A Logistic Regression Model with a Hierarchical Random Error Term for Analyzing the Utilization of Public Transport

Chong Wei, Tingting Lu, and Xuedong Yan

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Chong Wei; chwei@bjtu.edu.cn

Received 10 March 2015; Revised 22 June 2015; Accepted 2 July 2015

Academic Editor: Dong Ngoduy

Copyright © 2015 Chong Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Logistic regression models have been widely used in previous studies to analyze public transport utilization. These studies have shown travel time to be an indispensable variable for such analysis and usually consider it to be a deterministic variable. This formulation does not allow us to capture travelers’ perception error regarding travel time, and recent studies have indicated that this error can have a significant effect on modal choice behavior. In this study, we propose a logistic regression model with a hierarchical random error term. The proposed model adds a new random error term for the travel time variable. This term structure enables us to investigate travelers’ perception error regarding travel time from a given choice behavior dataset. We also propose an extended model that allows constraining the sign of this error in the model. We develop two Gibbs samplers to estimate the basic hierarchical model and the extended model. The performance of the proposed models is examined using a well-known dataset.

1. Introduction

Understanding the utilization of public transport is important for policy design and urban traffic planning. From a behavior analysis perspective, such utilization can be analyzed in terms of a binary choice problem in which the traveler must choose between public transit and a private mode of transport. Previous studies usually employ logistic regression models to discuss this binary choice problem. These models can be used to predict choice probability and to evaluate the effect of various attitudes on the utilization of public transport. The effects of demography, travel cost, travel time, and accessibility on such utilization can be analyzed through estimating the parameters of these models.

McGillivray [1] proposed the binary choice problem in the two-modal (public and private) case and applied a logistic model to investigate the dependence of modal choice on the individual values of time and cost by modal. Gebeylehu and Takano [2] showed that the logistic model is a useful tool for evaluating transit systems; their study developed a logistic model to investigate citizens’ perceptions of bus conditions in Addis Ababa. Hess [3] attempted to apply a logistic regression model to assist planners and policymakers in expanding the mobility and accessibility of public transport for older adults by analyzing the influence of the accessibility of public transport on ridership for older people in California and New York. Johansson et al. [4] further formulated the binary logit model with latent variables; this structure can examine the effects of attitudes and personality traits on mode choice. Muley et al. [5] and Badland et al. [6] focused on microscopic behavior in the binary choice problem. Muley et al. [5] employed a binary logistic regression model to explore the impact of personal and transit characteristics on the utilization of public transport. Badland et al. [6] adopted a logistic regression model to discuss how parking availability and public transport accessibility influenced the split between uptakes of the two modes. However, Buehler and Pucher [7] studied the binary choice problem from a macroscopic perspective and employed a logistic model to compare utilization of public transport in the US and Germany.

One benefit of using a logistic regression model to analyze the utilization of public transport is that the model can consider the combined effects of attributes through a linear combination and can easily evaluate the contribution of various attributes to such utilization. Previous studies agree that travel time is an important variable for formulating a logistic
regression model to analyze this binary choice problem. Although previous studies usually prefer to treat travel time as a deterministic variable in the model, one cannot neglect that travelers’ perception error can prevent travelers from accurately evaluating their actual travel time. On the other hand, travelers also cannot say exactly how long a given travel time, for instance, 10 minutes, actually is. Carrion [8] provided a case analysis for this error using self-reported as well as GPS measured travel times and suggested that perception error regarding travel time can influence travel behavior. Cheng and Tsai [9] studied travelers’ perceptions of travel time through a questionnaire survey. They found that travel time perceptions can be influenced by personal characteristics such as gender and age.

As aforementioned, researchers have recognized the effect of perception errors regarding travel time on travelers’ choice behavior. However, the classic formulation of a random error term in logistic regression models cannot reflect this perception error appropriately (Chen et al. [10]). Analyzing travel behavior using the classic logistic regression model might therefore prevent us from looking for further insights into the utilization of public transport. Hence, this study proposes a hierarchical logistic regression model to fill the gap. Previous studies have successfully applied hierarchical error terms to model heterogeneous unobservable utility in logistic regression models (e.g., Tilahun et al. [11] and Czado and Prokopenko [12]). Inspired by these studies, we adopt hierarchical error terms to solve the problem. However, we add these terms after important attributes such as travel time rather than after the entire utility function in order to capture travelers’ perception error regarding the attributes. This study is not merely an effort to improve the performance of the regression model; more importantly, the proposed model offers an alternative way to estimate the statistical characteristics of perception error regarding travel time and allows us to explore the property of this error. Although this study prefers to focus on modeling perception error regarding travel time, we believe that individuals might have perception errors on other attributes such as travel cost. The proposed model can be easily used to estimate other such errors as well.

We first propose a basic hierarchical model and then develop an extended model in this study. The extended model allows us to constrain the sign of perception error regarding travel time. Correspondingly, this study also develops two Gibbs samplers to estimate the parameters of the proposed models. We evaluate the performance of the proposed models using a well-known dataset provided by Horowitz [13, 14].

2. The Hierarchical Logistic Regression Model

For simplicity, we describe the proposed model based on a binary choice problem that was provided by Horowitz [13]. In this choice problem, an agent faces making a choice decision between using a private car or public transport. The attributes considered in this choice problem are CARS, DCOST, DOVTT, and DIVTT where

\[ \text{CARS} \] is “private cars owned by the traveler’s household,”
\[ \text{DOVTT} \] is “public transport out-of-vehicle time minus private car out-of-vehicle time,”
\[ \text{DIVTT} \] is “public transport in-vehicle time minus private car in-vehicle time.”

We formulate the choice problem through a logistic regression model. If we present this logistic regression model as a latent-variable model, then the model can be obtained as follows:

\[
Y_{ik} = \begin{cases} 
1 & \text{if } U_{ik} > 0 \\
0 & \text{otherwise}
\end{cases}
\]

\[
U_{ik} = \beta_0 + \beta_1 \cdot \text{CARS}_i + \beta_2 \cdot \text{DCOST}_{ik} + \beta_3 \cdot \text{DOVTT}_{ik} + \beta_4 \cdot \text{DIVTT}_{ik} + \tau_{ik} 
\]

\[
\tau_{ik} \sim \text{Logistic}(0, 1),
\]

where \( i \) and \( k \) are the indexes of individual and survey questions, respectively. \( Y_{ik} = 1 \) denotes that the choice result is private car and is otherwise public transport. The logistic regression model uses a random error term \( \tau_{ik} \) to present stochastic behavior; however, this formulation does not allow us to investigate the perception error of attributes. In this study, we introduce a random variable \( \Delta T_i \) into the logistic regression model. Here \( \Delta T_i \) denotes the random perception error on DIVTT for individual \( i \). We consider that the value of \( \Delta T_i \) can be different among individuals and that \( \Delta T_i \) follows a probability distribution. The logistic regression model with a hierarchical random error term can be obtained as follows:

\[
Y_{ik} = \begin{cases} 
1 & \text{if } U_{ik} > 0 \\
0 & \text{otherwise}
\end{cases}
\]

\[
U_{ik} = \beta_0 + \beta_1 \cdot \text{CARS}_i + \beta_2 \cdot \text{DCOST}_{ik} + \beta_3 \cdot \text{DOVTT}_{ik} + \beta_4 \cdot \text{DIVTT}_{ik} + \tau_{ik} + \Delta T_i 
\]

\[
\tau_{ik} \sim \text{Logistic}(0, 1), \quad \Delta T_i \sim N(\mu, 1), \quad \mu \sim \pi(\mu), \quad \beta \sim \zeta(\beta).
\]

In the above equation, we use \( \beta \) to denote the parameter vector \( [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4] \). In the proposed model, \( \Delta T_i \) is a latent random variable, which serves as the random error term for DIVTT rather than the random term of the utility function. This structure enables us to investigate the perception error of DIVTT. We assume that \( \Delta T_i \) follows the normal distribution with unknown mean \( \mu \) and specify the variance of the normal distribution as 1 for identification purposes. \( \pi(\mu) \) and \( \zeta(\beta) \) denote the prior distributions of \( \mu \) and \( \beta \), respectively. We wish to estimate the value of \( \mu \) along with \( \beta \). In this hierarchical model, DIVTT is divided between
DIVTT, and ΔT; this leads to DIVTT no longer being a deterministic variable in the model.

The logistic regression model shown by (2) can be further extended. Here, we consider that ΔT < 0 indicates that individual i subjectively weakens the difference of in-vehicle travel time when the agent makes the choice decision; on the other hand, ΔT > 0 suggests that the individual subjectively enlarges the difference of in-vehicle travel time. In addition, we impose the value of DIVTT, ΔT, to be a nonnegative (nonpositive) value if DIVTT > 0 (≤0). Following these considerations, we rewrite U, of (2) as

\[ U_{ik} = V_{ik} + \tau_{ik} \]

\[ = \beta_0 + \beta_1 \cdot \text{CARS}_i + \beta_2 \cdot \text{DCOST}_{ik} + \beta_3 \cdot \text{DOVTT}_{ik} + \tau_{ik} \]  

(3)

where

DIVTT is given as follows:

\[ \text{DIVTT} = \phi(\text{DIVTT}) \min(0, \text{DIVTT} - \Delta T_i) \]

\[ + [1 - \phi(\text{DIVTT})] \max(0, \text{DIVTT} + \Delta T_i). \]  

(4)

In (4), if DIVTT < 0 then the value of φ(DIVTT) is 1 or else 0.

As shown by (2) and (3), the combination of ΔT and τ leads to a hierarchical random error term structure; this structure can be presented as a Directed Arc Graph (see Figure 1).

### 3. Estimation Methods

First, we discuss how to estimate parameters of the model shown by (2). Stefanski [15] indicated that the logistic distribution can be represented as a normal scale mixture. Accordingly, Holmes and Held [16] suggested an auxiliary variable method to present the logistic regression model.

Along the same lines, the regression model presented by (2) can also be derived in the following form:

\[ Y_{ik} = \begin{cases} 1 & \text{if } U_{ik} > 0 \\ 0 & \text{otherwise} \end{cases} \]

\[ U_{ik} = \beta_0 + \beta_1 \cdot \text{CARS}_i + \beta_2 \cdot \text{DCOST}_{ik} + \beta_3 \cdot \text{DOVTT}_{ik} + \epsilon_{ik} \]

\[ \epsilon_{ik} \sim N(0, \lambda) \]

\[ \lambda_{ik} = 4\psi_{ik}^2 \]

\[ \psi \sim \text{KS} \]

\[ \Delta T_i \sim N(\mu, 1) \]

\[ \mu \sim \pi(\mu) \]

\[ \beta \sim \zeta(\beta). \]

μ and β are the parameters of the proposed model. U = [U_i ∀i, k], ΔT = [ΔT_i ∀i], and λ = [λ_i ∀i, k] are the latent parameters of the proposed model. P(β, U, ΔT, μ, λ | X, Y) is the joint distribution of these parameters, which can be uniquely identified by (2). μ(μ) and ζ(β) are the prior distributions for μ and β, respectively. KS denotes the Kolmogorov-Smirnov distribution.

We develop a Gibbs sampler to draw the samples of β, U, λ, ΔT, and μ from the joint distribution and calculate the estimates of the parameters through aggregating the samples. Let X, X, and Y denote the vector [CARS, DCOST, DOVTT, DIVTT], X = [X_i ∀i, k], and Y = [Y_i ∀i, k], respectively. We use t to denote the index of the sample. The outline of the Gibbs sampler can then be given as follows:

**Algorithm 1.**

*Step 0 (initialization).* Set initial values: β^0 = 0, U^0 = 0, λ^0 = 1, ΔT^0 = 0, and μ^0 = 0.

Set t ← 1.

*Step 1.* Draw tth sample of β

\[ \beta^t \sim P(\beta | U^{t-1}, \lambda^{t-1}, \Delta T^{t-1}, \mu^{t-1}, X, Y). \]  

(6)

*Step 2.* Draw tth sampler of U

\[ U^t \sim P(U | \beta^t, \lambda^{t-1}, \Delta T^{t-1}, \mu^{t-1}, X, Y). \]  

(7)

*Step 3.* Draw tth sampler of λ

\[ \lambda^t \sim P(\lambda | \beta^t, U^t, \Delta T^{t-1}, \mu^{t-1}, X, Y). \]  

(8)

*Step 4.* Draw tth sample of ΔT

\[ \Delta T^t \sim P(\Delta T | \beta^t, U^t, \lambda^t, \mu^{t-1}, X, Y). \]  

(9)
Step 5. Draw \( t \)th sample of \( \mu \)

\[
\mu^t \sim P \left( \mu \mid \beta^t, U^t, \lambda^t, \Delta T^t, X, Y \right). \tag{10}
\]

If \( t < T \) then go to Step 1; otherwise, stop the sampling procedure.

The set of conditional distributions \( \beta, U, \lambda, \Delta T, \) and \( \mu \) is referred to as the full set of conditional distributions of \( P(\beta, U, \Delta T, \mu, X, Y) \). Sampling \( \beta, U, \lambda, \Delta T, \) and \( \mu \) in turn from the full set of conditional distributions is equivalent to sampling these parameters from \( P(\beta, U, \Delta T, \mu, X, Y) \) simultaneously. We then illustrate how the Gibbs sampler draws \( t \)th samplers for \( \beta, U, \lambda, \Delta T, \) and \( \mu \) in detail.

Algorithm 2.

Step 1. Update \( X_{ik}^t \) for drawing \( t \)th samplers

\[
\text{for } i = 1 \text{ to } I \text{ do}
\]

\[
\text{for } k = 1 \text{ to } K_i \text{ do}
\]

\[
\text{DIVTT}_{ik,t-1} \leftarrow \text{DIVTT}_{ik} + \Delta T_{ik}^{t-1}
\]

\[
X_{ik,t-1}^t \leftarrow \left[ \text{CARST}_i, \text{DCOST}_{ik}, \text{DOVT}_{ik}, \text{DIVTT}_{ik} \right]
\]

end

end

Step 2. Draw \( t \)th sample for \( \beta \).

The conditional distribution \( P(\beta \mid U^{t-1}, \lambda^{t-1}, \Delta T^{t-1}, \mu^{t-1}, X, Y) \) is equivalent to \( P(\beta \mid U^{t-1}, \lambda^{t-1}, X^t, Y) \). On replacing \( X \) in Albert and Chib [17] by \( X^t \), the sample of \( \beta \) can be generated through the sampling scheme proposed by Albert and Chib [17]. As shown by them, \( \zeta(\beta) \) can be specified as a diffuse distribution, and \( P(\beta \mid U^{t-1}, \lambda^{t-1}, X^t, Y) \) can be derived as a normal distribution.

Step 3. Draw \( t \)th sample for \( U \).

The conditional distribution \( P(U \mid \beta^t, \lambda^{t-1}, \Delta T^{t-1}, \mu^t, X, Y) \) is equivalent to \( P(U \mid \beta^t, \lambda^{t-1}, X^t, Y) \). Albert and Chib [17] indicated that \( P(U \mid \beta^t, \lambda^t, X^t, Y) \) is a truncated normal distribution.

Step 4. Draw \( t \)th sample for \( \lambda \).

The conditional distribution \( \lambda^t \sim P(\lambda \mid \beta^t, U^t, \Delta T^{t-1}, \mu^{t-1}, X, Y) \) is equivalent to \( \lambda^t \sim P(\lambda \mid \beta^t, U^t, X^t) \). Holmes and Held [16] proposed a rejection sampling scheme that can be used to draw samples from \( P(\lambda \mid \beta^t, U^t, X^t) \).

Step 5. Draw \( t \)th sample for \( \Delta T \).

\( \Delta T^t \) can be generated through sampling \( \Delta T_{ij}^t \) \( i = 1, \ldots, I \) in turn. The conditional distribution of \( \Delta T_{ij}^t \) is equivalent to \( P(\Delta T_{ij} \mid \beta^t, X, Y, U^t, \lambda^t) \). The probability density function of \( \Delta T_{ij} \) is a normal distribution and is defined by (11)–(13).

Step 6. Draw \( t \)th sample for \( \mu \).

The conditional distribution \( P(\mu \mid \beta^t, U^t, \lambda^t, \Delta T^t, X, Y) \) is equivalent to \( P(\mu \mid \Delta T^t) \). The probability density function of \( P(\mu \mid \Delta T^t) \) is a normal distribution and is given by (16).

Applying the Bayesian theorem, we obtain \( P(\Delta T_{ij}^t \mid \beta^t, X, Y, \lambda^t) \) as a normal distribution. Without loss of generality, we only present how to obtain \( P(\Delta T_{ij} \mid \beta^t, X, Y, U^t, \lambda^t) \) for the case \( K_j = 1 \) for \( i = 1, \ldots, I \):

\[
P \left( \Delta T_{ij} \mid \beta^t, X, Y, U^t, \lambda^t \right) = \frac{1}{a^{\frac{\Delta T_{ij}}{2}}} e^{-\frac{\Delta T_{ij}}{a^2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu - \mu^t)^2}{2a^2}} \tag{11}
\]

The above equation shows that the conditional probability of \( \Delta T_{ij} \) is a normal distribution. The mean of the distribution is

\[
A = \left( \frac{U_{ik}^t / \beta_{ik}^t}{1 + (\lambda_{ik}^t / \beta_{ik}^t)^2} \right)^2
\]

and the variance of the distribution is

\[
B = \left( \frac{(\lambda_{ik}^t / \beta_{ik}^t)^2}{1 + (\lambda_{ik}^t / \beta_{ik}^t)^2} \right)^2.
\]

In (12), \( U_{ik}^t \) represents

\[
U_{ik}^t = U_{ik}^t - \beta_{ik}^t \cdot \text{CARST}_i - \beta_{ik}^2 \cdot \text{DCOST}_{ik} - \beta_{ik}^3 \cdot \text{DOVT}_{ik} - \beta_{ik}^3 \cdot \text{DIVTT}_{ik}.
\]

We consider the prior distribution of \( \mu \) as a normal distribution with mean 0 and variance \( \sigma^2 \). If we further consider the prior distribution as a diffuse prior then this probability function of the prior can be given as

\[
P(\mu) = \lim_{\sigma \to \infty} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\mu - 0)^2}{2\sigma^2}}.
\]

Accordingly, the conditional distribution \( P(\mu \mid \Delta T_{ij}^t) \) \( i = 1, \ldots, I \) can be obtained as

\[
P \left( \mu \mid \Delta T_{ij}^t \right) = \frac{1}{\left( 1/\sqrt{I} \right) \sqrt{2\pi}} e^{-\frac{(\mu - \Sigma_{i=1}^I \Delta T_{ij}^t / I)^2}{(2/I)}}.
\]

This conditional distribution is a normal distribution too; the mean of the distribution is \( \Sigma_{i=1}^I \Delta T_{ij}^t / I \) and the variance of the distribution is \( 1/I \).

Now, we discuss how to estimate the parameters of the extended model described by (3). In this case, we can also employ Steps 1 to 3 and 5 of Algorithm 2 to draw the samples of \( \beta, U, \lambda, \) and \( \mu \). We need only to use \( \text{DIVTT}_{ik,t-1} \) and \( X_{ik,t-1}^t \) to replace \( \text{DIVTT}_{ik,t-1}^* \) and \( X_{ik,t-1}^* \) in Step 1, respectively, where \( \text{DIVTT}_{ik,t-1}^* \) is defined by (4) and \( X_{ik,t-1}^* \) is defined as \( [\text{CARST}_i, \text{DCOST}_{ik}, \text{DOVT}_{ik}, \text{DIVTT}_{ik}^*] \). On the other hand, instead of Step 4, we develop an M-H sampling scheme to draw \( \Delta T_{ij} \) \( i = 1, \ldots, I \).
We derive the formulation of $P(\Delta T_i | \beta, X_i, U_i, \lambda_i, \mu^{-1})$ as

$$P(\Delta T_i | \beta, X_i, U_i, \lambda_i, \mu^{-1}) = \frac{\exp (V_{ik})}{1 + \exp (V_{ik})},$$

where $V_{ik}$ is defined by (3). The H-M sampling scheme is as follows.

Algorithm 3.

Step 0 (initialization). Set $i \leftarrow 1$.

Step 1. Generate a candidate value for $\Delta T^*_i$. Draw $\Delta T^*_i$ from a normal distribution: $\Delta T^*_i \sim N(\Delta T_i^{-1}, 1)$, where $\Delta T^*_i$ denotes the candidate value for $\Delta T_i$.

Step 2. Calculate the value of $\eta$

$$\eta = \min \left[ \frac{P(\Delta T_i^* | \mu^{-1}) \prod_{k=1}^{K} P(Y_{ik} | X_{ik}, \beta, \Delta T^*_i)}{P(\Delta T_i^{-1} | \mu^{-1}) \prod_{k=1}^{K} P(Y_{ik} | X_{ik}, \beta, \Delta T_i^{-1})}, 1 \right].$$

Step 3. Get $\Delta T_i^*$

$\eta \sim U(0, 1)$, if $\eta > \eta$ then $\Delta T^*_i = \Delta T_i$ else $\Delta T^*_i = \Delta T_i^{-1}$.

If $i < I$ then $i \leftarrow i + 1$ and go to Step 1.

The proposed model structure also allows us to further investigate traveler’s perception error regarding both DIVTT and DOVTT. To do this, we just need to modify (3) as follows:

$$U_{ik} = \beta_0 + \beta_1 \cdot \text{CARS} + \beta_2 \cdot \text{DCOST}_{ik} + \beta_3 \cdot \text{DOVTT}_{ik} + \beta_4 \cdot \text{DIVTT}_{ik} + \tau_{ik},$$

where

$$\text{DOVTT}_{ik} = \phi (\text{DOVTT}_{ik}) \min (0, \text{DOVTT}_{ik} - \Delta S_i)$$

$$+ \max (0, \text{DOVTT}_{ik} + \Delta S_i) \cdot \phi (\text{DOVTT}_{ik}).$$

We use $\Delta S_i$ to denote the traveler’s perception error on DOVTT and use $\nu$ to denote the mean of $\Delta S_i$. Let $\Delta S$ be $[\Delta S, \forall i]$; the outline of the sampling scheme for $r$th sample of $\beta, U, \lambda, \Delta T, \Delta S, \mu$, and $\nu$ can be described as follows.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARS</td>
<td>0–7</td>
</tr>
<tr>
<td>DCOST</td>
<td>−111–89</td>
</tr>
<tr>
<td>DOVTT</td>
<td>−6–48</td>
</tr>
<tr>
<td>DIVTT</td>
<td>−59–102</td>
</tr>
<tr>
<td>CHOICE</td>
<td>CAR = 1, TRANSIT = 0</td>
</tr>
<tr>
<td>Number of samples</td>
<td>842</td>
</tr>
</tbody>
</table>

Algorithm 4.

Step 1. Consider $\beta^* \sim P(\beta | U^{t-1}, \lambda^{t-1}, \Delta T^{t-1}, \mu^{t-1}, \Delta S^{-1}, \nu^{-1}, X, Y)$.

Step 2. Consider $U^t \sim P(U | \beta^*, \lambda^{t-1}, \Delta T^{t-1}, \mu^{t-1}, \Delta S^{-1}, \nu^{-1}, X, Y)$.

Step 3. Consider $\lambda^t \sim P(\lambda | \beta^*, U^t, \Delta T^{t-1}, \mu^{t-1}, \Delta S^{-1}, \nu^{-1}, X, Y)$.

Step 4. Consider $\Delta T^t \sim P(\Delta T | \beta^*, U^t, \lambda^t, \mu^{t-1}, \Delta S^{-1}, \nu^{-1}, X, Y)$.

Step 5. Consider $\Delta S^t \sim P(\Delta S | \beta^*, U^t, \lambda^t, \Delta T^t, \mu^{t-1}, \nu^{-1}, X, Y)$.

Step 6. Consider $\mu^t \sim P(\mu | \beta^*, U^t, \lambda^t, \Delta T^t, \Delta S^t, \nu^{-1}, X, Y)$.

Step 7. Consider $\nu^t \sim P(\nu | \beta^*, U^t, \lambda^t, \Delta T^t, \mu^{t-1}, \Delta S^t, X, Y)$.

$\beta$, $U$, and $\lambda$ can be generated along the same line of Algorithm 2. Since DOVTT and DIVTT hold the same formulation, therefore, $\Delta S$ and $\nu$ can be sampled the same as $\Delta T$ and $\mu$ (see Algorithm 2 and (16)).

To estimate the parameters of the proposed model, we can apply the sampling algorithm to draw random samples for $\beta$, $U$, $\lambda$, $\Delta T$, $\Delta S$, and $\mu$. These samples can be denoted as $\beta^*, U^t, \lambda^t, \Delta T^t, \Delta S^t, \nu^t$, and $\mu^t$ for $t = 1$ to $M$, where $M$ is the number of drawings. The estimates of the parameters can be obtained through averaging the samples. For example, the estimates of $\nu$ and $\mu$ can be obtained as follows:

$$\bar{\nu} = \frac{\sum_{t=1}^{M} \nu^t}{M},$$

$$\bar{\mu} = \frac{\sum_{t=1}^{M} \mu^t}{M}.$$
Table 2: Estimates of the parameters of the basic model defined by (2).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>SD</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>$-1.1731$</td>
<td>$-1.179$</td>
<td>$-1.1671$</td>
<td>$0.3023$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_1$ (CARS)</td>
<td>$2.3149$</td>
<td>$2.305$</td>
<td>$2.3193$</td>
<td>$0.2233$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_2$ (DCOST)</td>
<td>$0.0172$</td>
<td>$0.0171$</td>
<td>$0.0173$</td>
<td>$0.0038$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_3$ (DOVTT)</td>
<td>$0.0620$</td>
<td>$0.0617$</td>
<td>$0.0624$</td>
<td>$0.0182$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_4$ (DIVTT)</td>
<td>$0.0096$</td>
<td>$0.0095$</td>
<td>$0.0098$</td>
<td>$0.0096$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$-3.5245$</td>
<td>$-3.5521$</td>
<td>$-3.4970$</td>
<td>$1.9869$</td>
<td>$&lt;0.0001$</td>
</tr>
</tbody>
</table>

DIC | 465.764 |
Number of observations | 894 |

We first estimate the hierarchical model defined by (2) using the dataset. We conduct the estimation using the Gibbs samplers. The Gibbs samplers generate 20,000 samples for each parameter. The first 5,000 samples are dropped as the burn-in procedure. The remaining samples are used to aggregate the estimates of the parameters. We use Table 2 to show the estimation results. The sign of the estimates of $\beta_2$, $\beta_3$, and $\beta_4$ suggests that DCOST, DOVTT, and DIVTT have a positive effect on the choice probability of car (and have a negative effect on the utilization of public transport). The value of DIC of the model is 465.764.

We estimate the parameter of the extended model through the sampling scheme with H-M step. We draw 20,000 samples of the parameter and also treat the first 5,000 samples as the burn-in procedure. The DIC of the extended model is 461.687. This result indicates that the performance of the extended model is better than that of the basic model defined by (2). We also estimate the parameters of the logistic regression model defined by (1). The value of DIC of this model is 465.848. This value is slightly higher than the DIC of the basic model defined by (2) and clearly higher than the DIC of the extended model defined by (3).

The samples of the parameters are used to investigate the shapes of the distributions of the parameters. Table 3 reports the estimates of the parameters of the extended model. Figures 2 and 3 provide the histogram for $\beta_4$ and $\mu$ of the extended model. The distributions of the parameters have symmetrical sides and one single peak.

In addition to the improvement of the performance, the main contribution of the proposed model is that it can allow us to capture the perception error on travel time and analyze the characteristics of the perception error. Let us look at the extended model. As shown by Table 3, the estimate of $\mu$ is $-1.022$; this result implies that people tend to weaken the difference of in-vehicle travel time when making the choice decision. The sampling schemes also draw samples of $\Delta T_i$ for $i = 1 \text{ to } I$. The latent parameter $\Delta T_i$ directly reflects the perception error on DIVTT of individual $i$.

To further investigate the property of $\Delta T_i$ of the extended model, we group these samples according to the value of DCARS, which is related to the demography of an individual. We calculate the mean and variance of the samples of $\Delta T_i$ for each group. Figure 4 plots the mean of the samples, and Figure 5 shows the variances of the samples versus the value of DCARS. Since the values of DCARS in the dataset do not contain 6 and only one-record DCARS is equal to 7, we study the case where the values of DCARS range from 1 to 5 to avoid bias.

One can find that the mean of the samples of $\Delta T_i$ tends to be close to 0 as DCARS increases. On the other hand, the variance decreases with the value of DCARS. This result implies that travelers who hold more cars will be more sensitive to the difference of travel time between public transport and private car. For example, the mean of $\Delta T_i$ for travelers who own 5 cars is $-0.357$; this value can change to $-1.5623$ for travelers who do not own any cars. As mentioned
Table 3: Estimates of the parameters of the extended hierarchical model defined by (3).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>SD</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>$-1.2490$</td>
<td>$-1.2531$</td>
<td>$-1.2448$</td>
<td>$0.3012$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_1$ (CARS)</td>
<td>$2.3322$</td>
<td>$2.3292$</td>
<td>$2.3352$</td>
<td>$0.2161$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_2$ (DCOST)</td>
<td>$0.0174$</td>
<td>$0.0174$</td>
<td>$0.0175$</td>
<td>$0.0038$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_3$ (DOVTT)</td>
<td>$0.0590$</td>
<td>$0.0587$</td>
<td>$0.0592$</td>
<td>$0.0190$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_4$ (DIVTT)</td>
<td>$0.0156$</td>
<td>$0.0155$</td>
<td>$0.0158$</td>
<td>$0.0101$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$-1.022$</td>
<td>$-1.0234$</td>
<td>$-1.0206$</td>
<td>$0.1026$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>DIC</td>
<td>$461.687$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>$894$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Estimates of the parameters of the extended hierarchical model defined by (20).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>SD</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>$-1.3607$</td>
<td>$-1.8495$</td>
<td>$-0.8590$</td>
<td>$0.3005$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_1$ (CARS)</td>
<td>$2.2569$</td>
<td>$1.9034$</td>
<td>$2.6305$</td>
<td>$0.2131$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_2$ (DCOST)</td>
<td>$0.0149$</td>
<td>$0.0086$</td>
<td>$0.0213$</td>
<td>$0.0039$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_3$ (DOVTT)</td>
<td>$0.0907$</td>
<td>$0.0569$</td>
<td>$0.1250$</td>
<td>$0.0209$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\beta_4$ (DIVTT)</td>
<td>$0.0082$</td>
<td>$-0.0072$</td>
<td>$0.0247$</td>
<td>$0.0097$</td>
<td>$0.1973$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$-0.4310$</td>
<td>$-0.5229$</td>
<td>$-0.3363$</td>
<td>$0.0559$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\nu$</td>
<td>$-0.5801$</td>
<td>$-0.6804$</td>
<td>$-0.4785$</td>
<td>$0.0558$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>DIC</td>
<td>$436.240$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>$894$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Mean of the samples of $\Delta T_i$ grouped by the value of CARS.

Figure 5: Variance of the samples of $\Delta T_i$ grouped by the value of CARS.

above, $\Delta T_i < 0$ indicates that individual $i$ subjectively weakens the difference of travel time when he/she makes the choice decision; therefore, if $\Delta T_i < 0$, a smaller $\Delta T_i$ can reduce the contribution of the difference of travel time on the travel mode choice decision even more.

To investigate traveler's perception error regarding both DIVTT and DOVTT, we also use Algorithm 4 to estimate the model defined by (20). Table 4 shows that the DIC of the model is significantly smaller than the standard logistic model defined by (1). In the model defined by (3), traveler's perception error regarding travel time can be characterized by the estimate of $\mu$. On the other hand, in the model defined by (20), traveler's perception error regarding travel time should be characterized by the summation of the estimates of $\mu$ and $\nu$. Looking at Table 4, one can find that the summation of the estimates of $\mu$ and $\nu$ is $-1.011$. This value is closed to the estimate of $\mu$ ($-1.022$) in the model defined by (3).

5. Conclusions

This study proposes a logistic regression model with a hierarchical random error term to analyze the binary choice problem. The proposed model can account for travelers' perception errors regarding attributes. Since a number of studies have shown perception error regarding travel time to have a significant impact on modal choice, this study focuses in particular on how to capture this error from behavior data.
We construct a hierarchical random error term structure in the logistic regression model though introducing a random error term for the travel time variable. In the proposed model, travel time is no longer a deterministic variable. To make the model more sensible, we also propose an extended model, where the sign of the DIVTT and DOVTT variables can be constrained. We develop a Gibbs sample to estimate the basic hierarchical model, while developing a Gibbs sampler with M-H step to estimate the parameters of the extended model. The binary choice dataset provided by Horowitz [13] is used to examine model performance and demonstrate the proposed model's benefit. The numerical example shows that the proposed model can estimate the perception error regarding travel time exactly and that the samples of $\mu$ and $\nu$ can be used to further analyze the characteristics of this error. Although this study prefers to look at the perception error regarding travel time, the proposed methodology can also be used to investigate travelers' perception error regarding travel cost, among others.

Conflict of Interests
The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments
This work is jointly supported by the National Basic Research Program of China (no. 2012CB725403), the Fundamental Research Funds for the Central Universities (no. 2014JBM056), and the National Natural Science Foundation of China (no. 51408035).

References
Submit your manuscripts at http://www.hindawi.com