Discussion Papers
Department of Economics
University of Copenhagen

No. 17-21

Travel time variability and rational inattention

Mogens Fosgerau and Gege Jiang

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Travel time variability and rational inattention

Mogens Fosgerau
University of Copenhagen

Gege Jiang
Hong Kong University of Science and Technology

October 24, 2017

Abstract

This paper sets up a rational inattention model for the choice of departure time for a traveler facing random travel time. The traveler chooses how much information to acquire about the travel time outcome before choosing departure time. This reduces the cost of travel time variability compared to models in which the information is exogenously fixed .

Keywords: rational inattention; random travel time variability; value of reliability; discrete choice
JEL codes: D1, D8, R4

1 Introduction

This paper connects the literature on rational inattention to the issue of valuing travel time variability.

Traffic congestion imposes very significant costs on developed economies, with estimates typically in the range of 1 percent of GDP. Traffic congestion not only leads to delays but also causes travel times to be variable and unpredictable from the point of view of travelers. Nonrecurring traffic congestion (due to accidents, bad weather, special events, and other shocks) contributes more than half of total delay in US urban areas (Schrank et al., 2011). Travel time variability thus adds significant economic costs to the costs of traffic congestion. Valuation of travel time variability is clearly very important for the evaluation of transport policy.

The standard theory used to establish a value of travel time variability is firmly neoclassical (Fosgerau, 2015). Travelers are equipped with scheduling preferences over departure and arrival time outcomes. They face a random travel time, but know the statistical distribution of travel time. They choose a departure time optimally to maximize expected scheduling utility given the travel time distribution.

It is, however, not clear in this neoclassical model what the relevant travel time distribution actually is from the point of view of travelers. There are many possibilities, depending on what information travelers are thought to take into account and how much effort they are likely to invest in processing that information. Transportation research has so far had little to say about this.

There are many information sources that travelers may or may not take into account. First, of course, they know the time of day and the day of week. They could plausibly have some idea about how the travel time distribution varies across those dimensions. Second, some days are special due to holidays or special events that the traveler might be aware of and take into account. Current weather is observable and forecasts are easily available so travelers may be able to factor in the impact of weather in their anticipated travel time distribution. Travelers may or may not know about roadworks. Finally, real time traffic information is available from various sources and travelers may also be able to factor this in, depending on whether these sources provide information about incidents and delays or they directly inform about travel times.

Previous research has considered various kinds of information provision to travelers but has mostly assumed that travelers effortlessly incorporate all the provided information into their decisions. Paper in transportation that deal with information include the following. Yang (1998) proposes a equilibrium model with endogenous market penetration of ATIS which is dependent on the private travel time saving due to the possession of ATIS. Avineri and Prashker (2006) compare the route choice behavior with and without prior static information and concludes that the static information about expected travel times increases the heterogeneity of travelers' choices. Ettema and Timmermans (2006) set up a model that captures the cost due to bias in the prior knowledge of travel time and demonstrates that providing information can reduce the bias and improve the expected utility of travelers. Paz and Peeta (2009) develop a behavior-consistent approach for network control which takes the compliance rate of travel information into consideration. Xiao and Lo (2016) investigate the effect of travel time information shared by social networks on travelers' departure time choices. Ben-Elia et al. (2013) study the influence of information accuracy on route choice. Lindsey et al. (2014) propose a theoretical model of the relationship between pre-trip information and the route choice problem under stochastic traffic conditions. Bifulco et al. (2016) model the impact of ATIS on network equilibrium and stability.

De Borger and Fosgerau (2012) consider a public transport firm deciding on provision of time table information, which reduces the travelers' cost of planning for a specific departure. In their model, travelers choose whether or not to plan and hence they decide whether or not to use the information provided.

It seems unrealistic to assume that travelers make use of all the available information as that would require considerable effort that could easily outweigh the gains from improving the timing of trips. On the other hand, it is reasonable to expect that travelers take some information into account. It also seems clear that the travelers' choice of how much information to take into account is a choice. It is entirely plausible that this choice depends on the cost and benefits of information acquisition.

We would therefore like to have a theory underlying the valuation of travel time variability that takes into account that the information processing strategy of the traveler is a choice that depends on preferences as well as the distribution of travel times and the cost of information. The broader consequences of decreases in the cost of information is, of course, a timely issue with the emergence of near universal access to real time traffic information through various services.

So it is very relevant to have a theory for the value of travel time variability that takes information acquisition into account. The micro-economic theory of rational inattention is a natural choice of a theoretical framework for this endeavour (Sims, 2003, 2010). It is a theory that recognizes that people have limited information processing capability and that the cost of information processing influences how people behave under uncertainty. Under rational inattention, people choose optimally how much information to take into account, balancing the cost of information against the gains from better informed expectations.

Recently, Matějka and McKay (2015) have shown that rational inattention can be used to provide a foundation for the multinomial logit model that does not assume that decision makers have precise evaluations of all available choice alternatives. Along with most of the previous literature on rational inattention, they use the Shannon entropy to define the cost of information.[1]

This paper sets up a rational inattention model for the traveler choice of departure time under travel time variability (TTV), where the traveler also chooses an information strategy. This is in contrast to the neoclassical model which takes the traveler's information as given. Here, the traveler's information strategy is endogenously chosen, taking the cost of information into account.

This paper sets up a rational inattention model for a traveler's choice of departure time when the travel time is random. We find that the marginal cost of TTV is always positive and that the traveler's payoff decreases with the unit cost of information. The traveler forms a consideration set of departure times that are chosen with positive probability. A small consideration set is cheap in terms of information but offers limited possibilities for compensating travel time variability through varying the departure time. A large consideration set has the converse implications. The consideration set will then typically have gaps, as we also find in our numerical simulations. If the scheduling utility is very concave, then the consideration set has no gaps but then it comprises at most two departure times. In the case where the traveler has linear scheduling utility rates as in Fosgerau and Engelson (2011), the marginal cost of TTV found by Fosgerau and Engelson emerges as a limiting case, with the marginal cost of TTV in the rational inattention model always being smaller.

Our simulation results also indicate that the cardinality of the consideration set decreases as the information cost increases. The marginal cost of TTV increases as the unit information cost increases.

Section 2 reviews the neoclassical model for departure time choice under

[1]Fosgerau et al. (2017) generalize this result to arbitrary additive random utility discrete choice models, using a generalized entropy to define the cost of information.

travel time variability and derives the marginal value of travel time variability. This is a point of reference for Section 3 that sets up a rational inattention model for the context of a traveler deciding the timing of a trip with uncertain duration. Section 4 derives the marginal value of travel time variability in the rational inattention model and specializes to the slope and step models of scheduling preferences. Section 5 presents some numerical experiments. Section 6 concludes.

2 The neoclassical model

Before presenting the new rational inattention model, we review the neoclassical model in which a traveler chooses departure time to maximize expected scheduling utility, see Engelson and Fosgerau (2016).

We consider then a traveler about to undertake a trip of uncertain duration. He cares about the timing of the trip, with preferences given by a scheduling utility $u(a,t)$, where a is the departure time and t is the travel time, such that $a+t$ is the arrival time.[2]

Travel time T is random, taking values in a finite set $\mathcal{T}$. We write $p(t) = P(T=t)$, assuming without loss of generality that $\forall t \in \mathcal{T} : p(t) > 0$. To talk about travel time variability, we parametrize travel time as $T = \mu + \sigma X$, where $X \in \mathcal{X}$ is standardized travel time with $EX = 0$ and $EX^2 = 1$.

2.1 Scheduling preferences

The traveler's scheduling preferences are given by a scheduling utility that is specified in terms of utility rates h, w as follows (Vickrey, 1973; Tseng and Verhoef, 2008; Engelson and Fosgerau, 2016).

$$u(a,t) = \int_0^a h(t')dt' + \int_{a+t}^0 w(t')dt'. \tag{1}$$

We assume that $h, w > 0$ such that travel time is always costly. Then the traveler prefers to depart later and arrive earlier, ceteris paribus. We assume also that $w(t') < h(t')$ if and only if $t' < 0$. This implies that a traveler with zero travel time will prefer to travel at time 0 and that a traveler with positive travel time will prefer to depart before time 0 and

[2]Gender of representative person determined at random by flipping a Danish 20 kroner coin.

arrive after time 0. Finally, we assume that h is weakly decreasing, while w is strictly increasing.

We shall refer to the case in which the utility rate at the destination is linear as in the slope model (Fosgerau and Engelson, 2011), i.e.

$$h\left(t'\right) = \beta_0 + \beta_1 t' \tag{2}$$

$$w\left(t'\right) = \gamma_0 + \gamma_1 t' \tag{3}$$

Without loss of generality, we assume $\beta_0 = \gamma_0$, $\beta_1 = \gamma_1 - 1$.

Another important case is the step model (Fosgerau and Karlstrom, 2010)

$$h\left(t'\right) = \alpha, w\left(t'\right) = \begin{cases} \alpha - \beta, t' < 0 \\ \alpha + \gamma, t' \geq 0 \end{cases} \tag{4}$$

2.2 Optimal behavior

The traveler chooses departure time to maximize his expected utility.

$$a^* = \underset{a}{\operatorname{argmax}}\{Eu(a, a+T)\},$$

which has first-order condition

$$h(a^*) = Ew(a^* + T).$$

The second-order condition is always satisfied.

Insert the optimal departure time into utility to find the optimal expected utility $u^* = Eu(a^*, a^* + T)$. Then use enveloping to find the value of travel time (VTT) as

$$-\frac{\partial u^*}{\partial \mu} = h(a^*),$$

and the value of travel time variability (VTTV) as

$$-\frac{\partial u^*}{\partial \sigma} = E\left(Xw(a^* + \mu + \sigma X)\right)$$

In the case of the slope model (Vickrey, 1973; Tseng and Verhoef, 2008; Fosgerau and Engelson, 2011), the VTT is

$$-\frac{\partial u^*}{\partial \mu} = \beta_0 + \frac{\beta_1 \gamma_1}{\beta_1 + \gamma - 1}\mu,$$

and the VTTV is expressed in terms of variance as

$$-\frac{\partial u^*}{\partial \sigma^2} = \beta_1/2.$$

In the case of the step model (Vickrey, 1969; Small, 1982; Fosgerau and Karlstrom, 2010), the VTT is α while the value of standard deviation is

$$-\frac{\partial u^*}{\partial \sigma} = (\beta + \gamma) \int_{\frac{\gamma}{\beta+\gamma}}^{\infty} F^{-1}(x)dx,$$

where F is the CDF of the standardized travel time X.

3 A rationally inattentive traveler

Now we present the rational inattention model, in which the traveler is allowed to choose a signal, a random variable that is informative about the random travel time. It is possible in principle to extend the neoclassical model with an exogenous signal. The crucial difference between the neoclassical model and the rational inattention model to be formulated is that the rational inattention model allows the signal to be endogenously chosen.

For simplicity we assume that the traveler chooses his action, the departure time, from a finite set $\mathcal{A}$. In the rational inattention model, the action is not a single departure time, but a random variable, denoted by A, where $A \in \mathcal{A}$.

The traveler maximizes his net payoff, which is the expected scheduling utility less the cost of information,

$$\Lambda(P) = E(u(A,T)) - \lambda I(A,T), \tag{5}$$

where λ is the cost per unit of information and $I(A,T)$ is the amount of information about travel time comprised in the action, which will be specified below.

We note that there is no signal evident in this specification. As discussed in the next section 3.1, the underlying model is that the traveler chooses an information strategy, a signal, and then chooses how to respond to the signal. In the formulation here we skip the intermediate step of designing the signal, recognizing that the signal is implicit in the conditional distribution $p(A = a|T = t)$ that is chosen by the traveler. We discuss the information

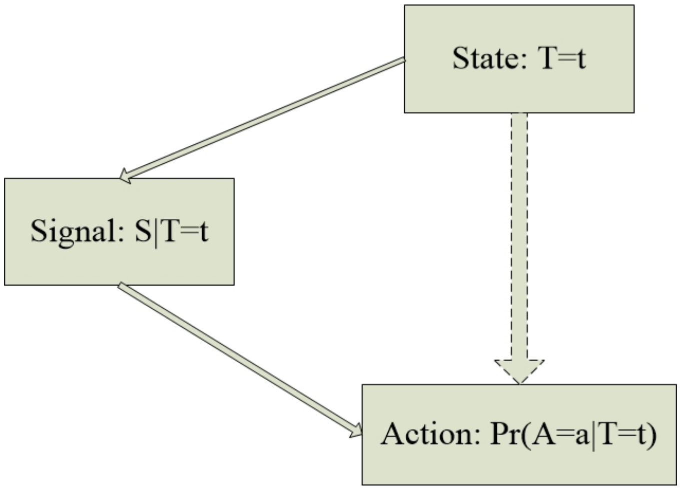

Figure 1: Decision process

cost $I(A, T)$ in Section 3.2 below.

3.1 Information strategy

The rationally inattentive traveler chooses a signal, a random variable that contains information about travel time. The traveler is able to design his signal, which means that he is able to select the distribution of the signal conditional on the travel time. He is, however, not completely free in his choice since he must pay a cost that increases as the signal becomes more informative. The cost is not only monetary but includes any effort made by the traveler.

Every day, the traveler observes that day's realization of the signal and makes inference about that day's travel time distribution. Based on that conditional distribution, he chooses the optimal departure time. That makes the departure time a random variable with conditional distribution $\Pr(A = a|T = t)$ and the departure time will contain information about travel time.

It is a standard result in the rational inattention literature (Hébert and Woodford, 2016) that it is wasteful to obtain a signal that is more informative than the action, and we may therefore simplify, at no loss of generality, by identifying the signal with the random action. Our rationally inattentive traveler will then, effectively, be choosing the conditional distribution $\Pr(A = a|T = t)$. This is illustrated in Figure 1.

Write

$$p(a|t) = \Pr(A = a|T = t)$$

as short-hand for the conditional probability and let

$$p(a) = \Pr(A = a) = \sum_{t \in \mathcal{T}} p(a|t)\, p(t) \tag{6}$$

be the unconditional probability. Let

$$P = \{p(a|t) : a \in \mathcal{A}, t \in \mathcal{T}\}$$

be the matrix of conditional probabilities. This matrix constitutes the *information strategy* chosen by the traveler.

3.2 The cost of information

We now discuss the cost of information. The simplest measure of information is the mutual Shannon information, which is defined in terms of the Shannon entropy. If a discrete random variable Z has density $q()$, then its Shannon entropy $H(Z) = -E \log q(Z)$ captures the amount of information encoded in Z.

The mutual Shannon information between random variables A and T is

$$\begin{aligned} I(A,T) &= H(A) - E(H(A|T)) \\ &= E\left(\log \frac{p(A,T)}{p(A)\,p(T)}\right) \\ &= -\sum_{a \in \mathcal{A}} p(a) \log p(a) + \sum_{t \in \mathcal{T}, a \in \mathcal{A}} p(a|t)\, p(t) \log p(a|t). \end{aligned}$$

The mutual Shannon information $I(A,T)$ measures the information about T present in A (or vice versa) (Cover and Thomas, 2006). For example, if A and T are independent, then knowing A does not give any information about T and vice versa, so their mutual information is zero. At the other extreme, if A is a deterministic and invertible function of T then all information conveyed by A is shared with T: knowing A determines the value of T and vice versa. As a result, in this case the mutual information is the same as the entropy of T (which is the same as the entropy of A).[3]

[3]Fosgerau et al. (2017) propose a natural generalization of the Shannon entropy, based

3.3 Optimal rationally inattentive behavior

The rationally inattentive traveler finds the optimal strategy that maximizes his payoff (5) subject to the consistency constraint (6). In the case when the information cost is the mutual Shannon information, the optimal conditional choice probabilities are

$$p\,(a|t) = \frac{\exp\left(u\,(a,t)\,/\lambda\right)p\,(a)}{\sum\limits_{a'}\exp\left(u\,(a',t)\,/\lambda\right)p\,(a')}, \tag{7}$$

where $p\,(a)$ then depend on $p\,(a|t)$ still through (6) (Matějka and McKay, 2015; Fosgerau et al., 2017). This may be recognized as a logit model in which actions with $p\,(a) = 0$ are omitted (Matějka and McKay, 2015).

The actions for which $p\,(a) > 0$ may be thought of as a *consideration set*, and the decision maker chooses among these actions according to a plain logit model with alternative specific constants $\log p\,(a)$. There are two extreme cases of information cost. If information is costless, $\lambda = 0$, then travelers will simply choose the departure time producing maximum utility with probability 1 under each realization of travel time. The matrix of conditional probabilities $p(a|t)$ then only contains zeros and ones. If, on the other hand, information is infinitely costly, $\lambda = \infty$, then travellers will acquire no information and will choose one departure time with probability 1 as in the neoclassical model and where the departure time will maximize the expected utility under the prior for the travel time distribution.

Let $\hat{P}$ be the optimally chosen information strategy. Caplin et al. (2016) derive a necessary and sufficient condition for $\hat{P}$.

Proposition 1 *(Caplin et al., 2016) The information strategy $\hat{P}$ is optimal if and only if for all $a \in A$,*

$$\sum_{t\in T}\frac{\exp\left(u\,(a,t)\,/\lambda\right)p\,(t)}{\sum\limits_{a'}\exp\left(u\,(a',t)\,/\lambda\right)p\,(a')} \leq 1 \tag{8}$$

with equality if $p(a) > 0$.

Caplin et al. (2016) propose an algorithm for the optimal solution. It

on duality in relation to additive random utility discrete choice models. Replacing the Shannon entropy by their generalized entropy in the definition of mutual information leads to generalized entropy rational inattention models. In this study, we consider the information cost as Shannon entropy for simplicity.

starts with a guess for $P_0(a)$. The update of $P(a)$ follows

$$P_{n+1}(a) = E\left(\frac{\exp\left(u\left(a,T\right)/\lambda\right)P_n(a)}{\sum\limits_{a'}\exp\left(u\left(a',T\right)/\lambda\right)P_n(a')}\right). \tag{9}$$

The iteration of (9) has converged when the conditions in Proposition 1 are satisfied (with some chosen precision). Caplin et al. (2016) prove that the solution generated by this algorithm is optimal if the initial consideration set with $P_0(a) > 0$ is sufficiently large that it includes all possible choice options.

We expect that the solution is unique in most cases. Matějka and McKay (2015) provide necessary and sufficient conditions for uniqueness, showing that non-uniqueness is a very special case. We restate their sufficiency result translated to the present context.

Lemma 2 (Matějka and McKay (2015)) *If the random vectors* $(exp(u(a,t)/\lambda), a \in \mathcal{A})$ *are linearly independent with unit scaling, i.e. if there does not exist a vector* $(\alpha_a, a \in \mathcal{A})$ *such that ,*

$$\forall t \in \mathcal{T} : 0 = \sum_a \alpha_a \exp\left(u(a,t)/\lambda\right),$$

then the traveler's problem has a unique solution.

It is intuitively clear that it is impossible to find such a vector $(\alpha_a, a \in \mathcal{A})$ in most cases. The next proposition establishes a slightly weaker version of that impossibility.

Proposition 3 *Let* w *be linear. There does not exist weights* $\alpha_a, a \in \mathcal{A}$ *such that*

$$0 = \sum_a \alpha_a \exp\left(\int_0^a h(t')dt' + \int_{a+t}^0 w(t')dt'\right),$$

for all t *in an open set.*

Fosgerau et al. (2017) show the following result, which is useful for reducing the number of options that need to be considered.

Proposition 4 *(Fosgerau et al., 2017) Suppose that option* a *is dominated by option* d *in the sense that* $\forall t \in \mathcal{T} : u\left(a,t\right) \leq u\left(d,t\right)$ *with strict inequality for some* t*. Then* $p\left(a\right) = 0$*.*

In the context of random travel time, Proposition 2 means that very early or late departure times are excluded from consideration.

One may wonder (and we have) whether gaps in the consideration set are possible. Let $a_1, a_2, a_3 \in A$, $a_1 < a_2 < a_3$. Then the question is whether it is possible that $p(a_1) > 0, p(a_2) = 0$, and $p(a_3) > 0$. It turns out this is possible, this is evident in the simulations provided below in this paper. The next proposition provides a strong condition under which there can be no gaps in the consideration set. This result conversely contributes to generating some intuition regarding how gaps in the consideration set can occur.

We say that a function f is *superconcave*, if $exp(f)$ is concave. This means that f is concave enough to overcome, in a sense, the convexity of the exponential function. This is analogous to the concept of superconvexity.

Proposition 5 *If the utility function is weakly superconcave as a function of a for all $t \in \mathcal{T}$, then there can be no gaps in the consideration set. If the utility function is a strictly superconcave function of a, then the cardinality of the consideration set is no larger than* 2.

Proposition 5 suggests that we should expect gaps in the consideration set to occur in most cases. This is intuitive since a consideration set with low cardinality is cheap in terms of information. Spreading out the points in the consideration set allows the traveler to compensate for a wider range of travel times. Put in another way, we see that if a departure time is in the consideration set, then nearby departure times may be omitted to reduce the cost of information and losing only a little of the expected scheduling utility.

In the step model of Fosgerau and Karlstrom (2010) and the slope model of Fosgerau and Engelson (2011), the utility functions are linear and concave functions of a, respectively. As the exponential function is convex, the utility function in the step model is not superconcave. Whether the slope model is superconcave depends on the unit information cost λ. For a fixed set of departure times $\mathcal{A}$ and travel times $\mathcal{T}$, then it can be shown with a straightforward calculation that the utility function is superconcave if the unit information cost λ is sufficiently large. That represents the situation when information is very expensive, people will get very limited information and only consider one or two options.

Let $\hat{P}$ be the optimally chosen information strategy. Fosgerau et al. (2017) also show, in the case when the information cost is the mutual Shannon information, that the optimal payoff is

$$\Lambda(\hat{P}) = \lambda \sum_{a,t} p\left(a|t\right) p(t) \log\left(\sum_{a'} \exp\left(u\left(a', t\right)/\lambda\right) p\left(a'\right)\right).$$

Using the envelope theorem on the payoff (5), with the optimally chosen information strategy, we immediately obtain the following proposition.

Proposition 6 *The traveler's optimal payoff decreases with the unit information cost, i.e.*

$$\frac{\partial \Lambda\left(\hat{P}\right)}{\partial \lambda} < 0.$$

4 The marginal cost of travel time variability for a rationally inattentive traveler

Consider the marginal change in payoff (5) following a change in the standard deviation of travel time σ. As $\hat{P}$ is optimally chosen, the envelope theorem implies that it may be considered constant. Moreover, σ does not affect the constraints on P that probabilities must be positive and sum to 1. We need then only consider the change in the expected scheduling utility $Eu\left(A, T\right)$. Holding P constant at $\hat{P}$, we have that the cost of a marginal increase in σ is

$$\begin{aligned}
-\frac{\partial \Lambda\left(\hat{P}\right)}{\partial \sigma} &= -\frac{\partial \Lambda\left(P\right)}{\partial \sigma}|_{P=\hat{P}} \\
&= -\frac{\partial}{\partial \sigma} E\left(\int_0^A h\left(s\right) ds + \int_{A+\mu+\sigma X}^0 w\left(s\right) ds\right) \\
&= E\left(X w\left(A + \mu + \sigma X\right)\right).
\end{aligned}$$

This expression is similar to what is found in the neoclassical scheduling model with an exogenously known travel time distribution and a fixed departure time choice (Engelson and Fosgerau, 2016).

By Lemma 9 in the Appendix, this cost is necessarily positive.

Proposition 7 *The cost of travel time variability is positive,*

$$-\frac{\partial \Lambda\left(\hat{P}\right)}{\partial \sigma} > 0.$$

The expression for the marginal cost of standard deviation of travel time simplifies considerably in the case when w is linear as in the slope model (2) and (3), since then

$$\begin{aligned}
\frac{\partial \Lambda\left(\hat{P}\right)}{\partial \sigma} &= -E\left(X\left(\gamma_0+\gamma_1\left(A+\mu+\sigma X\right)\right)\right) \\
&= -E\left(\left(\gamma_0+\gamma_1 A+\gamma_1 \mu\right) X+\gamma_1 \sigma X^2\right) \\
&= -E\left[E\left[\left(\gamma_0+\gamma_1 A+\gamma_1 \mu\right) X+\gamma_1 \sigma X^2 | X\right]\right] \\
&= -E\left[\left(\gamma_0+\gamma_1 E\left(A|X\right)+\gamma_1 \mu\right) X+\gamma_1 \sigma X^2\right] \\
&= -\gamma_1 E\left[E\left(A|X\right) X\right]-\gamma_1 \sigma E\left[X^2\right] \\
&= -\gamma_1 E\left(AX\right)-\gamma_1 \sigma,
\end{aligned}$$

by the law of iterated expectations and using that $EX = 0, EX^2 = 1$.

Denote by $|A| = \#\{a \in \mathcal{A} | p(a) > 0\}$ the number of elements in the optimally chosen consideration set. In the case when the consideration set consists of just one point, $|A| = 1$, then A and X are independent which implies that $E\left(AX\right) = EA \cdot EX = 0$. In this case, the marginal cost of travel time variance is constant

$$-\frac{\partial \Lambda\left(\hat{P}\right)}{\partial \sigma^2} = -\frac{\partial \Lambda\left(\hat{P}\right)}{\partial \sigma}\frac{1}{2\sigma} = \frac{\gamma_1}{2}.$$

This is exactly the result of Fosgerau and Engelson (2011). It emerges from the rational inattention model as a special case when the optimally chosen consideration set consists of just one point, and this corresponds to the assumption in the Fosgerau and Engelson (2011) model that the traveler chooses a single departure time. The Fosgerau and Engelson (2011) result is very convenient when it applies, since then the marginal cost of travel time variance depends only on the slope of the utility rate and not on the distribution of travel time or the information cost.

Let us now consider the more general case when there are at least two points in the consideration set. By Lemma 12 in the Appendix, if $|A| >$

1, then $E(AX) < 0$ and hence the cost of travel time variability for the rationally inattentive traveler is strictly smaller than $\gamma_1\sigma$ in the case of a linear w. The conditional choice of departure time A will be dependent on T and the marginal cost of travel time variance will be

$$-\frac{\partial\Lambda\left(\hat{P}\right)}{\partial\sigma^2} = \frac{\gamma_1}{2} + \frac{\gamma_1 E(AX)}{2\sigma} < \frac{\gamma_1}{2}.$$

Altogether, we have established the following result.

Proposition 8 *If the utility rate w is linear, then $0 < -\frac{\partial\Lambda(\hat{P})}{\partial\sigma^2} \leq \frac{\gamma_1}{2}$. If there are at least two points in the consideration set, $|A| > 1$, then the latter inequality is strict.*

In the case of the step model, we find that

$$\frac{\partial\Lambda\left(\hat{P}\right)}{\partial\sigma} = -(\beta+\gamma)\, E\left(X1_{\{A+T\geq 0\}}\right),$$

which is essentially the same result as in Fosgerau and Karlstrom (2010). The term $E\left(X1_{\{A+T\geq 0\}}\right)$ is observable if scheduling preferences are known. It is the mean standardized lateness. A difference from Fosgerau and Karlstrom (2010) is that the event of being late now depends on the distribution of $A+T$, which is influenced by the choice of the traveler in the present model.

5 Numerical example

This section provides a numerical example to illustrate the behavior of the rationally inattentive traveler facing random travel time variability. The travel time distribution used in the example is normal with mean $\mu = 4$ and standard deviation $\sigma = 1$, it is shown in Figure 2.[4]

The traveler has linear utility rates (2) and (3) with parameters $\beta_0 = \gamma_0 = -1$, $\beta_1 = -0.5$ and $\gamma_1 = 0.5$. Both the travel times and departure times are discretized into intervals of width 0.05.

[4] We have also conducted simulations where travel time follows a triangular or a log-normal distribution. The simulation results were qualitatively the same as those shown here.

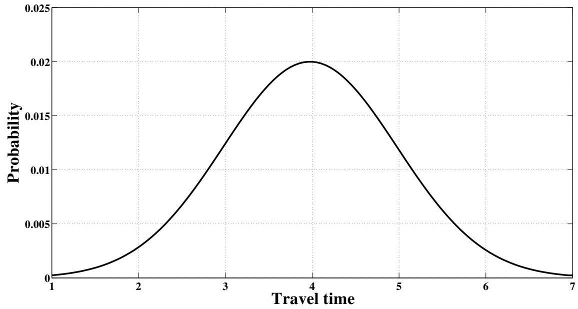

Figure 2: Travel time distribution

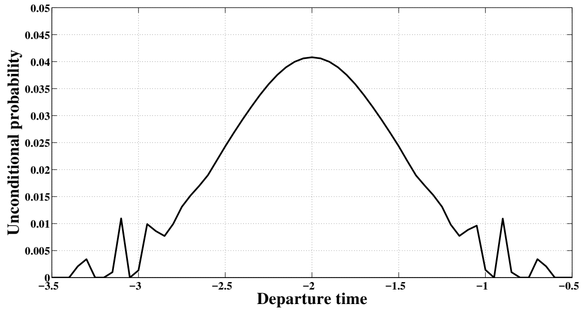

Figure 3: Unconditional probability with $\lambda = 0.01$

Figures 3-6 depict the unconditional departure time choice probabilities $p(a)$ under different values of the unit information cost λ. The consideration set is the set of departure times with $p(a) > 0$. As expected from Proposition 4, very early or late departure times are never chosen. Comparing Figures 3 - 6, it can be seen that the cardinality of consideration sets decreases with λ.

The shape of $p(a)$ varies quite a lot with λ. In Figure 3, the highest probability occurs in the middle of the consideration sets, while in Figures 4 and 6, the opposite is the case. We can explain the results as follows.

The unit information cost has two impacts on the departure time choice. First of all, it determines the size of consideration sets with $p(a) > 0$. The options with lower probability to produce a high payoff are ignored. That is the reason the departure time choices far away from the center are ignored in Figures 4 - 6, even if they could yield a high payoff under some specific realizations of t. Second, they will become able to choose conditional departure times producing larger utility with higher probability in each state

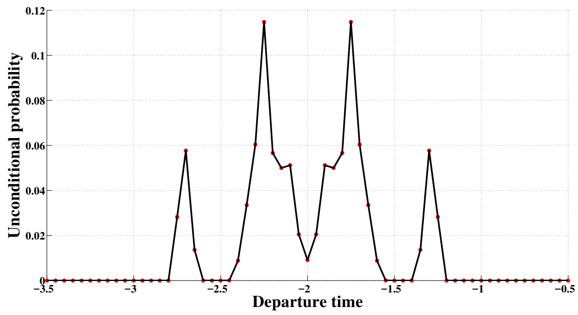

Figure 4: Unconditional probability with $\lambda = 0.1$

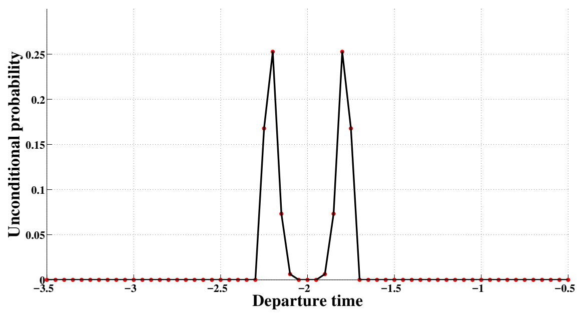

Figure 5: Unconditional probability with $\lambda = 0.2$

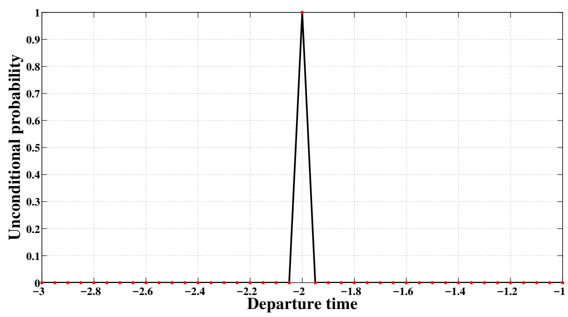

Figure 6: Unconditional probability with $\lambda = 0.25$

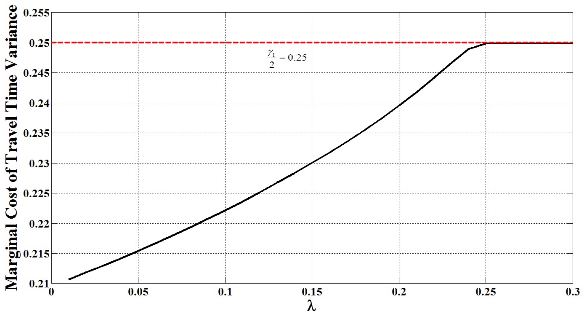

Figure 7: Marginal cost of travel time variability with different λ

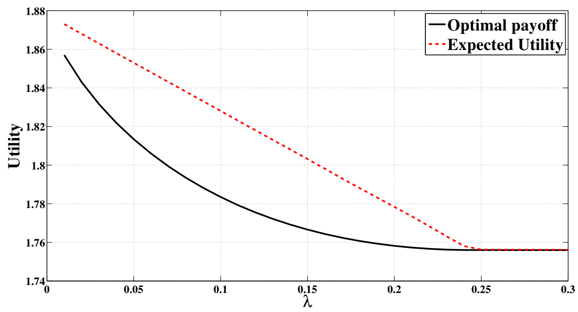

Figure 8: Optimal payoff and expected utility with different λ

t. According to the analysis of entry test for each choice in Caplin et al. (2016), an option is more likely to be chosen when its utility is high in states where other options yield much lower utility. That can explain the shape of Figures 3 and 4. When one departure time a is chosen with high unconditional probability, the unconditional probability of departure times nearby may drop as they produce similar utility as a. The optimal information strategy balances the desire to achieve high utility in very different states with the desire for parsimoniousness. As shown in Figures 4 and 5, rationally inattentive travelers will consider fewer options when information is expensive.

As the unit information cost increases, the consideration set shrinks and comprises fewer and fewer points. This implies that $p(a)$ for any $a \in \mathcal{A}$ depends non-monotonically on λ (except if $p(a)$ is always zero).

According to Proposition 8, the marginal effect of travel time variance is smaller than the case without information. Figure 7 shows the result. In all cases considered here, $-\frac{\partial \Lambda(P)}{\partial \sigma^2} < \frac{\gamma_1}{2} = 0.25$. It is evident that $-\frac{\partial \Lambda(P)}{\partial \sigma^2}$

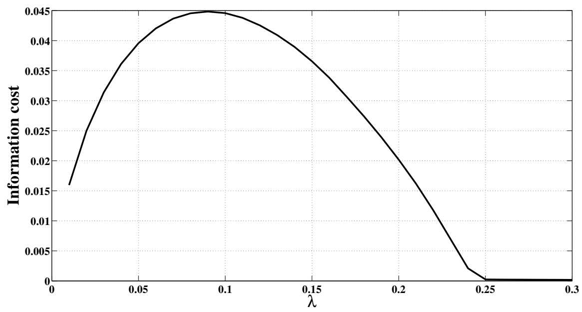

Figure 9: Information cost with different λ

increases with unit information cost λ. As information becomes more expensive, travelers acquire less information and hence become more influenced by travel time variability. It seems it should be possible to establish this formally, but we have not been able to do this.

Figure 8 and 9 shows the information cost, expected utility and optimal payoff for different values of λ. We have established in Proposition 6 that the optimal payoff is decreasing in λ. It is intuitive that the expected utility also decreases with λ because travellers are less able to respond to random travel time with a less informative information strategy.

Figure 9 shows that the (total) information cost is non-monotonic with respect to λ. For low values of λ, the change in λ dominates the change in the quantity of information acquired, $I(A, T)$, and hence the information cost increases. For high values of λ, the opposite is the case. At $\lambda = 0.25$ as in Figure 6, there is only one option left in the consideration set and the information cost can reduce only very little. Comparison of Figures 3-6 shows that the information cost has a large impact on the departure time choice, even if it only accounts for a small share of the optimal payoff.

6 Conclusion

This paper has set up a rational inattention model for the traveler choice of departure time under random travel time variability. The model takes into account that travelers may choose to be more or less informed about the variation in travel time, using the information to condition their choice of departure time. Implicit in the model is a signal that intermediates between travel time and departure time choice, which makes the departure time choice a random variable that carries information about the random travel time. The model generates predictions that all seem very reasonable.

Going beyond the present results, we may ask if travelers actually do acquire information. This is possible to observe in principle. We would just need to observe the distribution of (A, T) for groups of travelers and check if the mutual information is positive. To do this, a complication is that it is necessary to ensure that travelers are comparable, i.e. that they face the same travel time distribution and have the same (or similar) scheduling preferences.

The rational inattention model might be useful for studying consideration sets. It is a very attractive feature of that model that it generates consideration sets that are formed endogenously based on preferences.

7 Acknowledgements

Gege Jiang is supported by General Research Fund #615712 and #616113 of the Research Grants Council of the HKSAR Government and the Hong Kong PhD Fellowship. Mogens Fosgerau is supported by the Danish Innovation Fund under the URBAN project.

References

Avineri, E. and Prashker, J. N. (2006) The impact of travel time information on travelers learning under uncertainty *Transportation* **33**(4), 393–408.

Ben-Elia, E., Di Pace, R., Bifulco, G. N. and Shiftan, Y. (2013) The impact of travel informations accuracy on route-choice *Transportation Research Part C: Emerging Technologies* **26**, 146–159.

Bifulco, G. N., Cantarella, G. E., Simonelli, F. and Velonà, P. (2016) Advanced traveller information systems under recurrent traffic conditions: Network equilibrium and stability *Transportation Research Part B: Methodological* **92**, 73–87.

Caplin, A., Dean, M. and Leahy, J. (2016) Rational Inattention, Optimal Consideration Sets and Stochastic Choice *Working Paper* .

Cover, T. M. and Thomas, J. A. (2006) *Elements of information theory* 2nd edn Wiley-Interscience.

De Borger, B. and Fosgerau, M. (2012) Information provision by regulated public transport companies *Transportation Research Part B: Methodological* **46**(4), 492–510.

Engelson, L. and Fosgerau, M. (2016) The cost of travel time variability: Three measures with properties *Transportation Research Part B: Methodological* **91**, 555–564.

Ettema, D. and Timmermans, H. (2006) Costs of travel time uncertainty and benefits of travel time information: conceptual model and numerical examples *Transportation Research Part C: Emerging Technologies* **14**(5), 335–350.

Fosgerau, M. (2015) The Valuation of Travel Time Variability *International Transport Forum* (November).

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

Fosgerau, M. and Karlstrom, A. (2010) The value of reliability *Transportation Research Part B: Methodological* **44**(1), 38–49.

Fosgerau, M., Melo, E., de Palma, A. and Shum, M. (2017) Discrete Choice and Rational Inattention: a General Equivalence Result.

Hébert, B. and Woodford, M. (2016) Rational Inattention with Sequential Information Sampling.

Lindsey, R., Daniel, T., Gisches, E. and Rapoport, A. (2014) Pre-trip information and route-choice decisions with stochastic travel conditions: Theory *Transportation Research Part B: Methodological* **67**, 187–207.

Matějka, F. and McKay, A. (2015) Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model *American Economic Review* **105**(1), 272–298.

Paz, A. and Peeta, S. (2009) Behavior-consistent real-time traffic routing under information provision *Transportation Research Part C: Emerging Technologies* **17**(6), 642–661.

Schrank, D., Eisele, B. and Lomax, T. (2011) Urban Mobility Report *Technical Report September* Texas Transportation Institute, Texas A&M University College Station. **URL:** mobility.tamu.edu

Sims, C. A. (2003) Implications of rational inattention *Journal of Monetary Economics* **50**(3), 665–690.

Sims, C. A. (2010) Chapter 4 Rational Inattention and Monetary Economics *Handbook of Monetary Economics* Vol. 3 pp. 155–181. **URL:** http://www.sciencedirect.com/science/article/pii/B9780444532381000041

Small, K. A. (1982) The Scheduling of Consumer Activities: Work Trips *American Economic Review* **72**(3), 467–479.

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

Vickrey, W. S. (1969) Congestion Theory and Transport Investment *The American Economic Review* **59**(2), 251.

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

Xiao, Y. and Lo, H. K. (2016) Day-to-day departure time modeling under social network influence *Transportation Research Part B: Methodological* **92**, 54–72.

Yang, H. (1998) Multiple equilibrium behaviors and advanced traveler information systems with endogenous market penetration *Transportation Research Part B: Methodological* **32**(3), 205–218.

A Proofs

Proof of Proposition 3.

Let w be linear. Assume for all t in some neighborhood that

$$0 = \sum_a \alpha_a \exp\left(\int_0^a h(t')dt' + \int_{a+t}^0 w(t')dt'\right),$$

where $a \in \mathcal{A}$ are all different. We will arrive at a contradiction.

Index $\mathcal{A} = \{a_1, \ldots, a_K\}$. Differentiating repeatedly and using that w' is a constant shows that

$$0 = \sum_{k=1}^{K} w^n \left(a_k + t\right) \alpha_k \exp\left(\int_0^{a_k} h(t')dt' + \int_{a_k+t}^{0} w(t')dt'\right) \tag{10}$$

for all integer $n \geq 0$. But the matrix

$$V = \begin{pmatrix} 1 & w\left(a_1+t\right) & w^2\left(a_1+t\right) & \cdots & w^{K-1}\left(a_1+t\right) \\ 1 & w\left(a_2+t\right) & w^2\left(a_2+t\right) & \cdots & w^{K-1}\left(a_2+t\right) \\ 1 & w\left(a_3+t\right) & w^2\left(a_3+t\right) & \cdots & w^{K-1}\left(a_3+t\right) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & w\left(a_K+t\right) & w^2\left(a_K+t\right) & & w^{K-1}\left(a_K+t\right) \end{pmatrix}$$

is a Vandermonde matrix and hence is invertible. We have then from (10) that

$$\left(\alpha_1 \exp\left(\int_0^{a_1} h(t')dt' + \int_{a_1+t}^{0} w(t')dt'\right), \ldots, \alpha_K \exp\left(\int_0^{a_K} h(t')dt' + \int_{a_K+t}^{0} w(t')dt'\right)\right) V = 0$$

and hence that

$$\left(\alpha_1 \exp\left(\int_0^{a_1} h(t')dt' + \int_{a_1+t}^{0} w(t')dt'\right), \ldots, \alpha_K \exp\left(\int_0^{a_K} h(t')dt' + \int_{a_K+t}^{0} w(t')dt'\right)\right) = 0V^{-1} = 0,$$

which is a contradiction. ■

Proof of Proposition 5. Suppose $a_1, a_2, a_3 \in A$, $a_2 = \epsilon a_1 + (1-\epsilon)a_3$, $0 < \epsilon < 1$. If $p(a_1) > 0$ and $p(a_3) > 0$, then

$$\begin{aligned} & \sum_{t \in T} \frac{\exp\left(u\left(a_2, t\right)/\lambda\right) p\left(t\right)}{\sum_{a'} \exp\left(u\left(a', t\right)/\lambda\right) p\left(a'\right)} \\ & \geq \sum_{t \in T} \frac{\left(\epsilon \exp\left(u\left(a_1, t\right)/\lambda\right) + (1-\epsilon)(\exp\left(u\left(a_3, t\right)/\lambda\right))\right) p\left(t\right)}{\sum_{a'} \exp\left(u\left(a', t\right)/\lambda\right) p\left(a'\right)} \\ & = \epsilon \sum_{t \in T} \frac{\exp\left(u\left(a_1, t\right)/\lambda\right)) p\left(t\right)}{\sum_{a'} \exp\left(u\left(a', t\right)/\lambda\right) p\left(a'\right)} + (1-\epsilon) \sum_{t \in T} \frac{\exp\left(u\left(a_3, t\right)/\lambda\right) p\left(t\right)}{\sum_{a'} \exp\left(u\left(a', t\right)/\lambda\right) p\left(a'\right)} = 1 \end{aligned}$$

Therefore $p(a_2) > 0$. If the utility function is strictly superconcave, the equality will never hold, which contradicts Proposition 1, implying that the cardinality of the consideration set can be no larger than 2. ■

Lemma 9 *Let X be a non-degenerate random variable with mean zero. Let f be an increasing function. Then $E[Xf(X)] > 0$.*

Proof. Note that

$$\begin{aligned} E[Xf(X)] &= E[X(f(X) - f(0))] \\ &= E\left[1_{\{X<0\}}X(f(X) - f(0))\right] + E\left[1_{\{X\geq 0\}}X(f(X) - f(0))\right]. \end{aligned}$$

Since the event that $X < 0$ occurs with positive probability, the first term is the product of two negative numbers, while the second term is the product of two positive numbers. Hence the conclusion follows. ■

We will sign $E(AX)$ through a series of lemmas.

Lemma 10 *Suppose $|A| > 1$. For a general utility rate w, let $0 < p(a|x) < 1$ for all $(a, x) \in \mathcal{A} \times \mathcal{X}$. Then for every $x \in \mathcal{X}$ there is a point a_x such that $\frac{\partial p(a|x)}{\partial x} > 0$ if and only if $a < a_x$.*

Proof. We are holding the distribution of X constant and hence the probabilities $p(a)$ do not change. Use (6) and (1) to find that

$$\frac{\partial p(a|x)}{\partial x} = -\frac{\sigma}{\lambda} w(a + \mu + \sigma x) p(a|x) + p(a|x) \sum_{a' \in \mathcal{A}} \frac{\sigma}{\lambda} w\left(a' + \mu + \sigma x\right) p\left(a'|x\right).$$

Then $\frac{\partial p(a|x)}{\partial x} > 0$ if and only if

$$\sum_{a' \in \mathcal{A}} w\left(a' + \mu + \sigma x\right) p\left(a'|x\right) > w(a + \mu + \sigma x).$$

The LHS does not depend on a, while the LHS is monotonically increasing as a function of a. The inequality is satisfied at $a = \min \mathcal{A}$, since $0 < p(\min \mathcal{A}|x) < 1$. The inequality is not satisfied at $a = \max A$ by the parallel argument. The conclusion follows by monotonicity of w.

■

Lemma 11 *For a general utility rate w, holding the distribution of X constant.*

$$\frac{\partial E(A|X = x)}{\partial x} < 0.$$

Proof. Let $\mathcal{B} = \{a \in \mathcal{A} : p(a) = 0\}$. Compute

$$\frac{\partial E\left(A|X=x\right)}{\partial x} = \sum_{a\in\mathcal{A}\backslash\mathcal{B}} a\frac{\partial p\left(a|x\right)}{\partial x}+\sum_{a\in\mathcal{B}} a\frac{\partial p\left(a|x\right)}{\partial x} = \sum_{a\in\mathcal{A}\backslash\mathcal{B}} \left(a-a_x\right)\frac{\partial p\left(a|x\right)}{\partial x} < 0,$$

where we have used the previous lemma and that $\sum\limits_{a\in\mathcal{A}} \frac{\partial p(a|x)}{\partial x} = 0$. ■

Lemma 12 *Suppose that* $|A| > 1$*. Then* $E\left(AX\right) < 0$*.*

Proof. Let $\mathcal{X}$ be the set of support points for the standardized travel time distribution. Let $f\left(x\right) : [\min \mathcal{X}, \max \mathcal{X}] \to \mathbb{R}$ be the continuous extension of $E\left(A|X=x\right) : X \to \mathbb{R}$. that linearly interpolates between points. Then f is decreasing by Lemma 11 and defined at $x = 0$. Note also that $f < 0$ everywhere. Now,

$$\begin{aligned} E\left(AX\right) &= E\left(E\left(A|X\right)X\right) \\ &= \sum_x p\left(x\right)xf\left(x\right), \end{aligned}$$

and this is negative by Lemma 9. ■