any set $\Phi \subset \mathbb{R}$

$$\frac{dE\left(\frac{\partial T(t_d,\phi)}{\partial t_d}\middle|\,\phi\in\Phi\right)}{dE\left(\frac{\partial T(t_d,\phi)}{\partial t_d}\right)} > 0.$$

Next, note $\hat{\phi}(t_d)$ is implicitly defined by

$$t_d + T(t_d,\hat{\phi}(t_d)) = t^*,$$

and by the implicit function theorem

$$\frac{d\hat{\phi}(t_d)}{dt_d} = -\left(1+\frac{\partial T(t,\hat{\phi}(t_d))}{\partial t_d}\right)\left(\frac{\partial T(t,\hat{\phi}(t_d)}{\partial \phi}\right)^{-1} \leq 0.$$

By the implicit function theorem and (6)

$$\frac{d^2E\left(T(t_d,\phi)\right)}{dt_d^2} = (\beta+\gamma)\left(1+\frac{\partial T}{\partial t_d}\right)\frac{d\hat{\phi}}{dt_d} \times \left[(\alpha-\beta)F\left(\hat{\phi}(t_d)\right)\frac{dE\left(\frac{\partial T(t_d,\phi)}{\partial t_d}\middle|\,\phi\leq\hat{\phi}(t_d)\right)}{dE\left(\frac{\partial T(t_d,\phi)}{\partial t_d}\right)} +(\alpha+\gamma)\left[1-F\left(\hat{\phi}(t_d)\right)\right]\frac{dE\left(\frac{\partial T(t_d,\phi)}{\partial t_d}\middle|\,\phi>\hat{\phi}(t_d)\right)}{dE\left(\frac{\partial T(t_d,\phi)}{\partial t_d}\right)}\right]^{-1} \leq 0. \quad (9)$$

Thus expected travel times are concave, and there are not just two slopes. Combining this with (8) shows that if the last agent to arrive is not always late, then the slope of the travel time profile is less than it is in the model without uncertainty.

For it to be plausible that no agents are always late, agents must have heterogeneous desired arrival times.

While a structural congestion model has yet to be solved with uncertainty in travel times and heterogeneous desired arrival times, it seems unlikely that there are no agents who are always late unless the cost of being late is very large. But if the cost of being late was large, the slope after the peak of rush hour would be steep. Thus it seems unlikely that adding uncertainty resolves the problem of the slope after the peak being small relative to the slope before the peak.

4.2.3. *Possible solution: All drivers are inframarginal.* As discussed in Section 4.1.7, if all agents are inframarginal then the first-order conditions imply that the travel time profile only provides bounds on agents' preferences. However, as discussed earlier, the data suggests many drivers are marginal.

4.2.4. *Comparing solutions.* My preferred solution is to use a low value for γ/β as it is tractable without compromising results. It is uncertain whether adding uncertainty resolves the problem, and adding uncertainty makes the model significantly less tractable.

4.3. **Slope flat at peak.** The third problem is that the slope is relatively flat at the peak of rush hour, while the model implies it should reach a sharp peak.

4.3.1. *Possible solution: Add value of reliability.* Adding the value of reliability by allowing for uncertainty solves this problem. As implied by (9), near the peak, small changes in departure time can lead to imperceptible changes in expected schedule delay costs as decreases in expected schedule delay early are offset by increases in schedule delay late, or vice versa. This leads to agents being relatively indifferent over a variety of arrival times near the peak, and thus a relatively flat peak.

4.3.2. *Possible solution: Only measuring marginal inflexibility.* Allowing for non-piecewise linear schedule delay costs also solves this problem. With non-linear schedule delay costs agents do not find it particularly costly to arrive 10 minutes early or late, even though they hate being an hour early or late. However, for non-linear schedule delay costs with homogeneous agents to explain the travel times in Figures 1, 2, and 3 requires assuming agents do not mind arriving thirty minutes early or late. This is implausible. However, with heterogeneous desired arrival times it is possible for the window of indifference to be much smaller.

A similar solution would be to keep the piecewise linear schedule delay costs, but add the assumption that agents are indifferent between arrivals in a small window around their desired arrival time. As with non-piecewise linear schedule delay costs, in order to have a reasonably small window of indifference we must combine it with heterogeneous desired arrival times.

4.3.3. *Possible solution: All agents at peak are inframarginal.* Again, allowing for inframarginal agents can help. A less strict version of the solution in Section 4.1.7, would be to assume all those arriving at the peak of rush hour are inframarginal with respect to the choice of when to arrive. As before, this requires heterogeneous desired arrival times so a positive measure of agents can arrive exactly on-time.

In contrast to the other possible solutions, this solution does depend on the congestion technology. For those arriving at the peak to all be inframarginal, the congestion technology must allow for greater throughput when travel times are high. This could happen either because of the traditional trade-off between throughput and speed on a single route, or because, as Akbar and Duranton (2017) argue, as travel times climb, additional routes are used, and so system capacity increases.[15]

Combining heterogeneous desired arrival times with a congestion technology that allows greater throughput when travel times are high solves the problem of the flat peak because it allows almost everyone who wishes to arrive at the peak of rush hour (say, 8–9 a.m.) to do so. Thus, at the peak of rush hour the marginal

[15]This point is implicit in models of route choice when the routes do not have identical free flow travel times, such as the models of Pigou (1920), Arnott et al. (1990), and Liu and Nie (2011).

agent who is early or late is very flexible, and so the slope of the travel time profile is very flat. In contrast, at other times system capacity is lower and so more agents must arrive early or late. The marginal agent who is arriving is less flexible and so the slope is steeper.

4.3.4. *Comparing solutions.* All the proposed solutions work in this case, though allowing for agents to be indifferent over a small window and having heterogeneous desired arrival times is likely the most tractable. Fortunately, after fixing the first and second problems, the magnitude of error introduced by the third error is small. To see this, consider drawing the triangle that best fits Figures 1, 2, or 3. While it misses the flatness of the peak, it matches significantly better than the model does under the current assumptions.

5. Why these problems matter

These failures of the standard structural congestion models to fit the data limit our ability to use these models to quantify magnitudes and change qualitative results with heterogeneous agents. In this section I give examples of both.

Both of the examples use the congestion technology of the bottleneck model. In the bottleneck model, travel along the road is uncongested, except for a single bottleneck through which at most s vehicles can pass per unit time. When the departure rate, $\rho(t)$, exceeds s, a queue develops. Denoting queue length as $Q(t)$, variable travel times are given by, $T^v(t) = Q(t)/s$.

As an example of how our model's failure to match the data using reasonable parameter values affects our ability to quantify magnitudes, consider the task of measuring the social welfare gains from congestion pricing. In the bottleneck model with homogeneous agents the social welfare gains from adding first best tolls equal the total variable travel time multiplied by the value of time (Arnott et al. 1993). As Figure 1 vividly demonstrates, our model fails to accurately predict total variable travel time, and as a result, we overestimate the welfare gains by a factor of 23.[16]

[16] In Figure 1, total variable travel time is the area under the curve. Arnott et al. (1993) show the welfare gains equal $(1/2)(N^2/s)[\beta\gamma/(\beta+\gamma)]$. I use this equation to calculate the welfare gains using (1) the standard assumptions for β/α and γ/β, and choosing N/s to match (a) the length of rush hour or (b) peak travel time, and (2) choosing β/α, γ/β, and N/s to match the length of rush hour, peak travel time, and when the peak occurs. In the text I report the comparison of (1a) to (2), if instead we compare (1b) to (2) we *underestimate* the welfare gains by a factor of 0.55.

Furthermore, inasmuch as these empirical problems imply the standard structural models of congestion are wrong in important ways, and inasmuch as changing our model changes our qualitative results, then these empirical problems change our qualitative results. As an example, consider the distributional effects of congestion pricing with heterogeneous agents. When we add inframarginal agents by assuming a continuum of desired arrival times, we add agents who strongly prefer their current arrival times. A policy that changes when they arrive hurts these agents significantly; and as Hall (2017b) shows, adding inframarginal agents shrinks the set of parameter values for which congestion pricing generates a Pareto improvement.

Were we to change the model by allowing for non-piecewise linear schedule delay costs we would likely get the same outcome. This modeling change preserves the intuition behind the result in Hall (2017b). It is still the case that we have added inflexible agents who strongly prefer their current arrival times, and so a policy that changes arrival times hurts these agents significantly.

6. Conclusion

We need structural models of traffic congestion to answer a wide variety of policy relevant questions, and improving their ability to fit the data will help us answer these questions better.

This paper shows that while these models generate a travel time profile that has the right general shape, they fail to match observed travel times. I identify three problems with matching these models to the data: (1) travels times climb and fall far slower than the models predict, (2) travel times fall too slowly after the peak relative to how quickly they climb before the peak, and (3) travel times are essentially flat at the peak.

These problems are a consequence of how we model agent preferences, rather than the model of congestion itself, and so affect the bottleneck model, cell transmission model, hydrodynamic traffic flow model, and the models of Henderson (1974) and Chu (1995).

While we do not expect models to be perfect, I show these problems matter, and affect quantitative results with homogeneous or heterogeneous agents, as well as qualitative results with heterogeneous agents. Thus fixing them matters for our ability to meaningfully answer a wide variety of questions.

My recommendations for improving the fit of structural models of traffic congestion are as follows. If the goal is to only quantify some outcome with homogeneous agents, then it is easier to simply use significantly lower parameter values for agents' schedule delay costs than are typically assumed. If the goal is a quantitative or qualitative result with heterogeneous agents, then add the assumptions that (1) agents have a continuum of desired arrival times, which allows some agents to be inframarginal, (2) agents differ in the flexibility of their schedules, so some agents actually are inframarginal, and (3) use parameter values that imply the cost of being late is less than the cost of being early. These solutions do not fix the problem of travel times being essentially flat at the peak, however, its magnitude is small after fixing the other two. If it is to be fixed, it is most tractably fixed by allowing for agents to be indifferent between arrivals in a small window around their desired arrival time.

References

Akbar, Prottoy A. and Gilles Duranton. 2017. "Measuring Congestion in a Highly Congested City: Bogotá." *Working Paper*.

Arnott, Richard, Andre de Palma, and Robin Lindsey. 1990. "Departure Time and Route Choice for the Morning Commute." *Transportation Research Part B: Methodological* 24:209–228. doi:10.1016/0191-2615(90)90018-T.

Arnott, Richard, André de Palma, and Robin Lindsey. 1993. "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand." *American Economic Review* 83:161–179.

——— 1994. "The Welfare Effects of Congestion Tolls with Heterogeneous Commuters." *Journal of Transport Economics and Policy* 28:139–161. http://www.jstor.org/stable/20053032.

——— 1999. "Information and Time-of-Usage Decisions in the Bottleneck Model with Stochastic Capacity and Demand." *European Economic Review* 43:525–548. doi:10.1016/S0014-2921(98)00013-0.

California Department of Transportation. 2014. *Performance Measurement System*. Sacramento, California. http://pems.dot.ca.gov.

Chu, Xuehao. 1995. "Endogenous Trip Scheduling: The Henderson Approach Reformulated and Compared with the Vickrey Approach." *Journal of Urban Economics* 37:324–343. doi:10.1006/juec.1995.1017.

Daganzo, Carlos F. 1994. “The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory.” *Transportation Research Part B: Methodological* 28:269–287. doi:10.1016/0191-2615(94)90002-7.

De Borger, Bruno and Mogens Fosgerau. 2012. “Information Provision by Regulated Public Transport Companies.” *Transportation Research Part B: Methodological* 46: 492–510. doi:10.1016/j.trb.2011.11.007.

de Palma, André and Robin Lindsey. 2002. “Comparison of Morning and Evening Commutes in the Vickrey Bottleneck Model.” *Transportation Research Record: Journal of the Transportation Research Board* 1807:26–33. doi:10.3141/1807-04.

de Palma, André, Robin Lindsey, and Guillaume Monchambert. 2017. “The Economics of Crowding in Rail Transit.” *Journal of Urban Economics* 101:106–122. doi:10.1016/j.jue.2017.06.003.

de Palma, André, Robin Lindsey, and Emile Quinet. 2004. “Time-Varying Road Pricing and Choice of Toll Locations.” In *Road Pricing: Theory and Evidence* 9 of Research in Transportation Economics. Amserdam: Elsevier:107–131.

Fosgerau, Mogens and André de Palma. 2013. “The Dynamics of Urban Traffic Congestion and the Price of Parking.” *Journal of Public Economics* 105:106–115. doi:10.1016/j.jpubeco.2013.06.008.

Fosgerau, Mogens and Leonid Engelson. 2011. “The Value of Travel Time Variance.” *Transportation Research Part B: Methodological* 45:1–8. doi:10.1016/j.trb.2010.06.001.

Fosgerau, Mogens and Anders Karlström. 2010. “The Value of Reliability.” *Transportation Research Part B: Methodological* 44:38–49. doi:10.1016/j.trb.2009.05.002.

Fosgerau, Mogens and Robin Lindsey. 2013. “Trip-Timing Decisions with Traffic Incidents.” *Regional Science and Urban Economics* 43:764–782. doi:10.1016/j.regsciurbeco.2013.07.002.

Friedman, Milton. 1953. “The Methodology of Positive Economics.” In *Essays in Positive Economics*. Chicago: University of Chicago Press:3–43.

Friesz, Terry L., Ke Han, Pedro A. Neto, Amir Meimand, and Tao Yao. 2013. “Dynamic User Equilibrium Based on a Hydrodynamic Model.” *Transportation Research Part B: Methodological* 47:102–126. doi:10.1016/j.trb.2012.10.001.

Gubins, Sergejs and Erik T. Verhoef. 2014. “Dynamic Bottleneck Congestion and Residential Land Use in the Monocentric City.” *Journal of Urban Economics* 80: 51–61. doi:10.1016/j.jue.2013.09.001.

Hall, Jonathan D. 2017a. “Estimating Size and Distribution of Welfare Gains from Value Pricing.” *Working Paper*.

——— 2017b. "Pareto Improvements from Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways." *Working Paper*.

Hall, Jonathan D. and Ian Savage. 2017. "Anticipatory Pricing to Manage "Flow Breakdown"." *Working Paper*.

Han, Lanshan, Satish Ukkusuri, and Kien Doan. 2011. "Complementarity Formulations for the Cell Transmission Model Based Dynamic User Equilibrium with Departure Time Choice, Elastic Demand and User Heterogeneity." *Transportation Research Part B: Methodological* 45:1749–1767. doi:10.1016/j.trb.2011.07.007.

Henderson, Vernon J. 1974. "Road Congestion: A Reconsideration of Pricing Theory." *Journal of Urban Economics* 1:346–365. doi:10.1016/0094-1190(74)90012-6.

Hendrickson, Chris and George Kocur. 1981. "Schedule Delay and Departure Time Decisions in a Deterministic Model." *Transportation Science* 15:62–77. doi:10.1287/trsc.15.1.62.

Illinois Department of Transportation. 2017. *Travel Midwest Stats*. www.travelmidweststats.com.

Kraus, Marvin and Yuichiro Yoshida. 2002. "The Commuter's Time-of-Use Decision and Optimal Pricing and Service in Urban Mass Transit." *Journal of Urban Economics* 51:170–195. doi:10.1006/juec.2001.2242.

Lighthill, M. J. and G. B. Whitham. 1955. "On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads." *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 229:317–345. doi:10.1098/rspa.1955.0089.

Lindsey, Robin. 2004. "Existence, Uniqueness, and Trip Cost Function Properties of User Equilibrium in the Bottleneck Model with Multiple User Classes." *Transportation Science* 38:293–314. doi:10.1287/trsc.1030.0045.

Liu, Yang and Yu (Marco) Nie. 2011. "Morning Commute Problem Considering Route Choice, User Heterogeneity and Alternative System Optima." *Transportation Research Part B: Methodological* 45:619–642. doi:10.1016/j.trb.2010.11.003.

Liu, Yang, Yu (Marco) Nie, and Jonathan Hall. 2015. "A Semi-Analytical Approach for Solving the Bottleneck Model with General User Heterogeneity." *Transportation Research Part B: Methodological* 71:56–70. doi:10.1016/j.trb.2014.09.016.

Newell, Gordon F. 1987. "The Morning Commute for Nonidentical Travelers." *Transportation Science* 21:74–88. doi:10.1287/trsc.21.2.74.

Nie, Yu (Marco). 2015. "A New Tradable Credit Scheme for the Morning Commute Problem." *Networks and Spatial Economics* 15:719–741. doi:10.1007/s11067-013-9192-8.

Peer, Stefanie, Erik Verhoef, Jasper Knockaert, Paul Koster, and Yin-Yen Tseng. 2015. "Long-Run Versus Short-Run Perspectives on Consumer Scheduling: Evidence from a Revealed-Preference Experiment Among Peak-Hour Road Commuters." *International Economic Review* 56:303–323. doi:10.1111/iere.12103.

Pigou, Arthur Cecil. 1920. *The Economics of Welfare*. London: Macmillan and co., ltd.. 1st edition.

Richards, P. I. 1956. "Shock Waves on the Highway." *Operations research* 4:42–51.

Silva, Hugo E., Erik T. Verhoef, and Vincent A. C. van den Berg. 2014. "Airlines' Strategic Interactions and Airport Pricing in a Dynamic Bottleneck Model of Congestion." *Journal of Urban Economics* 80:13–27. doi:10.1016/j.jue.2013.08.002.

Small, Kenneth A. 1982. "The Scheduling of Consumer Activities: Work Trips." *American Economic Review* 72:467–479. http://www.jstor.org/stable/1831545.

Small, Kenneth A., Clifford Winston, and Jia Yan. 2005. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica* 73: 1367–1382. doi:10.1111/j.1468-0262.2005.00619.x.

Sullivan, Edward. 1999. *State Route 91 Impact Study Datasets*. San Luis Obispo, California: California Polytechnic State University. http://ceenve3.civeng.calpoly.edu/sullivan/sr91/.

Takayama, Yuki and Masao Kuwahara. 2017. "Bottleneck Congestion and Residential Location of Heterogeneous Commuters." *Journal of Urban Economics* 100: 65–79. doi:10.1016/j.jue.2017.05.001.

Tian, Li-Jun, Hai Yang, and Hai-Jun Huang. 2013. "Tradable Credit Schemes for Managing Bottleneck Congestion and Modal Split with Heterogeneous Users." *Transportation Research Part E: Logistics and Transportation Review* 54:1–13. doi:10.1016/j.tre.2013.04.002.

Ukkusuri, Satish V., Lanshan Han, and Kien Doan. 2012. "Dynamic User Equilibrium with a Path Based Cell Transmission Model for General Traffic Networks." *Transportation Research Part B: Methodological* 46:1657–1684. doi:10.1016/j.trb.2012.07.010.

U.S. Census Bureau. 2016. *S0801: Commuting Characteristics by Sex*. 2011-2015 American Community Survey 5-Year Estimates. https://factfinder.census.gov/bkmk/table/1.0/en/ACS/15_5YR/S0801/0100000US.

U.S. Department of Transportation. 2009. *2009 National Household Travel Survey*. Washington, D.C. http://nhts.ornl.gov.

van den Berg, Vincent and Erik T. Verhoef. 2011. "Winning or Losing from Dynamic Bottleneck Congestion Pricing?: The Distributional Effects of Road Pricing with Heterogeneity in Values of Time and Schedule Delay." *Journal of Public Economics* 95:983–992. doi:10.1016/j.jpubeco.2010.12.003.

Vickrey, William S. 1969. "Congestion Theory and Transport Investment." *American Economic Review* 59:251–260. http://www.jstor.org/stable/1823678.

——— 1973. "Pricing, Metering, and Efficiently Using Urban Transportation Facilities." *Highway Research Record* 476:36–48. http://pubsindex.trb.org/view.aspx?id=92531.

Xiao, Ling-Ling, Ronghui Liu, and Hai-Jun Huang. 2014. "Stochastic Bottleneck Capacity, Merging Traffic and Morning Commute." *Transportation Research Part E: Logistics and Transportation Review* 64:48–70. doi:10.1016/j.tre.2014.02.003.