Endogenous scheduling preferences and congestion[1]

Mogens Fosgerau[2]
Technical University of Denmark
Centre for Transport Studies, Sweden
Ecole Normale Supériore, France

Kenneth Small[3]
University of California, Irvine
Resources for the Future, Washington, D.C.

May 1, 2012

[1]We are grateful to Richard Arnott and Clifford Winston for comments. We also wish to thank conference participants at the Kuhmo-Nectar Conference in Valencia 2010.
[2]mf@transport.dtu.dk
[3]ksmall@uci.edu

Abstract

The location of activities in time is an under-researched question throughout economics. We present a dynamic model of commuting under urban congestion in which workers care about leisure and consumption. Implicit preferences for the timing of the commute form endogenously due to temporal agglomeration economies at home and work. Equilibrium exists uniquely and is indistinguishable from that of a generalized version of the classical Vickrey bottleneck model, based on exogenous trip-timing preferences; but optimal policies cannot be predicted correctly from the classical model. Tolling may make travelers better off, even without considering the use of revenues.

Keywords: urban congestion; agglomeration; endogenous preferences; scheduling preferences; bottleneck
JEL codes: D11, R41

1 Introduction

The scheduling of people's activities determines many economic actions and investment needs. In particular, humans' annoying habit of all wanting to do similar things at the same time results in many significant expenses, whether it be sports stadia, concert halls, convention centers, or peak-capacity communications links and traffic arteries. Yet standard economic tools do not deal with these activities well. Just as Starrett (1974) and Krugman (1991) have famously demonstrated that by breaking the conventional convexity assumptions of economic theory, spatial choices create new roles for such phenomenon as product differentiation, pecuniary externalities, and spatial concentration, so the inability to smooth the flow of activities across a typical day creates new phenomena.

In particular, congestion in transportation results because the demands for moving people and goods are agglomerated in space and time. Thus understanding it requires being able to model explicitly how those demands, and the congestion resulting from them, are simultaneously determined. Of the two types of agglomeration, the spatial type is much better understood (see e.g. Rosenthal and Strange, 2004). Researchers have developed many theoretical and empirical results about the strength of such spatially agglomerating forces as labor market pooling, knowledge transmission, and building of trust — each at a variety of geographical levels including regions, metropolitan areas, urban subcenters, and even small industrial districts.[1] In some cases, explicit models can be solved to explain complex equilibrium spatial patterns, such as those studied by Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002) concerning the internal structure of metropolitan areas. Most of this work is concerned with productivity at the workplace; some, such as Glaeser et al. (2001), also considers the value of a location to consumers.

Temporal agglomeration, by contrast, has elicited a much sparser literature. Henderson (1981) shows that if workers are more productive when large numbers are at work simultaneously, and wages reflect those productivity differences, then workers are induced to find an equilibrium that produces temporal clustering and hence traffic congestion. Wilson (1988) provides supporting empirical evidence that wages do in fact vary with work-start time as predicted by Henderson's model.[2] Vovsha and Bradley (2004) show that the timing of work trips involves preferences related to the workplace and also preferences related to the home, for example an aversion to departing from home too early in the day or returning too late. Thus, there is some support for workplace agglomeration based on productivity and some suggestive evidence that people also care about the timing of their activities at home.

However, there has been only limited success with modeling the simultaneous formation of these scheduling preferences and of congestion. The problem is difficult because it is inherently dynamic as well as nonlinear. For example, Henderson (1981) derives an equilibrium pattern from his model of workplace productivity, but he is forced to assume travel times are determined in such as way that allows for overtaking of earlier vehicles by ones departing later.

[1] See for example Chinitz (1961) and Scott (1988); Anas and Kim (1996) and Helsley and Sullivan (1991); Glaeser and Gottlieb (2009) on metropolitan areas; and Krugman (1991) on regions.

[2] Arnott (2007) reviews these papers and further applications, while adding his own innovation (still within a static framework) by allowing aggregate labor supplied to be affected by congestion tolls via a reduction in their net wage. Gutierrez-i-Puigarnau and Van Ommeren (2012), on the other hand, presents evidence that indicates that the relationship may be quite weak.

The most successful theoretical models of equilibrium temporal aggregation rely instead on exogenous scheduling preferences. Vickrey (1969, 1973) and many successors such as Arnott et al. (1990, 1993), assume that each worker has a predetermined preferred work arrival time and suffers disutility from deviating from that time. These papers describe congestion as a deterministic queue behind a bottleneck, and this description has enabled them to shed light on numerous questions including the effects of heterogeneous users, parallel and serial routes, and various pricing and investment strategies. For useful reviews, see Arnott et al. (1998) or Small and Verhoef (2007).

This paper returns to the problem of understanding the origin of scheduling preferences. We address agglomeration in time not only in the workplace but also at another location, here described as "home," where non-work activities ("leisure") take place. The result is a firmer microfoundation for the demand for travel, based on a few simple technological relationships along with the assumption that people choose schedules to maximize their combined utility of work and leisure. We are able to show that equilibrium exists and explore its properties.

We do so by making strong simplifying assumptions about the nature of the agglomerative forces and of the travel network connecting the locations where they occur. Specifically, we assume that worker productivity increases with the number of people simultaneously at work, and that the utility from leisure similarly increases in the number of people simultaneously at the non-work location. The former assumption is conventional; the latter is motivated by observing that people are inherently social and in fact go to great lengths to do things together. We show that these assumptions alone are sufficient to produce temporal clustering of both work and leisure, with congested travel connecting them. The actual clock time for this clustering is indeterminate; in that sense there is an infinity of equilibria, all basically identical except for a single clock time marking the start of the first such clustering. To choose among these equilibria, we could add some small absolute advantage to certain times over others, e.g. working during daylight and sleeping during darkness. This would not be difficult, but we adopt the simpler procedure of defining a fixed time interval over which effective leisure and output are computed. This, it turns out, is sufficient to uniquely identify the times at which people travel.

We find that congestion on the trip to work depends positively on the strength of the work agglomeration effect and negatively on that of the home agglomeration effect; and vice versa for the trip home. Thus, differences in morning and afternoon congestion patterns can be explained by different strengths of the productivity advantages of agglomeration at work and at home.

We carry out a detailed comparison of our results with those of the most similar model using exogenous preferences, namely Vickrey (1973) - which is somewhat more general than the more commonly used "bottleneck model" of Vickrey (1969). It turns out that in our model, scheduling preferences of the kind assumed by Vickrey arise endogenously in equilibrium. That is, an individual taking the equilibrium pattern of work and leisure productivities as given will appear to have scheduling preferences in the form of a utility function that depends on when the commute starts and ends. These scheduling preferences belong to a general class that, as far as we are aware, comprises all specifications considered by Vickrey and later authors in the context of the bottleneck model. We derive some properties of Nash equilibrium for this general class, in order to compare its predictions to those of our model, where the scheduling preferences arise endogenously. This allows us to evaluate the errors that result if policies aimed at regulating congestion are developed assuming (incorrectly) that scheduling preferences are exogenous.

We find that the assumption of exogenous scheduling preferences would lead an analyst to underestimate the benefit of congestion tolling. Using numerical simulations, we also find that in most of the cases we analyze, such an analyst would underestimate the direct benefits to travelers (not counting toll revenues); this means that people can in fact be made better off by pricing, even when not accounting for how toll revenues are used. This suggests that one key to gaining acceptance of congestion pricing might be to explain intuitively how it would help people achieve the benefits of agglomeration at work and home.

Humans are social animals and so it is entirely natural that the scheduling preferences of one individual should depend on the scheduling choices of others. Indeed, traffic congestion may be viewed as an example of the many social interactions that shape economic behavior (Manski, 2000). In our model the interaction occurs roughly at the level of a city, but there are many smaller-scale interactions that may also lead to situations where people must trade off a desire to be together against resource barriers to doing so. For example, one could consider the problem of arranging appointments between small groups of people who compete with other groups for meeting space: their apparent preferences over meeting schedules might arise from factors similar to those analyzed here.

2 Model with endogenous scheduling preferences

We consider a continuum of N homogenous workers. Utility is a differentiable and strictly increasing function of "effective leisure" H produced at home and of "output" W produced at work, $U(H, W)$, with the interpretation that work output is exchanged for consumption at a constant price normalized to one. We impose the following conditions on utility:

Condition 1 $U(H, W) \to -\infty$ *as either* $H \to 0$ *or* $W \to 0$.

Condition 2 *If* $H_a < H_b$ *and* $W_a > W_b$*, then*

$$\frac{U_H(H_a, W_a)}{U_W(H_a, W_a)} > \frac{U_H(H_b, W_b)}{U_W(H_b, W_b)}.$$

Condition 1 is needed only to rule out the extreme outcomes $H = 0$ or $W = 0$.[3] Condition 2 will ensure that Nash equilibrium is unique. It states that the marginal rate of substitution between leisure and consumption decreases as leisure increases and consumption decreases; together with assumptions that utility is strictly increasing and differentiable, this condition implies strict concavity of the utility function.

We now describe agglomeration. Let $[0, \Omega]$ denote the available time interval defining the morning. Worker productivity (aggregate output per worker per unit time) at any time t is positively related to the number of workers at work at that time. Since we are not trying to analyze the productivity effects of absolute city size, we will take agglomeration to depend on the *share* of work-

[3]At the cost of some inconvenience, this condition can be weakened (and cardinal utility avoided) by assuming that both $U(0, W)$ and $U(H, 0)$ are less than any value of $U(H, W)$ with $H, W > 0$.

ers who are at work at time s, $N_W(s)/N$, through an *agglomeration function* $g_W(\cdot)$. That is, each worker who is at work at time s produces output at rate $w(s) \equiv g_W[N_W(s)/N]$. We assume g_W is differentiable (hence continuous) and increasing. A worker who arrives at work at time a therefore produces total output

$$W(a) = \int_a^\Omega g_W\left(\frac{N_W(s)}{N}\right) ds. \tag{1}$$

We assume $g_W(0) = 0$, which implies that no production takes place when a single atomistic worker is alone at work. This is a strong assumption but maintained for reasons of analytical tractability; we discuss relaxing it in the Conclusion. Without loss of generality, we normalize $g_W(1) = 1$.

Similarly, effective leisure is produced in a social context, with an increasing and differentiable agglomeration function $g_H(\cdot)$, where $N_H(s)$ is the number of people at home at time s. Effective leisure for a worker is produced at time s at rate $h(s) \equiv g_H[N_H(s)/N]$, so that a worker who remains at home until time t produces effective leisure

$$H(t) = \int_0^t g_H\left(\frac{N_H(s)}{N}\right) ds. \tag{2}$$

We assume $g_H(0) = 0$ and normalize $g_H(1) = 1$. Leisure can be interpreted as home production subject to agglomeration economies: for example, families wanting to have discussions while eating dinner. Alternatively, it can be interpreted as social activities requiring travel that is uncongested and so is omitted from the model for simplicity: for example, a group of people seeing a show together at an uncongested location.

For later convenience, we define the cumulative agglomeration functions $G_W(z) = \int_0^z g_W(x)\,dx$ and $G_H(z) = \int_0^z g_H(x)\,dx$. Our definitions and normalizing assumptions imply that that $0 < G_H(1) < 1$ and $0 < G_W(1) < 1$.

We are interested in commuting between home and work. Given the symmetry with which we model effective leisure and output, it should not matter whether we consider the trip to or from work; for simplicity we consider just the first.[4] Thus

[4]We make no assumptions about when a given worker would return home. Notice we put no structure on the model that would cause an interdependence between the times at which the trip to

we require each worker to start the morning at home and end it at work. Travel between home and work occurs through a one-way bottleneck with a capacity of ψ workers per time unit. Travel time outside the bottleneck is identical for everyone at a constant value, which we normalize to zero. This normalization comes at no loss of generality so long as there is sufficient capacity for everyone to complete their travel by the end of the morning and still have time left over for leisure and work; thus we assume $\psi\Omega > N$. The queuing technology is as described by Vickrey (1969): A bottleneck is served at the rate ψ and a vertical queue (i.e., one whose physical extent can be ignored) with first-in-first-out queue discipline builds up whenever travelers arrive at the bottleneck at a rate faster than ψ.

At any time t, the number of workers at home, traveling, and at work are $N_H(t)$, $N_T(t)$, and $N_W(t)$, respectively, such that $N_H(t)+N_T(t)+N_W(t) = N$ for all $t \in [0, \Omega]$. Each worker remains at work until time Ω.

2.1 Nash equilibrium

Nash equilibrium occurs when no worker can gain from a unilateral change of departure time. With identical workers, this translates into the condition that utility achieved by choosing a given departure time is constant over the set of those times when departures actually occur, and that utility would be smaller for a departure time outside this set.

Workers depart from home (and arrive at the bottleneck) during an endogenously determined interval $[t_0, t_1]$. We will make use of the relative departure time, $t_d = t - t_0$. Let $\rho(t_d|t_0)$ be the departure rate from home at time $t_0 + t_d$, so that the cumulative number of departures is $R(t_d|t_0) \equiv \int_0^{t_d} \rho(s|t_0)\, ds$. The following lemma establishes the basic characteristics of the payoff functions the individual traveler faces as he chooses a departure time. (All proofs are given in the appendix.)

Lemma 3 *Suppose departures take place during an interval $I = [t_0, t_1] \subset [0, \Omega]$, where $t_1 = t_0 + N/\psi$, and where there is queue for all $t \in int(I)$. Then the rate of arrivals at work is ψ during interval I and zero elsewhere. For a worker*

work and the return home are undertaken; thus, if we were to model the latter, it would be a mirror image of the trip to work.

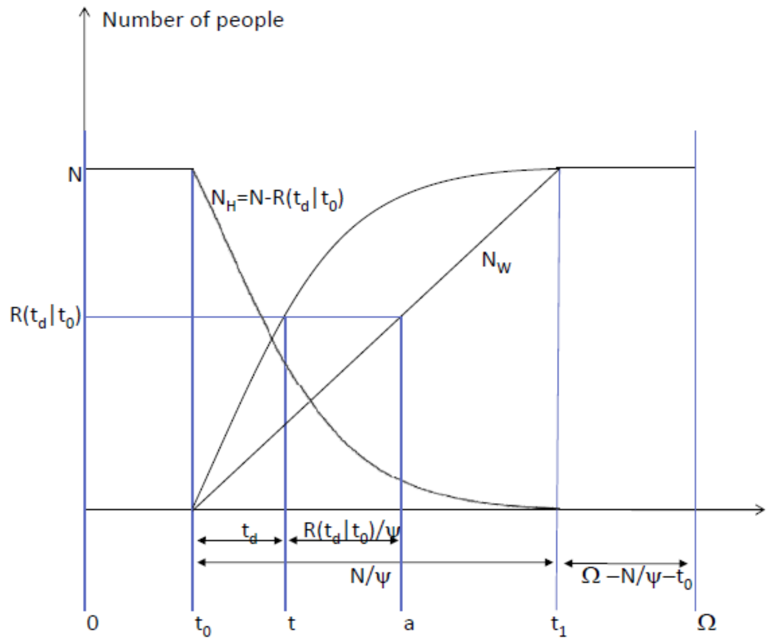

Figure 1: N_H and N_W

choosing departure time t, effective leisure is

$$H(t) = \begin{cases} t & \textit{if } t < t_0 \\ t_0 + \int_0^{t-t_0} g_H \left(1 - \frac{R(t_d|t_0)}{N}\right) dt_d & \textit{if } t \geq t_0 \end{cases} \tag{3}$$

which is (weakly) increasing and concave in t; and output at work is

$$W(a) = \begin{cases} \Omega - t_0 - \frac{N}{\psi} + \frac{N}{\psi} G_W(1) & \textit{if } a < t_0 \\ \Omega - t_0 - \frac{N}{\psi} + \frac{N}{\psi} \cdot \left[G_W(1) - G_W\left(\frac{\psi \cdot (a-t_0)}{N}\right)\right] & \textit{if } t_0 \leq a \leq t_1 \\ \Omega - a & \textit{if } a > t_1 \end{cases} \tag{4}$$

which is (weakly) decreasing and concave in a. These properties are strict for t and a in the open interval (t_0, t_1).

Lemma 3 may be visualized through Figures 1 and 2. Figure 1 shows the

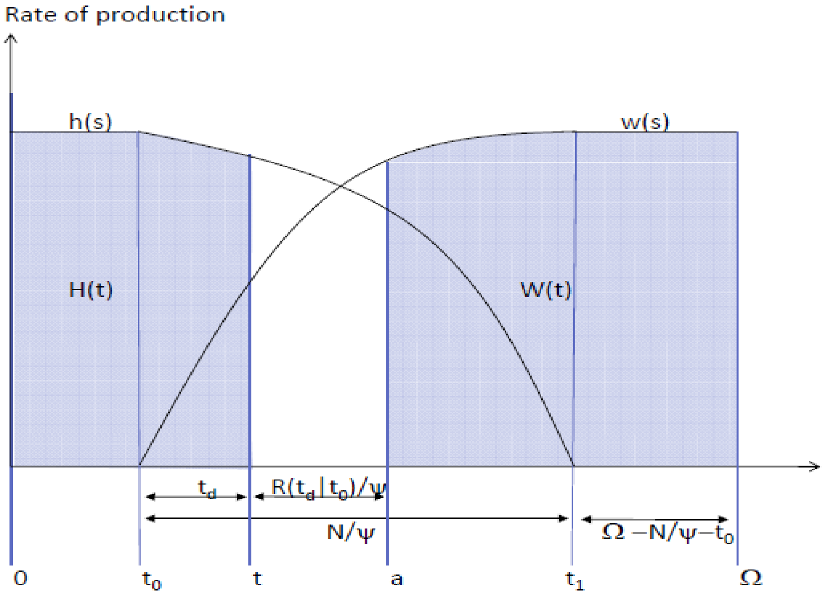

Figure 2: $h\left(s\right)$ and $w\left(s\right)$

number of workers at home and at work as a function of time, along with cumulative departures. Figure 2 depicts the resulting accumulation of work and leisure for a traveler who chooses to depart at some time $t \in [t_0, t_1]$. Effective leisure accumulates at rate $h(s) = g_H \left[1 - R\left(s - t_0 | t_0\right)/N\right]$, starting at time $s = t_0$ and continuing until $s = t$ (or until all other travelers have departed, if earlier). Work output accumulates at rate $w(s) = g_W \left[\psi \cdot (s - t_0)/N\right]$, starting at time $s = a$ when the traveler arrives (or the time the first other person arrives, if later) and continuing until the end of the morning, $s = \Omega$. The area under $w(s)$ between times t_0 and t_1 is $\int_0^{N/\psi} g_W\left(\psi x/N\right) dx = \left(N/\psi\right) G_W\left(1\right)$; that between times t_0 and a is $\left(N/\psi\right) \int_0^{\psi \cdot (a - t_0)/N} g_W\left(\psi x/N\right) dx = \left(N/\psi\right) G_W\left[\psi \cdot \left(a - t_0\right)/N\right]$. The former area gives the value of W accumulated during time interval $[t_0, t_1]$ for someone who arrives at work before or at t_0; while the difference between the two areas gives that value for someone arriving at a time $a > t_0$ such as shown in the figure. In both cases, this person also receives a contribution to W equal to the the rectangle to the right of t_1: that is, $\Omega - t_0 - N/\psi$.

The following theorem establishes key properties of equilibrium. In particular, it establishes that the departure interval has duration just long enough to allow the bottleneck to deliver all N workers to their destinations, and that it begins at a unique point in time that allows utility to be equalized for the first and last travelers.

Theorem 4 *(Nash equilibrium). Nash equilibrium exists uniquely. In Nash equilibrium, departures and arrivals take place during an interval* $I = [t_0, t_1] \subset (0, \Omega)$ *satisfying*

$$t_1 = t_0 + N/\psi. \tag{5}$$

Utility is constant on this interval and in particular is equal at the end points:

$$U[H(t_0), W(t_0)] = U[H(t_1), W(t_1)]. \tag{6}$$

The departure schedule $R(\cdot|t_0)$ *is strictly concave on* $[0, N/\psi]$ *and there is always a strictly positive queue on the interior of this interval.*

Note that the existence of agglomeration at home, as well as at work, is what enables both the location and the duration of the departure interval I to be determined uniquely. Furthermore, it is the extreme undesirability of either zero output or zero effective leisure, assumed in Condition 1, that bounds I away from the ends of the available time space $[0, \Omega]$.

3 Model with exogenous scheduling preferences

For comparison, we now describe a more conventional model where scheduling choices arise from fixed scheduling preferences – that is, preferences that are functions of clock time. Such preferences have typically been constructed around an assumed ideal work start time, with utility penalties for arriving earlier or later than that time. Occasionally, such preferences have instead been built from an assumed schedule giving instantaneous rates of accumulation of utility as functions of clock time. Here, we introduce a generalized version that incorporates either of these motivations. The key point is that preferences are defined in terms of specific times of day. In this section, we derive properties of this model; in the next section, we contrast them with our model of endogenous scheduling.

Let utility be a function $V(t,a)$ of departure and arrival times that is strictly concave, increasing in t, and decreasing in a. Define $\tilde{V}(t) = V(t,t)$, which gives utility for someone suffering no queuing delay, and assume that $\tilde{V}(t)$ attains a maximum at some value of t. Again we assume N identical users and a deterministic bottleneck with capacity ψ, and we consider Nash equilibrium in departure times. We call this model "the general Vickrey model".

In one special case, $V(\cdot)$ is derived as the integral of instantaneous utility rates $v_H(t)$ and $v_W(t)$ for time spent at home and work, respectively; see Vickrey (1973), Tseng and Verhoef (2008), and Fosgerau and Engelson (2011). In a further special case, actually a limiting case, $V(t,a)$ is linear in $(a-t)$ with slope, $-\alpha$, interpreted as unit disutility of travel time; and it is also piecewise linear in a in two segments with slopes, β and $-\gamma$, interpreted as unit disutilities of early or late arrival. See Vickrey (1969) and Arnott et al. (1993); this is sometimes called the "bottleneck model" or the "α,β,γ model".[5]

In the next theorem, we establish properties of the general Vickrey model that parallel those of our model as given in Theorem 4. We also state some additional results, which generalize results already known for the "bottleneck model".

Theorem 5 *(Properties of the general Vickrey model).*

(a) In the general Vickrey bottleneck model described by utility $V(t,a)$, Nash equilibrium occurs when departures and arrivals take place during the interval $[t_0,t_1]$ determined uniquely by (5) and by

$$\tilde{V}(t_0) = \tilde{V}(t_1) \tag{7}$$

where $\widetilde{V}(t) \equiv V(t,t)$. We write this equilibrium utility level, which is a function of ψ and N, as

$$\tilde{\Delta} = \tilde{\Delta}(\psi,N) = \tilde{V}\left[t_0(\psi,N)\right]. \tag{8}$$

[5]Vickrey (1973) showed that his specification of scheduling utility in terms of utility rates is consistent with the more conventional α,β,γ utility specification. He did not attempt to derive a scheduling equilibrium.

(b) The marginal utility gain from a capacity improvement is

$$\frac{\partial \tilde{\Delta}}{\partial \psi} = -\frac{N}{\psi^2} \frac{\tilde{V}'(t_0)\,\tilde{V}'(t_1)}{\tilde{V}'(t_0) - \tilde{V}'(t_1)} > 0 \tag{9}$$

and the marginal external utility loss from an additional traveler is

$$-\frac{\partial \tilde{\Delta}}{\partial N} = -\frac{1}{\psi} \frac{\tilde{V}'(t_0)\,\tilde{V}'(t_1)}{\tilde{V}'(t_0) - \tilde{V}'(t_1)} > 0. \tag{10}$$

(c) An optimal toll schedule $\widetilde{\tau}(t)$ (in utility units) satisfies

$$\tilde{V}(t) - \widetilde{\tau}(t) = \tilde{V}(t_0) - \widetilde{\tau}(t_0) \qquad \forall t \in [t_0, t_1]\,. \tag{11}$$

One such toll schedule also has $\widetilde{\tau}(t_0) = 0$, in which case everyone receives the same utility (before any revenue distribution) as without the toll.

(d) The welfare gain from the toll schedule with $\widetilde{\tau}(t_0) = 0$ described in part (c) is equal to the toll revenue (again in utility units), namely

$$\psi \int_{t_0}^{t_1} \left(\tilde{V}(t) - \tilde{\Delta}\right) dt. \tag{12}$$

4 Vickrey meets endogenous scheduling preferences

Individuals in our model have preferences defined only over leisure and consumption; they care indirectly about the timing of work trips only because it affects production at work and at home. Production is in turn affected by the scheduling of work trips of all other individuals through the agglomeration effects specified in our model.

However, a single individual, taking equilibrium as given, will appear to have preferences concerning the scheduling of his commute. This section will show that scheduling preferences like those specified in the general Vickrey model can appear to explain individual behavior when taking the equilibrium departure pattern as given, even if actual preferences are as we posit.

A naïve analyst observing an equilibrium departure pattern generated by our model may feel justified in applying the general Vickrey model to determine the effect of policies, since the general Vickrey model is able to generate exactly the observed equilibrium. However, a change in capacity or some other aggregate parameter will lead to a change in the equilibrium departure pattern and hence in the apparent scheduling preferences. Therefore such a naïve analyst will not be able to predict correctly the effects of a capacity change. A general result does not seem available concerning the direction of the mistakes such an analyst would make.[6] The same conclusion applies regarding the marginal external cost of additional travelers. Concerning the optimal toll, however, we can make a more definite statement: Theorem 7 shows that the naïve analyst would underestimate the welfare gain available from this toll.

Consider an individual in our endogenous scheduling model who, in Nash equilibrium, departs from home at time $t \geq t_0$ and arrives at work at time $a \leq t_1$. Inserting the appropriate portions of (3) and (4) into utility $U(H, W)$ shows that utility achieved is

$$V(t,a) = U\begin{bmatrix} t_0 + \int_0^{t-t_0} g_H\left(1 - \frac{R(s|t_0)}{N}\right) ds, \\ \Omega - t_0 - \frac{N}{\psi} \cdot \left[1 - G_W(1) + G_W\left(\frac{\psi(a-t_0)}{N}\right)\right] \end{bmatrix}. \tag{13}$$

We now consider the behavior of such an individual who regards the cumulative departure pattern $R(\cdot|t_0)$ as exogenous.

Theorem 6 *Consider a single individual who is part of a Nash Equilibrium in our endogenous scheduling model. If this individual takes the cumulative departure pattern as exogenous, he or she would act according to a utility function $V(t,a)$, for all t, a in the open interval (t_0, t_1), which meets the requirements of the general Vickrey model.*

We are then able to provide a strong result about the effect of the optimal toll in the true model compared to what a naïve analyst would predict.

Theorem 7 *(Optimal toll) Suppose the untolled equilibrium results in travel during interval $[t_0, t_1]$, so that the Vickrey model predicts an optimal toll that is zero*

[6] The numerical examples in the next section show situations where the general Vickrey model would lead to underestimates of the benefit of capacity expansion compared to the true model. We have not been able to find conditions under which such underestimation would always occur.

at t_0 and t_1 and positive elsewhere on $[t_0, t_1]$. Under endogenous scheduling preferences:

(a) The following toll produces a larger welfare gain than the maximum possible gain predicted by the the general Vickrey model:

$$\hat{\tau}(t) = \hat{V}(t,t) - \tilde{V}(t_0, t_0) \tag{14}$$

where $\tilde{V}(\cdot)$ is the Vickrey utility calculated from the observed untolled equilibrium, and $\hat{V}(\cdot)$ is the true utility when travelers depart at rate ψ starting at t_0. This toll is larger than the Vickrey toll on $(t_0, t_1]$.

(b) The optimal toll produces a welfare gain that is at least as large as that from the toll given in (14).

Part (b) of the theorem follows directly from part (a). The proof of the theorem considers a toll that removes queueing in the true model while maintaining the first departure time of the Nash equilibrium to be that in the solution in the Vickrey model. Then it is shown that this toll is larger than in the Vickrey model and hence that the efficiency gain is larger. This toll is not necessarily optimal, because it does not re-optimize the initial departure time t_0; an optimal toll would by definition result in an efficiency gain at least as large as the one just described.

We have not found an analytical solution for the fully optimal toll. We conjecture that if one exists, it contains terms related to the Pigou subsidies correcting the positive externalities of agglomeration, integrated over the period of travel.[7] In the numerical example in the next section, we determine the optimal toll numerically and find that in contrast to the Vickrey toll, it can shift t_0 substantially from its value in the untolled equilibrium.

5 Numerical example

We introduce for convenience the following notation.

[7] A simple condition on marginal utility at time t_0 does not work because as t_0 is shifted, revenues also change.

Notation 8

$$
\begin{aligned}
U^0 &= U\left[t_0, \Omega - t_0 - \frac{N}{\psi} + \frac{N}{\psi} G_W(1)\right] \\
U^1 &= U\left[t_0 + \int_0^{\frac{N}{\psi}} g_H\left(1 - \frac{R(s|t_0)}{N}\right) ds, \Omega - t_0 - \frac{N}{\psi}\right] \\
U_H^0 &= U_H\left[t_0, \Omega - t_0 - \frac{N}{\psi} + \frac{N}{\psi} G_W(1)\right]
\end{aligned}
$$

The derivatives U_H^1, U_W^0, and U_W^1 are similarly defined.

These quantities give the values of U and its derivatives for the first and last travelers, given that the first departs at t_0. Earlier, we denoted the equilibrium utility under the Vickrey interpretation of scheduling preferences by $\tilde{\Delta}$. Denote the true equilibrium utility by Δ. By definition, $\Delta = \tilde{\Delta}$ for the set of parameters under which both models are initially calibrated. But there is no reason why they should remain identical under hypothetical changes in ψ and N. The following lemma will be useful for the numerical example provided in this section.

Lemma 9 *Suppose the true model is our model of endogenous scheduling preferences. Then the Vickrey utilities for the first and last traveler vary according to $\tilde{V}'(t_0) = U_H^0$ and $\tilde{V}'(t_1) = -U_W^1$. Interpreting the equilibrium as coming from the general Vickrey model leads to apparent equilibrium utilities which would be thought to vary as:*

$$
\frac{\partial \tilde{\Delta}}{\partial \psi} = \frac{N}{\psi^2} \cdot \frac{U_H^0 U_W^1}{U_H^0 + U_W^1}, \qquad \frac{\partial \tilde{\Delta}}{\partial N} = -\frac{1}{\psi} \cdot \frac{U_H^0 U_W^1}{U_H^0 + U_W^1}. \tag{15}
$$

In order to facilitate interpretation, we compute some numerical results using a Cobb-Douglas utility function and simple power functions for $g_W(\cdot)$ and $g_H(\cdot)$:

$$
\begin{aligned}
U(H, W) &= \alpha \ln H + \ln W, && (16) \\
g_H(x) &= x^{\pi_H}, \\
g_W(x) &= x^{\pi_W}, && (17)
\end{aligned}
$$

with $\pi_W, \pi_H > 0$. We require $\pi_H < 1$ because otherwise, as we show in Appendix B, there is no solution. We set $\Omega = 5$, $N = 1$, and $\psi = 1$; thus the congested period is exactly hour long. (We then vary ψ in order to calculate the marginal benefit of capacity.)

Given some initial departure time t_0, equations (3) and (4) then give the following analytical expressions for effective leisure and work production for any departures time t$\in [t_0, t_0 + N/\psi]$:

$$H(t) = t_0 + \int_0^{t_d} \left(1 - \frac{R\left(t'|t_0\right)}{N}\right)^{\pi_H} dt'$$

$$W\left[a(t)\right] = \Omega - t_0 - \frac{1}{\psi}\left[\frac{\pi_W + R\left(t_d|t_0\right)^{\pi_W+1}}{\pi_W + 1}\right],$$

where $a(t) = t_0 + R\left[t_d|t_0\right]/\psi$ is the arrival time of the traveler departing at time t, and where $t_d = t - t_0$. If we substitute these expressions into the utility function in (16), we get utility as a function of departure time, which we write as $\widetilde{U}(t)$.

Our problem is twofold: (a) to find the cumulative departure function $R\left(t_d|t_0\right)$ that holds utility constant, and (b) to find the initial departure time t_0 that satisfies the boundary condition $R\left[(1/\psi)\,|t_0\right] = 1$. Task (a) is accomplished by setting the total derivative of $\widetilde{U}(t)$ equal to zero, yielding:

$$\frac{\alpha}{H(t)}\left[1 - R\left(\cdot\right)\right]^{\pi_H} - \frac{1}{W\left[a(t)\right]} R\left(\cdot\right)^{\pi_W} \frac{\rho\left(\cdot\right)}{\psi} = 0$$

where $(\cdot)$ is shorthand for $(t - t_0|t_0)$, and ρ is the derivative of R. We can rewrite this condition as

$$\rho\left(\cdot\right) = \frac{\psi\alpha W\left[a\left(t\right)\right]}{H(t)} \frac{\left[1 - R\left(\cdot\right)\right]^{\pi_H}}{R\left(\cdot\right)^{\pi_W}}. \tag{18}$$

Since $\rho = R'$, this is a differential equation in R, which can be solved numerically given a value of t_0.

For task (b), an outer iterative procedure is employed to find the unique value of t_0 for which total cumulative departures are exactly N. The value is unique because $H\left(t\right)$ and $W(t)$ depend positively and negatively, respectively, on t_0, so that ρ in (18) will be initially larger the larger is t_0. More details are in Appendix B.

Table 1: Simulation results

Simulation number	1	2	3	4	5	6
Parameters:						
(1) alpha	1	1	1	1	0.5	2
(2) π_H	0.5	0.5	0.5	0.75	0.5	0.5
(3) π_W	0.5	1	3	0.5	0.5	0.5
Nash equilibrium:						
(4) t_0	1.473	1.641	2.039	0.819	0.818	2.233
(5) t_d at max travel time	0.312	0.254	0.151	0.250	0.305	0.312
(6) Max travel time	0.416	0.480	0.626	0.580	0.427	0.417
(7) Total travel time	0.281	0.321	0.396	0.370	0.288	0.282
(8) $R(t_d = .1)$	0.410	0.513	0.716	0.581	0.423	0.411
Rel. components of $dU/d\psi$:						
(9) $dt_0 * (U_{H0} - U_{W0})/d\psi$	0.609	0.419	0.058	0.747	0.584	0.661
(10) $d \ln W/d\psi$	0.391	0.581	0.942	0.253	0.416	0.339
(11) Vickrey - Actual $dU/d\psi$	-0.017	-0.051	-0.111	-0.093	-0.0005	-0.058
Social optimum:						
(12) t_0	2.000	1.971	1.908	2.016	1.213	2.787
(13) $U^1 - U^0$	0	0.071	0.187	-0.040	0.004	-0.009
(14) $\Delta\hat{U}_{\min}$	0.125	0.061	-0.009	0.488	0.089	0.176
(15) Toll revenue	0.070	0.111	0.164	0.094	0.056	0.112
(16) Welfare gain	0.195	0.172	0.156	0.581	0.145	0.288

Results are shown in rows 4-8 of Table 1. Consider simulation #1, for which $\alpha = 1$ and $\pi_H = \pi_W = 1/2$. The congested period extends between times 1.473 and 2.473 in our 5-hour "morning"; that is, it is shifted earlier by 0.527 hours compared to one that is centered during the morning. Forty-one percent of travelers depart during the first six minutes, resulting in congestion reaching a peak travel time of 0.416 hours. Cumulative departures and arrivals are shown as the two left-most curves in the top panel of Figure 3.

By varying capacity ψ, we can calculate numerically the marginal value of capacity in an untolled equilibrium. The Vickrey analyst would predict a marginal value (in utility terms) as given by the first of equations (15), which is easy to calculate. But the true marginal value, obtained by differentiating $\widetilde{U}(t_0) \equiv U\left[H\left(t_0\right), W\left(t_0\right)\right]$ with respect to ψ, is

$$\frac{\partial \widetilde{U}\left(t_0\right)}{\partial \psi} = \left[U_H\left(t_0\right) - U_W\left(t_0\right)\right] \frac{\partial t_0}{\partial \psi} + \frac{\partial \ln W\left(t_0\right)}{\partial \psi}. \tag{19}$$

This equation states that the marginal utility of capacity arises from two sources: (i) the net utility change from more production and less leisure as t_0 is shifted later, and (ii) the change in production due to more people being at work at any given time relative to t_0. Rows 9 and 10 of Table 1 show the values of the two terms in this equation, relative to their sum; for simulation #1, the first dominates but not by a large amount. Row 11 shows how much this calculation differs from the standard Vickrey calculation of the value of capacity. Our main interest is in it sign: the negative values for all simulations indicate that the Vickrey model underpredicts the true value of a capacity increase.

We also compute the Vickrey toll (11). Its two terms can be calculated, using (13) and (16), as

$$\begin{aligned} \widetilde{V}\left(t\right) &= \alpha \ln\left[H\left(t\right)\right] + \ln\left[W\left(a\left(t\right)\right)\right] \\ \widetilde{V}\left(t_0\right) &= \alpha \ln\left(t_0\right) + \ln\left(\Omega - t_0 - \frac{1}{\psi} \cdot \frac{\pi_W}{\pi_W + 1}\right). \end{aligned}$$

Recall that the Nash equilibrium departure pattern $R\left(\cdot\right)$ is chosen to make $\widetilde{V}\left(t_1\right) = \widetilde{V}\left(t_0\right)$; thus the Vickrey toll begins and ends at a value of zero, because the Vick-

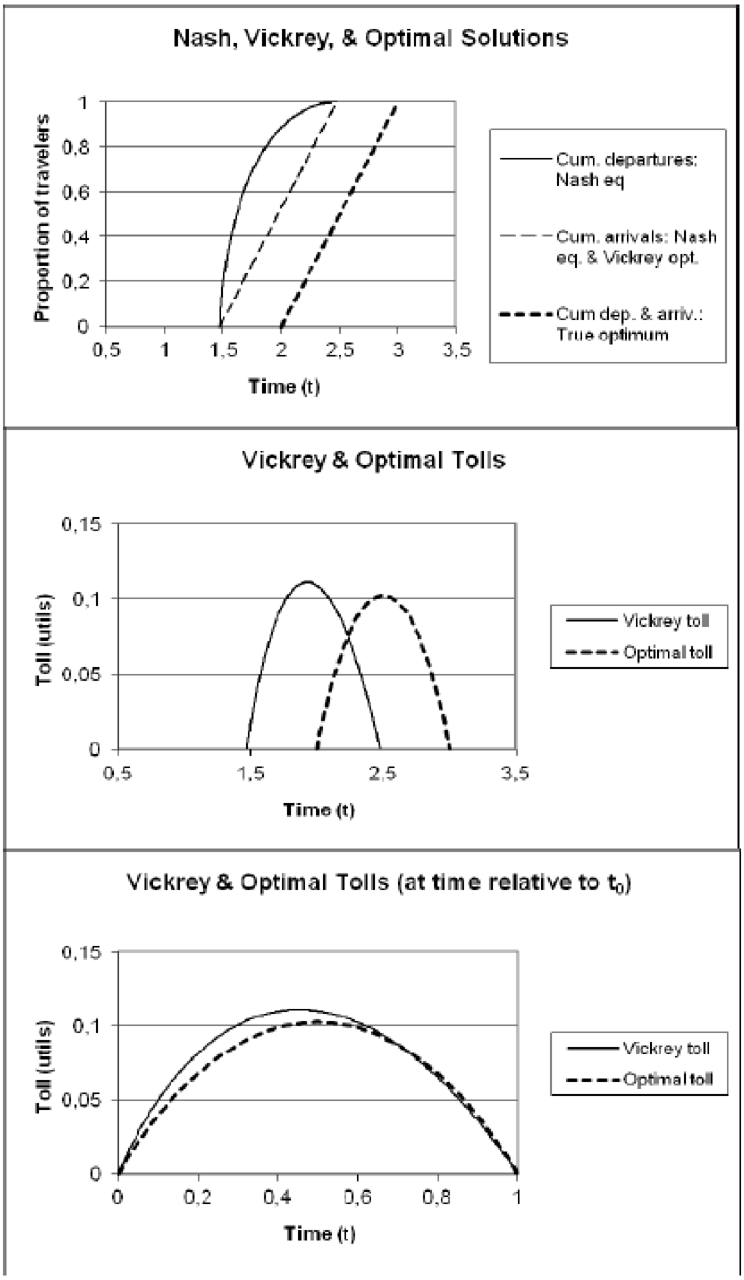

Figure 3: Graphs for Simulation #1

rey analyst assumes that the first and last travelers (who avoid queuing) care only about their schedules and therefore achieve the same utility as without the toll. The resulting toll for simulation #1 is shown as the solid curve in the the second panel of Figure 3; it rises rapidly after time t_0 and then falls more gradually from its peak value.

Finally, we have carried out a simulation of the social optimum. It is achieved by applying the optimal toll which, like the Vickrey toll, is set to maintain a constant departure rate $\rho = \psi$. In contrast to the Vickrey toll, however, the optimal toll actually achieves this departure rate. In this calculation, t_0 is set so as to maximize the average of travelers' gross utilities (i.e., their utilities before paying tolls). That value is

$$\widehat{\widehat{U}}(t) = \max_{t_0} \psi \int_{t_0}^{t_0+1/\psi} U\left[\widehat{H}(t), \widehat{W}(t)\right] dt, \tag{20}$$

where $\widehat{H}(t)$ is computed numerically from (3) but with $R(t_d|t_0) = \psi t_d$, and where $\widehat{W}(t)$ is computed as before except now t_0 is different. Details of calculating (20) are given in Appendix B. Since $U\left[\widehat{H}(t), \widehat{W}(t)\right]$ is a concave function of t, it attains its minimal value $\widehat{U}_{\min}$ at either t_0 or t_1; we fix the toll $\hat{\tau}$ to be zero at this point, which implies that the $\hat{\tau}$ is non-negative at all times in $[t_0, t_1]$. With an optimum toll, all travelers achieve the same utility, and this is equal to the net utility (utility minus toll) of any traveler and in particular it is equal to $\widehat{U}_{\min}$. We report on both $\widehat{U}_{\min}$ and $\widehat{\widehat{U}}$ in our simulation results.

Rows 12-16 provide results concerning the social optimum, and the second panel of Figure 3 shows the optimal toll itself. The optimal t_0 in simulation #1 is exactly 2.0, so that the departure interval is centered in the available time universe [0,Ω] = [0,5], reflecting the symmetry of leisure and work in this simulation (since $\pi_H = \pi_W$ and $\alpha = 1$). This optimal value of t_0 is substantially later than the Nash equilibrium value. The symmetry of this simulation causes the optimal toll to be perfectly symmetric about the midpoint of the morning (2.5); it also causes the utilities (net of toll) for the first and the last travellers to be equal, as shown in row 13, just as they are when the Vickrey toll is applied.

Simulations #2 and #3 explore successively larger values of π_W,whereas simulation #4 looks at a larger value of π_H. Simulations #5 and #6 consider a smaller and larger value, respectively, for α.

The shape of the equilibrium departure pattern, as a function of $t - t_0$, is governed mainly by the relative values of π_H and π_W. Smaller values result in the least congestion and the lowest fraction of travelers who depart early in the departure interval. The location of the rush hour within the available time shifts in opposite directions depending on which agglomeration is increased: earlier for increasing π_H and later for increasing π_W. Increasing α, the weight on household production in the utility function, makes little difference to the pattern of departures or to the amount of travel delay; but it shifts the start of the departure interval to a later time in order to restore balance between U_H and U_W.

Increasing π_W, which controls agglomeration at work, causes two changes (simulations #2 and #3). First, the optimal first departure time shifts earlier, whereas the Nash equilibrium value shifts later. In simulation #3, these opposing shifts are so strong that the socially optimal rush hour starts *earlier* than the Nash equilibrium rush hour, in contrast to all other simulations. Second, the utility difference for the last traveler, relative to the first, becomes positive, meaning the optimal toll is now higher for a departure at time t_1 than one at time t_0. These effects are reversed when it is π_H rather than π_W that is raised (simulation #4).

Simulations #5 and #6 change the relative importance of work through its weight α in the utility function. The optimal first departure is quite sensitive to α, in roughly the same way as the Nash equilibrium first departure, with the optimum about 0.4 to 0.5 hours later than the equilibrium. Increasing α, which controls the marginal utilityof leisure, shifts both the equilibrium and optimal first departure times later, allowing people to spend more time at home (simulation #6).

Row 14 reports information about $\widehat{U}_{\min}$, the minimum utility obtained in social optimum. As already noted, $\widehat{U}_{\min}$ is the utility of the traveler who is charged a toll of zero, and is therefore the utility net of toll realized by all travelers. Row 14 shows the difference between $\widehat{U}_{\min}$ and the utility obtained by travelers in Nash equilibrium. It is positive in most cases, indicating that the optimal toll leads to a strict Pareto improvement even if toll revenues are not returned to travelers. However, simulation #3 shows a contrary case, where travelers will lose in the social optimum (compared to Nash equilibrium) if toll revenues are not returned.

Figures 3-5 present graphs of the departure pattern $R(\cdot)$ and of the Vickrey and optimal tolls, for selected simulations. In all three cases, the Vickrey toll differs greatly from the optimal toll, though for different reasons. In simulation #3, with high work agglomeration, the Vickrey analyst predicts the optimal departure pattern quite closely, but is wildly off on the size and shape of the toll that will achieve it. In simulation #4, with high leisure agglomeration, it is just the opposite: the Vickrey analyst gets the shape of the toll roughly correct, but its position in time is much too early, resulting in people spending too little time at home and thus not taking advantage of the high marginal utility of leisure time with these parameters.

6 Conclusion

This paper has presented a dynamic model of traffic congestion in which scheduling preferences arise endogenously. A naïve Vickrey-like analyst — observing one equilibrium and assuming scheduling preferences to be exogenously given — would find that to be in accordance with his model. But because he ignores the positive temporal agglomeration externalities associated with being at home or at work, this analyst will underestimate the benefits of a toll that reduces queueing during the commute and will make potentially very large errors in predicting the effect of policies such as capacity expansion and tolling. For some parameter sets, such an analyst would apply a toll schedule and/or aim for a departure pattern that is quite far removed from the optimal one. So a policy conclusion is that a gradual approach to introducing a policy such as road pricing is advisable, since that allows the consequences to be observed as one goes along.

The general conclusions depend on the result of ignoring an externality (agglomeration) and therefore seem likely to be quite robust. Hence we may expect that these conclusions would survive if we relax the assumptions that we have made for the sake of analytical tractability.

We have assumed, first, that the productivity of effective leisure and work depend solely on the share of workers at home and at work at each point in time. We have furthermore assumed that a worker working alone is completely unproduc-

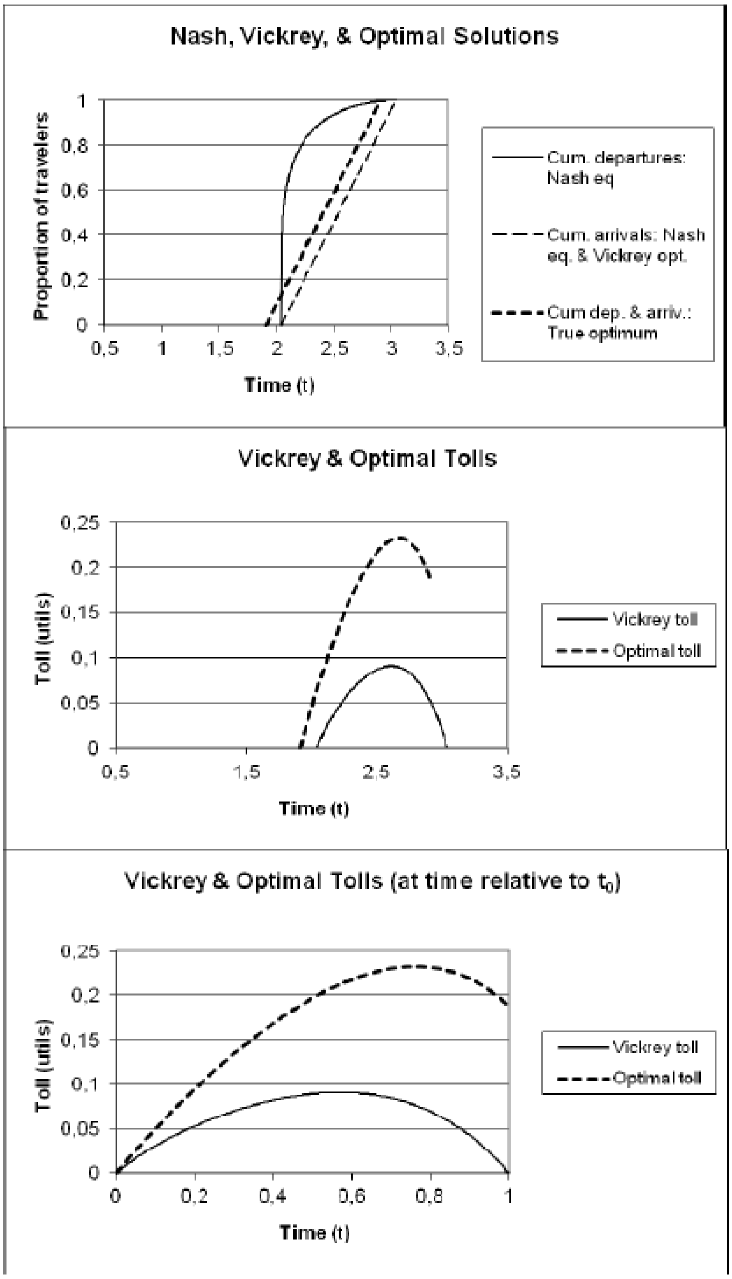

Figure 4: Graphs for Simulation #3

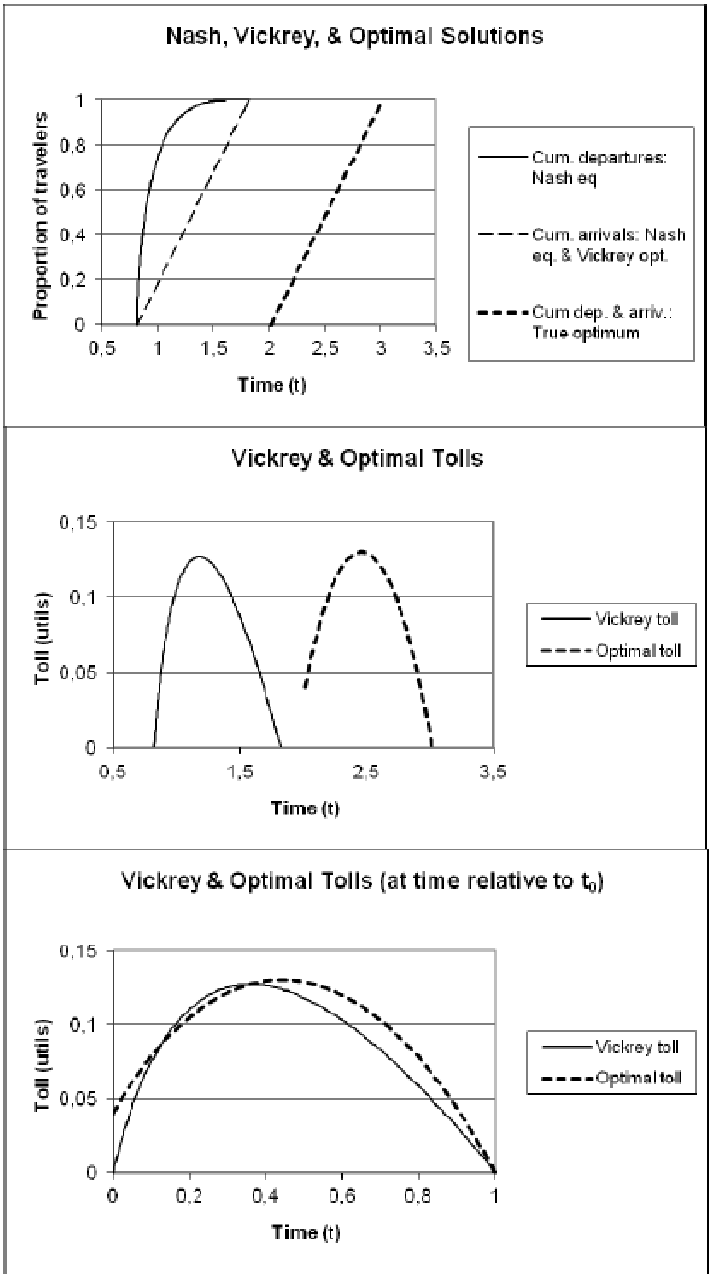

Figure 5: Graphs for Simulation #4

tive and similarly that a worker derives no effective leisure from solitary leisure time. The essential mechanism here is the positive externality associated with being at home or at work. As long as this remains, it seems possible to relax assumptions regarding productivity without affecting the validity of our general conclusions.

Second, we have made assumptions that allow us to ignore the evening commute. The model describes people as staying at work until some common time and is silent about what happens thereafter. Incorporating the evening commute would not affect the conclusions that depend on a Vickrey-like analyst ignoring agglomeration externalities.

Third, we have assumed that the demand for travel is completely inelastic. As has been noted earlier, this reduces the importance of the assumption that productivity depends on the share of all workers present at home or at work. Allowing for elastic demand would likely enable us to separate the problem into two parts, similarly to Arnott et al. (1993), including removing the indeterminacy on the level of the toll. However we would then have to deal explicitly with city-size agglomeration effects.

Fourth, the paper assumes that individuals are identical. Allowing for heterogeneity would raise the issue of how workers sort in equilibrium, ie. in which sequence they pass the bottleneck. The presence of heterogeneity would add a possibility for efficiency gain from a toll that causes workers to sort in a more efficient way. It seems likely that our general conclusions would remain valid due to the remaining central premise that positive agglomeration externalities shape scheduling preferences.

There are no firms in our model. Firms could seek to internalize agglomeration externalities by paying a wage that depends on the time a worker arrives at the work place (instead of the integral of the productivity rate over time spent at work, as implicitly assumed here). It would be an interesting extension of the present analysis to include such aspects of firm behavior, but probably also very difficult except in the trivial case of a single firm fully internalizing the externality.

The model with endogenous scheduling preferences generates an equilibrium that is indistinguishable from a model with exogenous scheduling preferences. It is hence not possible to falsify the latter model using only observation of individual choices in a single equilibrium; rather, in order to identify endogeneity, it is necessary to compare different equilibria. It may be possible to employ such an identification strategy empirically, for example by using capacity expansion or the introduction of a road pricing scheme as an exogenous instrument in an empirical investigation explaining variations in the temporal shape of the morning peak.

References

Anas, A. and Kim, I. (1996) General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration *Journal of Urban Economics* **40**(2), 232–256.

Arnott, R. A. (2007) Congestion tolling with agglomeration externalities *Journal of Urban Economics* **62**(2), 187–203.

Arnott, R., de Palma, A. and Lindsey, R. (1990) Economics of a bottleneck *Journal of Urban Economics* **27**(1), 111–130.

Arnott, R., de Palma, A. and Lindsey, R. (1993) A structural model of peak-period congestion: A traffic bottleneck with elastic demand *American Economic Review* **83**(1), 161–179.

Arnott, R., de Palma, A. and Lindsey, R. (1998) Recent developments in the bottleneck model *in* K. Button and E. T. Verhoef (eds), *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility* Edward Elgar Cheltenham, UK pp. 79–110.

Chinitz, B. (1961) Contrasts in Agglomeration: New York and Pittsburgh *The American Economic Review* **51**(2), 279–289.

de Palma, A. and Fosgerau, M. (2011) Dynamic and static congestion models: a review *in* A. de Palma, R. Lindsey, E. Quinet and R. Vickerman (eds), *A Handbook of Transport Economics* Edward Elgar chapter 9.

Fosgerau, M. and Engelson, L. (2011) The value of travel time variance *Transportation Research Part B: Methodological* **45**(1), 1–8.

Fujita, M. and Ogawa, H. (1982) Multiple equilibria and structural transition of non-monocentric urban configurations *Regional Science and Urban Economics* **12**(2), 161–196.

Glaeser, E. L. and Gottlieb, J. D. (2009) The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States *Journal of Economic Literature* **47**(4), 983–1028.

Glaeser, E. L., Kolko, J. and Saiz, A. (2001) Consumer City *Journal of Economic Geography* **1**(1), 27–50.

Gutierrez-i-Puigarnau, E. and Van Ommeren, J. N. (2012) Start Time and Worker Compensation Implications for Staggered-Hours Programmes *Journal of Transport Economics and Policy (JTEP)* **46**(2), 205–220.

Helsley, R. W. and Sullivan, A. M. (1991) Urban subcenter formation *Regional Science and Urban Economics* **21**(2), 255–275.

Henderson, J. V. (1981) The economics of staggered work hours *Journal of Urban Economics* **9**, 349–364.

Krugman, P. (1991) Increasing Returns and Economic Geography *Journal of Political Economy* **99**(3), 483–499.

Lucas, R. E. and Rossi-Hansberg, E. (2002) On the Internal Structure of Cities *Econometrica* **70**(4), 1445–1476.

Manski, C. F. (2000) Economic Analysis of Social Interactions *The Journal of Economic Perspectives* **14**(3), 115–136.

Rosenthal, S. S. and Strange, W. C. (2004) Evidence on the nature and sources of agglomeration economies *in* J. V. Henderson and J. Thisse (eds), *Cities and Geography* Vol. Volume 4 Elsevier pp. 2119–2171.

Scott, A. J. (1988) *Metropolis: from the division of labor to urban form* University of California Press.

Small, K. A. and Verhoef, E. T. (2007) *Urban transportation economics* Routledge London and New York.

Starrett, D. A. (1974) Principles of optimal location in a large homogeneous area *Journal of Economic Theory* **9**(4), 418–448.

Tseng, Y. Y. and Verhoef, E. T. (2008) Value of time by time of day: A stated-preference study *Transportation Research Part B: Methodological* **42**(7-8), 607–618.

Vickrey, W. (1969) Congestion theory and transport investment *American Economic Review* **59**(2), 251–261.

Vickrey, W. (1973) Pricing, metering, and efficiently using urban transportation facilities *Highway Research Record* **476**, 36–48.

Vovsha, P. and Bradley, M. (2004) Hybrid Discrete Choice Departure-Time and Duration Model for Scheduling Travel tours *Transportation Research Record* **1894**, 44–56.

Wilson, P. W. (1988) Wage variation resulting from staggered work hours *Journal of Urban Economics* **24**(1), 9–26.

A Proofs

Proof of Lemma 3. Since the queue exists throughout I, and queuing time constitutes all travel time, the queue discharge rate ψ is also the rate of arrivals at work. Then for any $t_d \in [0, N/\psi]$ we have $N_H(t_0 + t_d) = N - R(t_d|t_0)$ and $N_W(t_0 + t_d) = \psi t_d$. The expression for H then follows immediately from (2); note that it is constant for all $t \geq t_1$. The expression for W follows by calculating (1), evaluated for each of the three possible cases shown in (4).

Next, consider derivatives. First, $H'(t) = 0$ for $t \leq t_0$. For $t \geq t_0$, $H'(t) = g_H\left(1 - \frac{R(t-t_0|t_0)}{N}\right)$, which is strictly positive until time t_1 and zero afterwards – the latter because $R(s - t_0|t_0) = N$ for $s \geq t_1$. It is also continuous at t_1 because $g_H\left(1 - \frac{R(t_1-t_0|t_0)}{N}\right) = g_H(0) = 0$ by the definition of t_1 and the properties of

$g_H(\cdot)$. Next, $H''(t) = -g'_H\left(1 - \frac{R(t-t_0|t_0)}{N}\right) \cdot \frac{\rho(t-t_0|t_0)}{N}$, from which we see $H'' < 0$ for $t \in int(I)$ and $H'' = 0$ outside I – the latter because N_H is constant there and thus so is g_H.

Similarly, $W'(a) = 0$ for $a < t_0$, $W'(a) = -g_W[\psi \cdot (a - t_0)/N]$ for $a \in I$, and $W'(a) = -1$ for $a > t_1$; note W' is continuous everywhere because $\psi \cdot (t_1 - t_0)/N = 1$ (by the definition of t_1) and $g_W(1) = 1$. Furthermore, $W''(a) = -(\psi/N)g'_W[\psi \cdot (a - t_0)/N] < 0$ for $a \in int(I)$ and $W'' = 0$ outside I. Because H, W, H', and W' are all continuous at the boundaries of I, and H' and W' are non-increasing there, concavity holds at these boundaries as well. ■

Proof of Theorem 4. Consider a Nash equilibrium given in terms of a departure schedule $R(\cdot|t_0)$. The result that $I \subset (0, \Omega)$ follows from condition 1. A queue must begin forming with the first departure since otherwise the first person to depart could postpone departure and gain. The queue cannot start earlier than the first departure. The queue cannot end before the last departure, since then the last person to depart could depart earlier and gain work production without losing any leisure production. There is no residual queue, since then the last person to depart could postpone departure and gain. These arguments establish that $t_1 = t_0 + N/\psi < \Omega$. Utility must be constant on I by the definition of Nash equilibrium.

Appendix A.1 establishes unique existence of Nash equilibrium. This appendix also shows that the equilibrium departure rate $\rho(\cdot|t_0)$ is strictly decreasing on $(0, N/\psi]$ and hence that $R(\cdot|t_0)$ is strictly concave on $[0, N/\psi]$. ■

Proof of Theorem 5. (a) Nash equilibrium requires equal utilities, since all workers are identical, and thus in particular it requires (7). The proofs of existence and uniqueness are similar to those for our model of endogenous scheduling, and also to those for the bottleneck model, so are omitted here.

(b) First we differentiate equation (7), using (5) for t_1, to obtain:

$$\tilde{V}'(t_0)\frac{\partial t_0}{\partial \psi} = \tilde{V}'(t_1)\left(\frac{\partial t_0}{\partial \psi} - \frac{N}{\psi^2}\right); \qquad \tilde{V}'(t_0)\frac{\partial t_0}{\partial N} = \tilde{V}'(t_1)\left(\frac{\partial t_0}{\partial N} + \frac{1}{\psi}\right).$$

Solving,

$$\frac{\partial t_0}{\partial \psi} = -\frac{N}{\psi^2} \frac{\tilde{V}'(t_1)}{\tilde{V}'(t_0) - \tilde{V}'(t_1)} > 0; \qquad \frac{\partial t_0}{\partial N} = \frac{1}{\psi} \frac{\tilde{V}'(t_1)}{\tilde{V}'(t_0) - \tilde{V}'(t_1)} < 0.$$

Applying these results, we can differentiate (8) to obtain (9) and (10).

(c) The optimal toll eliminates congestion but never leaves the bottleneck delivering less than its full capacity until the end of the departure period. The logic here is the same as in Arnott et al. (1993): any congestion would cause a loss of scheduling utility without a compensating gain; but any unused capacity would permit reallocating someone to a departure time with a higher utility. That this toll is the one satisfying (11) can be seen by noting that with no queue, $\widetilde{V}(t)$ is the actual scheduling utility received by a traveler departing at t (in contrast to the situation before tolling, where scheduling utility was lower because arrival a differed from departure t for all but the first and last traveler). This traveler's utility net of toll payment is the left-hand side of (11), which therefore states that this net utility is constant in time, making the new departure and arrival schedule a Nash equilibrium. In the case $\tilde{\tau}(t_0) = 0$, (11) shows that this net utility is identical to that received before the toll was introduced.

(d) Total welfare change for a worker is defined as the change in that worker's net utility, plus toll revenues received from that worker. Each worker achieves the same utility net of toll payments as before the toll is imposed. Therefore, the net utility change to all workers is zero, and the total welfare change is equal to total revenues. ■

Proof of Theorem 6. For any $t, a \in (t_0, t_1)$, we have $0 < 1 - R(t - t_0|t_0)/N < 1$ and $0 < \psi \cdot (a - t_0)/N < 1$. We need to show that $V(t, a)$ is increasing in departure time t, decreasing in arrival time a, and strictly concave; and that $\tilde{V}(t) = V(t, t)$ attains a maximum. The first two statements can be shown by signing the derivatives of $V(t, a)$:

$$V_1(t, a) = U_H \cdot g_H \left[1 - \frac{R(t - t_0|t_0)}{N}\right] > 0 \tag{21a}$$

$$V_2(t, a) = -U_W \cdot g_W \left[\frac{\psi \cdot (a - t_0)}{N}\right] < 0, \tag{21b}$$

where U_H and U_W are evaluated at $H(t)$ and $W(a)$. To establish concavity, we compute the second derivatives of V:

$$\begin{aligned}
V_{11}(t,a) &= U_{HH}\cdot g_H^2 - U_H\cdot g_H'\frac{\rho(t-t_0|t_0)}{N} < 0 \\
V_{22}(t,a) &= U_{WW}\cdot g_W^2 - U_W\cdot g_W'\frac{\psi}{N} < 0 \\
V_{12}(t,a) &= -U_{HW}\cdot g_H\cdot g_W < 0,
\end{aligned}$$

where the arguments of g_H and g_W are the same as in (21). Strict concavity of V is equivalent to its Hessian being negative definite. That is, the following quadratic form must be strictly negative for any real numbers $x_1, x_2 \neq 0$:

$$\begin{aligned}
& x_1^2\cdot V_{11}(t,a) + 2x_1x_2\cdot V_{12}(t,a) + x_2^2\cdot V_{22}(t,a) \\
= \; & x_1^2\cdot U_{HH}\cdot g_H^2 - x_1^2\cdot U_H\cdot g_H'\cdot\frac{\rho(t-t_0|t_0)}{N} \\
& -2x_1x_2\cdot U_{HW}\cdot g_H\cdot g_W \\
& +x_2^2\cdot U_{WW}\cdot g_W^2 - x_2^2\cdot U_W\cdot g_W'\cdot\frac{\psi}{N} \\
< \; & (x_1\cdot g_H)^2\cdot U_{HH} \\
& -2(x_1\cdot g_H)\cdot(x_2\cdot g_W)\cdot U_{HW} \\
& +(x_2\cdot g_W)^2\cdot U_{WW}.
\end{aligned}$$

This expression is indeed strictly negative because U is strictly concave.

It remains to show that $\tilde{V}(t) = V(t,t)$ attains a maximum. This follows since $\tilde{V}$ inherits concavity from V and since $\tilde{V}(t_0) = \tilde{V}(t_1)$. ∎

Proof of Theorem 7. A Vickrey-like analyst would set a toll schedule $\tilde{\tau}(t)$ aimed at maintaining a departure schedule starting at t_0 which just eliminates queueing: namely $\tilde{R}(s|t_0) = \psi s$. He does so by calculating this schedule based on the assumed scheduling utility function $V(t,a)$ given by (13) with $a = t$ (de Palma and Fosgerau, 2011). This utility can be written as follows:

$$\tilde{V}(t) \equiv V(t,t) = U\left[t_0 + \int_0^{t-t_0} g_H\left(1-\frac{R(s|t_0)}{N}\right)ds,\ \int_{t-t_0}^{t_1-t_0} g_W\left(\frac{\psi s}{N}\right)ds + \Omega - t_1\right], \tag{22}$$

with $t_1 = t_0 + N/\psi$ and with $R(\cdot)$ following its no-toll equilibrium path as depicted in Fig. 1. We can assume the arbitrary toll constant would be chosen so that $\tilde{\tau}(t_0) = 0$. The toll would thus be:

$$\tilde{\tau}(t) = \tilde{V}(t) - \tilde{V}(t_0), \tag{23}$$

since this would assure that the anticipated utility net of toll would be identical for everyone under the desired departure pattern $\tilde{R}(\cdot)$.

As an intermediate step toward a truly optimal toll, consider now an analyst who also aims to achieve departure pattern $\tilde{R}(\cdot)$, but who knows the true model. This analyst would similarly use (23) except with $\tilde{V}(t)$ replaced by the true utility given that departure pattern. That utility accounts for how H changes as a result of the change in departure pattern: it is calculated as in (22) but with new cumulative departures ψs replacing the original cumulative departures $R(s|t)$ in the argument of g_H. Denoting the result by $\hat{V}(t)$, we have:

$$\hat{V}(t) \equiv U\left[t_0 + \int_0^{t-t_0} g_H\left(1 - \frac{\psi s}{N}\right) ds,\ \int_{t-t_0}^{t_1-t_0} g_W\left(\frac{\psi s}{N}\right) ds + \Omega - t_1\right]. \tag{24}$$

The toll $\widehat{\tau}(t)$ (in utility units) would thus be set to maintain utility net of toll at its original constant value $\tilde{\Delta} = \tilde{V}(t_0)$, i.e. it is given by (14). Because $\psi s < R(s|t_0)$ for every value of s in the arguments of g_H, the value of H (i.e., the first argument of $U[\cdot]$ in these equations) is greater in (24) than in (22). Therefore $\hat{V}(t) > \tilde{V}(t)$ and as a consequence $\hat{\tau}(t) > \tilde{\tau}(t)$, for every $t > t_0$.

Now consider the welfare gains from toll $\hat{\tau}(t)$. Vickrey would believe the welfare gain to be equal to the toll revenue from $\tilde{\tau}(\cdot)$, with constant departure rate $\rho = \psi$: that is, $\psi \int_{t_0}^{t_1} \tilde{\tau}(t)\, dt$. But we have just seen that the congestion-removing toll that really leaves workers' utility unaffected, namely $\hat{\tau}(t)$, is larger than $\widetilde{\tau}(t)$, and strictly larger for $t > t_0$. The departure rate with that toll is again $\rho = \psi$, so its welfare gain (the sum of all travelers' utilities) is equal to toll revenue $\psi \int_{t_0}^{t_1} \hat{\tau}(t)\, dt$. Hence the toll revenue and welfare gain are both strictly larger than believed by Vickrey. ∎

Proof of Lemma 9. From the definition of $\widetilde{V}$ and eq. (13), we have $\tilde{V}'(t_0)$

$= dV(t_0, t_0)/dt_0 = V_1(t_0, t_0) + V_2(t_0, t_0) = U_H^0 \cdot g_H(1) - U_W^0 \cdot g_W(0) = U_H^0$. Similarly, $\tilde{V}'(t_1) = V_1(t_1, t_1) + V_2(t_1, t_1) = U_H^1 \cdot g_H(0) - U_W^1 \cdot g_W(1) = -U_W^1$. Then use Theorem 5. ■

A.1 Unique existence of Nash equilibrium

The aim of this appendix is to show that Nash equilibrium exists uniquely, and that in Nash equilibrium the departure rate is strictly decreasing in time. Our strategy is to write utility as a function of departure time, measured as time after a given initial departure time, and to consider the implications of meeting a "total population" condition that the departure rate must integrate to N, the total number of workers over a duration of N/ψ time units. We do this by defining a function $\overline{d}(t)$ relating the last departure time to the first departure time so as to meet the total population condition. We show that as t covers its allowed values, $\overline{d}(t)$ first falls short of, then exceeds, the value N/ψ that represents the minimum time over which the bottleneck can deliver travelers to their destination. We then show that given the properties of utility and of the production functions for leisure and work, the departure rate must be smaller as initial departure time becomes larger, which enables us to locate a single departure time that meets both the total population condition and the Nash equilibrium condition.

We begin by defining abstract functions ρ, R, and u whose properties we will gradually restrict so that they can represent the departure rate, cumulative departures, and utility for a Nash equilibrium.

Definition 10 *An **anchored function** $R(d|t)$ is a real function defined for $d \geq 0$ and $t \in [0, \Omega]$, with $R(0|t) = 0$. It is associated with a utility profile defined as*

$$u(d|t) = U\left(t + \int_0^d g_H\left(1 - \frac{R(s|t)}{N}\right) ds, \Omega - \left(t + \frac{N}{\psi}\right) + \frac{1}{\psi} \int_{R(d|t)}^N g_W\left(\frac{s}{N}\right) ds\right).$$

Note that the anchored utility profile depends only on $R(\cdot)$ and on the utility and agglomeration functions.

Remark 11 *The term "anchored" means simply that it is required to start at zero, i.e., time t represents an initial departure.*

Definition 12 *A **candidate cumulative departure function** is an anchored function $R(\cdot|t)$ whose utility profile is constant:*

$$0 = \frac{\partial u(d|t)}{\partial d} = U_H \cdot g_H\left(1 - \frac{R(d|t)}{N}\right) - U_W \cdot g_W\left(\frac{R(d|t)}{N}\right)\frac{\rho(d|t)}{\psi}, \quad (25)$$

where

$$\rho(d|t) = \frac{\partial R(d|t)}{\partial d}.$$

*is called the **candidate departure rate profile**.*

Remark 13 *Equation (25) and the properties of U, g_H, and g_W, along with the anchoring requirement that $R(0|t) = 0$, imply that $\rho(\cdot|t) > 0$ and*

$$\rho(d|t) \underset{d \to 0^+}{\longrightarrow} \infty.$$

.

Definition 14 *Given a candidate departure rate profile, a **candidate departure duration** $\overline{d}(t)$ is defined implicitly by the total population condition*

$$\int_0^{\overline{d}(t)} \rho(s|t)\, ds = N.$$

Lemma 15 *Conditions 1 and 2 imply that a candidate departure duration brackets the value N/ψ, as follows:*

$$\overline{d}(0) < \frac{N}{\psi} < \overline{d}\left(\Omega - \frac{N}{\psi}\right).$$

Proof. Consider the first inequality and assume on the contrary that $\overline{d}(0) \geq \frac{N}{\psi}$. Note that $\rho(0|0) = \infty$, which rules out that $R(d|0) < \psi d$ for all $d > 0$ by the definition of ρ as the derivative of R. Then $\exists d > 0$ s.t. $R(d|0) = \psi d$. There is queue for all departures in $[0, d]$ and so utility is constant for departures in this interval. But $H = 0$ for departure at time 0 and so utility is $-\infty$ at this departure time while utility is finite for departure at time d. This is a contradiction and we conclude that $\overline{d}(0) < \frac{N}{\psi}$. Consider now the second inequality and assume on the contrary that $\frac{N}{\psi} \geq \overline{d}\left(\Omega - \frac{N}{\psi}\right)$. This implies that there is queue for all departures in the interval $\left[\Omega - N/\psi, \Omega - N/\psi + \overline{d}(\Omega - N/\psi)\right]$ and hence utility is constant for departure times in this interval. For the last possible departure time we have $W = 0$ and hence $U = -\infty$, while U is finite for the first departure time. This is a contradiction and we conclude that $\frac{N}{\psi} < \overline{d}\left(\Omega - \frac{N}{\psi}\right)$ as desired. ■

By continuity, Lemma 15 ensures that there is at least one t with $\overline{d}(t) = N/\psi$. Since $\rho(\cdot|t)$ is decreasing by Lemma 16, $R(\cdot|t)$ is concave. Hence $R(d|t) \geq \psi d$ for all $d \in]0, N/\psi[$. Then there will be queue from time t to time $t + N/\psi$ and so $u(d|t)$ describes the evolution of utility under the departure schedule $R(\cdot|t)$, which then describes a Nash equilibrium. Thus existence of Nash equilibrium is established.

It remains to show that Nash equilibrium is unique. Consider two Nash equilibria indexed by a and b, starting departures at time t_a and t_b, respectively, where $t_a < t_b$. Consider then the inequalities

$$H_a\left(R^{-1}(x|t_a)\right) < H_b\left(R^{-1}(x|t_b)\right), W_a\left(R^{-1}(x|t_a)\right) > W_b\left(R^{-1}(x|t_b)\right). \quad (26)$$

Then (26) is valid at $x = 0$. It follows by continuity that there exists $x' > 0$ such that (26) is valid for all $x < x'$. Then by Condition 2,

$$\rho\left(R^{-1}(x|t_a)|t_a\right) > \rho\left(R^{-1}(x|t_b)|t_b\right)$$

for all such x. This implies that $R^{-1}(x|t_a) < R^{-1}(x|t_b)$ and so (26) holds also at $x = x'$. This argument shows that (26) holds at all $x \in [0, 1]$, since there can be no first x where it fails. But this is a contradiction since

$$\frac{N}{\psi} = \int_0^N \frac{1}{\rho\left(R^{-1}(x|t_a)|t_a\right)} dx < \int_0^N \frac{1}{\rho\left(R^{-1}(x|t_b)|t_b\right)} dx = \frac{N}{\psi}.$$

This establishes that Nash equilibrium is unique.

The following lemma establishes that the equilibrium departure rate is decreasing.

Lemma 16 *Consider Nash equilibrium with first departure at time* t_0. *Then* $\rho'(d|t_0) < 0$ *for* $d \in int(I)$.

Proof. The departure rate $\rho(d|t_0)$ satisfies

$$0 = U_H \cdot g_H\left(1 - \frac{R(d|t_0)}{N}\right) - U_W \cdot g_W\left(\frac{R(d|t_0)}{N}\right)\frac{\rho(d|t_0)}{\psi}.$$

Differentiate and rearrange slightly to find that

$$\begin{aligned} & U_W \cdot g_W\left(\frac{R(d|t_0)}{N}\right)\frac{\rho'(d|t_0)}{\psi} \\ = & (U_{HH}H' + U_{HW}W') \cdot H' + (U_{HW}H' + U_{WW}W') \cdot W' \\ & -U_H \cdot g'_H\left(1 - \frac{R(d|t_0)}{N}\right)\frac{\rho(d|t_0)}{N} - U_W \cdot g'_W\left(\frac{R(d|t_0)}{N}\right)\frac{\rho(d|t_0)^2}{N\psi}. \end{aligned}$$

The RHS of this equation is strictly negative since U is strictly concave and $U_H > 0$, $U_W > 0$. The desired conclusion follows. ■

B Numerical simulation details

B.1 Nash equilibrium

For given parameters ψ, α, π_W, and π_H, the numerical solution proceeds by first choosing a trial value of t_0 and dividing the time axis into many small periods starting at t_0.We then compute ρ and R stepwise. For the first few periods, cumulative departures are computed from (29) as explained below; after that they are computed by cumulating values of departure rate ρ, which is computed in each time interval from (18) using values of R, H, and W as determined in the previous period. When R is near one, we apply (30) to find the end of the departure interval. We compare the duration of this departure interval with its required value, N/ψ, to determine whether the trial value of t_0 is consistent with equilibrium; if not, we adjust t_0 iteratively until the correct interval length is achieved. Finally, we check the accuracy by comparing the values of U computed at all the time periods; they should be identical, and if not we make the time steps smaller in order to increase the accuracy of the calculation. We typically get satisfactory results with 500 time periods, achieving identical utilities to within about 0.1 percent, and within about 0.03 percent for all but the first 10 time periods.

The procedure for values of t near t_0 or t_1 is different because equation (18) is numerically unsatisfactory there. First, that the departure rate ρ is infinite at $R = 0$, the beginning of the rush hour. Furthermore, as R approaches 1 (the end of the departure interval), $\rho \to 0$, causing that approach to possibly occur very slowly since R is the integral of ρ. These extremes make a simple stepwise numerical procedure inaccurate, so instead we calculate R in those two regions by solving an approximate version of (18) for constant U_H/U_W. This is quite accurate because H and W contain constants at the boundaries, which allows the marginal utilities α/H and $1/W$ to be nearly constant there. Furthermore, near the first boundary we can approximate $[1 - R(\cdot)]$ in (18) as a constant equal to unity, and near the second boundary we can approximate $R(\cdot)$ as a constant, also equal to unity. The approximate differential equations are then

$$\rho = K_0 \cdot R^{-\pi_W} \qquad \text{near } t = t_0, \tag{27}$$

$$\rho = K_1 \cdot (1 - R)^{\pi_H} \qquad \text{near } t = t_1, \tag{28}$$

where $K_0 = \psi U_H(0)/U_W(0)$ and $K_1 = \psi U_H(1)/U_W(1)$ with notation (0) and (1) indicating values at the endpoints of the departure interval and with boundary conditions $R(0) = 0$ and $R(1/\psi) = 1$. We seek the solutions to these differential equations, with boundary conditions $R(t_0) = 0$ for (27) and $R(t_1) = 1$ for (28). The solutions are:

$$R = [(1 + \pi_W) \cdot K_0 \cdot (t - t_0)]^{\frac{1}{\pi_w + 1}} \qquad \text{near } t = t_0, \tag{29}$$

$$1 - R = [(t_1 - t)(1 - \pi_H) K_1]^{\frac{1}{1 - \pi_h}} \qquad \text{near } t = t_1. \tag{30}$$

Note that (30) requires $\pi_H < 1$ to be valid, i.e. to give $(1 - R)$ as a finite real number. Basically, this is because for larger values of π_H, the departure rate (28) becomes so small as $R \to 1$ that the limiting value $R = 1$, indicating the end of the departure period, can never be reached.

B.2 Optimum

The maximization in (20) is performed by writing out the arguments of $U\left[\widehat{H}(t), \widehat{W}(t)\right]$ in the integrand as functions of $t_d = t - t_0$:

$$
\begin{aligned}
\widehat{H}(t_d|t_0) &= t_0 + \int_0^{t_d} (1-\psi s)^{\pi_H} ds = t_0 + \frac{1}{\psi} \cdot \frac{1-(1-\psi t_d)^{\pi_H+1}}{\pi_H+1} \\
W(t_d|t_0) &= \Omega - t_1 + \frac{1}{\psi} \cdot \int_{t_d}^{1/\psi} (\psi s)^{\pi_W} ds \;=\; \Omega - t_0 - \frac{1}{\psi} + \frac{1-(\psi t_d)^{\pi_W+1}}{\pi_W+1}
\end{aligned}
$$

where now we have explicitly indicated in the notation that these arguments depend on t_0. (There is not carat on W when written as a function of t_d because, conditional on t_0, it is the same function as in the Nash equilibrium calculation.) The average value of this utility is

$$
\overline{U}(t_0) = \psi \int_0^{1/\psi} U\left[\widehat{H}(t_d|t_0), W(t_d|t_0)\right] dt_d
$$

and the first-order condition for maximizing it is

$$
0 = \frac{d\overline{U}}{dt_0} = \psi \int_0^{1/\psi} U_H \frac{\partial \widehat{H}}{\partial t_0} dt_d + \psi \int_0^{1/\psi} U_W \frac{\partial W}{\partial t_0} dt_d = \psi \int_0^{1/\psi} \frac{\alpha}{\widehat{H}} dt_d - \psi \int_0^{1/\psi} \frac{1}{W} dt_d \tag{31}
$$

since $\partial \widehat{H}/\partial t_0 = 1$ and $\partial W/\partial t_0 = -1$. Intuitively, (31) states that t_0 is increased until the marginal benefit of increased home time, measured by U_H, is just matched by the marginal disbenefit of decreased work time, U_W. The internal dynamics, i.e. the shape of the departure pattern within the congested period, play no role in this calculation because they are unaffected by changes in t_0 given that the toll has been adjusted to make this departure pattern simply a constant at rate ψ.