A route guidance system considering travel time unreliability

Haengju Lee, Saerona Choi, Hojin Jung, Byungkyu Brian Park & Sang H. Son

To cite this article: Haengju Lee, Saerona Choi, Hojin Jung, Byungkyu Brian Park & Sang H. Son (2019): A route guidance system considering travel time unreliability, Journal of Intelligent Transportation Systems, DOI: 10.1080/15472450.2018.1542303

To link to this article: https://doi.org/10.1080/15472450.2018.1542303

Published online: 23 Jan 2019.

Submit your article to this journal

Article views: 44

View Crossmark data
A route guidance system considering travel time unreliability

Haengju Lee\textsuperscript{a}, Saerona Choib, Hojin Jung\textsuperscript{c}, Byungkyu Brian Park\textsuperscript{d}, and Sang H. Son\textsuperscript{c}

\textsuperscript{a}School of Business, Pusan National University, Busan, Republic of Korea; \textsuperscript{b}Transportation Safety Research and Development Institute, Korea Transportation Safety Authority, Gimcheon, Republic of Korea; \textsuperscript{c}Department of Information and Communication Engineering, DGIST, Daegu, Republic of Korea; \textsuperscript{d}Link Lab & Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA

ABSTRACT
Under a stochastic roadway, drivers need a route guidance system incorporating travel time variability. To recommend a customized path depending on the trip purpose and the driver’s risk-taking behavior, various path ranking methods have been developed. Unlike those methods, our proposed disutility method can easily incorporate a target arrival time in the ranking process by measuring how late the travel is and by penalizing it depending on the severity of lateness. In addition, the disutility-based route guidance system can properly address travel time unreliability that causes unacceptable disruptions to the driver’s schedule (i.e., unexpected long delay). We compare the disutility-based path ranking method with other ranking methods, the percentile travel time, the mean excess travel time, and the on-time arrival probability. We show that the disutility has stronger discriminating power and requires less solution space to find an optimal path. The most important advantage is that it can estimate a driver’s risk-taking behavior for each trip purpose by using the discrete choice analysis. We construct a simulation framework to acquire the travel time data on a hypothetical roadway. We analyze the data and show how various ranking methods recommend a customized path. Using the data, we show the advantage of the disutility method over the other methods, which is generating a customized path with respect to the target arrival time by properly penalizing the travel time lateness.

ARTICLE HISTORY
Received 13 March 2016
Revised 26 October 2018
Accepted 26 October 2018

KEYWORDS
Delay penalty cost; route guidance system; target arrival time; travel time disutility; travel time unreliability

Introduction
The travel time on a roadway is subject to a significant degree of uncertainty that arises from both supply side and demand side. The uncertainty introduced by supply-side sources (e.g., weather conditions, traffic incidents, work zones, and traffic signal controls) results in roadway capacity variation, which typically leads to non-recurrent congestion (Chen & Zhou, 2010). On the other hand, the uncertainty introduced by demand-side sources (e.g., special events, population characteristic, and temporal factors) results in travel demand fluctuation, which usually leads to recurrent congestion (Clark & Watling, 2005). There are also complex interactions between the supply-side and demand-side sources of uncertainty. For example, inclement weather may reduce roadway capacity and change the travel demand pattern (e.g., departure time change or trip cancellation) at the same time (Chen & Zhou, 2010).

Under the travel time uncertainty, drivers are not able to foretell exactly when they will arrive at their destinations. This uncertainty is considered as a risk by drivers. Thus, a route guidance system generating a path merely minimizing the expected (or average) travel time may not be suitable to drivers who incorporate travel time variability in their route decisions. Furthermore, if a driver has a target arrival time to a destination, a route guidance system needs to model a lateness factor measuring how late a trip is from the target arrival time in ranking multiple alternative paths. The lateness factor should penalize excessively late arrival times. Recent empirical studies (Srinivasan, Prakash, & Seshadri 2014; van Lint, van Zuylen, & Tu, 2008) revealed that the travel time distribution is heavily skewed with a long fat tail over a wide range. This means that there is a considerable number of unlucky drivers experiencing unexpected long delay, and the consequence of it is much more serious than those of modest delays. Hence, it is necessary to consider travel time unreliability on the long tail of its distribution. Most papers dealing with travel time...
uncertainty have focused on travel time \textit{reliability} to determine an optimal path (see Literature review section). However, we explicitly take into account travel time unreliability on the long tail of the travel time distribution to determine an optimal path.

Another important characteristic to be modeled under travel time uncertainty is heterogeneous delay penalty rates. The heterogeneity originates from various trip purposes and different drivers' risk-taking behaviors. A driver tries to reduce risk of arriving late if the trip is for an important event. For example, a trip for a typical grocery shopping has a negligible penalty cost for being late. In contrast, an airport trip catching a reserved flight or a trip for a job interview can entail a significant penalty cost. In addition, drivers may impose different penalties depending on their risk-taking behaviors. For a trip with an even same purpose (say it is for a job interview), a driver may impose a much higher penalty cost for being late due to his higher degree of risk aversion than other drivers. de Palma and Picard (2005) analyzed survey data in Paris to understand route choice behavior, and showed that risk aversion depends on several key socio-economic variables (i.e., larger risk aversion for transit users, blue collars, and for business appointments).

In this paper, we propose a route guidance system based on disutility. The aforementioned travel time unreliability and heterogeneous penalty rates for a given target arrival time can be modeled in the disutility formulation. For a pair of origin and destination, multiple alternative paths may exist especially in a large-scale roadway. We rank those paths based on the disutility to determine the optimal path (i.e., the one minimizing the disutility). The disutility is the sum of two terms. The first one is an operating cost associated with travel time (e.g., fuel consumption). The second one is a penalty cost on late arrival (beyond the target arrival time). Intrinsically, there are two objectives to be minimized in the disutility function: average travel time and unexpected long delay. In some cases, the objectives may conflict with each other. This means that an optimal solution for one objective may be a very undesirable solution for the other objective. For example, assume that Path A consists of mainly local roads with minimum average travel time, but with high uncertainty. Thus, a driver has a high probability of experiencing a long delay due to traffic signals and congestion. On the other hand, assume that Path B is a highway with a longer average travel time, but with less uncertainty. Path B may be an optimal choice for a risk-averse driver, whereas Path A may be an optimal solution for a risk-prone driver. The risk-averse driver is willing to accept a slightly longer average travel time in order to avoid encountering a very significant delay.

There have been many prior studies employing different reliability indices (e.g., percentile travel time, travel time budget, and mean excess travel time). We provide a complete review on those indices in Literature review section. We describe the disutility model in Disutility model description section and compare it with other ranking methods in Comparison with PTT, METT, and OTP section. We show that disutility has stronger discriminating power than the other ranking methods. We also compare them in the stochastic dominance framework. We show that the disutility method requires smaller solution space to find an optimal path. One of main advantages of the disutility method is the ability to estimate a driver's risk-taking tendency that varies depending on the trip purpose. Because the choice behavior is modeled by decision maker's disutility, we can incorporate discrete choice analysis in the literature. It estimates the penalty cost from the historical choice data, and the estimated cost is especially useful when a driver travels to an unknown destination (where a route guidance system plays an important role). We detail the estimation method in Estimation of normalized penalty cost section.

Computing the expected disutility of a path requires full knowledge of the travel time distribution. In practice, it is often difficult to completely characterize the travel time distribution, particularly for a fast-changing and large-scale roadway. In fact, the whole historical travel time data may have not been stored in the database (maybe due to the limited sample size along arterial roadways such as current INRIX data at \url{http://inrix.com}), so it is difficult to estimate the distribution. We study the route choice problem assuming partial information about the travel time distribution, minimum, mean, and variance. As mentioned before, travel time distribution is skewed with a long tail (Srinivasan et al., 2014; van Lint et al., 2008). For this type of distribution, a shifted log-normal distribution is adequate to approximate the non-negative random variable. The log-normal distribution is widely adopted to describe natural phenomena. Examples include the blood pressure of adult human in Makuch, Freeman, and Johnson (1979), the income in Clementi and Gallegati (2005), the stock market indices in Black and Scholes (1973), time to repair a maintainable system in O’Connor and Kleyner (2012), and travel time on a roadway in Srinivasan et al. (2014).
The random travel time in the shifted log-normal distribution is composed of two parts: the minimum-possible travel time and the excess travel time over the minimum-possible travel time. The excess travel time (i.e., travel time minus the minimum-possible travel time) is a random variable following the log-normal distribution with a location parameter and a scale parameter. Srinivasan et al. (2014) adopt the shifted log-normal distribution to compute the on-time arrival probability. In contrast, we adopt the distribution to compute the expected disutility. In Shifted log-normal approximation section, we show a nice structural property of the computed expected disutility. That is the expected disutility is increasing in the minimum-possible travel time, the location parameter, and the scale parameter. Hence, we only consider paths on the efficient frontier in the three-dimensional graph of those three parameters after eliminating dominated paths. In Simulation study section, we construct a simulation framework to acquire the travel data on a hypothetical roadway. We analyze the data and apply the percentile arrival time, the mean excess travel time, the on-time arrival probability, and the proposed disutility to rank paths. All four ranking methods generate a customized path, but only the disutility model explicitly considers a target arrival time while penalizing the travel time lateness. The main advantage of the disutility model is that we model only one variable with explicit economic meaning, i.e., the penalty rate, to generate a path of driver’s interest.

The route guidance system ranks multiple alternative paths sharing the same origin and the destination. However, enumerating all the paths is not practical especially for a large-scale roadway. Complexity can be managed by constructing a reasonable subset of paths for each origin and destination pair. The paths in the subset are then ranked according to their expected disutility values. If we know the distributions of all the links and paths, we can employ the second-order stochastic dominance (SSD) approach (Nie et al., 2012; Wu & Nie, 2011). The SSD-admissible path set, generated by a label-correcting algorithm, serves well as a subset of paths. However, for a fast-changing and large-scale roadway, it is not easy to update and study the travel time distribution. Even in this case, we should have key statistical variables such as mean and variance to characterize the travel time. Assuming that it is the case, we may resort to the k-shortest path algorithm (Eppstein, 1998; Yen, 1971). However, it has a disadvantage of often not generating acceptable alternative paths because of too many overlapping links among paths. To overcome this shortcoming, a link weighting based k-shortest path algorithm can be applied (Lee & Park, 2008; Park & Rilett, 1997). An overlapping penalty function in the algorithm prevents the overlap among the paths and topologically disperses them. Therefore, for such a large-scale network, the disutility-based path ranking algorithm can be applied after a subset of paths is generated by one of previously developed algorithms (Eppstein, 1998; Lee & Park, 2008; Park & Rilett, 1997; Yen, 1971).

The main contribution of this work is to propose the route guidance system based on disutility, which considers the travel time unreliability and heterogeneous penalty rates while explicitly incorporating target arrival time. A customized path depending on the trip purpose and the driver’s risk-taking behavior can be recommended by using one variable (i.e., the penalty rate). In addition, the penalty rate reflecting driver’s decision behavior can be estimated from the historical data and can be applied to systematically generate a path for a new destination. We illustrate one possible implementation of the proposed disutility-based ranking in Figure 1 (any variations to this flowchart can be made to serve its own purpose). All the proofs of the results are in Appendix.

**Literature review**

When the travel time on a roadway is highly variable, a route guidance system that recommends a path merely minimizing the expected travel time may fail to assist a driver well. In fact, the recent route guidance system becomes more intelligent by providing real-time traffic information (Gkiotsalitis & Stathopoulos, 2015; He, Zheng, Guan, & Mao, 2016; Liao & Chen, 2015; Ma, Zhou, & Lee, 2016).

A risk-averse driver may consider that reliability of a path to arrive on time is important (de Boer, Snelder, van Nes, & van Arem, 2017). In this regard, there are various optimal path problems considering the travel time reliability. Wu and Nie (2011) used stochastic dominance to characterize admissible paths in a route choice model with uncertain travel times and analyzed the relationship between the stochastic dominance approach and other route choice models employing a variety of reliability indices. Lam and Chan (2005) argued that travelers should consider both travel time index and travel time reliability for their route choices. They proposed a path preference index by combining the path travel time and the path travel time reliability indices. The major reliability indices include the percentile travel time (PTT), the
travel time budget (TTB), the mean excess travel time (METT), and the expected utility. However, these indices do not explicitly consider the target arrival time that the driver attempts to meet. In contrast, we explicitly incorporate the target arrival time in our disutility formulation and adjust the penalty cost to generate customized paths. We review the major reliability indices.

To rank a set of paths with random travel times, Frank (1969) assumed that drivers maximize the probability of completing a trip within a given time, which is called on-time arrival probability (OTP). This is equivalent to minimizing the PTT for a desired OTP. Nie and Wu (2009a) showed how to find a minimum PTT path by finding all non-dominated paths under the first-order stochastic dominance rule. They developed a general label-correcting algorithm to find the non-dominated paths and conducted a case study of reliable route guidance using the Chicago regional transportation network. Nie and Wu (2009b) extended their previous work (Nie & Wu, 2009a) by considering the correlations between travel times on adjacent links.

Another method of ranking paths is to use the TTB (also called effective travel time). Hall (1983) assumed that drivers tend to reserve a safety margin to hedge against variation of travel times. Specifically, drivers typically add an extra buffer to the expected travel time to ensure on-time arrival when planning important trips. The sum of the expected travel time and the safety margin is called the TTB. The safety margin is the product of the standard deviation of travel time and a scalar called punctuality parameter. Uchida and Lida (1993) defined travel time variation as a risk and proposed two risk assignment models. Both models are built using the TTB concept. Nikolova and Stier-Moses (2014) established conditions that characterize an equilibrium traffic assignment and found the existence conditions of the equilibrium assuming that strategic travelers select the path minimizing the TTB.
Lo, Luo, and Siu (2006) formulated a multi-class mixed-equilibrium mathematical program using the TTB. In their work, travelers are in different classes with different degrees of risk aversion.

Chen and Zhou (2010) argued that neither PTT nor TTB can properly account for unreliability aspect of travel time variability. To avoid unacceptable disruptions to their schedule, they defined a new criterion considering both reliability and unreliability of travel time. The criterion is called the mean excess travel time (METT). This is an extension of the PTT and is the sum of the PTT (reliability part) and a tardy time measure (unreliability part). Their approach is similar to our work in that it considers unreliability of travel time in the route choice model. However, the distinguished feature of our model is the consideration of the target arrival time in the path ranking process. We compare PPT and METT with the disutility method in Comparison with PTT, METT, and OTP section.

Some researchers employed expected utility theory assuming that each decision maker selects an alternative maximizing the expected utility. Various functional forms can be defined to represent traveler’s utility. However, effective algorithms to find an optimal path exist only for linear or exponential utility functions (Loui, 1983; Mirchandani & Soroush, 1987). So, the standard shortest path algorithm cannot be applied to find an optimal path for the quadratic utility function, which is defined to capture risk-averse behavior (Mirchandani & Soroush, 1987). The optimal solution lies in the set of non-dominated paths by simultaneously minimizing the mean and variance of travel times (Loui, 1983), which is named as the mean-variance rule.

For a complete review, we introduce another research stream in characterizing the route choice, which is the robust optimization. In this framework, decision-makers are assumed to optimize returns in the worst scenario, and they find a best path minimizing the worst-case travel time. Bell and Cassir (2002) assumed that travel times are set by an adversary picking the worst-possible travel time for the traveler. Bertsimas and Sim (2003) studied a robust optimization model without considering correlations between travel times, and proposed a polynomial algorithm to find a robust shortest path. They defined the robust shortest path as the path having the best worst-case travel time, and the worst-case travel time is obtained assuming that the number of links deviating from the nominal value does not exceed the budget of uncertainty. Note that the deviation from the nominal value is maximal and the budget of uncertainty is a model parameter capturing the driver’s degree of risk aversion. Ordoñez and Stier-Moses (2010) applied the robust shortest path algorithm to derive the robust equilibrium in a traffic assignment problem. In the robust equilibrium, travelers select paths by solving a robust optimization problem that imposes a limit on the number of links that can deviate from the mean.

Disutility model description

We consider a directed and connected road network, $G(N, A)$, where $N$ represents a set of nodes and $A$ a set of links. We assume that a driver intends to travel from an origin of $o \in N$ to a destination of $d \in N$ with a certain travel purpose. The total number of alternative paths from $o$ to $d$ is denoted by $N$. Each path is indexed by $j$. Let $\pi_j$ be a path from $o$ to $d$ and $T_j$ be the random travel time of $\pi_j$. We denote the mean of $T_j (= E[T_j])$ by $\mu_j$ and the standard deviation of $T_j (= \sqrt{Var(T_j)})$ by $\sigma_j$. Assume that the driver has a target time to arrive at destination $d$, which is denoted by $\tau$. Without loss of generality, the current time is set to be zero. Let $c$ be the unit travel time cost that is incurred to the driver (i.e., unit operating cost such as fuel consumption) and $p$ be a unit penalty cost pertinent to late arrival (i.e., penalty cost for travel time beyond the target arrival time of $\tau$) for the driver. Note that the driver has his/her own value of $p$, which depends on the travel purpose and his risk-taking behavior. In Estimation of normalized penalty cost section, we discuss how to estimate the penalty cost by using the historical choice data of the driver. We denote the ratio of $p$ to $c$ by $\bar{p} (= \frac{p}{c})$, which measures the severity of the penalty cost compared with the travel time cost.

The expected disutility of the driver associated with driving on path $\pi_j$ is given by

$$u_j = c \cdot \mu_j + p \cdot E[{T_j - \tau}]^+, \tag{1}$$

where $X^+ \equiv \max\{X, 0\}$ and $j = 1, \ldots, N$. The first part is the total expected travel time cost, and the second term is the total expected penalty cost for being late. The objective is to find an optimal path, denoted by $\pi^*$, minimizing the expected disutility. This is equivalent to minimizing the following normalized expected disutility (NED). It is $u_j$ divided by $c$.

$$\text{NED}_j = \mu_j + \bar{p} \cdot E[{T_j - \tau}]^+,$$

Mathematically, the problem is to find a path of $\pi^* = \pi_j$, such as

$$j^* \in \arg\min_{j=1, \ldots, N} \text{NED}_j.$$
In an extreme case of no late penalty cost (i.e., \( p = 0 \) or \( \bar{p} = 0 \)), it is simply the traditional shortest path problem, finding a path with the minimum expected travel time. Since the penalty cost is non-negative, the zero value is the lower bound of the cost. The upper bound of the cost can also be determined. As the value of \( \bar{p} \) increases (i.e., \( \bar{p} \to \infty \)), the second term gets more weights in determining the minimum value of the disutility. Based on this intuition, we define a path of \( \pi_p \) that minimizes the second term (i.e., the penalty part) of the disutility. It is given by

\[
\bar{p}^0 = \min \left\{ \bar{p}^* | \arg \min_{j=1, \ldots, N} \text{NED}_j \right\}.
\]

**Proposition 3.1.** For all \( \bar{p} \geq \bar{p}^0 \), the optimal path is \( j^0 \). Thus, it is enough to consider \( \bar{p} \in [0, \bar{p}^0] \) to determine the optimal path.

We explain the determination of \( \bar{p}^0 \) with an example. We consider three paths as in Figure 2 where the NED of each path is a function of \( \bar{p} \). In the graph, \( \mu_j \) corresponds to the y-intercept and \( E[T_j - \tau] \) corresponds to the slope by Equation (1).

Therefore, Path 2 has the minimum expected travel time, whereas Path 1 has the minimum expected penalty (i.e., minimum slope). This means that \( j^0 = 1 \). Accordingly, the value of \( \bar{p}^0 \) is the minimum value of \( \bar{p} \) at which Path 1 minimizes NED as marked in the graph.

**Comparison with PTT, METT, and OTP**

In this section, we compare the major reliability indices, PTT, METT, and OTP, with NED. To prompt the comparison, we define those indices. The most common reliability index is the PTT. For each path \( \pi_j \), we let \( t_j(z) \) be the \( x \)-percentile travel time (\( x \)-PTT). If we use the \( x \)-PTT to rank paths, we find a path with smallest \( x \)-PTT. Mathematically,

\[
\begin{align*}
\bar{j}^{\text{PTT}}(x) &\in \arg\min_{j=1, \ldots, N} t_j(x). \\
\end{align*}
\]

On the other hand, METT is the conditional excessive delay expectation measuring travel time reliability and unreliability (Chen & Zhou, 2010). We let \( f_j \) and \( F_j \) be the probability distribution function (PDF) and the cumulative distribution (CDF) of \( T_j \) respectively. A finite upper bound of the support of the travel time is denoted by \( T \) (i.e., \( F_j(T) = 1 \)). The METT of path \( \pi_j \) is the sum of the \( x \)-PTT and a tardy time measure.

\[
\text{METT}_j(z) = t_j(z) + \frac{1}{1 - z} \int_{t_j(z)}^{T} (t - t_j(z)) \cdot f_j(t) \, dt.
\]

The second terms are an expected excessive delay conditional on the choice of \( x \)-PTT. With this ranking criterion, we find a path minimizing METT.

\[
\begin{align*}
\bar{j}^{\text{METT}}(x) &\in \arg\min_{j=1, \ldots, N} \text{METT}_j(x). \\
\end{align*}
\]

While both PTT and METT do not explicitly consider the target arrival time (i.e., \( \tau \)) in their formulations, OTP considers it because it is the probability of completing a trip within a given time (Nie & Wu, 2009a). The given time is the target arrival time plus an extra buffer, \( \tau + b \). A traveler adjusts the extra buffer for each trip. The formulation is given by \( \text{OTP}_j(b) = F_j(\tau + b) \). We let \( \bar{\text{OTP}}_j(b) = 1 - \text{OTP}_j(b) = 1 - F_j(\tau + b) \). In this ranking criterion, we find a path maximizing OTP, which is in turn minimizing \( \bar{\text{OTP}} \). The minimization problem is selected to make it consistent with the other formulations.

\[
\begin{align*}
\bar{j}^{\text{OTP}}(b) &\in \arg\min_{j=1, \ldots, N} \bar{\text{OTP}}_j(b). \\
\end{align*}
\]

In Graphical comparison section, we compare four ranking methods graphically. We then compare them based on the stochastic dominance theory in Stochastic dominance-based comparison section.
**Graphical comparison**

We define closely-related indices to facilitate graphical representation. Those are the normalized PTT (NPTT) and the normalized METT (NMETT) and obtained by multiplying $1/C_0$ such as $NPTT_j(x) = (1-x) \cdot t_j(x)$ and $NMETT_j(x) = (1-x) \cdot METT_j(x)$. It is clear that a path minimizing PTT (or METT) also minimizes NPTT (or NMETT). If the CDF of $T_j$ is depicted as in Figure 3(a), we can easily represent the NPTT in the graph, which is the rectangular area connecting the points A, B, C, and E. Hence,

$$NPTT_j = \text{Area}_{ABCE}$$

To derive mathematical formulations for NMETT and NED, we apply the following functional equality.

The equation measures the expected late time over time $a$. For any $a \geq 0$, using integration by part, we have

$$\int_a^T (t-a) \cdot f_j(t) \, dt = (t-a) \cdot F_j(t) |_a^T - \int_a^T f_j(t) \, dt = T-a - \int_a^T f_j(t) \, dt.$$  

Note also that the above equation is $E[T_j-a]^+$ because $E[T_j-a]^+ = \int_0^T (t-a)^+ \cdot f_j(t) \, dt = \int_0^T (t-a) \cdot f_j(t) \, dt$. We then obtain the equation for $NMETT_j(x)$ as follows.

$$NMETT_j(x) = (1-x) \cdot t_j(x) + T-t_j(x) - \int_{t_j(x)}^T f_j(t) \, dt.$$  

The first part is the NPTT, and the rest is the area connecting E, C, and D in Figure 3(a). We obtain

---

**Figure 3.** Graphical Comparison of NPTT, NMETT, and NED. (a) NPTT and NMETT (b) NOTP and NED (c) Same NPTT and NMETT (d) Different NED.
Similarly, applying the above equality to the NED, we have
\[
\text{NED}_j = E[T_j] + \bar{p} \cdot E[T_j - \tau]^+
= T - \int_0^T F_j(t)dt + \bar{p} \cdot \left( T - \tau - \int_0^T F_j(t)dt \right).
\]

We use \( E[T_j] = E[T_j - 0]^+ \) in the second equality. Referring to Figure 3(b), we have
\[
\text{NED}_j = \text{Area}_{ABCD} + \bar{p} \cdot \text{Area}_{ECD}.
\]

We also normalize \( \text{OTP} \) to have \( \text{NOTP} = (\tau + b) \cdot \text{OTP} \). According to Figure 3(b), \( \text{NOTP}_j \), which is \( (\tau + b) \cdot (1 - F_j(\tau + b)) \), is the area connecting A, G, I, and H.
\[
\text{NOTP}_j(b) = \text{Area}_{AGIH}.
\]

Travelers adjust the value of \( \tau \) for both NPTT and NMETT depending on different trip purposes, while they adjust the value of \( \bar{p} \) for NED and the value of \( b \) for \( \text{NOTP} \). We explain the inherent choice ambiguity in using NPTT, NMETT, and \( \text{NOTP} \) to rank paths. This ambiguity results from that the three ranking methods consider only part of the distribution. NPTT does not consider the shape of \( F_j \) other than the location of point B. Similarly, \( \text{NOTP} \) only considers the location of point G. NMETT considers the shape of \( F_j \) for \( t \geq t_j(\tau) \), but not for \( t < t_j(\tau) \). In contrast, NED incorporates the shape of the distribution over the entire region. This means that NED may result in choice ambiguity, but with less frequency. In that sense, NED is a better scoring criterion because it has stronger discriminating power.

We illustrate one example in Figure 3(c). We add the CDFs of other path travel times, \( F_i \) (with dashed lines) and \( F_k \) (with dotted lines). The CDFs of \( F_i, F_k, \) and \( F_j \) overlap when \( t \geq t_j(\tau) \). This means that the three paths have the same values of NPTT and NMETT (The shaded region corresponds to NMETT), leading to a difficulty in ranking those paths for this \( \tau \) value. The NED can resolve this choice ambiguity. Figure 3(d) illustrates the difference of the expected travel times of \( \pi_i \) and \( \pi_j \) (both paths have the same tardy time). The difference of the NED is the area between \( F_i \) and \( F_j \). (Similarly, we can compare them with \( \pi_k \)) Hence, the NED ranks \( \pi_i \) higher than \( \pi_j \) and \( \pi_k \) without any ambiguity.

**Stochastic dominance-based comparison**

First, we briefly introduce the stochastic dominance (SD) theory in Wu and Nie (2011). The SD theory is used to compare random variables when their distributions are known. Wu and Nie (2011) adopted this theory to compare random travel times. The first and second order SD are defined as follows.

**Definition 4.1.** (First order SD) We write \( T_j \succ_1 T_i \) if \( F_j(t) \geq F_i(t), \forall t, \) and \( \exists \) an open interval \( \Lambda \in [0, T] \) with nonzero Lebesgue measure such that \( F_j(t) > F_i(t), \forall t \in \Lambda \). (Second order SD) We write \( T_j \succ_2 T_i \) if \( \int_0^T F_j(t)dt \geq \int_0^T F_i(t)dt, \forall t, \) and \( \exists \) an open interval \( \Lambda \in [0, T] \) with nonzero Lebesgue measure such that \( \int_0^T F_j(t)dt > \int_0^T F_i(t)dt, \forall t \in \Lambda \).

These SD rules impose a partial order, so they are applied to eliminate paths that are dominated by others. The index set of paths that are not dominated is useful because it helps reduce the solution space for a decision maker. We define those two sets.

**Definition 4.2.** Let \( \Gamma_{SSD} \) be the index set of paths that are not dominated by other paths in the first order SD and \( \Gamma_{SSD} \) be the index set of paths that are not dominated by other paths in the second order SD.

We summarize the results in Wu and Nie (2011) that are useful for our comparison. The first one states that the set based on the first order SD includes the set based on the second order SD. The second one is that to find an optimal path according to PPT or METT or OTP, it is enough to observe paths in the set of \( \Gamma_{SSD} \).

**Lemma 4.1.** (Wu & Nie, 2011) The relationship between the sets is
\[
\Gamma_{SSD} \supseteq \Gamma_{SSD}.
\]

For a given \( \tau \), the optimal path based on PPT or METT or OTP can be found in \( \Gamma_{SSD} \).

In contrast, we show that the optimal path can be found in \( \Gamma_{SSD} \) if paths are ranked by NED in Proposition 4.1. This means that finding the optimal path under NED requires smaller solution space (i.e., less number of candidate paths), compared with PTT, METT, OTP. We state this result in Proposition 4.2.

**Proposition 4.1.**
\[
\min_{j=1,...,N} \text{NED}_j = \min_{j \in \Gamma_{SSD}} \text{NED}_j.
\]

**Proposition 4.2.** Because \( \Gamma_{SSD} \supseteq \Gamma_{SSD} \), we can further reduce the solution space by eliminating paths that are dominated in the second order SD under NED.

**Estimation of normalized penalty cost**

If the route guidance system collects all the past decisions for trips that the driver is familiar with, the set
of penalty costs categorized according to the trip purpose can be estimated. This set of costs is used for other unfamiliar trips for the driver. Since the route guidance system is especially useful to guide the driver traveling to unknown destinations, the NED’s ability to estimate the driver’s risk-taking behavior depending on trip purpose is an essential addition to the system. We categorize various trip purposes and label them according to the importance of trip. We let those levels be \( l = 1, \ldots, L \) such as \( l = 1 \) represents the least important trip (e.g., grocery shopping) and \( l = L \) the most important trip (e.g., job interview). Then our objective is to estimate \( \overline{p}_l \) for \( l = 1, \ldots, L \).

We apply the maximum likelihood estimation method in the discrete choice model. A logit choice model is used to have a closed-form solution for choice probabilities. Train (2009) provides an excellent editorial work on the logit based choice model. For each \( l = 1, \ldots, L \), we go through the following process to estimate \( \overline{p}_l \). For notational simplicity, we drop the subscript \( l \) as far as it is obvious. Let \( \pi_1^{(1)}, \pi_2^{(1)}, \ldots, \pi_{N_l}^{(1)} \) be the alternative paths with a target arrival time \( \tau^{(1)} \) in the first choice data with label \( l \). The corresponding travel times are denoted by \( T_1^{(1)}, T_2^{(1)}, \ldots, T_{N_l}^{(1)} \). We let his/her choice be \( \pi_r^{(1)} \). Generally, for the \( k^{th} \) choice data with label \( l \), we denote alternative paths by \( \pi_1^{(k)}, \pi_2^{(k)}, \ldots, \pi_{N_l}^{(k)} \), the target arrival time by \( \tau^{(k)} \), and his/her path choice by \( \pi_r^{(k)} \). Let the total number of choice data with label \( l \) be \( M \). The expected utility in choosing \( \pi_r^{(k)} \) is the negative expected disutility, i.e., \( -\mu_r^{(k)} - \overline{p}_l \cdot E[\tau^{(k)} - \tau^{(k)}]^+ \). In the multinomial logit choice model, the probability of choosing \( \pi_r^{(k)} \) in the \( k^{th} \) choice data is given by

\[
q_r^{(k)} = \frac{e^{-\mu_r^{(k)} - \overline{p}_l \cdot E[\tau^{(k)} - \tau^{(k)}]^+}}{\sum_{j=1}^{N_l} e^{-\mu_j^{(k)} - \overline{p}_l \cdot E[\tau^{(k)} - \tau^{(k)}]^+}}.
\]

We obtain the log-likelihood function by adding the log of the choice probabilities,

\[
\sum_{k=1}^{M} \log q_r^{(k)}
\]

Applying the maximum likelihood estimation method, we find the estimation of \( \overline{p}_l \) that maximizes the above log-likelihood.

### Shifted log-normal approximation

The computation of the expected late penalty cost, \( E[T_j - \tau]^+ \) in Equation (1), requires full knowledge of the travel time distribution. In a fast-changing and large-scale roadway, it may be difficult to completely characterize and update the travel time distribution. However, it can be feasible to keep and update several key statistical variables for each path: mean, variance, and minimum travel time. Note that the mean, variance, and minimum travel time can be updated recursively as a new observation is available (Knuth, 1997; Welford, 1962).

If the distribution of \( T_j \) is approximated by the shifted log-normal distribution, the travel time is composed of two parts, the minimum-possible travel time and the random travel time in excess of the minimum. Srinivasan et al. (2014) also adopted the shifted log-normal distribution to compute the reliability of travel time (i.e., on-time arrival probability). In their model, the free-flow travel time is used for the minimum-possible travel time. Because the free-flow travel time (or the travel time at the speed limit) may not be reached when the road congestion always exits, we believe that the minimum-possible travel time is more appropriate to describe the travel time characteristic. The travel time is written as

\[
T_j = \gamma_j + \exp (\eta_j + \nu_j \cdot Z).
\]

The term of \( \gamma_j \) represents the minimum-possible travel time of path \( \pi_j \), and \( \exp (\eta_j + \nu_j \cdot Z) \) is the excess travel time where \( Z \) is the standard normal random variable. This means that the logarithm of the excess travel time (i.e., \( \ln (T_j - \gamma_j) \)) follows the normal distribution with mean \( \eta_j \) and standard deviation \( \nu_j \).

In the log-normal distribution, \( \eta_j \) is called the location parameter, and \( \nu_j \) is called the scale parameter. We denote \( T_j \sim \text{SLN}(\eta_j, \nu_j, \gamma_j) \). The PDF of \( T_j \) is then given by

\[
f_T(T_j = t|\eta_j, \nu_j, \gamma_j) = \frac{1}{\nu_j \sqrt{2\pi}} \exp \left( \frac{-(\ln(t - \gamma_j) - \eta_j)^2}{2\nu^2_j} \right).
\]

The parameters, \( \eta_j \) and \( \nu_j \), can be estimated from the mean and the standard deviation of \( T_j \) (i.e., \( \mu_j, \sigma_j \)). Those are \( \eta_j = \ln (\mu_j - \gamma_j) + \ln \left( \frac{\sigma^2_j}{\sigma^2_j + (\mu_j - \gamma_j)^2} \right) \) and \( \nu_j = \sqrt{\ln \left( 1 + \frac{\sigma^2_j}{(\mu_j - \gamma_j)^2} \right)} \). Reversely, the mean and the standard deviation of \( T_j \) can be derived from \( \eta_j \) and \( \nu_j \), which are \( \mu_j = \gamma_j + \exp (\eta_j + \nu_j^2) \) and \( \sigma^2_j = (\exp (\nu^2_j) - 1) \cdot (2\eta_j + \nu^2_j) \). We estimate \( \gamma_j \) by the minimum travel time of \( T_j \). We denote the PDF and CDF of \( Z \) by \( \phi(\cdot) \) and \( \Phi(\cdot) \), respectively. We
introduce two properties of the shifted log-normal distribution in the following two lemmas for the later NED computations. Lemma 6.1 is the computation of the expected delay. Lemma 6.2 is utilized in the proof of Proposition 6.1.

**Lemma 6.1.** If the distribution of $T_j$ is approximated by the shifted log-normal distribution with parameters, $\eta_j, \nu_j$ and $\gamma_j$ (i.e., $T_j \sim \text{SLN}(\eta_j, \nu_j, \gamma_j)$), the expected delay is

$$E[T_j - \bar{\tau}]^+ = \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln (\tau - \gamma_j) + \nu_j}{\nu_j} \right)$$

$$- (\tau - \gamma_j) \cdot \Phi \left( \frac{\eta_j - \ln (\tau - \gamma_j)}{\nu_j} \right).$$

**Lemma 6.2.** We have the following equality relationship.

$$\exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \phi \left( \frac{\eta_j - \ln (\tau - \gamma_j) + \nu_j}{\nu_j} \right)$$

$$= (\tau - \gamma_j) \cdot \phi \left( \frac{\eta_j - \ln (\tau - \gamma_j)}{\nu_j} \right).$$

Using Lemma 6.1, we can rewrite the NED in Equation (1) in the closed-form solution. It is

$$\bar{u}_j = \exp \left( \eta_j + \frac{\nu_j^2}{2} + \gamma_j + \bar{\rho} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \right)$$

$$\cdot \Phi \left( \frac{\eta_j - \ln (\tau - \gamma_j) + \nu_j}{\nu_j} \right)$$

$$- \bar{\rho} \cdot (\tau - \gamma_j) \cdot \Phi \left( \frac{\eta_j - \ln (\tau - \gamma_j)}{\nu_j} \right).$$

(2)

The following proposition shows that the NED function is increasing in $\gamma_j, \eta_j$, and $\nu_j$. This means that if Path A has a larger minimum-possible travel time, a larger location parameter and a larger scale parameter than Path B does, the NED of Path A is bigger than that of Path B. In this case, we call that Path A is dominated by Path B. Hence, we can only consider paths lying on the efficient frontier in the threedimensional graph of the three parameters after eliminating dominated paths. We illustrate this using an example. Assume that we have six alternative paths, which are denoted by P1, P2, P3, P4, P5, and P6. For simplicity, we assume that the minimum-possible travel time of all the six paths are zero. Figure 4 is a scatter plot of the location and scale parameters. All of the paths are represented as cross marks with labels of their path names. The two numbers in parenthesis are $\eta_j$ and $\nu_j$. For example, P1 has $\eta_1 = 1.77$ and $\nu_1 = 0.48$. Note that P5 is dominated by P1, and P6 is dominated by P3 and P4. After eliminating P5 and P6, we can construct the efficient frontier connecting P1, P2, P3, and P4. Therefore, we can efficiently reduce the solution space to the efficient frontier.

**Proposition 6.1.** If the distribution of $T_j$ is approximated by $\text{SLN}(\eta_j, \nu_j, \gamma_j)$, then the expected normalized disutility, $\bar{u}_j$, is increasing in $\gamma_j, \eta_j$, and $\nu_j$.

Note that Proposition 6.1 states the monotonicity in terms of $\eta_j$ and $\nu_j$, not in terms of $\mu_j$ and $\sigma_j$. Note that a random variable with greater mean and standard deviation may not have greater mean and standard deviation for the logarithm of it. To show this, assume that Path A has mean of 800 and standard deviation of 10. We can easily see that Path A has greater mean and greater standard deviation than Path B does. However, the logarithm of travel time of Path A has mean 6.68 and standard deviation 0.02, and that of Path B has mean 5.99 and standard deviation of 0.03. Hence, the standard deviation of the logarithm of travel time of Path A is smaller than that of Path B. This can be

![Figure 4](image-url)
explained because the coefficient of variation \( \frac{\sigma_j}{\mu_j} \) determines the magnitude of the standard deviation of the logarithm of travel time. Therefore, according to Proposition 6.1, Path A (or Path B) is not dominated by Path B (or Path A).

**Simulation study**

In this section, we analyze the simulated travel time data acquired on a hypothetical roadway as shown in Figure 5(a), which was developed through the component object model, or COM interface of VISSIM (Manual, 2006a, 2006b). The roadway consists of one pair of origin and destination, 52 links (4.5 km of freeway, 3.8 km of main arterial, and minor streets), and 9 traffic signalized intersections. The level of service D (Manual, 2010) was selected as a background traffic condition for traffic signalized intersections and the freeway. During the 1,800 seconds of simulation time, 800 vehicles were generated and their routes (or paths) were tracked. The dynamic assignment procedure in VISSIM was simulated not once but repetitively, and drivers chose their paths through the roadway based on the travel cost they have experienced during the preceding simulations as if those travel times were experienced from previous days. As in Figure 5(b), there exist three paths going from the origin (a circular point located on the bottom in the right side) to the destination (a circular point located on the top in the left side) for the simulation result. Path 1 is the freeway denoted by the dashed lines. Path 2 is the main arterial denoted by the solid lines. Path 3 is a composite of main arterial and minor streets, which is denoted by the dotted lines.

**Data-based estimation**

The descriptive statistics of the simulated travel times are summarized in Table 1. For example, the mean travel time of Path 1 is 446 seconds, and its standard deviation is 222 seconds. The minimum travel time of Path 1 is 222 seconds, and the maximum travel time is 1435 seconds. The total number of vehicles traveling on Path 1 is 310. Path 1 has the largest mean travel time but the smallest standard deviation among the three paths. Therefore, it is highly possible that Path 1 turns out to be the most reliable path. The total number of vehicles tracked is 797. This is because the three vehicles did not complete their trips within the simulation time (i.e., 1,800 seconds).

We estimate the reliability indices, PTT, METT, OTP, and NED, using the simulated travel time data. For a given \( \alpha \) value, the \( \alpha \)-PTT is computed from the data and is shown in Figure 6(a). The figure illustrates \( \alpha \)-PTT for \( \alpha \) values from 0.90 to 0.99 with increment of 0.01. When \( \alpha = 0.9 \), the PTT of Path 1 is 695, that of Path 2 is 767, and that of Path 3 is 604. Therefore, Path 3 has the minimum PTT. In contrast, when \( \alpha = 0.99 \), the PTT of Path 1 is 1295, that of Path 2 is 1343, and that of Path 3 is 1534. In this case, Path 1 has the minimum PTT. Based on the PTT analysis,

**Table 1. Descriptive statistics in simulation study (time unit in seconds).**

<table>
<thead>
<tr>
<th>Path 1 (( j = 1 ))</th>
<th>Path 2 (( j = 2 ))</th>
<th>Path 3 (( j = 3 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean (( \mu_j ))</strong></td>
<td>446</td>
<td>436</td>
</tr>
<tr>
<td><strong>Standard deviation (( \sigma_j ))</strong></td>
<td>222</td>
<td>259</td>
</tr>
<tr>
<td><strong>Min Travel Time</strong></td>
<td>222</td>
<td>261</td>
</tr>
<tr>
<td><strong>Max Travel Time</strong></td>
<td>1,435</td>
<td>1,377</td>
</tr>
<tr>
<td><strong>No. of vehicles</strong></td>
<td>310</td>
<td>379</td>
</tr>
</tbody>
</table>

Figure 5. Roadway and paths in simulation study. (a) Road network (b) Resulting Paths
the optimal path is Path 3 if $\alpha = 0.90, 0.91, \ldots, 0.97$, and it is Path 1 if $\alpha = 0.98$ and 0.99. This means that Path 1 is chosen by only extremely-risk-averse drivers with $\alpha = 0.98$ or 0.99 (i.e., drivers considering the worst case of travel times). To alleviate this extremity, the METT analysis is performed. Both METT and NED require the computation of $E[T_j - a]^+$, for $a > 0$. This is estimated from the data by

$$\tilde{E}[T_j - a]^+ = \frac{\sum_{i=1}^{n_j} (T_{j(i)} - a)^+}{n_j},$$

Figure 6. Path ranking for simulation data. (a) PTT analysis (b) METT analysis (c) $\bar{OTP}$ analysis with $\tau = 900$ (d) $\bar{OTP}$ analysis with $\tau = 1,000$ (e) NED analysis with $\tau = 900$ (f) NED analysis with $\tau = 1,000$
where \( n_j \) is the number of data points of path \( \pi_j \) and \( T_j^{(i)} \) is \( i \)-th travel time data of path \( \pi_j \). Then the METT of path \( \pi_j \) is estimated by \( t_j(x) + E[T_j - t_j(x)] \), which is depicted in Figure 6(b). The optimal path is Path 3 if \( x = 0.90, 0.91, \) and \( 0.92 \), but it is Path 1 if \( x \geq 0.93 \). Compared with the PTT analysis, the METT recommends Path 1 (the reliable path) for both moderately-risk-averse and extremely-risk-averse drivers. This is because the METT penalizes the late travel time with the tardy time measure.

Now we consider the case where a driver has an explicit target arrival time. Directly mapping the target arrival time to a proper value of \( x \) in PTT and METT is not an easy task. However, it is possible for both \( \text{OTP} \) (=1-OTP) and NED. The CDF is estimated from the data as follows.

\[
\tilde{F}_j(\tau + b) = \frac{\sum_{i=1}^{n_j} I(T_j^{(i)} \leq \tau + b)}{n_j},
\]

whereas \( I(\cdot) \) is the indicator function. The \( \text{OTP} \) of path \( \pi_j \) is estimated by \( 1 - \tilde{F}_j(\tau + b) \). We consider two values for the target arrival time, which are \( \tau = 900 \) and \( \tau = 1,000 \). Assume first that \( \tau = 900 \). We increase \( b \) from zero with increment of 20. The \( \text{OTP} \) is illustrated in Figure 6(c). If \( b \leq 260 \), the optimal path is Path 3. However, if \( b \geq 280 \), the optimal path is Path 1. A driver with a higher safe margin of \( b \) tends to prefer the reliable path, Path 1. We now assume that \( \tau = 1,000 \). This case is illustrated in Figure 6(d). If \( b \leq 160 \), the optimal path is Path 3. If \( 180 \leq b \leq 360 \), the optimal path is Path 1. However, it is interesting to observe that if \( b = 380, 400 \), the optimal path is Path 2. This is because the maximum travel time of Path 2 is 1,377. Hence, \( \text{OTP} \) of Path 2 in those cases becomes zero. This results from that \( \text{OTP} \) considers only part of the distribution, as pointed out in the previous graphical comparison.

We perform the NED analysis for each target arrival time. The NED of path \( \pi_j \) is estimated by \( \tilde{E}[T_j] + \tilde{p} \cdot \tilde{E}[T_j - \tau] \) from the simulated travel time data. We first consider \( \tau = 900 \) and increase the value of \( \tilde{p} \) from zero with increment of 2. If \( \tilde{p} = 0, 2, \ldots, \) and 22, the optimal path is Path 3. If \( \tilde{p} \geq 24 \), it is Path 1. Figure 6(e) shows the NED in terms of \( \tilde{p} \). A driver with a higher value of \( \tilde{p} \) tends to prefer a reliable path, Path 1. In contrast, for a driver with a lower value of \( \tilde{p} \), the risky path but with the minimum average travel time (Path 3) is preferred. We then consider \( \tau = 1,000 \), which is shown in Figure 6(f). If \( \tilde{p} = 0, 2, \ldots, \) and 6, the optimal path is Path 3. If \( \tilde{p} \geq 8 \), it is Path 1. Compared with the case of \( \tau = 900 \), a driver with \( \tilde{p} = 8, \ldots, \) and 22 selects the safer path (i.e., Path 1). This is because Path 3 has a longer tail than Path 1 beyond \( \tau = 1000 \) (i.e., the maximum travel time of Path 3 is 1872 whereas that of Path 1 is 1435).

We articulate the benefits of NED compared with PTT, METT, and \( \text{OTP} \). All of four ranking methods recommend a different path depending on trip purposes and risk-taking behaviors, as seen in the simulation study. This verifies that the proposed NED can generate a customized path very well, as the other ranking methods do. However, only NED explicitly considers a target arrival time while penalizing the travel time lateness. By doing this, NED considers the shape of the distribution over the entire region while avoiding the unusual results (e.g., Path 2 recommendation of \( \text{OTP} \) for \( b = 380, 400 \) and \( \tau = 1,000 \)). Another distinct advantage of NED is that the rank-control lever has an economic meaning to travelers. The lever of \( \tilde{p} \) is simply the ratio of the unit penalty cost to the unit travel time cost. Hence, unlike \( x \) for PTT and METT and \( b \) for \( \text{OTP} \), it is straightforwardly computed by users’ cost analysis. Finally, NED first addresses the question of how to systematically estimate the customized preference (i.e., normalized penalty cost \( \tilde{p} \)) from the historical choice data.

### Shifted log-normal approximation

While assuming that the only available data is Table 1 (without the whole historical travel times), we fit the shifted log-normal distribution. The estimation result is summarized in Table 2. We depict the histograms of travel time data and overlap the approximated shifted log-normal distributions in Figure 7. We can observe that all three histograms have the skewness to the right (i.e., there are significant drivers experiencing long delays), and the shifted log-normal distribution explains the skewness well. Even though there are some areas showing mismatches between the histograms and the log-normal distributions, the approximation based on the shifted log-normal distribution is effective for a fast-changing and large-scale roadway. Indeed, the route guidance system must generate an optimal path within a small processing time window. Otherwise, impatient drivers may be reluctant in using the system. In this regard, both speed and

<table>
<thead>
<tr>
<th>Path 1 (( j = 1 ))</th>
<th>Path 2 (( j = 2 ))</th>
<th>Path 3 (( j = 3 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_j )</td>
<td>4.7767</td>
<td>5.1308</td>
</tr>
<tr>
<td>( \nu_j )</td>
<td>0.9427</td>
<td>0.8643</td>
</tr>
<tr>
<td>( \tau_j )</td>
<td>261</td>
<td>190</td>
</tr>
</tbody>
</table>
performance are important factors in the route guidance system. By adopting the shifted log-normal approximation, the system can generate an optimal path without any need to estimate the travel time distributions for all alternative paths for each route request. Furthermore, it allows closed-form solutions for the disutility model (i.e., no need to use Monte Carlo simulation). In Table 3, we compare the estimation of the expected delay, \( E[T_{ij}/C_{0s}/C_{138}] \) for \( s = 900 \) and \( 1000 \). The third row is for the estimation from the simulated travel time data and the last row is for the estimation based on the shifted log-normal distribution. We can again see that the shifted log-normal distribution explains the skewness to the right. The three distributions are compared in Figure 7(d). The peak of Path 1 is located in the highest travel time, but it is tallest. (Again, Path 1 may be the most reliable path.)

**Figure 7.** Histograms and shifted log-normal distributions in simulation study. (a) Path 1 (b) Path 2 (c) Path 3 (d) Fitted SLN distributions

| Table 3. Estimation of expected delay. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( \tau \)      | 900              | 1000            | 900             | 1000            | 900             | 1000            |
| \( T_{ij}/C_{0s}/C_{138} \) | 18               | 25              | 19              | 11              | 16              | 14              |
| Data            | Path 1           | Path 2          | Path 3          | Path 1          | Path 2          | Path 3          |
Assume that a route guidance system ranks the paths using the NED. The NED with respect to $\bar{p}$ for the case of $\tau = 900$ is depicted in Figure 8(a). When $\bar{p} = 0, 2, \ldots, 10$, the optimal choice is Path 3. However, when $\bar{p} \geq 12$, the optimal choice is Path 1. Likewise, the NED for the case of $\tau = 1,000$ is depicted in Figure 8(b). If $\bar{p} = 0, 2, \ldots, 14$, the optimal choice is Path 3. If $\bar{p} \geq 16$, it is Path 1. Differently from the NED analysis based on the simulated data in Data-based estimation section, a driver with $\bar{p} = 12$ and 14 selects the riskier path (Path 3) for $\tau = 1,000$ (i.e., $\bar{p} = 12$) in the shifted log-normal approximation. This is because the maximum travel times are smoothed out in the shifted log-normal approximation.

Conclusions and future research

The proposed disutility-based route guidance system considers the travel time unreliability and recommends an optimal path that is customized to the travel purpose and drivers’ risk-taking behavior. This is especially effective for a roadway on which the travel time distribution is skewed with a long fat tail over a wide range. In this type of distribution, there are considerably many drivers experiencing unexpected long delays. Hence, it is important to consider the travel time unreliability on the long tail of the travel time distribution to decide an optimal path. The problem setting that we consider in the route choice problem is that a driver has a target arrival time and a delay cost for the late arrival. The purpose of travel and drivers’ risk-taking behavior can be represented by a single variable of the delay cost. If the historical choice data is available, then the disutility model can estimate the delay cost using the discrete choice analysis. This is a potential strong point for a route guidance system compared with other ranking methods. Because each trip has a different purpose and each driver possesses a heterogeneous risk-taking behavior, it is proper for a route guidance system to generate a customized path for each route request. This paper achieves this objective.

To compute the disutility of a path, we need to have full knowledge of the travel time distribution. However, it is sometimes not possible to characterize it in fast-changing and large-scale roadways. In this case, the shifted log-normal distribution is adopted because it explains the skewness of travel time distribution well. The shifted log-normal distribution is often used to model non-negative random variables when the coefficient of variation is large. In this case, the normal distribution is not appropriate to use because it assigns a significant probability to negative travel time. Additionally, it allows closed-form solutions for the disutility model. For this reason, the optimization is achieved effectively without any need to resort to a computationally expensive Monte Carlo simulation to estimate the expected delay. Note that the simulation method is unfortunately not viable to generate an optimal path within a small processing time window. Therefore, the shifted log-normal approximation is attractive for practical applications. Because the disutility is increasing in the three parameters of the shifted log-normal distribution (i.e., minimum-possible travel time, location parameter, and scale parameter), we only consider paths on the efficient frontier of the graph of the three parameters.

We apply the disutility-based model to the simulated travel times on a hypothetical roadway. Diverse
optimal paths are generated depending on the trip purpose and drivers’ risk-taking behavior. For a driver with a low penalty cost, he may be willing to take a risky path. For a driver with a high penalty cost, he may prefer a safe path. We show that the disutility-based model serves this tendency well. We also validate this tendency with the shifted log-normal approximation using only the major statistics of the simulated travel times.

We discuss several potential future researches. To shed light on potentials of the disutility-based route guidance system in the field implementation, it is necessary to develop a simulation framework to test the effects of various factors. The factors include the market penetration rate of the proposed route guidance system, the drivers’ compliance rates, and the existence of other competing route guidance strategies. Another direction for future research is to study the interaction between the disutility-based route choice system and traffic flows in an equilibrium framework. The equilibrium analysis is to study the result of two competing mechanisms and to find the traffic flow pattern. The two mechanisms exist because drivers try to choose their paths minimizing their disutilities, and the disutilities vary depending on the usage of the roadway. Thus, it is meaningful to study the equilibrium traffic pattern by modeling the disutility-based travel decisions and congestion. In addition, real-time traffic information is available due to various traffic sensors (e.g., loop detectors, connected vehicles, video surveillance system). This information motivates the development of online routing algorithms (Du, Peeta, & Kim, 2013; Tian et al., 2013). We leave the development of the disutility-based route guidance system that incorporates the dynamic traffic information to future research.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03030126); and Global Research Laboratory Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT Future Planning (NRF-2013K1A1A2A02078326).

Disclosure statement

No potential conflict of interest was reported by the authors.

References


Transportation Research Part B, 36(8), 671–681. doi: 10.1016/S0191-2615(01)00022-4


Appendix

Proof of Proposition 3.1. Because the optimal path is \( p_0 \) if \( \bar{p} = \bar{p}_0 \), we have

\[
\mu_j + \bar{p}_0 \cdot E[T_j - \tau] \geq \mu_j + \bar{p}_0 \cdot E[T_j - \tau],
\]

for all \( j = 1, ..., N \). This can be rewritten as

\[
\mu_j - \mu_j + \bar{p}_0 \cdot (E[T_j - \tau] - E[T_j - \tau]) \geq 0.
\]

For any \( \bar{p} \geq \bar{p}_0 \),

\[
\mu_j - \mu_j + \bar{p} \cdot (E[T_j - \tau] - E[T_j - \tau]) \geq 0.
\]

The above equation follows because \( \bar{p} \) minimizes \( E[T_j - \tau] \) (i.e., \( E[T_j - \tau] - E[T_j - \tau] \geq 0 \)). Hence,

\[
\mu_j + \bar{p} \cdot E[T_j - \tau] \geq \mu_j + \bar{p} \cdot E[T_j - \tau].
\]

Therefore, \( \bar{p} \) minimizes the expected disutility if \( \bar{p} \geq \bar{p}_0 \).

Proof of Proposition 4.1. Let \( p_0 \) be a path minimizing the NED (i.e., \( k = \arg\min_{k \in NED} \text{NED}_k \)). If \( k \in \Gamma_{SDD} \), we have the result. So, we assume that \( k \notin \Gamma_{SDD} \) minimizes the NED. Then there is a path of \( \pi_k \) such that \( T_k \approx T_i \). Note that

\[
\text{NED}_i - \text{NED}_k = \int_{T_i} F_k(t) dt - \int_{T_i} F_i(t) dt + \bar{p} \cdot \left( \int_{T_i} F_i(t) dt - \int_{T_i} F_k(t) dt \right) \leq 0.
\]

Since \( \text{NED}_k \) is the minimum value over all the paths, \( \text{NED}_i < \text{NED}_k \) is not possible. Therefore, \( \text{NED}_i = \text{NED}_k \). Now we set \( k = i \) and repeat the whole process until \( k \in \Gamma_{SDD} \). This process ends in a finite number of steps because the number of paths are finite (i.e., \( N \)).
Proof of Lemma 6.1. The expected delay of path \( \pi_j \) is

\[
E[T_j] = \int_{-\infty}^{\infty} (x-\gamma_j) \cdot \frac{1}{(x-\gamma_j) \cdot \nu_j \sqrt{2\pi}} \cdot \exp \left(-\frac{(\ln(x-\gamma_j) - \gamma_j)^2}{2\nu_j^2}\right) dx
\]

We let \( x' = x - \gamma_j \) in the second equality. In the last equality, we use \(-1 - \Phi(x) = \Phi(-x)\). Let \( y = -\frac{x' + \gamma_j}{\nu_j} \). Then the first element in the above equation is

\[
\int_{-\infty}^{\infty} \frac{1}{\nu_j \sqrt{2\pi}} \cdot \exp \left(-\frac{(\ln x' - \gamma_j)^2}{2\nu_j^2}\right) dx'
\]

We use \(-1 - \Phi(x) = \Phi(-x)\) again in the last equality. Then we have

\[
E[T_j] = \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right) - (\tau - \gamma_j) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j)}{\nu_j} \right).
\]

Proof of Lemma 6.2. Note that

\[
\exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln (\tau - \gamma_j) + \nu_j}{\nu_j} \right)
\]

\[
= \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp \left(-\frac{(\eta_j - \ln(\tau - \gamma_j) + \nu_j)^2}{2}\right)
\]

\[
= (\tau - \gamma_j) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j)}{\nu_j} \right).
\]

Proof of Proposition 6.1. The first-order equation of \( \bar{u}_j \) in Equation (2) with respect to \( \eta_j \) is

\[
\frac{\partial \bar{u}_j}{\partial \eta_j} = \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) + \bar{p} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right)
\]

\[
+ \bar{p} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right) \cdot \frac{1}{\nu_j}
\]

\[
- \bar{p} \cdot (\tau - \gamma_j) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j)}{\nu_j} \right) \cdot \frac{1}{\nu_j}
\]

\[
= \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) + \bar{p} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right) \geq 0.
\]

Therefore, \( \bar{u}_j \) is increasing in \( \eta_j \). The second equality follows Lemma 6.2. The first-order equation of \( \bar{u}_j \) with respect to \( \nu_j \) is

\[
\frac{\partial \bar{u}_j}{\partial \nu_j} = \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \nu_j \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right)
\]

\[
+ \bar{p} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right) \cdot \frac{1}{\nu_j}
\]

\[
- \bar{p} \cdot (\tau - \gamma_j) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j)}{\nu_j} \right) \cdot \frac{1}{\nu_j}
\]

\[
= \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \nu_j + \bar{p} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \nu_j \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right)
\]

\[
+ \bar{p} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \nu_j \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right) \geq 0.
\]

We also use Lemma 6.2 to have the second equality. Thus, \( \bar{u}_j \) is increasing in \( \nu_j \). Finally, the first-order equation of \( \bar{u}_j \) with respect to \( \gamma_j \) is

\[
\frac{\partial \bar{u}_j}{\partial \gamma_j} = 1 + \bar{p} \cdot \exp \left( \eta_j + \frac{\nu_j^2}{2} \right) \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j) + \nu_j}{\nu_j} \right)
\]

\[
\cdot \frac{1}{\nu_j} \cdot (\tau - \gamma_j)
\]

\[
+ \bar{p} \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j)}{\nu_j} \right) \cdot \frac{1}{\nu_j}
\]

\[
= 1 + \bar{p} \cdot \Phi \left( \frac{\eta_j - \ln(\tau - \gamma_j)}{\nu_j} \right) \geq 0.
\]