# The Continuous Cross-Nested Logit Model: Formulation and Application for Departure Time Choice

Jason D. Lemp
Graduate Student Researcher
The University of Texas at Austin – 6.508, E. Cockrell Jr. Hall
Austin, TX 78712-1076 jdl880@mail.utexas.edu

Kara M. Kockelman
(Corresponding author)
Professor and William J. Murray Jr. Fellow
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin – 6.9 E. Cockrell Jr. Hall Austin,
TX 78712-1076
kkockelm@mail.utexas.edu
Phone: 512-471-0210 & FAX: 512-475-8744

Paul Damien
B.M. Rankin Jr. Professor of Business
McCombs School of Business The
University of Texas at Austin
paul.damien@mccombs.utexas.edu

The following paper is a pre-print and the final publication can be found in
*Transportation Research Part* B 44 (5): 646-661, 2010.
Presented at the 89th Annual Meeting of the Transportation Research Board,
January 2010

Key Words: Travel behavior modeling, departure time modeling, continuous logit, continuous cross-nested logit, Bayesian estimation

ABSTRACT

Discrete choice models, like the multinomial logit (MNL), have long been recognized for their ability to capture a wide array of transport-related choice phenomena. However, a number of choices are continuous response variables (e.g., location choice, departure time choice, activity duration, and vehicle usage). In this paper, the continuous cross-nested logit (CCNL) model is introduced. The CCNL model results from generalizing the discrete cross-nested logit (CNL) model for a continuous response variable, much like the continuous logit model emerges by generalizing the MNL. The model is formulated and shown to come from the generalized extreme value (GEV) class of models. In addition, the structure of utility correlations is presented. The model's parameters are estimated for a work-tour departure time context using Bayesian estimation techniques and San Francisco Bay Area data. Empirical results suggest model predictions that are very similar to the continuous logit, but it out-performs the continuous logit in terms of out-of-sample prediction with these data. Theoretically, however, the CCNL allows a more flexible choice behavior to emerge. Finally, a simple welfare example is illustrated and a number of model extensions are presented

## 1. INTRODUCTION

Supporting applications of random utility maximization theory, the GEV class of models (McFadden 1978) has become a mainstay in travel behavior analysis of discrete choice behaviors. McFadden's (1973) multinomial logit (MNL) model represents the most familiar and straightforward of these models. However, the MNL model suffers from the independence of irrelevant alternatives (IIA) property, which results in equivalent cross-elasticities across each pair of choice alternatives[1].

The nested logit model (Williams 1977, Daly and Zachary 1979, and McFadden 1978) relaxes this assumption, allowing correlations to emerge across similar alternatives. However, choice alternatives in common nests still retain the IIA property. Vovsha (1997) introduced the cross-nested logit (CNL) model (later generalized by Ben-Akiva and Bierlaire [1999], Wen and Koppelman [2001] and Papola [2004][2]), which allows choice alternatives to appear in multiple nests, thereby, offering more flexible correlation patterns than the nested logit. Small's (1987) ordered GEV (OGEV) model represents a special case of the CNL, where alternatives are ordered in nature (e.g., departure time choice). Thus, each nest in the OGEV contains consecutive alternatives in a sequence.

While the models described above all have important applications in travel behavior research, each is only applicable for discrete choice analysis. A number of travel-related decisions are inherently continuous in nature. For example, residential location choice and travel destination choice can be viewed as continuous variables in two dimensions. Other examples include annual vehicle-miles traveled in a particular vehicle and the timing of travel. Of course, one could discretize these choices, but doing so requires rather artificial setting of interval boundaries. If boundaries are changed, different results may emerge in model calibration and application[3]. Moreover, model application often requires a point prediction, whereas discrete models only provide interval predictions. Smaller intervals can alleviate this difficulty to some extent, but issues will remain unless continuous models are pursued. In addition, discretization always results in points that are very close lying on either side of an interval boundary. Such points may be viewed as very similar options from the decisionmaker's perspective, but cannot be treated adequately with discrete choice methods.

Although discretization of continuous response variables may be inappropriate, one of the main advantages of the GEV class of models is that they are based in random utility theory. This provides a solid foundation for estimation of economic welfare, which can be used for project evaluation and policy analysis. Essentially a generalization of the MNL, the continuous logit model (McFadden 1976, Ben-Akiva and Watatanada 1981, Ben-Akiva et al. 1985) features the advantages of a random utility model, but in a continuous choice setting. However, as it is a generalization of the MNL, it also suffers from the IIA property, meaning that random errors at every infinitesimal unit are uncorrelated. Of course, one generally will expect relatively high correlation between proximate alternatives.

[1] In other words, if one alternative's attributes improve (e.g., travel time under the alternative decreases), the probability of that alternative draws probability away from each other alternative equally (in a relative sense).
[2] These generalizations reformulated the model to allow each nest to have its own nesting (or inclusive value) parameter.
[3] In a departure time setting, for instance, one could build an AM peak interval (from 6am to 9am). If instead, this interval was split into two separate intervals (e.g., one from 4:30am to 7:30am and another from 7:30am to 10:30am), one could not presume application results (and certainly not calibration results) would be the same.

This paper introduces the continuous cross-nested logit (CCNL) model, which offers the advantages of a random utility model in a continuous choice setting, while allowing for correlations across similar alternatives, namely those close on the continuous spectrum. The model is estimated in a work-tour departure time context using Bayesian estimation techniques. The remainder of the paper is structured as follows. Section 2 discusses the importance of departure time modeling and related literature. Section 3 presents the model formulation and the relationships that exist between it and the continuous logit and the discrete CNL. Section 4 illustrates the behavior of the model in comparison to the continuous logit for specified utility functions. Section 5 presents an empirical analysis and welfare application of work-tour departure time choice, and Section 6 offers some conclusions and extensions to this work.

**2. A REVIEW OF DEPARTURE TIME MODELING**

Activity scheduling represents an important component of travel behavior, though it is often greatly simplified in urban travel models, as compared to other choice dimensions (such as destination and mode). Over the past several years, activity scheduling has received greater attention, since planning and policy questions have shifted toward congestion and demand management. In addition, our understanding of travel behavior and the tools available for computing have progressed significantly. Even with recent advances in travel demand theory and applications, however, time-of-day (TOD) modeling remains a major weakness of most model systems (Vovsha et al. 2005, TRB 2007).

Some of the earliest models of travel timing were developed in trip departure time contexts, using the familiar MNL model (see, e.g., Abkowitz 1981, Small 1982, McCafferty and Hall 1982). However, the MNL has well-known limitations, including the assumption of independence across irrelevant alternatives. Other researchers have relaxed this assumption using the nested logit (Chin 1990), the ordered-GEV (Small 1987), and the error-components logit (de Jong et al. 2003 and Hess et al. 2007). While these models allow more flexible substitution patterns to emerge, applications have generally been constrained timing choice over a limited temporal spectrum (e.g., the morning peak).

Some recent discrete choice applications have occurred in the context of tour timing, which possesses (at least) two timing dimensions: the departure time and the return time. Vovsha and Bradley (2004), Abou Zeid et al. (2006), and Popuri et al. (2008) all modeled the two-dimensional tour timing choice in a joint MNL context using 30-min or 1-hour intervals. While these models do not recognize possible correlations across alternatives, they do offer advantages of computational tractability and incorporation of multiple timing dimensions across an entire day. Moreover, Abou Zeid et al. (2006) and Popuri et al. (2008) introduced continuous utility functions in their models, via sinusoidal functions interacted with covariates (though the final models only recognize this continuity through evaluation of the utility function at specific points in time).

Of course, none of the models mentioned above treat departure time in a continuous setting. Most literature on continuous departure time models utilize a hazard function, which defines the probability limit that an individual will departure between times $t$ and $t + h$, as $h$ tends toward zero. Wang (1996) used a parametric proportional hazard specification to model activity start times, while Bhat and Steed (2002) and Komma and Srinivasan (2008) allowed for unobserved heterogeneity in departure time chocies using Gamma mixing distributions in non-parametric hazard settings. In contrast, Gadda et al. (2009) used Bayesian estimation techniques and an accelerated hazard form.

The main disadvantage of such past models is that they offer no connection to random utility theory, which provides a basis for traveler welfare estimation. Of course, the failure of discrete choice models to recognize time's continuity represents a drawback of such models. In addition, the time interval boundary specification required by discrete choice methods is usually ad hoc and results in points close on the temporal spectrum being placed on different sides of interval boundaries even though those points may be perceived very similarly (Bhat and Steed 2002).

In contrast to these existing methods for modeling travel timing, the continuous cross-nested logit (CCNL) offers a random utility framework in a continuous choice setting. In addition, it offers added behavioral flexibility over the continuous logit by allowing correlations across alternatives close on the continuous spectrum, as one would expect for a continuous response variable. The following section details the CCNL model formulation.

## 3. MODEL FORMULATION

McFadden (1978) developed the GEV class of models for discrete choice applications to make use of random utility theory, where each agent is assumed to choose an alternative that maximizes its utility (or profit, for example). Every GEV model is derived from a function, $G(y_1,\ldots,y_J)$, where $j = 1,\ldots,J$ indexes the set of alternatives (with $G$ satisfying some regularity conditions [see, e.g., McFadden 1978 and Small 1987]). If random utility for any alternative $j$ is defined as a systematic component plus a random error component (where the joint density of all error components is distributed according to the extreme value distribution), as shown in equation (1), then the probability that alternative $k$ is chosen (i.e., alternative $k$ offers the maximum utility) is given by equation (2).

$$U_j = V_j + \varepsilon_j \tag{1}$$

$$P_k = \frac{y_k G_k(y_1,\ldots,y_J)}{G(y_1,\ldots,y_J)} \tag{2}$$

Here, $G_k(y_1,\ldots,y_J)$ denotes the partial derivative of $G$ with respect to $y_k$, and $y_j = e^{V_j}$, for all $j$. For the MNL model, $G$ is given by equation (3) and the choice probability for alternative $k$ takes equation (4)'s familiar form.

$$G(y_1,\ldots,y_J) = \sum_{j=1}^{J} y_j \tag{3}$$

$$P_k = \frac{\exp(V_k)}{\sum_{j=1}^{J}\exp(V_j)} \tag{4}$$

### 3.1 Continuous Logit

The continuous logit model represents a generalization of the MNL for a continuous response variable (see, e.g., McFadden 1976, Ben-Akiva and Watatanada 1981, and Ben-Akiva et al. 1985). Suppose the continuous response variable of interest, $t$, is bounded by $b_1 a$ and $b_2 b$. Further, discretize $t$ such that $t_j$ denotes the $j^{th}$ discrete alternative (where $j = 1,\ldots,J$), and let $t_1 = b_1 a$ and $t_J = b_2 b$. Now, suppose $J$ (the number of discrete alternatives) is computed as $J = 1 + \frac{b_2 - b_1}{s}$, where $s$ denotes the distance between each discrete alternative. Furthermore, let $y_j = y(t_j) = e^{V(t_j)}$. The model can now be written as an MNL, with generating function given by equation (3) and choice probabilities given by equation (4). As $s$

decreases in size, the number of discrete alternatives grows, but the generating function and choice probabilities can still be written in the same form. However, in the limit as $s \to 0$, one obtains the continuous logit generating function and choice density function shown in equations (5) and (6), respectively.

$$G(y_1, \dots, y_J) = \int_{b_1}^{b_2} y(t)dt \tag{5}$$

$$p_{t_k} = \frac{\exp(V(t_k))}{\int_{b_1}^{b_2} \exp(V(t))dt} \tag{6}$$

The choice density shown in equation (6) appears slightly different from that derived by McFadden (1976), Ben-Akiva and Watatanada (1981), and Ben-Akiva et al. (1985). Their derivations include an additional additive component in the utility equation, called the opportunity density (formulated as the natural logarithm of an attractiveness function), which defines the density of elemental alternatives at any particular point in the continuous spectrum. Essentially, this was included in their formulations as a way to inform the model of the supply of housing, employment, or other attributes in a location choice context. However, it can simply be viewed as one component in the systematic utility equations, and is, thus, unnecessary here. Their formulations also employ two-dimensional integrals, since the applications were all in location choice contexts.

**3.2 Cross-Nested Logit**

The cross-nested logit (CNL) model is for discrete alternatives and has been well documented in the literature (see, e.g., Small 1987, Vovsha 1997, Ben-Akiva and Bierlaire 1999, Wen and Koppelman 2001, Papola 2004) and offers a rather flexible correlation structure for discrete alternatives. The discrete CNL is formulated here to aid in the notation and formulation of the continuous cross-nested logit (CCNL) detailed in the next sub-section. The CNL's generating function and choice probabilities are shown in equations (7) and (8), respectively.

$$G(y_1, \dots, y_J) = \sum_{m=1}^{M} \left[ \sum_{j \in C_m} \left( \alpha_{jm} y_j \right)^{\rho_m} \right]^{\frac{1}{\rho_m}} \tag{7}$$

$$P_k = \frac{\sum_{m=1}^{M} \left( (\alpha_{km} y_k)^{\rho_m} \left[ \sum_{j \in C_m} \left( \alpha_{jm} y_j \right)^{\rho_m} \right]^{\frac{1}{\rho_m} - 1} \right)}{\sum_{n=1}^{M} \left[ \sum_{j \in C_n} \left( \alpha_{jn} y_j \right)^{\rho_n} \right]^{\frac{1}{\rho_n}}} \tag{8}$$

Here, *m* = 1,…,*M* indexes the set of nests, $\alpha_{jm}$ is an allocation parameter defining the degree to which alternative *j* is a member of nest *m*, $\rho_m$ denotes the inclusive value parameter for nest *m*, and $C_m$ is the subset of alternatives in nest *m*. To be consistent with random utility theory, $\rho_m \geq 1, \forall\, m$. Furthermore, $\alpha_{jm}$ should satisfy the condition $\alpha_{jm} \geq 0, \forall\, j, m$ and $\sum_{m=1}^{M} \alpha_{jm} = 1, \forall\, j$ (Wen and Koppelman 2001, Bierlaire 2006, Abbe et al. 2007, and Marzano and Papola 2008). As a practical matter, each subset of alternatives, $C_m$, should be distinct (i.e., $C_m \neq C_n, \forall\, m \neq n$) for model identification purposes.

### 3.3 Continuous Cross-Nested Logit

Similar to the way in which the MNL is generalized for a continuous response variable, the CNL can be generalized, though some additional support is needed, as discussed here. Of great importance is nest composition. It makes good sense to think of these nests as small, contiguous intervals of the continuous spectrum of alternatives. And, since the response variable is continuous, it seems reasonable to restrict our attention to the case of ordered alternatives. Thus, each nest should be constructed so that it contains a set of sequential elemental alternatives.

The set of nests could be structured in a couple different ways. For instance, one could construct a finite number of nests, similar to the discrete CNL model. However, a more general approach would be to consider the set of nests in the same manner as the set of alternatives, effectively infinite. Such treatment requires parameterization of the inclusive value and allocation parameters, as discussed in more detail below.

As before, suppose the continuous response variable of interest, *t*, is bounded by $b_1a$ and $b_2b$ (e.g., 0 and 24 hours for trip departure time); *t* is discretized so that $t_j$ denotes the $j^{\text{th}}$ discrete alternative; and *J* is the total number of choice alternatives, computed as $J = 1 + \frac{b_2 - b_1}{s}$. In addition, suppose the number of nests equals the number of alternatives (i.e., $J = M$), and the nest interval is given as 2*h* (i.e., each nest, *m*, is composed of elemental alternatives ranging from alternative $t_m - h$ to alternative $t_m + h$). Let $\alpha(t_j, t_m)$ denote the allocation parameter for alternative *j* in nest *m* and let $y_j = y(t_j) = e^{V(t_j)}$, the exponent of systematic utility for alternative *j*. Like the parallel between MNL and continuous logit, we can now write the generating function and choice probabilities for this discretized model as they appear in equations (7) and (8). As before, taking the limit, $s \to 0$ (and $J = M \to \infty$), results in the generating function and choice density function for the CCNL model, as shown in equations (9) and (10), respectively.

$$G(y_1, \dots, y_J) = \int_{b_1}^{b_2} \left( \int_{q-h}^{q+h} [\alpha(r,q)y(r)]^{\rho}\, dr \right)^{\frac{1}{\rho}} dq \tag{9}$$

$$p_{t_k} = \frac{\int_{t_k-h}^{t_k+h} [\alpha(t_k,w)y(t_k)]^{\rho} \left( \int_{w-h}^{w+h} [\alpha(r,w)y(r)]^{\rho} dr \right)^{\frac{1}{\rho}-1} dw}{\int_{b_1}^{b_2} \left( \int_{q-h}^{q+h} [\alpha(r,q)y(r)]^{\rho} dr \right)^{\frac{1}{\rho}} dq} \tag{10}$$

Here, a single inclusive value parameter, $\rho$, is considered instead of allowing for each nest to have a different parameter (which would require further parameterization). In the context of continuous response, this seems reasonable since one may expect similar correlations across alternative pairs separated by common distances (e.g., alternatives 5 min apart in the AM share the same amount of latent information as alternatives 5 min apart in the PM). Of course, the correlation structure may vary over times of day, particularly around one's preferred arrival time, which is a limitation of this specification. However, without preferred arrival time data (which is rarely available in revealed preference data), the current specification is adequate. As with the discrete CNL model, $\rho \geq 1$, to be consistent with random utility theory. Unlike $\rho$, the allocation parameters, $\alpha(r,q)$, are parameterized here, though they could be taken as constants over the nest interval.

Suppose that for any nest *m* and alternative *j*, the allocation parameter, $\alpha(t_j, t_m)$, is parameterized such that if *m* = *j*, then $\alpha(t_j, t_m)$ takes on its greatest value. And define *h* so that, if $|t_j - t_m| \geq h$, then

$\alpha(t_j, t_m) = 0$; otherwise $\alpha(t_j, t_m) > 0$. In other words, the allocation parameter is zero if alternative *m* and alternative *j* are more than *h* units apart (e.g., one hour apart), and strictly positive otherwise. (Remember that the number of nests equals the number of alternatives.) Finally, $\alpha(t_j, t_m)$ must be normalized as shown in equation (11), just as it was for the discrete CNL (for unbiased results [Abbe et al. 2007]).

$$\int_{b_1}^{b_2} \alpha(t_j, t_m)\, dt_m = 1, \forall\, j \tag{11}$$

With this constraint it can be shown that equations (9) and (10) reduce to equations (5) and (6) for $\rho = 1$ (i.e., the CCNL collapses to the continuous logit), similar to the manner in which the CNL collapses to the MNL for inclusive value parameters of 1. Even with these restrictions, the analyst retains a great deal of flexibility in the parameterization of the allocation parameters. Here, a simple triangular formulation is proposed, as shown in equation (12) and as illustrated in Figure 1.

$$\alpha(t_j, t_m) = \begin{cases} \frac{h-|t_j-t_m|}{h^2} & \text{if } |t_j - t_m| \le h \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

This specification allows time to continue through midnight, meaning that 0 hours and 24 hours are identical departure time choices. This ensures that condition (11) holds for $t_j$ “close to” (i.e., within $h$ units of) either limit, $b_1$ or $b_2$. In addition, it allows for correlations to emerge across these times, reflecting the cyclical nature of a day period. (Essentially, travelers view an 11:59 pm departure time virtually the same as 12:01 am.) By formulating $\alpha$ in this way, correlations across alternatives separated by a common distance are identical. This specification also allows a single parameter, $h$, to control behavior of the allocation parameters.

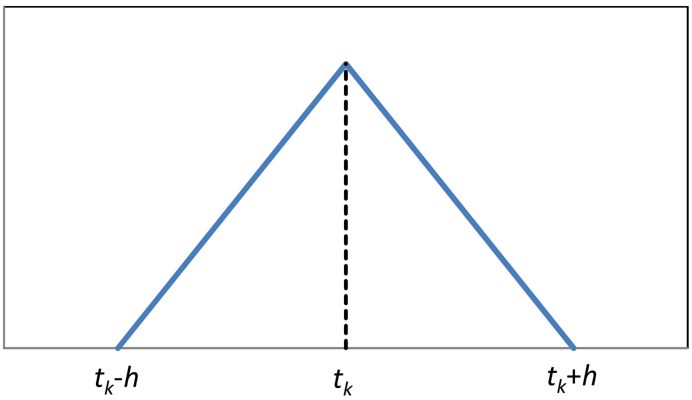


Figure 1: Allocation Parameter Illustration

Of course, other formulations could be used here as well, to allow for other shapes (e.g., normal or uniform) or for skew. However, adding further complexity to the model is probably unnecessary and can add estimation challenges. Another option is to specify $\alpha$ as constant across alternatives present in each nest. While this formulation seems simpler, it would not reduce the number of parameters to be estimated in the model (the nest size, *h*, must still be estimated) and allocation parameters for alternatives at nest boundaries would not be well-defined[4]. In the next section, the model behavior of the CCNL is compared to that of the continuous logit to illustrate the roles of the allocation and inclusive value parameters.

## 4. MODEL BEHAVIOR AND UTILITY

In this section, the utility function is formulated in the context of continuous choice, and some example parameter values are used to illustrate the shape of predictive densities for the continuous logit and CCNL. For the CCNL, *h* and $\rho$ are varied to demonstrate their roles in the choice behavior.

### 4.1 Utility Formulation

Without loss of generality, suppose one wishes to model home-to-work departure time choice, where the choice set includes any instant from 0 to 24 ($b_1$ to $b_2$) hours measured from midnight to midnight on consecutive days. One could imagine any number of continuous utility forms to use in this context. For instance, Abou Zeid et al. (2006) and Popuri et al. (2008) specified random utility in a continuous framework, both using sinusoidal functions of the departure time interacted with covariates. A similar systematic utility specification is adopted here, as shown in equation (13).

$$V_i(t) = X_i\beta\gamma(t) + \sum_{q=1}^{Q} \eta_q g_{iq}(t) \tag{13}$$

$$\gamma(t) = \left[\sin\left(\frac{2\pi t}{24}\right), \sin\left(\frac{4\pi t}{24}\right), \dots, \sin\left(\frac{2L\pi t}{24}\right), \cos\left(\frac{2\pi t}{24}\right), \cos\left(\frac{4\pi t}{24}\right), \dots, \cos\left(\frac{2L\pi t}{24}\right)\right]' \tag{14}$$

Here, *X* is a row vector of *K* individual-specific variables; $\beta$ is a matrix of parameters to be estimated with size *K* x 2*L*; and $\gamma(t)$ is a 2*L* x 1 column vector of a collection of cyclical functions of departure time *t*. Note that some covariates may be interacted with fewer than 2*L* cyclical functions by restricting the applicable elements of $\beta$ to be zero (where the integer value of *L* is chosen by the analyst). Time-varying variables (such as travel time, travel time variability, and cost) are represented by $g_{iq}(t)$, and there are *Q* of these variables. Since these variables vary over time, they need not enter the utility function in any special form. There are a couple of reasons for selecting this utility form. First, it allows utility to take on a variety of shapes, including multimodality. In addition (and as pointed out by Abou Zeid et al. [2006] and Popuri et al. [2008]), 24 hours is a multiple of the period of each cyclical function, which offers day-to-day consistency in the utility function (e.g., utility at 0 and 24 hrs is identical).

### 4.2 Model Behavior

Recently, several papers have investigated the correlation structure implied by the cross-nested logit model. Here, those results are applied in the continuous context to illuminate the correlation structure of this paper's CCNL model.

The covariance between any two discrete CNL random error components can be expressed as follows (for derivation, see, e.g., Papola and Marzano 2005 and Abbe et al. 2007):

---

[4] For instance, if the nest parameter *h* is 1, then the allocation parameter for the 8 am alternative in the nest centered at 9 am is ill-defined. From the left, it would take a value of zero, while from the right, it would take a value of $a$. Since numerical integration is required in model estimation, which relies on evaluation of discrete density values, either treatment of such boundary points will affect estimation results, and neither will be correct.

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon_i \varepsilon_j \frac{\partial F(\varepsilon_i, \varepsilon_j)}{\partial \varepsilon_i \partial \varepsilon_j} d\varepsilon_i d\varepsilon_j - (\rho_0 \varphi)^2 \quad (15)$$

Here, $F(\varepsilon_i, \varepsilon_j)$ represents the joint cumulative distribution function (CDF) for error terms, $\varepsilon_i$ and $\varepsilon_j$, $\varphi$ is Euler's constant, and $\rho_0$ is the scale parameter (set to one in this case as is typical to identify the model). This representation is rather complex, but Marzano and Papola (2008) have shown that this equation (15) can actually be rewritten as follows:

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[F(\varepsilon_i, \varepsilon_j) - F(\varepsilon_i)F(\varepsilon_j)\right] d\varepsilon_i d\varepsilon_j \quad (16)$$

While equation (16) still involves integration over the domain of the error terms, the partial derivative of the joint CDF has been eliminated. Here, $F(\varepsilon_i)$ and $F(\varepsilon_j)$ denote the marginal CDFs of the error terms, which, thanks to the normalization of the allocation parameters (so that they sum to 1 across all nests for each alternative), can be written as follows:

$$F(\varepsilon_i) = \exp\left(-\exp\left[-\left(\varepsilon_i - \ln\left(\sum_{m=1}^{M} \alpha_{im}\right)\right)\right]\right) = \exp(-\exp[-\varepsilon_i]) \quad (17)$$

The right-hand side of equation (17) results from the normalization of the allocation parameters. Note here that nests are indexed from 1 to *M*. The bivariate CDF of any two error terms can be written as follows:

$$F(\varepsilon_i, \varepsilon_j) = \exp\left(-\sum_{m=1}^{M}\left[(\alpha_{im}\mathrm{e}^{-\varepsilon_i})^{\rho_m} + \left(\alpha_{jm}\mathrm{e}^{-\varepsilon_j}\right)^{\rho_m}\right]^{\frac{1}{\rho_m}}\right) \quad (18)$$

Here, $\rho_m$ is the inclusive value parameter corresponding to nest *m*. Note that $\rho_m = \rho \ \forall \ m$ under the CCNL specification of Section 3. Moreover, the summation in equation (18) is analogous to an integral under the CCNL specification, which facilitates writing the bivariate CDF.

$$F(\varepsilon_i, \varepsilon_j) = \exp\left(-\int_{t_i-h}^{t_j+h}\left[(\alpha(t_i, t_m)\mathrm{e}^{-\varepsilon_i})^{\rho} + \left(\alpha(t_j, t_m)\mathrm{e}^{-\varepsilon_j}\right)^{\rho}\right]^{\frac{1}{\rho}} dt_m\right) \quad (19)$$

Here, the allocation parameters, $\alpha(q, r)$, take the functional form shown in equation (12). Remember that under the CCNL representation, $\varepsilon_i$ and $\varepsilon_j$ represent the error components related to elemental alternatives *i* and *j*, where alternatives *i* and *j* represent departure times of $t_i$ and $t_j$. Also, note that we have implicitly assumed $t_i < t_j$ (i.e., outbound trips occur before their associated return trips).

With this information, it is not difficult to numerically compute correlation coefficients across random error components of the CCNL. However, a couple of observations can be made without computation. First, suppose *j* = *i* + 1, and $t_j$ is one elemental time unit greater than $t_i$. In other words, suppose we take the limit as $t_j \to t_i$. The joint CDF then reduces to the following (thanks to the normalization of allocation parameters):

$$F(\varepsilon_i, \varepsilon_j) = \exp\left(-[\mathrm{e}^{-\rho\varepsilon_i} + \mathrm{e}^{-\rho\varepsilon_j}]^{\frac{1}{\rho}}\right) \quad (20)$$

This is exactly the joint CDF of nested logit random errors sharing a common nest, with correlation coefficient given by $(1-\rho^{-2})$. And, since the correlation across any other alternatives separated by a distance greater than zero must be smaller, $(1-\rho^{-2})$ represents the maximum correlation across any pair of alternatives under the CCNL specification. Of course, one would expect near (or even exactly) perfect correlation between alternatives separated by an infinitesimally small time step.

Another observation that can be made is that the time interval between alternatives need only be measured in units of *h* (where 2*h* represents the nest interval size). In other words, the correlation across any two alternatives separated by a distance *ah*, where *a* is some constant, will be the same, even if *h* changes. This results from the joint CDF (equation 19) depend only on the shared area between the respective allocation parameter functions.

To illustrate the range of correlation coefficients one can achieve under the CCNL model, numerical computation was used. Table 1 shows the correlation coefficient for alternatives separated by a variety of distances (measured in units of *h*) for a variety of inclusive value parameters, $\rho$.

Table 1: Correlation Coefficients for Varying $\rho$ and Distance Variables

| | | $\rho$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1.1 | 1.25 | 1.5 | 2 | 3 | 5 | 10 |
| Distance between Alternatives | 0 | 0.173 | 0.360 | 0.555 | 0.750 | 0.889 | 0.960 | 0.990 |
| | 0.2*h* | 0.165 | 0.341 | 0.524 | 0.705 | 0.831 | 0.894 | 0.920 |
| | 0.4*h* | 0.145 | 0.299 | 0.457 | 0.610 | 0.713 | 0.763 | 0.782 |
| | 0.6*h* | 0.119 | 0.245 | 0.372 | 0.491 | 0.571 | 0.607 | 0.622 |
| | 0.8*h* | 0.091 | 0.186 | 0.281 | 0.368 | 0.425 | 0.451 | 0.461 |
| | *h* | 0.064 | 0.129 | 0.194 | 0.254 | 0.292 | 0.309 | 0.315 |
| | 1.2*h* | 0.041 | 0.082 | 0.123 | 0.160 | 0.184 | 0.195 | 0.199 |
| | 1.4*h* | 0.023 | 0.046 | 0.069 | 0.089 | 0.102 | 0.108 | 0.110 |
| | 1.6*h* | 0.010 | 0.020 | 0.030 | 0.039 | 0.045 | 0.048 | 0.049 |
| | 1.8*h* | 0.002 | 0.005 | 0.008 | 0.010 | 0.011 | 0.012 | 0.012 |
| | 2*h* | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Clearly, as the distance between alternatives grows, the correlations shrink; and correlations grow with increasing $\rho$. One may note that even for very large values of $\rho$, error term correlations die off rather quickly as the distance between the alternatives exceeds *h* units. In addition, alternatives separated by zero distance are not perfectly correlated (except when $\rho$ goes to infinity). This is a product of the model specification. Of course, such alternatives under the continuous logit specification will have zero correlation. The following section details the empirical application of the CCNL, comparing it to the continuous logit.

**5. EMPIRICAL ANALYSIS**

In this section, work-tour departure time data is used to illustrate the CCNL estimation using Bayesian techniques. For comparison purposes, the same data is used to estimate a continuous logit model.

### 5.1 Data Description

The data used here come from the 2000 San Francisco Bay Area Travel Survey (BATS), which collected travel information for approximately 17,000 households over a 2-day period. The data was obtained in tour format (rather than trips), with each record coded with a variety of demographic and travel information. Here, only individual/day-specific variables were included in the analysis. Further, the sampling frame was restricted to the first home-based work tour made on a weekday for an individual (over the 48-hour survey period).[5] This limited the sample to approximately 18,000 tours, out of the 100,000 tours in the entire dataset. Unfortunately, model estimation is extremely computationally expensive, due to the numerical integration required to obtain likelihood values.[6] Thus, an = 997 random sample of tours was used for model estimation.

Unfortunately, the data do not contain time-varying travel time or cost variables that are shown in the systematic utility equation (13). Instead, speed regression equations (formulated as the natural logarithm of reported speed divided by network free-flow speed) were developed and their parameters estimated using a very similar methodology to that of Popuri et al. (2008). In addition, a measure of travel time variance was modeled (measured as the squared difference between reported and predicted travel times). Due to space constraints, the results of these models are not presented here, though it should be noted that the models appropriately predict the slowest speeds and the highest travel time variances during typical AM and PM peak periods. Lemp (2009) provides more details on these results.

### 5.2 Estimation Methods

The models were estimated using Bayesian methods. If we let the set of independent variables be denoted by *X*, the set of response variables be *Y*, and the parameters be $\theta$, the Bayesian posterior distribution can be written as follows (Gelman et al. 2004):

$$p(\theta|Y,X) \propto p(Y|\theta,X)p(\theta) \tag{21}$$

Here, $p(Y|\theta,X)$ represents the likelihood function and $p(\theta)$ is a prior distribution on the model parameters (reflecting the analyst's prior beliefs). In this work, the prior on $\beta$ and $\eta$ (from the utility equation [13]) were chosen to be independent and normally distributed with vague (i.e., large variance) prior parameters. For the CCNL, priors on *h* and $\rho$ were chosen to be independent gamma distributions (bounded *below* by 0.25 and 1, respectively), both with shape and inverse-scale parameter values of 1 and 0.5. These priors offer the model *some* information on *h* and $\rho$[7] (unlike the vague priors selected for other parameters), which pulls them closer to their respective left-side bounds. This is quite reasonable for *h*, since *h* specifies the minimum time interval between uncorrelated alternatives (and one would not expect correlations between alternatives a great distance apart). While similar expectations may not exist for $\rho$, its prior can be viewed as follows: unless the data offer significant proof for another

---

[5] Since work tours are usually assumed to be scheduled first (among all tour types), there was no need to restrict the times available for tour scheduling (i.e., the entire 24-day was assumed to be available to each worker).

[6] Each draw obtained for the CCNL estimation (on a PC with 2.66 GHz processor and 3.25 GB of RAM) takes roughly 3.5 to 5 seconds, depending on the parameter values. This amounts to approximately 50 hours for the 40,000 draws required for convergence.

[7] The priors on *h* and $\rho$ are relatively vague, as compared to the likelihood contribution of the *n* = 997 sample used here. In other words, the likelihood function plays a much more important role in posterior distribution of these parameters than the prior.

value, the prior guides $\rho$ to a value of one, thereby reducing the model to the continuous logit. Also, while $h$ is not required to be at least 0.25 in model formulation, this restriction aids in numerical integration computations. For small $h$, a large number of function evaluations would be required to obtain reasonable integral estimates using Simpson's rule (Press et al. 1989). Since values of $\rho$ close to 1 will effectively eliminate the impact of $h$ on the model, this should not be a sizable issue in model estimation, though this means that if correlations exist in departure time choice (i.e., $\rho > 1$), departure time alternatives a minimum of 1 hour apart will be correlated.

The Bayesian inference proceeds as follows. First, a Markov Chain is constructed so that its stationary distribution is equivalent to the posterior distribution. Second, samples are generated from this Markov Chain until draws converge to the posterior distribution. (See, e.g., Gelman et al. 2004 and Gamerman and Lopes 2006.) Here, an adaptive Metropolis-Hastings (MH) algorithm is employed to construct the Markov Chain. The adaptive MH algorithm uses a proposal density, denoted $q(\theta|\theta_i)$, where $\theta_i$ is the current value of the parameters in the Markov Chain, and proceeds as follows:

1. Draw a new set of parameters, $\theta_j$, from the proposal density $q(\theta|\theta_i)$.
2. Compute $\delta(\theta_i, \theta_j) = \min\{1, \frac{p(Y|\theta_j, X)p(\theta_j)q(\theta_i|\theta_j)}{p(Y|\theta_i, X)p(\theta_i)q(\theta_j|\theta_i)}\}$.
3. With probability $\delta$, let the new value of $\theta$ be $\theta_j$. With probability $1 - \delta$, let the new value of $\theta$ be $\theta_i$.
4. Repeat.

The algorithm works best when the proposal density, $q(\theta|\theta_i)$, is very close to the posterior distribution. Since the form of the posterior is unknown, $q(\theta|\theta_i)$ was chosen to be multivariate normal with mean $\theta_i$. Using this proposal density for the CCNL model estimation, it is possible to draw a value of $\rho$ less than 1 and/or a value of $h$ less than 0.25. In such cases, the entire set of parameters was re-drawn from the same proposal density until an acceptable set of parameters was obtained. The covariance matrix of the proposal density was initially set to be zero for all off-diagonal elements, with very small values on the diagonal. This helped ensure that proposals were often accepted at the beginning of the chain. After 1,000 draws were obtained, the covariance matrix of the proposal density was estimated from all previous draws and updated every 20$^{th}$ iteration. After 5,000 draws were obtained, only the previous 5,000 draws were used to update the covariance matrix (not including the current draw). As long as the current parameter values were not used in estimation of the covariance matrix, this approach has been found to converge to the proper posterior distribution (Holden et al. 2009).

After about 200,000 iterations, draws from the continuous logit posterior appeared to converge. Another 100,000 draws were obtained after convergence. To eliminate correlation between successive draws (which is inevitable using the MH algorithm described here), only every 50$^{th}$ draw (from these 100,000) was used for inference. Because draws from the CCNL were obtained much more slowly (30 times slower), the CCNL Markov chain was started at the mean parameter values obtained from the continuous logit estimation. This allowed the CCNL's posterior to converge after only 100,000 draws. An additional 50,000 draws were obtained after convergence, again using only every 50$^{th}$ draw (from these 50,000) for inference. Geweke's (1992) diagnostic was used to check for convergence and results were found to be acceptable. Results of model estimation are presented in the next section.

### 5.3 Empirical Results

Tables 2 and 3 show the estimation results for the two models[8]. Mean parameter estimates are generally consistent across the two models. The mean coefficient estimates on average travel time, travel time variance, and travel cost have negative signs for both models, as expected. Mean values of travel time (VOTTs) implied from the models' draws are $7.34 and $2.97 per hour for the continuous logit and CCNL models, respectively. These values are rather low compared to expectations, since VOTTs are often estimated to be much greater (e.g., around $8 to $25 per hour [Brownstone and Small 2005]). Mean values of reliability (VOR) for the two models are almost identical, with estimates of $0.028 and $0.029 per squared minute (or $9.95 and $10.30 per hour of travel time's standard deviation). The following paragraph discusses these results in more depth.

The intervals are rather wide on both travel cost parameters, which is probably due to the relative absence of variation in the travel cost variable across departure time alternatives, as compared to the average travel time and travel time variance attributes. This is particularly true for the continuous logit estimates, and results in 90% VOTT intervals that range from $0.03 to $30.00 per hour (with a median of $1.43 per hour) and VOR intervals from $0.05 to $22.13 per hour (with a median of $2.64 per hour of standard deviation). The CCNL's parameter estimates suggest tighter 90% intervals, ranging from $1.20 to $6.01 per hour for VOTT (with a median of $2.28 per hour) and from $6.58 to $14.04 per hour for VOR (with a median of $9.81 per hour). Of course, travelers may not trade off time and money (or even reliability so much) in their departure slot choices, due to ignorance or constraints on departure timing (like child drop off and/or fixed work-start times). So these VOTT and VOR estimates cannot really be compared to those coming from route or mode choice models. Moreover, these estimates are specific to the imputed network variables used here, which are imperfect (as discussed in detail by Lemp 2009).

[8] Note that each covariate is interacted with differing numbers of cyclical functions in the utility specification. While no formal methods for variable selection were used here, initial results from the continuous logit were examined to identify covariates with the most important effects.

Table 2: Model Estimation Results

| Variable | Continuous Logit | | CCNL | |
|---|---|---|---|---|
| LOS Variables | Mean | 95% Interval | Mean | 95% Interval |
| Average Travel Time | -0.00095 | (-0.0178, 0) | -0.0042 | (-0.0135, -0.0011) |
| Travel Time Variance | -0.00004 | (-0.0063, 0) | -0.0025 | (-0.0064, -0.0008) |
| Cost | -0.04754 | (-0.3633, -0.0038) | -0.0954 | (-0.1992, -0.0434) |
| Departure Time Functions | | | | |
| Sin(2*pi*t/24) | 3.6977 | (1.7867, 6.0541) | 2.1785 | (1.2983, 3.0339) |
| Sin(4*pi*t/24) | 2.4269 | (0.2672, 5.0664) | 0.6261 | (0.2759, 0.937) |
| Sin(6*pi*t/24) | 2.4645 | (0.8885, 4.1858) | 1.5397 | (1.1058, 1.9953) |
| Sin(8*pi*t/24) | 0.5528 | (-0.2218, 1.3142) | 0.4923 | (0.1411, 0.8538) |
| Cos(2*pi*t/24) | -4.7262 | (-8.0947, -1.9919) | -2.8292 | (-3.391, -2.3379) |
| Cos(4*pi*t/24) | -0.5141 | (-1.9731, 0.8283) | -0.3064 | (-0.8347, 0.2194) |
| Cos(6*pi*t/24) | 2.4113 | (1.3118, 3.5476) | 1.5987 | (1.0785, 2.0579) |
| Cos(8*pi*t/24) | 1.3434 | (0.5838, 2.1487) | 0.7941 | (0.3889, 1.2113) |
| Male Indicator Interactions | | | | |
| Sin(2*pi*t/24) | 0.7372 | (0.1533, 1.3411) | 0.7580 | (0.3246, 1.206) |
| Sin(4*pi*t/24) | 0.6996 | (0.2083, 1.1869) | 0.7346 | (0.4706, 1.0062) |
| Cos(2*pi*t/24) | -0.1548 | (-1.2793, 0.9883) | -0.3178 | (-0.7462, 0.1223) |
| Cos(4*pi*t/24) | 0.2382 | (-0.2816, 0.7504) | 0.1479 | (-0.2071, 0.4989) |
| Age Interactions | | | | |
| Sin(2*pi*t/24) | 0.0583 | (-0.0073, 0.1269) | 0.0735 | (0.0342, 0.1178) |
| Sin(4*pi*t/24) | 0.0600 | (-0.0213, 0.1427) | 0.0762 | (0.0302, 0.1281) |
| Sin(6*pi*t/24) | 0.0107 | (-0.0437, 0.0656) | 0.0193 | (-0.0126, 0.0565) |
| Sin(8*pi*t/24) | 0.0175 | (-0.0035, 0.0387) | 0.0138 | (0.0011, 0.0279) |
| Cos(2*pi*t/24) | -0.2124 | (-0.3635, -0.072) | -0.2027 | (-0.3129, -0.1094) |
| Cos(4*pi*t/24) | -0.1242 | (-0.2073, -0.0428) | -0.1037 | (-0.1676, -0.0493) |
| Cos(6*pi*t/24) | -0.0666 | (-0.1054, -0.0291) | -0.0433 | (-0.0698, -0.0197) |
| Cos(8*pi*t/24) | -0.0267 | (-0.046, -0.0081) | -0.0145 | (-0.0262, -0.0033) |
| Part-Time Indicator Interactions | | | | |
| Sin(2*pi*t/24) | -3.1017 | (-4.8199, -1.4657) | -2.4492 | (-3.032, -1.9058) |
| Sin(4*pi*t/24) | -2.3430 | (-4.1796, -0.4361) | -1.5309 | (-2.0617, -1.0047) |
| Sin(6*pi*t/24) | -0.7430 | (-2.0028, 0.5482) | -0.4138 | (-1.0535, 0.2374) |
| Cos(2*pi*t/24) | -0.3411 | (-4.946, 3.4978) | -1.8135 | (-2.3856, -1.324) |
| Cos(4*pi*t/24) | -0.8122 | (-3.7006, 1.5505) | -1.5343 | (-1.9832, -1.0824) |
| Cos(6*pi*t/24) | -0.6394 | (-1.9494, 0.5628) | -0.7680 | (-1.1835, -0.3127) |
| High Income HH Indicator Interactions | | | | |
| Sin(2*pi*t/24) | -0.1619 | (-0.6706, 0.3492) | -0.1105 | (-0.4649, 0.2569) |
| Sin(4*pi*t/24) | -0.6585 | (-1.0927, -0.2146) | -0.5793 | (-0.8003, -0.3479) |
| Cos(2*pi*t/24) | 0.3320 | (-0.7165, 1.3286) | 0.1772 | (-0.1772, 0.5688) |
| Cos(4*pi*t/24) | 0.0923 | (-0.3726, 0.5582) | 0.0300 | (-0.2049, 0.2567) |

Table 3: Model Estimation Results (Cont'd)

| Variable | Continuous Logit | | CCNL | |
|---|---|---|---|---|
| HH Size Interactions | Mean | 95% Interval | Mean | 95% Interval |
| Sin(2*pi*t/24) | -0.2642 | (-0.4788, -0.0432) | -0.2304 | (-0.3905, -0.0608) |
| Sin(4*pi*t/24) | -0.0519 | (-0.2373, 0.1416) | -0.0158 | (-0.1391, 0.1111) |
| Cos(2*pi*t/24) | 0.4706 | (0.039, 0.8958) | 0.3622 | (0.0602, 0.6185) |
| Cos(4*pi*t/24) | -0.0110 | (-0.1986, 0.1645) | -0.0382 | (-0.1763, 0.0965) |
| No. of Other Tours Interactions | | | | |
| Sin(2*pi*t/24) | -0.8870 | (-1.2404, -0.5448) | -0.8571 | (-1.1497, -0.5676) |
| Sin(4*pi*t/24) | -0.4643 | (-0.7912, -0.1508) | -0.3929 | (-0.6774, -0.104) |
| Cos(2*pi*t/24) | 0.7624 | (-0.0748, 1.6177) | 0.5722 | (-0.0927, 1.1643) |
| Cos(4*pi*t/24) | 0.2327 | (-0.1372, 0.6069) | 0.1470 | (-0.1048, 0.3997) |
| Travel Distance Interactions | | | | |
| Sin(2*pi*t/24) | 0.0241 | (-0.0126, 0.0683) | 0.0293 | (-0.0055, 0.072) |
| Sin(4*pi*t/24) | -0.0374 | (-0.088, 0.0167) | -0.0292 | (-0.079, 0.024) |
| Sin(6*pi*t/24) | -0.0053 | (-0.0329, 0.0233) | -0.0065 | (-0.0318, 0.0195) |
| Cos(2*pi*t/24) | 0.0644 | (-0.0359, 0.1499) | 0.0652 | (-0.0208, 0.1541) |
| Cos(4*pi*t/24) | -0.0035 | (-0.0606, 0.0495) | 0.0041 | (-0.0451, 0.0554) |
| Cos(6*pi*t/24) | -0.0489 | (-0.0695, -0.0288) | -0.0408 | (-0.0645, -0.0179) |
| CBD Destination Zone Indicator Interactions | | | | |
| Sin(2*pi*t/24) | -0.7472 | (-1.6973, 0.1987) | -0.6528 | (-1.1057, -0.2365) |
| Sin(4*pi*t/24) | -0.7457 | (-1.4699, 0.0221) | -0.7339 | (-1.0776, -0.3483) |
| Cos(2*pi*t/24) | 1.6963 | (0.3099, 3.0539) | 1.4181 | (0.8338, 2.0036) |
| Cos(4*pi*t/24) | 0.6503 | (0.0173, 1.3087) | 0.5738 | (0.3014, 0.8679) |
| CCNL Structural Parameters | | | | |
| $h$ | n/a | n/a | 0.7504 | (0.5, 1.4) |
| $\rho$ | n/a | n/a | 2.3958 | (1.0675, 4.6205) |
| | | | | |
| Observations | 997 | | 997 | |

The mean estimates of the CCNL's correlation parameters appear quite reasonable, and statistically significant. For instance, $h$'s mean estimate is about 0.75, indicating that the minimum time between uncorrelated alternatives is 1.5 hours. While there is nothing special about this particular value, one certainly does not expect correlations to exist across alternatives very far apart (e.g., several more hours in time). Of course, the mean estimate of $\rho$ (of 2.40) seems lower than expected. As stated earlier, the expression $(1-\rho^{-2})$ defines the correlation between alternatives separated by an infinitesimally small distance, and one would expect near-perfect correlation between such alternatives. The correlation for such alternatives implied by the mean estimate of $\rho$ is 0.83, not 1 (or something closer to one). However, given that $\rho$ (along with $h$) defines correlations between other alternatives as well, the estimate seems reasonable.

To better understand behavioral differences suggested by the two models and the effects each covariate has on departure time choice, Figures 2 and 3 show density profiles for example individuals. In each plot, covariate values are taken to be the average over the sample for all but one covariate. Two representative values are chosen for the final covariate to demonstrate that covariate's effect[9]. Level-of-service variable effects are identical across all plots.

Figure 2a shows the difference in predictive densities for males and females, Figure 2b illustrates the age effect on departure time, Figure 2c reveals predictive densities for full-time versus part-time workers, and Figure 2d demonstrates departure time choice differences between individuals from high- and low-income households (where high income is defined as $75K per year or more). Clearly, men and older individuals tend to depart earlier than women and younger persons. In the case of the age variable, similar results were noted by Komma and Srinivasan (2008) and Gadda et al. (2009). Komma and Srinivasan (2008) also found that higher income workers tend to depart later, similar to the results presented here.[10] Presumably this results from such individuals having more flexible work schedules. Finally, both models predict part-time workers to depart later, relative to full-time workers, consistent with the findings of Abou Zeid et al. (2006), Popuri et al. (2008), Gadda et al. (2009), and others. Since part-time workers schedules often do not conform to those of full-time workers, and they tend to work shorter shifts, this result seems very reasonable.

[9] For instance, if the final covariate is an indicator variable, the two representative values are 0 and 1. For other covariates, a low value and a high value are chosen to illustrate differences across individuals.

[10] Popuri et al. (2008) noted later departures for individuals from households with two or more vehicles, which tend to be higher income households. Conversely, Gadda et al. (2009) found the opposite effect for high income households, though parameter estimates in that case were not statistically significant for the variable.

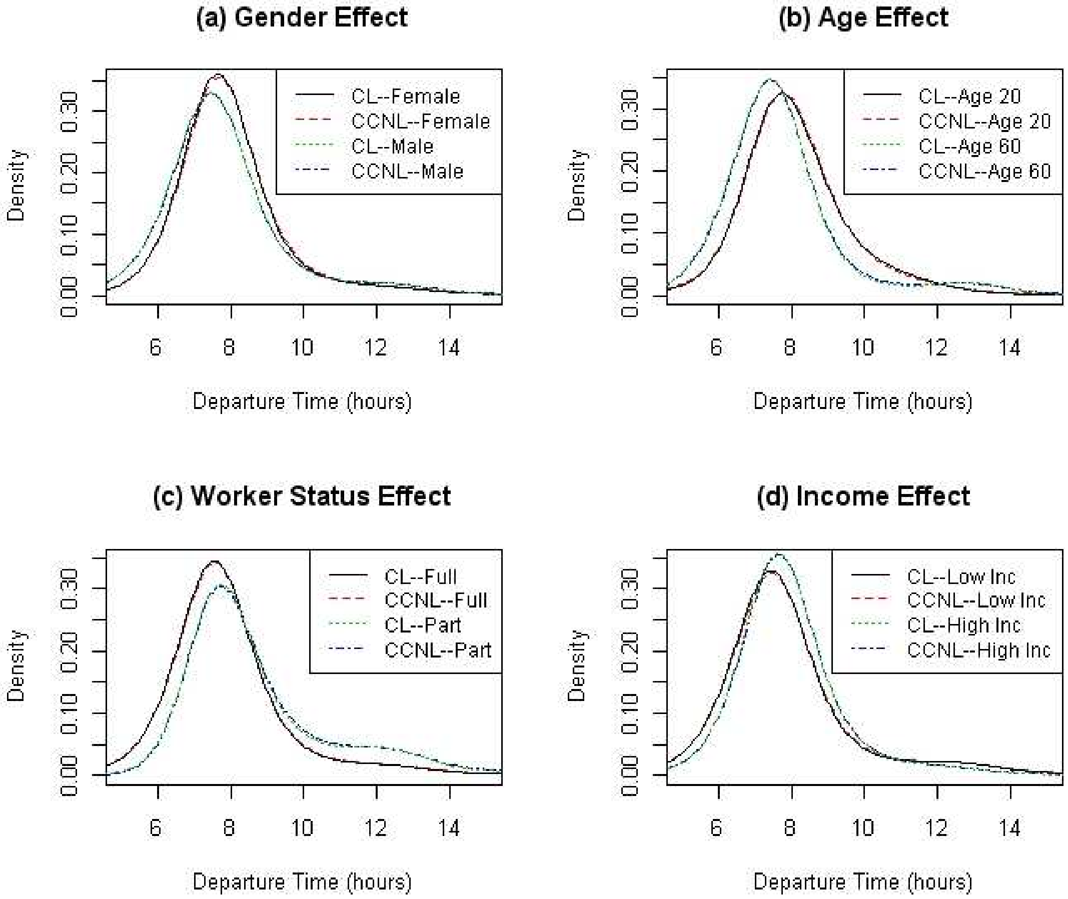


Figure 2: Gender, Age, Worker Status, and Income Effects on Average Individuals' Predictive Densities for Continuous Logit and CCNL

Figure 3 shows the effect of the four remaining covariates on departure time choice, including household size (1 versus 5 in Figure 3a), the number of other tours (0 and 3 in Figure 3b), the free-flow distance (2 mi versus 35 mi in Figure 3c), and the effect the destination being categorized as central business district (CBD). Individuals from larger households tend to depart earlier (similar results are noted by Komma and Srinivasan [2008]), which is not so surprising since they may have additional obligations, such as dropping off a child at school. The number of additional tours an individual undertakes tends to results in later departures, thanks to added scheduling constraints and shorter work-activity durations[11]. Not surprisingly, the longer the distance traveled, the earlier one departs[12], presumably in order to arrive on time (as shown in Figure 3c), and travelers destined for the CBD tend to depart a bit later (in contrast to the findings of Komma and Srinivasan [2008]). In fact, one may expect such individuals to depart earlier, since one would typically find higher congestion (and lower speed travel) in such areas during the AM peak. However, the model controls for congestion and speed effects. Those working in the CBD may have particular job types that allow for later work start times. Komma and Srinivasan (2008) controlled for two occupation types, finding that those in "professional" occupations depart a fair bit later than others; so those working in the CBD may largely be working in

[11] Note that average work-activity duration for those with additional tours is only about 6 hours, while duration for those with no additional tours exceeds 8 hours.
[12] Komma and Srinivasan (2008) and Gadda et al. (2009) obtained similar results.

"professional" occupations. Interestingly, the differences in predictive densities from the continuous logit and CCNL are negligible in most cases, which seem to indicate that the continuous logit is working nearly as well as the CCNL, though the continuous logit assumes iid error terms. Further comparisons between the two models, via out-of-sample predictive accuracies are detailed below. And these suggest that the CCNL offers superior predictive tendencies.

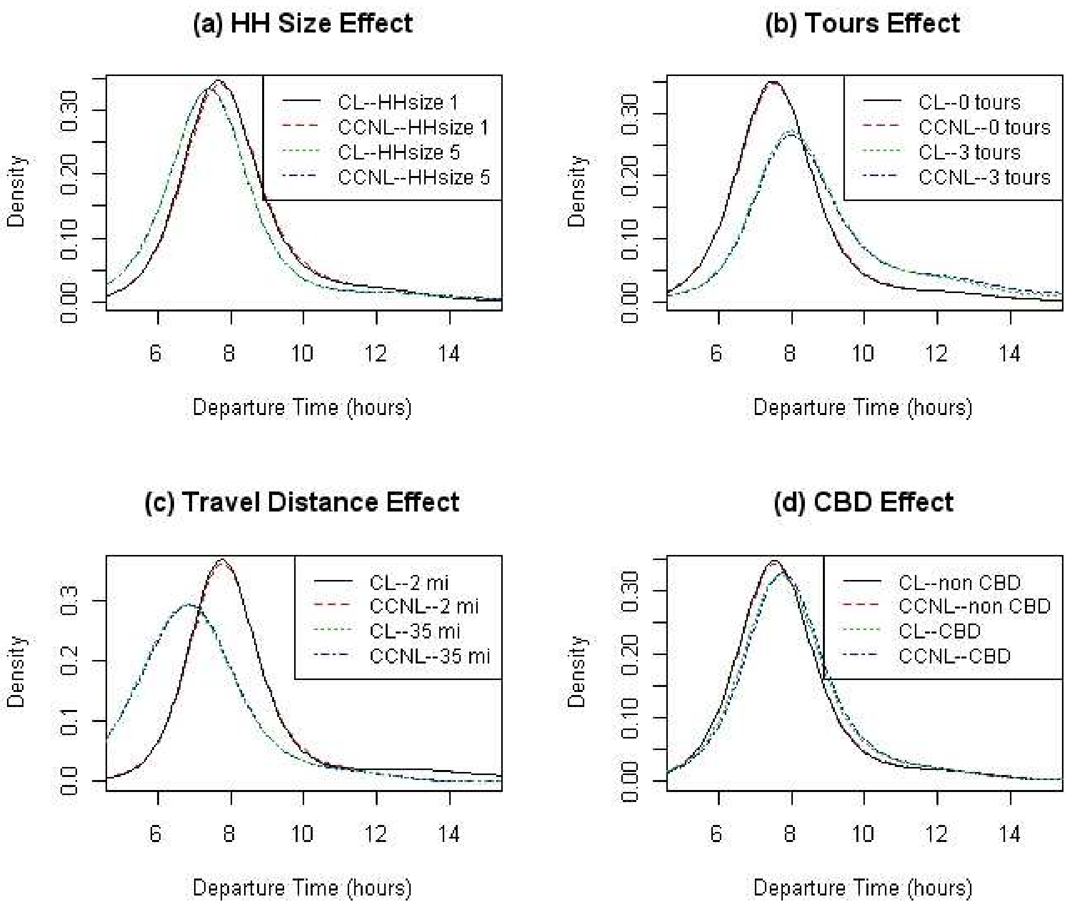


Figure 3: Example Individuals' Predictive Densities for Continuous Logit and CCNL

*5.3.1 Out-of-Sample Prediction*

In order to evaluate the models' abilities to capture variation in departure time choice, out-of-sample prediction provides a number of benefits. Furthermore, it aids in illustrating the merits of Bayesian methods. The out-of-sample data is composed of 3,550 records, representing about 20% of the total data. Since Bayesian estimation output offers a collection of parameter draws from the posterior distribution, each draw is used to compute the likelihood each model would predict the actual departure time outcome for each individual. The distribution of total log likelihoods can then be characterized. Here, each of the 2,000 continuous logit and 1,000 CCNL posterior draws are employed. Figure 4 shows the total log likelihood (across all individuals) distributions for the two models.

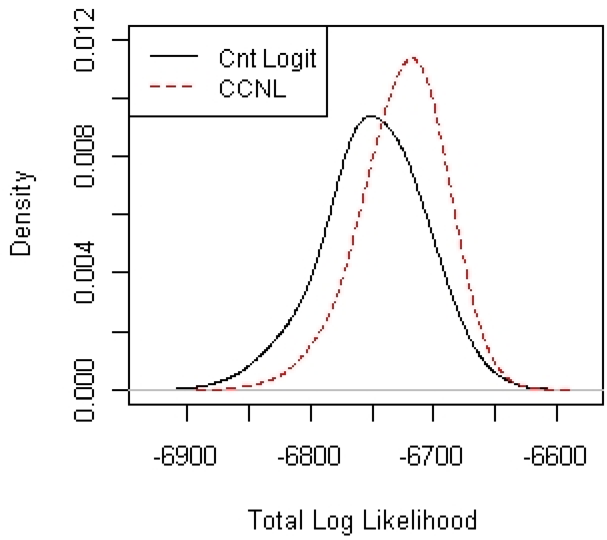


Figure 4: Out-of-Sample Likelihood Predictions for Continuous Logit and CCNL

As shown in Figure 4, the CCNL out-performs the continuous logit in this data context, though not by an overwhelming amount. The difference in mean log likelihoods is about 23. Pearson's chi-squared test can be used to test for statistical significance between these distributions. Using 11 bins, the value of the chi-squared test statistic was found to be over 1,000 with 10 degrees of freedom, suggesting with near certainty that the CCNL's log-likelihood distribution differs from that of the continuous logit.

Good (1958) proposed Bayes factors (*BF*s) for testing whether differences between two models are significant. Here, the *BF* is the exponent of the difference between Figure 4's mean log-likelihoods. Kass and Raftery (1995) proposed a test statistic of 2ln(*BF*), and they suggest that values between 2 and 6 provide positive evidence for rejecting the null hypothesis (i.e., rejecting the continuous logit model here), values between 6 and 10 provide strong evidence, and values over 10 offer very strong evidence. In this case, the test statistic takes a value of 46, offering very strong statistical evidence that the continuous logit model is inferior to the CCNL in terms of model prediction.

Alternatively, one could think of measuring how often the CCNL beats the continuous logit (in terms of its predictive ability). For instance, if one randomly draws a pair of parameter values from their respective posterior distributions (i.e., one from the continuous logit and one from the CCNL), the CCNL's corresponding likelihood beats the continuous logit's about 65% of the time. While the difference in overall predictive capability is statistically significant, the practical benefit of the CCNL over the continuous logit is not so evident. Of course, this is only one measure of model adequacy. In the next section, some example policy simulations are examined, demonstrating how economic welfare calculations can be made for the two models.

*5.3.2 Economic Welfare Example*

To illustrate how economic welfare change can be handled for the continuous logit and CCNL models, an example is formulated here. Ben-Akiva and Watanatada (1981) showed that consumer surplus for the continuous logit can be computed as the limiting formula for the MNL, as follows:

$$CS_i = \ln\left(\int_a^b \exp[V_i(t)]dt\right) \tag{22}$$

Moreover, one can convert differences in consumer surplus into monetary terms by dividing the term in equation (15) by the coefficient on cost. Consumer surplus for the discrete cross-nested logit is computed as follows (Hunt et al. 2007):

$$CS_i = \ln\left(\sum_{m=1}^{M}\left[\sum_{j\in C_m}\left(\alpha_{jm}y_{ij}\right)^{\rho_m}\right]^{\frac{1}{\rho_m}}\right) \tag{23}$$

It follows that the consumer surplus for the CCNL can be computed as follows:

$$CS_i = \ln\left(\int_a^b \left(\int_{q-h}^{q+h}[\alpha(r,q)y_i(r)]^{\rho}\, dr\right)^{\frac{1}{\rho}} dq\right) \tag{24}$$

As a base scenario, travelers (from the sample used in model estimation) are assumed to face the conditions provided in the data. Three tolling policy simulations are investigated here. In the first, it is assumed that \$0.15/mile tolls are assessed on all roads during the peak periods, resulting in (assumed) peak-period travel time *delay* reductions of 50%[13]. The \$0.15/mile toll represents a doubling in travel costs for most travelers (since operating costs are assumed to be \$0.15/mile), though some roadways in the San Francisco region are already tolled. So the additional \$0.15/mile toll represents less than a doubling of costs for travelers incurring other tolls.

In the second simulation, \$0.15/mile tolls are assessed during peak periods, with assumed peak-period travel time *delay* reductions of only 10%. And in the final simulation, \$0.30/mile tolls during peak periods are assumed to reduce peak-period travel time *delay* by 50%. These scenarios are distinctive, and not meant to be consistent; they should be viewed simply as potential outcomes of the tolling policies. In order to truly understand the tolls' effects, a feedback mechanism (with network loading) is required. Here, the purpose is to illustrate how CS-change calculations are performed and to appreciate the continuous logit and CCNL model behaviors.

Using these assumptions, an individual's CS change is measured by the difference in CS values across the base scenario and the tolling policy scenarios, divided (i.e., normalized) by the cost coefficient from the model. Since this is a Bayesian analysis, CS change is computed for each of 100 parameter draws from the posterior distribution for the two models. The same random draws are used in computations for each of the three simulations. The sample used here is identical to the sample used in model estimation (with 997 tour observations). Because individuals choosing transit (and non-motorized) modes are not expected to incur such tolls, only those sample tours made by the automobile mode are considered here (for 821 sample records).

Figure 5 shows the distribution of aggregate consumer surplus change (over all sample individuals) under each specification. Figure 5a shows the results for toll policy simulation 1, Figure 5b shows the results for simulation 2, and Figure 5c shows the results for simulation 3. In each policy simulation, the general shape of CS change distributions is very similar for the two models, though the CCNLs' distributions have smaller variances than those of the continuous logit. Average CS changes per traveler

[13] Note that delay is not the same as travel time. Routes enjoying little delay under the status quo scenario will see little change in travel time profiles when delays are reduced.

(across the 821 travelers) under the continuous logit are about -$0.88, -$1.04, and -$1.84 for simulations 1, 2, and 3, respectively, while under the CCNL, mean CS changes are about -$0.89, -$1.03, and -$1.81. Not surprisingly, these mean CS changes are very similar across the two models. However, standard deviations (at the traveler level) for the three policy simulations are $0.20, $0.05, and $0.22 under the continuous logit and $0.06, $0.03, and $0.10 under the CCNL. Thus, CS estimates under the continuous logit specification suffer from greater uncertainty, which is due to that model's larger interval estimates for network parameters (i.e., coefficients on travel time, variance, and cost) It should be noted that even though CS is negative under each tolling policy simulation, total economic welfare is not, since one must also consider toll revenues. After accounting for toll revenues, the mean of total economic welfare change at the traveler level is estimated to be $0.33, $0.17, and $0.40 for simulations 1, 2, and 3, respectively, under the continuous logit model; and $0.36, $0.22, and $0.41 for the three simulations under the CCNL model. Of course, these are only three examples, and drawing firm conclusions at this stage may be unwise; but the example does illustrate how CS is computed for the CCNL, while highlighting a key advantage of the Bayesian approach (i.e., obtaining the distribution of CS rather than simply a point estimate).

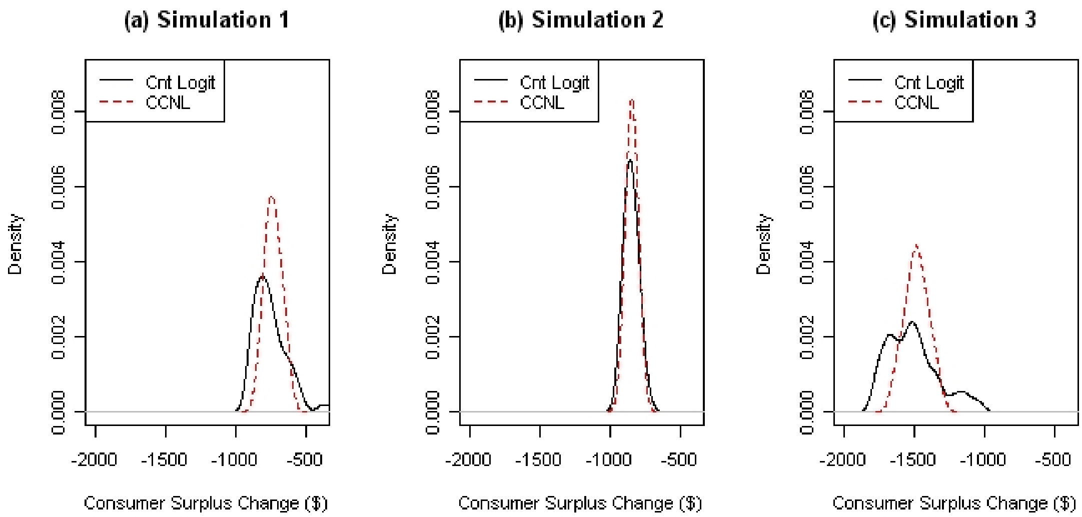


Figure 5: Distribution of Consumer Surplus Change for Three Tolling Policy Simulations under Continuous Logit and CCNL Specifications

Another interesting evaluation can be made with these simulations with regard to how the models predict departure time choice changes at an aggregate level. Figure 6 shows the aggregate (over all individuals) departure time choice distributions for each of the simulations along with the status quo (or baseline scenario). As expected, the predictive distributions are wider (and less peaked) under each of the simulations as compared to the status quo scenario for both models. In addition, there appears to be some evidence of grouping near the left- and right-side shoulders of the AM peak period, particularly under simulation 3. This seems very reasonable since these shoulders represent times right before and after tolling is assumed to begin, and given the choice between driving to work at 5:55 am with no tolls or at 6:05 am with tolls, most people would likely choose the earlier time. Under the CCNL specification, the height of these peaks appears to be a bit more pronounced than under the continuous logit specification. Since the correlations under the CCNL essentially allow areas of higher utility to draw probability away from areas of lower utility, it is not surprising to see such behavior.

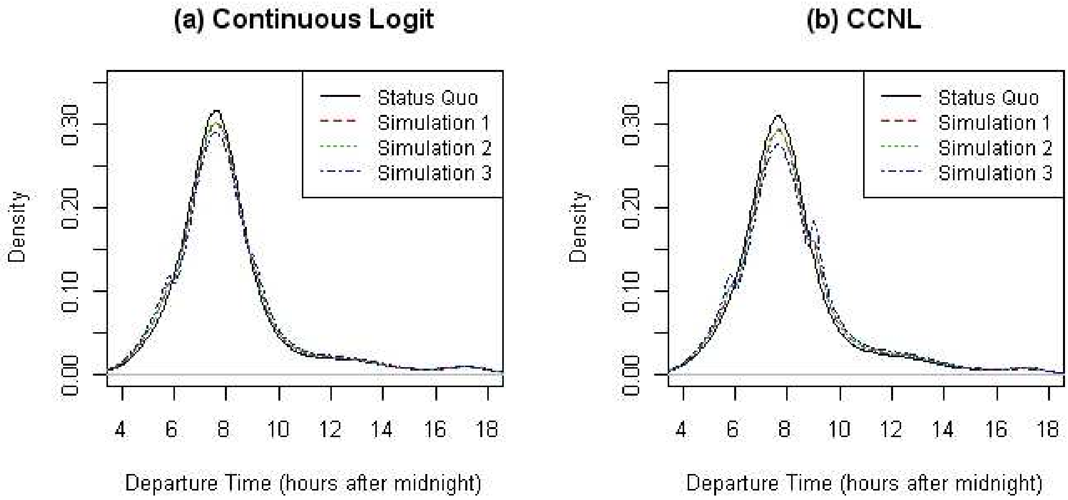


Figure 6: Distribution of Travelers' Departure Time Choices for Four Simulations under Continuous Logit and CCNL

To further investigate these predictive distributions, Table 4 presents the proportion of individuals choosing each of three TOD periods for each simulation and both models. The TOD periods include the off-peak (defined as before 5 am or after 10 am), 1-hour peak shoulders (defined as 5 to 6 am or 9 to 10 am), and the AM peak (defined as 6 to 9 am). These were specifically chosen to examine the shoulder periods one hour before and one hour after the tolled AM peak period. While the status quo simulation proportions are similar between the two models, the CCNL predicts larger percentage drops in AM peak period travelers in each of the three toll policy simulations, relative to the status quo. Even more important to notice are the predicted percentage changes relative to the peak change. These are computed as the proportion of travelers choosing the period under simulations 1, 2, and 3, divided by the proportion of travelers choosing the period under the status quo scenario. One would expect most peak-period travelers to shift to similar departure time alternatives (i.e., peak shoulders) in the face of peak-period tolls. As shown in Table 4, both models predict larger percentage increases in the shoulder periods compared to off-peak periods for each simulation, as expected. However, the relative differences between off-peak and shoulder period share changes are greater under the CCNL (except for simulation 1 where they are nearly identical for the two models). In other words, the CCNL tends to predict larger share increases for shoulder periods relative to the continuous logit. Of course, this comes from the CCNL's ability to capture correlations across departure time choices, and is reasonable and desirable.

Table 4: Predicted Departure Time Proportions for Three TOD Periods across Toll Policy SImulations

| | Proportion | | Percentage Increase from Base | |
|---|---|---|---|---|
| Status Quo | Cnt. Logit | CCNL | Cnt. Logit | CCNL |
| Off-Peak | 0.147 | 0.153 | n/a | n/a |
| 1-Hour Peak Shoulders | 0.153 | 0.157 | n/a | n/a |
| AM Peak | 0.700 | 0.691 | n/a | n/a |
| Simulation 1 | | | | |
| Off-Peak | 0.159 | 0.169 | 8.2 | 10.5 |
| 1-Hour Peak Shoulders | 0.170 | 0.179 | 11.1 | 14.0 |
| AM Peak | 0.671 | 0.652 | -4.1 | -5.6 |
| Simulation 2 | | | | |
| Off-Peak | 0.160 | 0.169 | 8.8 | 10.5 |
| 1-Hour Peak Shoulders | 0.169 | 0.181 | 10.5 | 15.3 |
| AM Peak | 0.670 | 0.651 | -4.3 | -5.8 |
| Simulation 3 | | | | |
| Off-Peak | 0.172 | 0.184 | 17.0 | 20.3 |
| 1-Hour Peak Shoulders | 0.185 | 0.204 | 20.9 | 29.9 |
| AM Peak | 0.643 | 0.613 | -8.1 | -11.3 |

While the added ability to capture correlations across choice alternatives under the CCNL model (in contrast to the continuous logit) appears to have only minor implications for predictive densities (as illustrated in Figures 2 and 3), out-of-sample predictions are better (and statistically significant) for the CCNL. Moreover, peak-toll policy simulations suggest the two specifications can have very different policy implications. While one cannot know which model's predictions are correct (since the simulations represent idealized scenarios), very sensible policy implications emerge under the CCNL in comparison to the continuous logit. Thus, the empirical evidence suggests the CCNL offers reasonable choice behavior in the departure time choice context while providing added model flexibility (and a more defensible behavioral basis) over the continuous logit specification.

## 6. CONCLUSIONS

This paper has presented a new GEV model for continuous response, the CCNL. The model represents a generalization of the discrete cross-nested logit for continuous response, much like the continuous logit represents a generalization of the MNL, but the model is much more behaviorally defensible than the continuous logit (with its independent and identically distributed error term assumption). And, like any in the GEV class, the model conforms to random utility theory, offering a strong basis by which to estimate the economic welfare implications for evaluation of policy alternatives. While empirical analysis was performed in a work tour departure time context, the model may be quite valuable for a variety of data contexts. Specifically related to transportation, a number of choice contexts are inherently continuous, like location and destination choices, activity duration, trip distances, and vehicle usage. But the applicability of the model is not limited to the field of transportation. For instance, size of home to build, amount to invest, and production/output models under random profit maximization (among many others) represent other contexts for which the CCNL may be well-suited.

Empirical results suggest that the CCNL performs better than the continuous logit model, in terms of out-of-sample prediction of departure times, and the CCNL offers more flexible choice behavior to emerge, along with welfare estimates. However, predictive densities for specific individuals are very similar across the two models (for this particular case of departure time data) and the CCNL model is much more computationally burdensome in estimation. Here, generating draws from the CCNL's posterior distribution took on the order of 30 times longer than for the continuous logit. Moreover, the numerical integration procedure for generating likelihood values suffered from more error in the CCNL context than in the continuous logit setting. Reducing such error is likely to result in even greater estimation time, though the benefits of the model may be more apparent.

Of course, the analysis provided here represents very specific utility and allocation parameter formulations. It is possible that utility could be better formulated in other ways. The allocation parameters represent another avenue of future research. In this paper, the allocation parameters were parameterized in terms of a single parameter and were represented by triangular functions. While it seems reasonable to require the allocation parameters to have a single peak, a number of shapes could be used, introducing skew, smoothness, and other potentially desirable features, including variations across time-of-day. The work undertaken here represents only a first step. It is clear that superior methods for handling continuous choice variables are needed. The CCNL offers one promising technique, but more research and experimentation is needed to fully appreciate its relative merits and flaws.

**ACKNOWLEDGEMENTS**

Funding was provided by the Strategic Highway Research Program (SHRP) II, Project CO4, titled "Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand". The authors would like to thank Mark Bradley for formatting the data used in this work and recognize the suggestions of Stephane Hess and those of several anonymous reviewers.

**REFERENCES**

Abbe, E., M. Bierlaire, and T. Toledo (2007) Normalization and Correlation of Cross-Nested Logit Models. *Transportation Research Part B*, 41, 795-808.

Abkowitz, M.D. (1981) An Analysis of the Commuter Departure Time Decision. *Transportation*, 10 (3), 283-297.

Abou-Zeid, M., T.F. Rossi, and B. Gardner (2006) Modeling Time of Day Choice in the Context of Tour and Activity Based Models. *Transportation Research Record*, 1981, 42-49.

Ben-Akiva, M. and T. Watanatada (1981) Application of a Continuous Spatial Choice Logit Model. In *Structural Analysis of Discrete Choice Data with Econometric Applications* (C.F. Manski and D. McFadden, eds.), MIT Press, Cambridge, MA, 320-343.

Ben-Akiva, M., N. Litinas, and K. Tsunokawa (1985) Continuous Spatial Choice: The Continuous Logit Model and Distributions of Trips and Urban Densities. *Transportation Research Part A*, 19 (2), 119-154.

Ben-Akiva, M. and M. Bierlaire (1999) Discrete Choice Methods and Their Applications to Short-Term Travel Decisions. In *Handbook of Transportation Science* (R. Hall ed.), Kluwer, 5-34.

Bhat, C.R. and J.L. Steed (2002) A Continuous-Time Model of Departure Time Choice for Urban Shopping Trips. *Transportation Research Part B*, 36 (3), 207-224.

Bierlaire, M. (2006) A Theoretical Analysis of the Cross-Nested Logit Model. *Annals of Operations Research*, 144, 287-300.

Chin, A.T.H. (1990) Influences on Commuter Trip Departure Time Decisions in Singapore. *Transportation Research Part A*, 24 (5), 321-333.

Daly, A.J. and S. Zachary (1979) Improved Multiple Choice Models. In *Identifying and Measuring the Determinants of Mode Choice* (D. Hensher, Q. Dalvi, eds.), Teakfield, London, 335-357.

de Jong, G., A. Daly, M. Pieters, C. Vellay, M. Bradley, and F. Hofman (2003) A Model for Time of Day and Mode Choice Using Error Components Logit. *Transportation Research Part E*, 39 (3), 245-268.

Gadda, S., K.M. Kockelman, and P. Damien (2009) Continuous Departure Time Models: A Bayesian Approach. Proceedings of the 88th Annual Meeting of the Transportation Research Board, January 2009, Washington, D.C.

Gamerman, D. and H.F. Lopes (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, FL.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin (2004) *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL.

Geweke, J. (1992) Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds.), Oxford University Press, Oxford, 169-193.

Good, I.J. (1958) Significance Tests in Parallel and in Series. *Journal of the American Statistical Association*, 53, 799-813.

Hess, S., J.W. Polak, A. Daly, and G. Hyman (2007) Flexible Substitution Patterns in Models of Mode and Time of Day Choice: New Evidence from the UK and the Netherlands. *Transportation*, 34, 213-238.

Holden, L., R. Hauge, and M. Holden (2009) Adaptive Independent Metropolis-Hastings. *The Annals of Applied Probability*, 19 (1), 395-413.

Hunt, L.M., P.C. Boxall, and B. Boots (2007) Accommodating Complex Substitution Patterns in a Random Utility Model of Recreational Fishing. *Marine Resource Economics*, 22, 155-172.

Kass, R.E. and A.E. Raftery (1995) Bayes Factors. *Journal of the American Statistical Association*, 90 (430), 773-795.

Komma, A. and S. Srinivasan (2008) Modeling Home-to-Work Commute-Timing Decisions of Workers with Flexible Work Schedules. Proceedings of the 87th Annual Meeting of the Transportation Research Board, January 2008, Washington, D.C.

Lemp, J.D. (2009) Capturing Random Utility Maximization Behavior in Continuous Choice Data: Application to Work Tour Scheduling. Doctoral Dissertation, Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin.

Marzano, V. and A. Papola (2008) On the Covariance Structure of the Cross-Nested Logit Model. *Transportation Research Part B*, 42, 83-98.

McCafferty, D. and F.L. Hall (1982) The Use of Multinomial Logit Analysis to Model the Choice of Time to Travel. *Economic Geography*, 58 (3), 236-246.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics* (P. Zaremmbka ed.), Academic Press, New York.

McFadden, D. (1976) The Mathematical Theory of Demand Models. In *Behavioral Travel Demand Models* (P.R. Stopher and A.H. Meyburg, eds.), Lexington Books, Lexington, MA, 305-314.

McFadden, D. (1978) Modeling the Choice of Residential Location. In *Spatial Interaction Theory and Planning Models* (A. Karlquist, L. Lundquist, F. Snickbars, and J.W. Weibull, eds.), North-Holland, Amsterdam, 75-96.

Papola, A. (2004) Some Developments on the Cross-Nested Logit Model. *Transportation Research Part B*, 38, 833-851.

Papola, A. and V. Marzano (2005) A Specification Procedure of the CNL Model Reproducing Any Homoskedastic Covariance Matrix. Proceedings of the European Transport Conference, Strasbourg, France.

Popuri, Y., M. Ben-Akiva, and K. Proussaloglou (2008) Time of Day Modeling in a Tour-Based Context: The Tel-Aviv Experience. *Transportation Research Record*, 2076, 88-96.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling (1989) *Numerical Recipes in Pascal: The Art of Scientific Computing*. Cambridge University Press.

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

Small, K.A. (1987) A Discrete Choice Model for Ordered Alternatives. *Econometrica*, 55 (2), 409-424.

Transportation Research Board (2007) Metropolitan Travel Forecasting: Current Practice and Future Direction. TRB Special Report 288, Committee for Determination of the State of the Practice in Metropolitan Area Travel Forecasting, Washington, D.C.

Vovsha, P. (1997) Application of Cross-Nested Logit Model to Mode Choice in Tel-Aviv, Israel, Metropolitan Area. *Transportation Research Record*, 1607, 6-15.

Vovsha, P. and M. Bradley (2004) A Hybrid Discrete Choice Departure Time and Duration Model for Scheduling Travel Tours. *Transportation Research Record*, 1894, 46-56.

Vovsha, P. (2005) Making the State of the Art the State of the Practice: Advanced Modeling Techniques for Road Pricing. Proceedings of the USDOT Expert Forum on Road Pricing and Travel Demand Modeling, November 2005, Alexandria, VA.

Wang, J.J. (1996) Timing Utility of Daily Activities and Its Impact on Travel. *Transportation Research Part A*, 30 (3), 189-206.

Wen, C.H. and F.S. Koppelman (2001) The Generalized Nested Logit Model. *Transportation Research Part B*, 35, 627-641.

Williams, H.C.W.L. (1977) On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A*, 9, 285-344.