The paired combinatorial logit model: properties, estimation and application

Frank S. Koppelman a,b,*, Chieh-Hua Wen c,1

a Department of Civil Engineering, Northwestern University, Evanston, IL 6208-3109, USA
b The Transportation Center, Northwestern University, Evanston, IL 6208-3109, USA
c Department of Shipping and Transportation Management, National Penghu Institute of Marine and Management Technology, Penghu, Makung, Taiwan, ROC

Received 9 December 1996; received in revised form 24 March 1999; accepted 15 April 1999

Abstract

The Independence of Irrelevant Alternatives (IIA) property of the multinomial logit (MNL) model imposes the restriction of zero covariance between the utilities of pairs of alternatives. This restriction is inappropriate for many choice situations; those in which some pairs or sets of alternatives share the same unobserved attributes. The nested logit (NL) model relaxes the zero covariance restriction of the MNL model but imposes the restriction of equal covariance among all alternatives in a common nest and zero covariance otherwise. The paired combinatorial logit (PCL) model relaxes these restrictions further by allowing different covariances for each pair of alternatives. This relaxation enables the estimation of differential competitive relationships between each pair of alternatives. The closed form of the PCL model retains the computational advantages of other logit models while the more flexible error correlation structure, compared to the MNL model and NL models, enables better representation of many choice situations. This paper describes the derivation, structure, properties and estimation of the PCL model. The empirical results demonstrate that the PCL model is statistically superior to the MNL and NL models and may lead to importantly different travel forecasts and policy decisions. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Discrete choice; Extreme value models; Intercity travel; Logit; Mode choice

*Corresponding author. Tel.: +847-491-8794; fax: +847-491-4011; e-mail: f-koppelman@nwu.edu
1 Tel.: +886-6-9264115; fax: +886-6-9264265; e-mail: chwen@npit.edu.tw
1. Introduction

Discrete choice analysis is used to model the choice of one among a set of mutually exclusive alternatives based on principles of utility maximization. Decision-makers are assumed to select the alternative with the highest utility. The utility of an alternative includes a deterministic portion which is a function of the attributes of the alternative and characteristics of the decision-maker and a random component which represents unknown and/or unobservable components of the utility function.

The distribution of error terms in discrete choice models is generally assumed to be either normal or Type I extreme value (Johnson and Kotz, 1970). The normal assumption results in the multinomial probit (MNP) model (Daganzo, 1979) which allows complete flexibility in the variance structure of the error terms. However, use of the MNP model has been limited due to computational difficulty, the requirement for numerical estimation of multivariate normal distributions, involved in its estimation and prediction. Development of new computational approaches (McFadden, 1989; Hajivassiliou and McFadden, 1990; Börsch-Supan and Hajivassiliou, 1993; Keane, 1994) increases the potential for greater use of the MNP. Still, widespread use is likely to be limited due to conceptual, computational and statistical problems, including difficulty in interpretation of covariance parameters and forecasting the effects of introducing new alternatives (Horowitz, 1991).

The multinomial logit (MNL) model (McFadden, 1973), the most widely used discrete choice model, has the advantage of a closed form mathematical structure which simplifies computation in both estimation and prediction. However, it imposes the constraint that the relative probabilities of each pair of alternatives are independent of the presence or characteristics of all other alternatives. This property, widely known as the Independence of Irrelevant Alternatives (IIA), follows directly from the constraint that random error terms are independently (no correlation) and identically (same variance) distributed. This property implies that the introduction or improvement of any alternative will have the same proportional impact on the probability of all other alternatives. This representation of choice behavior will result in biased estimates and incorrect predictions in cases that violate these conditions. While the MNL is widely believed to be robust to such mis-specification in many cases, important cases are likely to arise for which such mis-specification will substantially impact the accuracy and usefulness of estimated models in providing insight into choice behavior and prediction of likely future behavior.

The MNL model constraints can be relaxed by allowing the random components to be non-independent, non-identical or both. The most widely used relaxation of the independence assumption of the MNL model is the nested logit (NL) model, which allows dependence or correlation between the utilities of pairs of alternatives in common groups (McFadden, 1978; Williams, 1977; Daly and Zachary, 1978).

Other relaxations of the independence assumption of the MNL model; which can be derived from McFadden’s (McFadden, 1978) generalized extreme value (GEV) model; include the ordered generalized extreme value (OGEV) model (Small, 1987), the paired combinatorial logit (PCL) model (Chu, 1981, 1989) and the cross-nested logit (CNL) model (Vovsha, 1997). The OGEV

---

2 The MNL and NL models can be derived from the GEV model, also.
model allows correlation between pairs of alternatives in an ordered choice set based on their proximity in the order. However, the only reported empirical test of this model (Small, 1987) obtained results which were less satisfactory than the NL model and were not significantly different from the MNL model. The PCL model (Chu, 1981) allows differential correlation between pairs of alternatives. Initial application of the PCL model, to the estimation of mode choice models using 1970 census data for the Chicago Metropolitan Area, did not demonstrate statistical superiority to the MNL model, most likely due to small sample size and limited model specification. The CNL model allows differential similarities between pairs of alternatives through a structure that allows each mode to appear in multiple nests with different allocation coefficients (Vovsha, 1997). All of these models retain the advantage of a closed form mathematical structure.

Bhat (1995) and Recker (1995) proposed modifications of the MNL to relax the equal variance requirement of the MNL model; however, both require the use of numerical integration in estimation and prediction. Similarly, relaxation of both the independent and identical assumptions require the adoption of error components models (McFadden and Train, 1996) or the multinomial probit model (Daganzo, 1979) which require the use of numerical integration.

The purpose of this paper is to demonstrate the advantages of the PCL model, clarify its properties and illustrate its potential to provide useful additional understanding in applied work. Estimation is undertaken by use of a full information maximum likelihood procedure to simultaneously estimate both the utility function parameters and the similarity coefficients, constrained by the conditions of random utility maximization (Daly and Zachary, 1978). The estimation results of the PCL model are demonstrated and compared to the MNL and NL models using intercity mode choice data from the Toronto–Montreal corridor.

The remainder of this paper is organized as follows. Section 2 describes the formulation, structure, identification problems and estimation approach for the PCL model including a structural comparison to the MNL and NL models. Section 3 describes the data and estimation results and compares the estimated PCL model with the corresponding MNL and NL models. Section 4 provides a summary and implications for PCL model application.

2. The PCL model

2.1. Model formulation

The PCL model is a member of the generalized extreme value (GEV) family of models (McFadden, 1978). A GEV model can be derived from any function,

\[ G(Y_1, Y_2, \ldots, Y_n), \quad Y_1, Y_2, \ldots, Y_n \geq 0 \]

which is non-negative, homogeneous of degree one, approaches infinity with any \( Y_i, \ i = 1, 2, \ldots, n \) and has \( k \)th cross-partial derivatives which are non-negative for odd \( k \) and non-positive for even \( k \). Any function which satisfies these conditions defines a probability function for alternative \( i \) as

\[ P_i = \frac{Y_i G_i(Y_1, Y_2, \ldots, Y_n)}{G(Y_1, Y_2, \ldots, Y_n)}, \]

where

\[ G(Y_1, Y_2, \ldots, Y_n) \geq 0, \quad G(Y_1, Y_2, \ldots, Y_n) \to \infty \quad \text{as} \quad Y_1, Y_2, \ldots, Y_n \to \infty \]

for each \( i \).
where $G_i$ is the first derivative of $G$ with respect to $Y_i$. The transformation, $Y_i = \exp(V_i)$, where $V_i$ represents the observable components of the utility for each alternative $(U_i = V_i + e_i, \forall i)$ is used to ensure positive $Y_i$. The PCL model is obtained from the $G$ function:

$$G(Y_1, Y_2, \ldots, Y_n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left( Y_i^{1/1-\sigma_{ij}} + Y_j^{1/1-\sigma_{ij}} \right)^{1-\sigma_{ij}},$$

where the double summation includes all pairs of alternatives in the choice set and $\sigma_{ij}$ is an index of the similarity between alternatives $i$ and $j$. The PCL model is consistent with random utility maximization if the conditions, $0 \leq \sigma_{ij} < 1$, are satisfied for all $(i, j)$ pairs. If $\sigma_{ij} = 0$ for all $(i, j)$ pairs, the PCL collapses to the MNL.

Substituting equation Eq. (3) into Eq. (2) and using the transformation of $Y_i$ gives the probability of choosing alternative $i$ as

$$P_i = \frac{\sum_{j \neq i} e^{V_i/1-\sigma_{ij}} \left( e^{V_i/1-\sigma_{ij}} + e^{V_j/1-\sigma_{ij}} \right)^{-\sigma_{ij}}}{\sum_{k=1}^{n-1} \sum_{m=k+1}^{n} \left( e^{V_k/1-\sigma_{km}} + e^{V_m/1-\sigma_{km}} \right)^{-\sigma_{km}}}. \tag{4}$$

This expression can be rewritten as

$$P_i = \sum_{j \neq i} P_{ij} P_{ij}, \tag{5}$$

where

$$P_{ij} = \frac{e^{V_i/1-\sigma_{ij}}}{e^{V_i/1-\sigma_{ij}} + e^{V_j/1-\sigma_{ij}}}, \tag{6}$$

$$P_{ij} = \frac{\left( e^{V_i/1-\sigma_{ij}} + e^{V_j/1-\sigma_{ij}} \right)^{1-\sigma_{ij}}}{\sum_{k=1}^{n-1} \sum_{m=k+1}^{n} \left( e^{V_k/1-\sigma_{km}} + e^{V_m/1-\sigma_{km}} \right)^{-\sigma_{km}}}, \tag{7}$$

where $P_{ij}$ is the conditional probability of choosing alternative $i$ given the chosen binary pair $(i, j)$ and $P_{ij}$ is the unobserved probability for the pair $(i, j)$.

2.2. Behavioral interpretation

The PCL model has many potential applications in travel behavior analysis as well as other fields. For example, consider the choice situation among three overlapping routes as shown in Fig. 1. Route pairs (1,2) and (2,3) share common links but pair (1,3) does not. We expect some similarity of unobserved characteristics between paths 1 and 2 and between paths 2 and 3 due to

---

3 An alternative formulation of the PCL model (Chu, 1989); which adds $(1 - \sigma_{ij})$ as a multiplicative term for each pair of alternatives in the $G$ function (Eq. (3)) and, consequently, in the numerator and denominator terms in equation Eq. (7); obtained inferior goodness of fit results with the current data set.
the use of common links. Neither the MNL nor the NL can represent this structure of similarities. However, the PCL model, which allows differential similarities between all pairs of alternatives can be used to represent this route choice situation. The same issue may exist among urban or intercity modal alternatives where mode similarity may result from other common characteristics such as being private versus common carrier, exclusive use vehicles versus shared vehicles and/or air versus ground alternatives.

2.3. A structural comparison of the PCL and NL models

The MNL model is a restricted case of both the PCL model (when all its similarity parameters, $\sigma_{ij}$, equal zero) and the NL model (when all its similarity parameters, $\sigma_m$, equal zero). Neither the NL nor the PCL models are restricted cases of the other. The primary difference between these models is the manner in which they represent similarity between pairs of alternatives. In the PCL model, each pair of alternatives can take on a similarity relationship that is independent of the similarity relationship between other pairs of alternatives. In the NL model, all pairs of alternatives in a common grouping have the same similarity as all other pairs. In cases where the similarity structure of the PCL model is the same as the corresponding NL model, the models will be approximately, but not exactly, equivalent.

The PCL model can be formulated to closely replicate any NL structure by imposing restrictions among similarity parameters that match the similarity restrictions in the NL nest structure. For example, an NL model with a single nest in a two level structure (Fig. 2) can be approximated...
by a PCL model with positive similarity for pairs of alternatives in the NL nest (pair 1,2 in this case). Effectively, this leads to a two level structure with binary nests for all pairs included in the NL nest (1,2 in this case) and all other alternatives at the upper level (Fig. 3). However, in the PCL model each alternative appears once for its pairing with each other alternative (Fig. 4). As a result, the PCL cross-elasticities between pairs of alternatives include the relationships between the pair in its nest (determined by the value of the similarity parameter) and the relationship of each of these alternatives to the opposite alternative outside the nest (with zero similarity). This difference results in lower cross-elasticities for pairs of PCL alternatives with the same similarity and utility parameters as the corresponding NL model. This can be offset, to some extent, if the PCL similarity parameters are larger than the corresponding NL similarity parameters.  

Fig. 3. PCL model with similarity for single pair of alternatives.

Fig. 4. PCL model with similarity for single pair of alternatives showing those pairs of the nested alternatives which do not appear in the nest.

---

4 This property of the PCL model limits the maximum correlation between utilities to $1/(n-1)$, where $n$ is the number of alternatives.
2.4. Cross- and direct-elasticities

Another way to understanding the structure of the PCL model and its relationship to the MNL and NL models is by examination of direct- and cross-elasticities. These describe the effect of a change in the systematic utility of an alternative on the probability of choosing that alternative and the probability of choosing each other alternative, respectively. The direct- and cross-elasticities for the MNL, NL and PCL models are shown in Table 1.

The direct-elasticities for both the NL and PCL models are greater than for the MNL model for alternatives in the NL nest or which have positive similarity to any other alternatives in the PCL structure. These collapse to the MNL if the similarity parameter equals zero for the NL nest or for all pairs of alternatives in the PCL.

The cross-elasticities of the MNL model depend exclusively on the probability of the mode of change. These give the commonly observed equal proportional effect of the addition, deletion or change of any alternative on all other alternatives. The corresponding elasticities for the NL model are differentiated between pairs of alternatives that are in the same nest versus pairs which are not in a common nest. The elasticities for pairs of alternatives in which one or both is not in the nest are identical to those for the MNL. The cross-elasticity between pairs of alternatives which are in the same nest, $m$, is larger than for other pairs of alternatives. The magnitude of the cross-elasticity increases as $\sigma_m$, the similarity parameter, increases from zero; the cross-elasticity collapses to the MNL when $\sigma_m$ equals zero.

The cross-elasticities for the PCL model are more complex. If $\sigma_{ij}$ is equal to zero for any pair, the cross-elasticity for that pair is identical to that for the MNL model. The cross-elasticity increases in magnitude as $\sigma_{ij}$ increases from zero. The magnitude of the cross-elasticity is related to the probability of the $ij$ pair, $P_{ij}$, and the conditional probabilities of the alternatives in the nest.

2.5. The correlation of utilities between alternatives

A third perspective on the structure of the PCL model relative to the MNL and NL models is to examine the structure of covariances of the random errors between all pairs of alternatives. The

---

Table 1
Direct- and cross-elasticities of the MNL, NL and PCL models for change in attribute $k$ of alternative $i$, $X_{ik}$

<table>
<thead>
<tr>
<th></th>
<th>Direct-elasticity</th>
<th>Cross-elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNL model</td>
<td>$(1 - P_i)\beta_iX_{ik}$</td>
<td>$-P_i\beta_iX_{ik}$</td>
</tr>
<tr>
<td>NL model</td>
<td>$(1 - P_n)\beta_iX_{nk}$, $n$ not in nest</td>
<td>$-P_n\beta_iX_{nk}$</td>
</tr>
<tr>
<td></td>
<td>$(1 - P_n)\beta_iX_{nk}$, $n$ in nest $m$</td>
<td>$-P_n\beta_iX_{nk}$</td>
</tr>
<tr>
<td></td>
<td>$\left[ (1 - P_n)P_{nj</td>
<td>m} + \left( \frac{1}{1-\sigma_n} \right) (1 - P_{nh</td>
</tr>
<tr>
<td>PCL model</td>
<td>$\left{ \sum_{j\neq i} \frac{P_{ij}(P_{ij}(0))}{P_i} \left[ 1 - \sigma_{ij}(P_{ij}(0)) \right] - P_i \right} \beta_iX_{ik}$</td>
<td>$-P_i + \left( \frac{\sigma_{ij}}{1-\sigma_{ij}} \right) P_{ij}(P_{ij}(0)) \beta_iX_{ik}$</td>
</tr>
</tbody>
</table>
random error components, \( \epsilon_n \), in all three models have a marginal extreme value distribution. The distributions for alternatives in the MNL and NL models are:

\[
F_{\text{MNL}, \text{NL}}(\epsilon_n) = \exp\left[ -\exp(-\epsilon_n) \right].
\]

The corresponding marginal distribution for the PCL model is:

\[
F_{\text{PCL}}(\epsilon_i) = \exp\left\{ -(J - 1) \exp(-\epsilon_i) \right\},
\]

where \( J \) is the number of alternatives. The variance of this distribution is identical with that for the MNL and NL models.

The bivariate cumulative distribution function for any pair of random error components \( (\epsilon_i, \epsilon_j) \) is obtained by setting all other error components to infinity (Small, 1987; Daganzo and Kusnic, 1993) as

\[
F_{\text{PCL}}(\epsilon_i, \epsilon_j) = \exp\left\{ -\left[ \exp\left( \frac{-\epsilon_i}{1 - \sigma_{ij}} \right) + \exp\left( \frac{-\epsilon_j}{1 - \sigma_{ij}} \right) \right]^{1-\sigma_{ij}} \right\}
\times \exp\left\{ -(J - 2) \exp(-\epsilon_i) - (J - 2) \exp(-\epsilon_j) \right\}
= F_{\text{NL}}(\epsilon_i, \epsilon_j) \times F_{\text{PCL}}(\epsilon_i) \times F_{\text{PCL}}(\epsilon_j),
\]

where \( F_{\text{NL}}(\epsilon_i, \epsilon_j) \) is the bivariate c.d.f. for the NL model given by

\[
F_{\text{NL}}(\epsilon_i, \epsilon_j) = F_{\text{NL}}(\epsilon_{n/m}, \epsilon_{n'/m}) = \exp\left\{ -\left[ \exp\left( \frac{-\epsilon_{n/m}}{1 - \sigma_m} \right) + \exp\left( \frac{-\epsilon_{n'/m}}{1 - \sigma_m} \right) \right]^{1-\sigma_m} \right\},
\]

where \( \sigma_m \) is the similarity parameter for the nest including alternatives \( n \) and \( n' \), equivalent to \( i \) and \( j \), in nest \( m \) and is equal in this formulation to the corresponding \( \sigma_{ij} \) in the PCL. \( F_{\text{PCL}}(\epsilon_i) \) and \( F_{\text{PCL}}(\epsilon_j) \) are marginal extreme value distributions of alternatives \( i \) and \( j \), respectively. Because the PCL covariance is a negative function of the number of alternatives (\( J \) in Eq. (10)), the \( \sigma_{ij} \) in the PCL model must be greater than the corresponding \( \sigma_m \) in the NL model to obtain approximately the same value of correlation between the alternative utilities.

The covariance (or correlation) of utilities in the PCL model cannot be written in closed form but can be calculated numerically. This is analogous to Small’s (Small, 1987) observation for the OGEV model. The correlations between pairs of alternatives for the PCL model as a function of the similarity parameters and the number of alternatives in the choice set are reported in Table 2. The correlations increase with \( \sigma_{ij} \) and decrease with the number of alternatives. Further, as indicated earlier, the maximum utility correlation decreases with the number of alternatives. The corresponding correlations obtained from the NL model, for a range of similarity parameters, are reported in Table 3 for comparison. The values of the PCL similarity parameters to represent the same level of utility correlation between pairs of alternatives as implied by the NL model is illustrated in Table 4. The blanks in this table indicate that there is no value of the PCL similarity which will imply error correlation as large as that implied by the NL model similarity.

---

5 The alternatives and error terms for the MNL and NL models are indexed by \( n \) and \( n' \) and for the PCL model by \( i \) and \( j \).

6 The standard OGEV (Small, 1987) is a special case of the PCL model in which the correlations between alternatives are a function of their proximity in the ordering of alternatives.
2.6. Similarity parameter identification

The estimation in the PCL model involves utility and similarity parameters. The utility parameters are estimable and unique conditional on any set of similarity parameters (Daganzo and Kusnic, 1993). The number of similarity parameters, which can be uniquely estimated, is limited by the estimation structure for choice models; that is, the comparison between utilities for pairs of alternatives. The maximum number of similarity parameters is equal to the number of variance–covariance parameters that can be estimated for the MNP model. That is,

\[
\binom{n}{2} = \frac{n(n-1)}{2}
\]

similarity parameters can be identified out of the total number of possible parameters,

\[
\binom{n}{2} = \frac{n(n-1)}{2}
\]

Thus, it is necessary to impose at least one restriction on the similarity matrix. This can be done by setting one or more of the similarity parameters equal to zero, imposing constraints between similarity parameters or parameterizing the similarity parameters as functions of exogenous variables describing either the decision maker, the alternatives or both. For example, in the route choice case described earlier, the similarities can be formulated as a function of the portion of common links on different paths (Yai et al., 1997; Gliebe et al., 1999).

2.7. Estimation

Full information maximum likelihood must be used to estimate the utility function and similarity parameters because the choice of pairs of alternatives cannot be observed. The maximization is based on the general choice model log-likelihood function:
\[ L = \sum_q \sum_i \delta_{qi} \log P_{qi}, \]  

(12)

where \( \delta_{qi} \) is 1 if individual \( q \) chooses alternative \( i \) and 0 otherwise, and \( P_{qi} \) the estimated probability that individual \( q \) chooses alternative \( i \). Since the hessian of the log-likelihood function for the PCL model is not negative semi-definite over its entire range; repeated optimization with different starting values may be required to locate the global optimum. However, a local maximum, which is superior to the MNL model, is guaranteed (Chu, 1989). As for the NL model, PCL estimations can be started from the estimated MNL parameters with zero values for the similarity coefficients.

The compatibility of the PCL model with random utility maximization requires that the similarity parameters lie within the unit interval. This is accomplished by use of a constrained maximize log-likelihood function with each similarity parameter constrained to the unit interval; we do this using the GAUSS constrained maximum likelihood module (Aptech Systems, 1995).

One of the advantages of this approach is that use of the PCL model to identify similarities between alternatives avoids the tedious search among numerous NL nesting structures.

3. Empirical analysis

The data used in this study was assembled by VIA Rail in 1989 to estimate the demand for high-speed rail in the Toronto–Montreal corridor and to support future decisions on rail service improvements in the corridor (KPMG Peat Marwick and Koppelman, 1990). The data set contains four intercity travel modes of interest (air, train, bus and car). However, the small number of bus users, 18, makes it impossible to estimate reliable bus-specific parameters (both the three similarity parameters between bus and the other alternatives, the alternative specific constant and any alternative specific variables). Thus, we consider only those travelers who chose air, train or car and further limit the study to the 2769 travelers who have all three modes available to them. The distribution of choices in the data set is: train (463, 16.72%), air (1039, 37.52%), and car (1267, 45.76%).

The estimation results for the MNL, PCL and NL models are reported in Table 5. The utility function specification includes mode-specific constants, frequency, travel cost, and out-of-vehicle and in-vehicle travel times. \(^7\) Three PCL models had all the similarity parameters within the 0–1 range and were significantly different from the MNL. These are the PCL models with similarity parameters for the train–car pair only (column 2), for the air–car pair only (column 3) and for both the air–car and train–car pairs (column 4). All three models reject the MNL at high levels of significance and the PCL model with both similarity parameters rejects both of the other PCL models, also at a high level of significance.

\(^7\) A specification including selected alternative specific variables gives approximately equivalent results with respect to the PCL similarity coefficients.
Two NL models had nesting parameters in the 0–1 range and rejected the MNL model; these are train and car in a nest (column 5) and air and car in a nest (column 6). Comparison of the PCL and NL models with common similarity structure (train–car similarity in columns 2 and 5 and air–car similarity in column 3 and 6) provides useful insights. First, the PCL

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimated parameters (standard errors)</th>
<th>MNL model</th>
<th>PCL models</th>
<th>NL models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Train–car similarity</td>
<td>Air–car similarity</td>
<td>Train–car and air–car similarity</td>
</tr>
<tr>
<td>Mode constants</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Air</td>
<td>3.6643</td>
<td>2.9361</td>
<td>2.7558</td>
<td>1.9603</td>
</tr>
<tr>
<td></td>
<td>(0.431)</td>
<td>(0.321)</td>
<td>(0.398)</td>
<td>(0.396)</td>
</tr>
<tr>
<td>Train</td>
<td>1.6728</td>
<td>1.7782</td>
<td>1.2588</td>
<td>1.2942</td>
</tr>
<tr>
<td></td>
<td>(0.225)</td>
<td>(0.135)</td>
<td>(0.204)</td>
<td>(0.164)</td>
</tr>
<tr>
<td>Car (base)</td>
<td>Frequency</td>
<td>0.0944</td>
<td>0.0926</td>
<td>0.0683</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Travel cost</td>
<td>−0.0461</td>
<td>−0.0388</td>
<td>−0.0311</td>
<td>−0.0242</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>In-vehicle time</td>
<td>−0.0099</td>
<td>−0.0084</td>
<td>−0.0083</td>
<td>−0.0076</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Out-of-vehicle time</td>
<td>−0.0426</td>
<td>−0.0387</td>
<td>−0.0358</td>
<td>−0.0321</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Similarity parameters</td>
<td>Train–car</td>
<td>–</td>
<td>0.8788</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td>Air–car</td>
<td>–</td>
<td>–</td>
<td>07057</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
<td>(0.082)</td>
<td>(0.082)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>At convergence</td>
<td>−1919.8</td>
<td>−1912.1</td>
<td>−1911.7</td>
</tr>
<tr>
<td></td>
<td>At market share</td>
<td>−2837.1</td>
<td>−2837.1</td>
<td>−2837.1</td>
</tr>
<tr>
<td></td>
<td>At zero</td>
<td>−3042.1</td>
<td>−3042.1</td>
<td>−3042.1</td>
</tr>
<tr>
<td>L’hood ratio index</td>
<td>versus market share</td>
<td>0.3233</td>
<td>0.3260</td>
<td>0.3262</td>
</tr>
<tr>
<td></td>
<td>versus zero</td>
<td>0.3689</td>
<td>0.3715</td>
<td>0.3716</td>
</tr>
<tr>
<td>Value of time</td>
<td>In-vehicle time</td>
<td>C$13</td>
<td>C$13</td>
<td>C$16</td>
</tr>
<tr>
<td></td>
<td>Out-of-vehicle time</td>
<td>C$55</td>
<td>C$60</td>
<td>C$69</td>
</tr>
<tr>
<td>Log likelihood test</td>
<td>versus MNL</td>
<td>15.4 &gt; 3.8</td>
<td>16.2 &gt; 3.8</td>
<td>31.8 &gt; 6.0</td>
</tr>
<tr>
<td>Implied utility correlation</td>
<td>Train–car</td>
<td>0.000</td>
<td>0.448</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Air–car</td>
<td>0.000</td>
<td>0.000</td>
<td>0.422</td>
</tr>
</tbody>
</table>
similarity parameters and the NL nest parameters (which are very different) imply similar levels of utility correlation between alternatives (Table 5) and, second, the implied values of time are almost identical. These results are consistent with the discussion of the relationship between the NL and a PCL that has been constrained to match the NL structure as closely as possible.

Finally, the PCL model with similarity parameters for both air and car and for train and car (column 4), a structure that cannot be incorporated in the NL model, yields the highest log-likelihood and statistically rejects all the other models. These results demonstrate the statistical and structural superiority of the PCL model. The direct- and cross-elasticities of the MNL, NL and PCL models in response to a change in rail service (Tables 6 and 7) indicate the magnitude of prediction differences that would be obtained through use of the MNL or either NL model rather than the more complete PCL model. The direct elasticities for train (Table 6) differ substantially across the four models presented. This is highlighted by the ratios of the elasticities of each model to the PCL model elasticities, which show some dramatic differences in train direct-elasticities across models. The NL model, with the train–car nest overestimates the direct-elasticities while the model with the air–car nest underestimates the direct elasticities. This indicates the sensitivity of the choice of NL structure that must be made even if the differences in goodness of fit between (among) two (or more) NL structures are very small, as in this case.

The cross-elasticities of the air and car modes to changes in train service (Table 7) also differ substantially across models. The most interesting comparison is the relative cross-elasticities for the two different modes given by each model. The PCL model shows smaller cross-elasticities for

---

Table 6
Direct-elasticity of train ridership in response to improvements in train service

<table>
<thead>
<tr>
<th>Train level of service attribute</th>
<th>MNL model</th>
<th>NL model with train–car nested</th>
<th>NL model with air–car nested</th>
<th>Preferred PCL model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part a: Elasticity values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>0.333</td>
<td>0.401</td>
<td>0.260</td>
<td>0.312</td>
</tr>
<tr>
<td>Cost</td>
<td>2.068</td>
<td>2.331</td>
<td>1.427</td>
<td>1.473</td>
</tr>
<tr>
<td>In-vehicle time</td>
<td>1.788</td>
<td>2.039</td>
<td>1.685</td>
<td>1.862</td>
</tr>
<tr>
<td>Out-of-vehicle time</td>
<td>2.927</td>
<td>3.336</td>
<td>2.757</td>
<td>2.993</td>
</tr>
<tr>
<td><strong>Part b: Ratio of MNL and NL elasticities to PCL elasticities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>1.067</td>
<td>1.285</td>
<td>0.833</td>
<td>1.000</td>
</tr>
<tr>
<td>Cost</td>
<td>1.404</td>
<td>1.582</td>
<td>0.969</td>
<td>1.000</td>
</tr>
<tr>
<td>In-vehicle time</td>
<td>0.960</td>
<td>1.095</td>
<td>0.905</td>
<td>1.000</td>
</tr>
<tr>
<td>Out-of-vehicle time</td>
<td>0.978</td>
<td>1.115</td>
<td>0.921</td>
<td>1.000</td>
</tr>
</tbody>
</table>

---

8 The somewhat low value of intercity IVT and high value of OVT for business travelers can be addressed in all models by imposition of constraints between the IVT and OVT parameters.
9 Using the non-nested hypothesis test (Horowitz, 1983) between this PCL model and both NL models.
10 The emphasis in this table is placed on changes in train service because the original study was directed toward consideration of major improvements in rail service in the corridor.
air than for car reflecting the estimated similarity between car and train. The MNL model and the second NL model, air–car nested, have identical cross-elasticities for air and car, as expected. The NL model with train and car nested, shows a much greater effect on car than on air; again, resulting from train and car being in the same nest.

Overall, these results indicate that the differences in the estimated models may lead to substantial differences in forecasts of train ridership under proposed futures resulting in possible differences about whether and what improvements to make.

### 4. Conclusions

The PCL model, which allows different similarity, utility correlation and substitution between each pair of alternatives, adds useful flexibility to the family of logit models by providing relaxation of the independence between alternatives property beyond that provided by the NL model. Constrained full information maximum likelihood can be used to simultaneously estimate all model parameters taking account of necessary constraints on the similarity coefficients to ensure that the model satisfies the conditions of random utility maximization.

The PCL model significantly rejects the MNL and the best NL models in an application to intercity mode choice among three alternatives. Further, its use will produce forecasts which are importantly different with respect to both changes in ridership on the alternative under study and the alternatives from which those riders are likely to be drawn. These differences are substantial enough to lead to the adoption of different policy and implementation decisions.
Acknowledgements

The authors wish to acknowledge Chaushie Chu, Chandra Bhat, Christopher Forinash and unidentified referees for valuable comments and suggestions on earlier versions of this paper. This research was supported, in part, by NSF Grant DMS-9313013 to the National Institute of Statistical Sciences and, in part, by a Dissertation Year Fellowship to the second author from the Transportation Center, Northwestern University.

References


