2007

Conditional Logit, IIA, and Alternatives for Estimating Models of Interstate Migration

Christiadi

Brian Cushing

Conditional Logit, IIA, and Alternatives for Estimating Models of Interstate Migration

Christiadi and Brian Cushing

RESEARCH PAPER 2007-4

Christiadi
Department of Economics
University of the Pacific
3601 Pacific Avenue
Stockton, CA 95211
E-Mail: cchristiadi@pacific.edu

Brian Cushing
Department of Economics
West Virginia University
PO BOX 6025
Morgantown, WV 26506-6025
E-Mail: brian.cushing@mail.wvu.edu

Paper presented at the 46th annual meeting of the Southern Regional Science Association, Charleston, SC, March 29-31, 2007

Abstract: Many researchers have used the conditional logit model to examine migration. One common objection to this model is that it carries the independence from irrelevant alternatives (IIA) assumption, which may be too restrictive. This study compares the conditional logit with models that partially relax (nested logit) or fully relax (mixed logit) the IIA assumption. We will begin to learn whether assuming IIA holds poses serious estimation problems for migration modeling. Given the substantial computational cost of the more complex models, a finding that a well-specified, but computationally much simpler, conditional logit model may suffice would be useful.

1. Introduction

The increasing availability of individual data and the rapid advancement in computer technology have permitted researchers to analyze migration in new ways (see Cushing and Poot (2004)). In this respect, many recent migration studies have used the conditional logit model, which examines migration choice as a multinomial discrete choice. Unlike an aggregate migration model, the conditional logit model can focus on individuals, thus better representing migration as an individual's utility maximization decision. Moreover, this model allows analyses that are not possible with aggregate models, such as incorporating individual characteristics as explanatory variables and computing cross elasticities of choosing among alternatives.

The conditional logit model has only seen limited application in the migration literature. Mueller (1985) was among the first to apply a conditional logit model to migration, when he examined an individual's destination choice among states. Probably because of the state of computer technology, the conditional logit model did not resurface substantially in the migration literature for more than 15 years, until studies such as Davies et al. (2001).

The main concern about the conditional logit model is its assumption of independence from irrelevant alternatives (IIA). This assumption implies that the probability ratio of individuals choosing between two alternatives does not depend on the availability or attributes of the other alternatives. This assumption may be realistic in some situations. For example, people who move for a job transfer typically have fixed their destination, and retirees may consider only one or two possible destinations in which they want to live. For these people, any changes in the other destinations will not significantly affect their choice decision. In general, however, the IIA assumption is too restrictive, especially when the number of alternatives in the choice set is large, such as a in model of state destination choice for the United States.[1]

Violating IIA may lead a model to incorrectly predict the probability of destinations being chosen. The model may overestimate the probability of choosing California, while at the same time underestimating the probability of choosing another state. In light of this problem, several

[1] Statistically, the larger the number of alternatives, the higher the likelihood of finding at least one restricted model (excludes one or more alternatives), that is significantly different from the unrestricted model, which includes all alternatives. Thus the easier it is to violate the IIA assumption.

models have been developed to relax the IIA assumption, including nested logit, mixed logit, multinomial probit, and heteroscedastic extreme value models. They are more computationally complex than the conditional logit model, making them more difficult to estimate, which in turn requires more computer time and often results in a breakdown of the estimation procedure.

This study applies two of the above models: nested logit and mixed logit. In this study, while it took about 1.5 minutes for the conditional logit model to converge, it took more than 30 minutes to run the nested logit model, and nearly 10 hours to run the mixed logit model.[2] This essay examines to what extent the outcomes of these two models differ from those of the conditional logit model. Based on the comparison, this study then assesses whether relaxing the IIA assumption warrants the application of the more complex nested logit or mixed logit models.

The next section compares various discrete choice models, followed by a more detailed discussion of the nested logit and mixed logit models. Later sections describe the econometric specification applied in this study, then analyze empirical results, comparing how the outcomes from the nested logit and mixed logit models differ from those of the conditional logit model.

2. Discrete Choice Models

Discrete choice models are based on utility maximization. In a destination choice model, this means that the chosen destination must give the individual greater utility compared with other destinations. If the utility of individual i choosing state j is represented as U_{ij}, then location j will be chosen if and only if $U_{ij} > U_{il}$ for $j \neq l$.

Because researchers do not know U_{ij}, the individual's true utility, they cannot tell for sure which destination an individual will eventually choose. U_{ij} consists of two components, the observable and the unobservable components:

$$U_{ij} = V_{ij} + \varepsilon_{ij}. \qquad (1)$$

[2] Models that relax the IIA assumption other than these two are computationally more burdensome, thus have a significantly longer convergence time. Dahlberg and Eklöf (2003) found the convergence time for their multinomial probit models to be significantly longer than for their mixed logit models. This study attempted to apply a heteroscedastic extreme value model but it continually failed to converge.

U_{ij} consists of a predicted utility, V_{ij}, observable based on the choice's attributes, and an unobserved random component, ε_{ij}. If ε_{ij} were known, researchers would know U_{ij} and could tell for sure which destination would be chosen. Since researchers do not know ε_{ij}, the best they can do is predict the final outcome in terms of probability.

The probability of individual *i* choosing state *j* can be described as:

$$
\begin{aligned}
P_{ij} &= P(U_{ij} > U_{il}) \\
&= P((V_{ij} + \varepsilon_{ij}) > (V_{il} + \varepsilon_{il}) \\
&= P((\varepsilon_{il} - \varepsilon_{ij}) < (V_{ij} - V_{il})) \text{ for all } j \neq l. \qquad (2)
\end{aligned}
$$

To solve Equation (2) the researcher must impose a probability density function on ε_{ij}. Each type of probability distribution imposed on ε_{ij} leads to a different discrete choice model, as shown in Table 1.

Conditional Logit Model

The conditional logit model assumes that ε_{ij} exhibits the extreme value distribution. The probability density function takes the following form:

$$f(\varepsilon_{ij}) = e^{-\varepsilon ij}\, e^{-e^{-\varepsilon ij}} \qquad (3)$$

and its cumulative density function is expressed as:

$$F(\varepsilon_{ij}) = e^{-e^{-\varepsilon ij}} \qquad (4)$$

More importantly, this model restricts all ε_{ij} to be independent and identically distributed (iid). The probability of individual *i* choosing destination *j* can be solved as a closed-form expression of:

$$P_{ij} = \frac{e^{Vij}}{\Sigma_j\, e^{Vij}} = \frac{e^{\alpha' Zij}}{\Sigma_j\, e^{\alpha' Zij}} \qquad (5)$$

Z_{ij} represents all the observed factors or explanatory variables and α represents parameters obtained from the model.

With ε_{ij} being iid, Equation (5) imposes the IIA assumption. Consider the probability that individual i chooses state j versus state l:

$$P_{ij} = \frac{e^{\alpha' Z ij}}{\Sigma_j \, e^{\alpha' Z ij}} \quad ; \text{and } P_{il} = \frac{e^{\alpha' Z il}}{\Sigma_j \, e^{\alpha' Z ij}}$$

The probability ratio of choosing between j and l is:

$$\frac{P_{ij}}{P_{il}} = \frac{\Sigma_j \, e^{\alpha' Z ij}}{\Sigma_j \, e^{\alpha' Z ij}} \, / \, \frac{e^{\alpha' Z ij}}{e^{\alpha' Z il}} = \frac{e^{\alpha' Z ij}}{e^{\alpha' Z il}} \tag{6}$$

The probability ratio depends only on the attributes of j and l, and does not depend on the attributes of other destinations.

Nested Logit Model

A nested logit model relaxes the IIA assumption by allowing the unobserved factors, ε_{ij}, to be correlated. First, a nested logit model partitions choices into different subsets. Based on the partition, a nested logit model then allows ε_{ij} to have the same correlation within a nest, but maintains independence across nests. In other words, a nested logit partially relaxes the IIA assumption by maintaining IIA for choices within the same nest, but relaxing it for choices across nests.

Let the set of all alternatives j be partitioned into K subsets. Each subset is called a nest and denoted B. Thus, all the alternatives are partitioned into B_1, B_2, B_3…., and B_K, and each j is now an element of B_k, where k goes from 1 to K.

The utility of individual i choosing destination j is the same as that in the regular conditional logit model:

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

To incorporate the nesting, the observed utility can be represented as consisting of two components: (1) A, which is constant for all alternatives within a nest but varies across nests; and (2) B, which varies over all alternatives within a nest. That is,

$$U_{ij} = A_{ik} + B_{ij} + \varepsilon_{ij}, \text{ where } j \epsilon B_k. \qquad (7)$$

A_{ik} depends only on variables that describe nest k and B_{ij} depends on variables that describe the alternative j.

The nested logit model assumes that the ε_{ij} exhibit the generalized extreme value distribution with a cumulative joint distribution function described as:

$$F(\varepsilon_{ij}) = \exp(-\sum_{k=1}^{K} \{ \sum_{j \epsilon B_k}^{K} e^{-(\varepsilon ij/\lambda k)} \}) \qquad (8)$$

Equation (8) shows that the choices are partitioned into K subsets of B_k. λ_k is a parameter indicating the degree of substitutability between unobserved utility among choices in different nests. When λ_k equals one, choices across nests are statistically independent, thus nesting becomes unnecessary. In that case, the cumulative distribution of ε_{ij} (Equation (8)) collapses to that of a conditional logit model (Equation (5)).

With ε_{ij}'s cumulative distribution function following Equation (8), the probability of individual i choosing destination j can be solved as a closed-form expression of:

$$P_{ij} = \frac{e^{Vij/\lambda k} (\Sigma\, j \epsilon B_k\, e^{Vij/\lambda k})^{\lambda k-1}}{\sum_{k=1}^{K} (\Sigma\, j \epsilon B_k\, e^{Vij/\lambda k})^{\lambda k}} \qquad (9)$$

The probability ratio of individual i choosing between choice j and l is

$$\frac{P_{ij}}{P_{il}} = \frac{e^{V_{ij}/\lambda_k} (\Sigma\, j \epsilon B_k\, e^{V_{ij}/\lambda_k})^{\lambda_k - 1}}{e^{V_{il}/\lambda_d} (\Sigma\, j \epsilon B_d\, e^{V_{ij}/\lambda_d})^{\lambda_d - 1}} \tag{10}$$

Equation (10) shows that the probability ratio depends not only on attributes of choices *j* and *l* but also on those of the other choices. If both choices belong to the same nest (that is k=d), the probability ratio becomes:

$$\frac{P_{ij}}{P_{il}} = \frac{e^{V_{ij}}}{e^{V_{il}}} = \frac{e^{\alpha' Z_{ij}}}{e^{\alpha' Z_{il}}} \tag{11}$$

The ratio in Equation (11) depends only on the characteristics of choices *j* and *l*, which is the property of the IIA assumption (the same as Equation (5)).

Nesting Pattern in a Nested Logit Model

Developing the nesting pattern is an important element of a nested logit application. When researchers successfully set an acceptable nesting pattern, they obtain more information about the individuals' choice decisions than what they get from conditional logit parameter estimates. As an illustration, consider the nesting pattern to be used in this study, as shown in Figure 1.

The model nests destinations into two groups: warm states and cold states. Successfully running this nested logit model yields two sets of information. First, we learn how each explanatory variable affects the probability of a particular destination being chosen. Second, we learn something about unobserved factors that correlate with warmness of the destination. These unobserved factors are not yet captured by the variable, temperature, which is used to distinguish warm states from cold states. These unobserved factors could include items such as the love for (or desire to avoid) snow or preference for specific seasonal recreational activities.

Looking at the nesting pattern, researchers are often tempted to interpret the nesting as a representation of sequential choice decisions. Such an interpretation means that in choosing their destination, individuals would first make a selection based on a key attribute, which in this case is the warmness of the state. Afterwards, they choose their ultimate state destination only from the group they have selected in the first stage. This implies that states like Alabama or Arizona are not available once an individual selects cold states. That is incorrect. Regardless of their preference towards either cold or warm states, individuals still have some probability of choosing any of the states. It is true, however, that a strong preference toward cold states implies a lower probability of choosing any of the warm states. Thus, while nesting may seem to model sequential decision-making in some instances, it is not generally intended to represent sequential decisions.

The purpose of nesting is simply to categorize choices. When researchers suspect that the unobserved factors of a certain group of choices are correlated, they can categorize them into the same nest. The nesting basically puts alternatives with similar attributes into the same nest. The nesting pattern employed depends on the researcher's judgment, which could be based on natural consideration, casual observation, or theory. As long as the researcher suspects the error terms among certain choices are correlated, he can exercise the nesting accordingly (Hensher (1986), Train et al. (1987), and Train (2003)).[3] Since researchers may see different ways of how unobserved factors correlate, they may come up with more than one nesting pattern for the same choice decision, which in turn can yield different results.[4]

The data determine whether or not a nesting is appropriate. A nesting pattern is acceptable when the parameter, λ, falls between 0 and 1. The value of $1-\lambda$ measures the correlation among the unobserved components of utility within a nest (Train (2003)). When the value of λ falls between 0 and 1, the model is consistent with utility maximization for all possible values of explanatory variables (Ortuzar and Willumsen (1994) and Train (2003)). In this case, an improvement in the destination attribute will increase the probability of that destination being chosen. On the contrary, a negative λ is not consistent with utility maximization since it implies that an improvement in the destination attribute decreases the probability of that destination being chosen. If λ is greater than 1, an increase in the utility of a destination in the nest not only increases its selection probability but also the selection probability of the rest of the states in the

[3] Train et al. (1987) explains "As in all nested logit models, the direction of conditionality reflects the correlations among unobserved factors across alternatives; as such it arises from patterns in the researcher's lack of information, rather than from the household's decision process" (p113).
[4] The discrete choice literature has not developed a well-defined methodology to determine which of the nesting patterns best represents reality (Green (2000)).

same nest (Ortuzar and Willumsen (1994)). In this case, the model is only consistent with utility maximization for a certain range of values of explanatory variables (for details, see Herriges and Kling (1996), Kling and Herriges (1995), Lee (1999), and Train et al. (1987)). Ortuzar and Willumsen (1994) state that such a range is rarely available.

Mixed Logit Model

Unlike a nested logit, a mixed logit model fully relaxes the IIA assumption. This model is similar to a conditional logit model except that it allows parameter estimates to vary across individuals. Consider the utility function expressed in Equation (1):

$$
\begin{aligned}
U_{ij} &= V_{ij} + \varepsilon_{ij} \\
&= \alpha' Z_{ij} + \varepsilon_{ij}
\end{aligned}
$$

Like a conditional logit, a mixed logit assumes the error terms, ε_{ij}, are iid with extreme value distribution. A mixed logit, however, relaxes the restriction that α is the same for each individual, allowing it to be stochastic instead. In a mixed logit model, the person's utility is

$$U_{ij} = \alpha_i' Z_{ij} + \varepsilon_{ij}\,, \qquad (12)$$

where α now differs across individuals. Researchers can estimate U_{ij} if they know the probability density function (pdf) for α. Researchers can impose a certain type of distribution for α (e.g., normal, lognormal, uniform).[5] When α is assumed to be the same for all individuals, the probability of individual *i* choosing state *j* (P_{ij}) can be described exactly as in Equation (5), used for the conditional logit model. When α is not fixed, the probability of individual *i* choosing destination *j* (in this case, labeled as M_{ij}) can be estimated by estimating P_{ij} over all the possible values of α.

$$M_{ij} = \int P_{ij}(\alpha)\, f(\alpha)\, d\alpha$$

[5] When α is restricted to have a normal density function, the model becomes a close approximation of the multinomial probit model [Train (2003)].

$$= \int \left(\frac{e^{\alpha' Z_{ij}}}{\Sigma_j \, e^{\alpha' Z_{il}}} \right) f(\alpha) \, d\alpha \qquad (13)$$

Thus, a mixed logit probability is a weighted average of the logit formula evaluated at different values of α, with the weights given by the density, $f(\alpha)$. This equation is a multi-dimensional integral so that it does not have a closed-form solution. Consequently, it must be solved through simulation.

Another way to look at a mixed logit model is by representing the utility as an error component specification. α can be perceived as consisting of its mean, a, and a deviation around the mean, ξ, which differs across individuals. That is,

$$\begin{aligned} U_{ij} &= \alpha' Z_{ij} + \varepsilon_{ij} \\ &= (a + \xi_i) Z_{ij} + \varepsilon_{ij} \\ &= a' Z_{ij} + \xi_i' Z_{ij} + \varepsilon_{ij} \qquad (14) \end{aligned}$$

In this case, the ε_{ij} are still assumed to be iid. The unobserved components of utility are $\eta_{ij} = \xi_i' X_{ij} + \varepsilon_{ij}$. In the conditional logit model, $\xi_i' Z_{ij}$ are identically zero, implying no correlation in utility across alternatives. With nonzero error components, $\xi_i' Z_{ij}$, utility becomes correlated across alternatives, which relaxes the IIA assumption.

Now that we understand the mathematical representation of a mixed logit model, one may ask what this implies in real life. By allowing α to vary across individuals, a mixed logit model can represent variations in individuals' utility functions. Each individual now has his or her own value of α, implying that each person can have different weights for each destination attribute. A mixed logit model incorporates taste variations that exist across individuals.

Multinomial Probit Model

The multinomial probit model assumes that the vector of ε_{ij} , labeled ε_i, follows a multivariate normal distribution with covariance matrix Ω. That is,

$$\varepsilon_i \sim N(0, \Omega), \text{ with } \Omega = I_J * \textstyle\sum, \text{ for } j = 1, \ldots, J. \qquad (15)$$

In this case, I is an identity matrix, and $\sum$ is the covariance of ε_i or $E(\varepsilon_{ij}\,\varepsilon_{il})$. The density of ε_i is

$$f(\varepsilon_i) = \frac{1}{(2\pi)^{J/2}|\Omega|^{1/2}} e^{-1/2\varepsilon'_i\,\Omega^{-1}\varepsilon_i} \qquad (16)$$

For example, for $J = 3$, Ω is

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

If Ω is a diagonal matrix, that is σ_{jl} are zeros for all $j \neq l$, then the ε_{ij} are independent or uncorrelated. If all the nonzero components of Ω have the same value, then the ε_{ij} are identical or *homoscedastic*. A multinomial probit model does not have either of these restrictions, thus it fully relaxes IIA. Under this assumption, the probability of individual *i* choosing destination *j* can be expressed as:

$$\begin{aligned} P_{ij} &= \text{Prob}(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il}) \text{ for all } j \neq l \\ &= \int \boldsymbol{I}(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il} \text{ for all } j \neq l)\, f(\varepsilon_i)\, d\varepsilon_i. \qquad (17) \end{aligned}$$

$\boldsymbol{I}(.)$ is an indicator of whether the statement in the bracket is accepted or rejected, and the probability for an alternative being chosen is the integral of the conditions over all the values of ε_{ij} Since the components of Ω are not independent, Equation (17) is a J-dimensional integral.[6] Here lies the drawback of a multinomial probit model. Relaxing IIA entails significant computational burden, especially with a large number of alternative choices. Estimating a multinomial probit model must rely on simulation.

[6] In the estimation, the multinomial probit model measures utility in terms of differences in utility rather than level of utility. The error term is also represented as differences in the errors. By default, the difference of two normally distributed variables also has a normal distribution. Thus, instead of a *J*-dimensional integral, the estimation of a multinomial probit model is a *J-1* dimensional integral.

Researchers can also allow α to vary across individuals rather than be fixed. In this case, the normal distribution is imposed on α with a mean, a, and a deviation around the mean, ξ, which differs across individuals. Like the mixed logit model, this multinomial probit model can represent taste variations in individuals' utility functions.

$$\begin{aligned} U_{ij} &= \alpha' Z_{ij} + \varepsilon_{ij} \\ &= (a + \xi_i) Z_{ij} + \varepsilon_{ij} \\ &= a' Z_{ij} + \eta_{ij} \end{aligned} \tag{18}$$

Equation (18) is equivalent to Equation (14) of the mixed logit model, with $\eta_{ij} = \xi_i' Z_{ij} + \varepsilon_{ij}$. The main difference is that in a multinomial probit model, the η_{ij} are restricted to follow a joint normal distribution, while in a mixed logit, the ε_{ij} are restricted to be iid with extreme value distribution, but $\xi_i' Z_{ij}$ are allowed to follow any kind of probability distribution. When researchers restrict α in a mixed logit to follow the normal distribution, the model becomes a close approximation of the multinomial probit model (Dahlberg and Eklöf (2003) and Train (2003)).

Heteroscedastic Extreme Value Model

Like the conditional logit model, a heteroscedastic extreme value model restricts the errors, ε_{ij}, to follow an extreme value distribution. Unlike the conditional logit model, it assumes that the ε_{ij}, while independent, are heteroscedastic (not identical). This assumption fully relaxes IIA. Bhat (1995) argued that non-identical error variance is more realistic than identical variance. He used the transportation mode model to illustrate this point. If the unobserved factor in choosing the best transportation is the individual's level of comfort, then the variance of comfort from taking the train must differ from that of taking an automobile.

More specifically, Bhat (1995) restricted the density function for ε_{ij} to take the form of:

$$f(\varepsilon_{ij}) = \frac{1}{\theta_j} e^{-\varepsilon_{ij}/\theta_j} e^{-e^{-\varepsilon_{ij}/\theta_j}} \tag{19}$$

The θ_j are scale parameters that allow the value of ε_{ij} to differ across alternatives. If the θ_j are the same for all alternatives, then the model reverts to the conditional logit model.

Given this distribution, the probability of individual i choosing destination j is

$$
\begin{aligned}
P_{ij} &= \text{Prob}(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il}) \text{ for all } j \neq l \\
&= \text{Prob}(\varepsilon_{ij} > V_{il} - V_{ij} + \varepsilon_{il}) \text{ for all } j \neq l \\
&= \int \prod_{ij} \Lambda \left(\frac{V_{il} - V_{ij} + \varepsilon_{il}}{\theta_j}\right) \frac{1}{\theta_j} \lambda \left[\frac{\varepsilon_{ij}}{\theta_j}\right] d\varepsilon_{ij} \qquad (20)
\end{aligned}
$$

In this case, $\lambda(.)$ and $\Lambda(.)$ represent the density and cumulative density function of the extreme value distribution. That is,

$$\lambda(x) = e^{-x} e^{-e^{-x}}; \qquad \text{and } \Lambda(x) = e^{-e^{-x}}$$

Unlike a mixed logit or a multinomial probit, a heteroscedastic extreme value model requires only a one-dimensional integral. The estimation of this model is not stable, however, and computation of its likelihood function requires numerical integration (SAS Institute (2002)).

In summary, a researcher's distributional assumptions regarding the error components determine which discrete choice model is applied. Researchers can impose two types of characteristics on the error components: (1) their iid property and (2) their distribution function. Assuming the error components to be correlated, non-identical, or both relaxes the IIA assumption. Assumptions regarding the distribution function of the error components determine which discrete choice model should be used.

These models not only differ in terms of their mathematical representations. Different sets of characteristics of the error components imply different practical meaning. A mixed logit or a multinomial probit model allows choice to reflect taste variation. A nested logit model allows choice decisions to be categorized into different nests with each nest containing choices with a similar attribute. A heteroscedastic extreme value model allows choice decisions for which the variance of the unobserved factors of one alternative is different from the other alternatives. Three of the five possible models are successfully applied in this study: conditional logit, nested logit, and mixed logit models. The next section discusses the econometric specification for these three models.

3. Application of the Nested Logit and Mixed Logit Models to Migration

The conditional logit model has now been successfully applied in many migration studies, so that its use and interpretation are reasonably well understood. Application of models that relax IIA, such as the nested logit and mixed logit models estimated in this paper have been scarce in the migration literature, leaving much less understanding of their properties, problems, and interpretation.

A nested logit model has a few desirable properties: (1) it has a closed form solution; (2) it is computationally easier than other models that relax the IIA assumption; and (3) its extension from the conditional logit model can reveal an appealing story on how individuals make location choices. Newbold (1997), Frey et al. (1996), and White and Liang (1998) were among the first to apply nested logit models to migration research. However, they applied a limited instead of a full information nested logit model, by estimating the model sequentially. White and Liang (1998), for instance, first estimated a binomial logit model of move versus don't-move, and then estimated the movers' destination choice using a conditional logit model. This sequential nested logit model yields consistent but inefficient estimates (Green (2000)). Moreover, Hensher (1986) showed that its results are not comparable with those of the conditional logit model because they are derived from different sample sets. For example, White and Liang (1988) estimated the destination choice model only for those who moved, thus omitting nonmovers. A full information nested logit model would estimate the two decisions simultaneously, using the whole sample. This study applies a full-information nested logit model. Though requiring a more complex computer application, this model yields efficient and consistent estimates (Green (2000) and Hensher (1986)). In addition, the results can meaningfully be compared with those of the conditional logit model (Hensher (1986)). Knapp et al. (2001) were among the first to apply a full information nested logit model in a migration study. Unlike Knapp et al. (2001), whose two-branch and one-level nested logit model examines a choice among two alternatives, this study examines a choice of 48 alternatives. Given the complexity of the model, a valid empirical result of this nested logit model would by itself make advance the migration literature.

Nested logit models partially, but not fully, relax the IIA assumption. This essay also considers a mixed logit model, which fully relaxes IIA. Compared with a nested logit model, a mixed logit model is even less recognized in the migration literature, as well as in most other literatures. Mainly, this is because a mixed logit requires an evaluation of multiple integrals rather than a single integral. Moreover, since it does not have a closed form solution, a mixed logit model must be estimated through simulation. Only with major improvements in computer speed and in the understanding of simulation methods could we fully utilize this model. Dahlberg and Eklöf (2003) applied a mixed logit model to a study on intra-metropolitan migration. Examining migrants' choice among municipalities within a single metropolitan area, they compared conditional logit with mixed logit and multinomial probit models. They concluded that a well-specified conditional logit model can provide results that are qualitatively similar to those that relax the IIA assumption.

Although also comparing a conditional logit with other models that relax the IIA assumption, this study differs from that of Dahlberg and Eklöf (2003) in five respects. First, it compares the conditional logit with a nested logit and a mixed logit rather than with a multinomial probit and a mixed logit model. The nested logit can yield some different perspectives of the migration process. Second, this study examines an individual's state destination choice, which includes long distance or short distance migration rather than short distance intra-metropolitan residential choice. Third, the choice set consists of 48 choices rather than only 26 choices. Fourth, the sample size includes many more observations (11,431 compared with 1,444 individuals). Since a mixed logit model allows parameters to differ among individuals, differences in sample size significantly affect the complexity of the model. Finally, this study examines both nonmovers and movers instead of only movers. Excluding nonmovers in a migration model may cause a selection bias problem.

4. Model Specifications

This study examines models of individuals' destination choice among U.S. states. The models assume that an individual chooses a destination that maximizes utility. The models describe the individual's utility as a linear function of destination attributes as well as individual characteristics. In general, the models take the form:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \,.$$
$$= \alpha_i' Z_{ij} + \varepsilon_{ij}$$

where Z_{ij} represents both destination attributes and individual characteristics.

Three models are applied: a conditional logit model, a nested logit model, and a mixed logit model. The conditional logit model follows that applied by Davies et al. (2001). Our analysis excludes Hawaii and Alaska from the choice set, and combines the District of Columbia with Maryland. Thus, individuals have to choose from among 48 states available, including the state of origin.

Unlike Davies et al. (2001), the empirical model includes not only place attributes but also individual characteristics as explanatory variables. The place attributes include distance, employment growth, employment size, temperature, a dummy indicating the destination being adjacent to the origin, a dummy indicating the destination being a non-origin state, and dummies representing three U.S. regions (with the Northeast as reference), which attempts to account for some state fixed effects. This model includes two individual characteristics, age and education. To incorporate these variables, we create interaction terms between the individual characteristics and certain place attribute variables. More specifically, age interacts with the non-origin dummy and education interacts with distance. The first interaction term measures the effect of age on the likelihood of choosing a non-origin state (making an interstate move), while the second measures how education affects the likelihood of choosing more distant states. Appendix A provides a more complete description of all explanatory variables.

Based on theory and findings of previous studies, we expect that individuals will more likely choose destinations which are closer, have a larger labor market, experience high employment growth, and have milder winters. Likewise, we expect that older individuals are less likely to make an interstate move, and that more highly-educated individuals would be more willing to make a longer distance move. Thus, we expect the coefficient of the first interaction term to be negative and the second to be positive.

The second model applied is the nested logit model. After experimenting with different nesting patterns, two were acceptable. Recall that a nesting pattern is definitely acceptable when its inclusive value parameter, λ_k, falls between 0 and 1. Figure 1 shows the first nesting specification. It shows that in choosing their destination, individuals consider the mild winters to be an important destination attribute.[7] It also shows that some unobserved factors associated with milder winters are not captured by the January Temperature variable. These factors might include preference for snow, seasonal recreational activities, or natural landscape associated with milder winters. Those who prefer warmer temperature will have a higher probability of choosing Alabama or West Virginia over choosing Colorado, Connecticut, North Dakota, or Wyoming. We can draw similar types of conclusions for those who prefer colder states. This nesting specification neither implies a sequential decision process nor exclusion from the choice set of states not in the preferred nest, e.g., North Dakota not being an available choice for someone who prefers warmer winters. Instead it suggests that choices in the less-preferred nest are simply less likely to be chosen.

Figure 2 describes the second nesting specification, showing that individuals can partition the alternatives into coastal states and noncoastal states[8]. The nesting implies that those alternatives nested in the same nest, e.g., coastal states, exhibit correlated unobserved factors such as high preference for coastal related amenities.

The third model applied is a mixed logit model. In this model, the parameter estimates for all explanatory variables are allowed to vary across individuals, unlike in the conditional logit

[7] For this study, a state is considered a "warm state" if its average January temperature exceeds the sample average, and a cold state, otherwise.
[8] Coastal is defined as located on the Atlantic Ocean, the Gulf of Mexico, or the Pacific Ocean.

model. In other words, in this model each individual is allowed to have his/her own utility function. For the mixed logit model, we impose a normal distribution. Thus, the results should be a close approximation to those of the multinomial probit model.

5. Data

The individual data come from the one percent Public Use Microdata Sample (PUMS) of the *2000 Census of Population and Housing*. We extracted one subsample (one percent of the one percent PUMS) from the PUMS data, but excluded individuals who were enrolled in school or were less than 20 years old in 2000. In the end, the sample included 11,431 individuals. Using the PUMS data, interstate migration is defined as residing in a different state in 2000 than in 1995. Appendix B shows the average characteristics of individuals in the sample.

Data on employment size and employment growth come from the *Regional Economic Information System (REIS)*. January temperature data come from National Climatic Data Center (NCDC). The distance between origin and destination is based on longitude and latitude coordinates. If the destination is the same as the origin, then distance is zero. The remaining variables, dummy of non-origin, adjacency dummy, and region dummies were created based on U.S. maps.

6. Empirical Results

Table 2 shows the regression results of the three models. It took only 1 minute and 23 seconds for the conditional logit model to complete the regression, 35 minutes and 20 seconds for the nested logit model, and 9 hours, 52 minutes, and 20 seconds for the mixed logit model. As expected, the more complex models took longer to complete the estimation. Relaxing IIA entails a significant time/computational cost.

The results also show that the more complex (least restrictive) the model the more efficient is its estimation, yielding a higher likelihood value. The mixed logit yields the highest log likelihood value of –5,837, followed by the nested logit model at –6,019, and the conditional logit at –6,025. Based on the likelihood-ratio tests shown in Table 3, the likelihood values of both the nested logit and mixed logit models are statistically different from that of the conditional logit model. Thus, the restrictions imposed on the nested logit and mixed logit models to yield the conditional logit model are rejected.

The next step is to compare parameter estimates. Hensher (1986) and Train (2003) indicate that the results of these three models are directly comparable. We start by looking at the results for the conditional logit model. With the exception of the South Region dummy, all estimated coefficients are statistically significant with the expected sign. Individuals are more likely to choose destinations that are closer or adjacent to the state of origin, provide a large labor market, and experience high employment growth. The interaction terms suggest that older individuals are less likely to move to another state and that more highly-educated people are more willing to move to a distant state. The results for the region dummies indicate that people are more likely to choose the Midwest but less likely to choose the West than to choose the Northeast region, all else equal. People have no preference between the South and Northeast regions.

The results from the nested logit models look similar to those from the conditional logit model. In general, all estimated coefficients of the two nested logit models have the same sign and most have the same level of statistical significance as the corresponding coefficients in the conditional logit model. Compared with the first nested logit model (warm vs. cold), the second nested logit model (coastal vs. noncoastal) yields results that are closer to the standard conditional logit model. In line with that, the log likelihood value of the second model (-6,021) is also closer to the standard conditional logit (-6,025) than is the log likelihood of the first model (-6,019).

For the most part, the magnitudes of the coefficients are very similar. The only qualitative differences are the insignificance of the Midwest Region dummy and the somewhat lower significance level for January Temperature in the first nested logit model. The reduction in the magnitude and significance of the January-Temperature coefficient is possibly due to the inclusion of the nesting pattern, which in this case is explained by January Temperature as well. In other words, part of the January Temperature effect was picked up by the nesting pattern, with its residual effect becoming smaller.

Comparing the mixed logit model with the conditional logit model is more straightforward because the two models have exactly the same explanatory variables. Once again, all estimated coefficients have the same sign and same level of statistical significance as the corresponding coefficient in the conditional logit model, with the exception of a lower significance level for the Midwest Region dummy. The coefficients in the mixed logit model are consistently higher (in absolute terms) than those from the conditional logit model, with some quite a bit larger. Whether yielding larger coefficients is a norm or just a coincidence is a subject for further study. Dahlberg and Eklöf's study (2003) does not find such a pattern.

In summary, while statistically different, the results of the three models are qualitatively very similar. There are no conflicting signs and the magnitudes of the coefficients are very close, with just a few exceptions. These findings agree with those of Dahlberg and Eklöf (2003).

7. Conclusions

With improvements in computer speed and better understanding of simulation, researchers can now examine migration using much more complex models that allow researchers to better represent reality. Better computer technology has allowed this study to successfully estimate complex models that would have been unfeasible to estimate just a few years back. In addition to estimating the simpler conditional logit model, this study estimated nested logit and mixed logit models with 11,431 individuals in the sample and 48 alternatives in the choice set. The two more complex models required significant time and computational costs.

The three models differ in terms of their treatment of the IIA assumption. The conditional logit model carries the IIA assumption, the nested logit model partially relaxes it, and the mixed logit model fully relaxes it. This study has compared their results, then assessed whether the need to relax the IIA assumption warrants application of the more complex nested logit or mixed logit models.

The results of these three models, while statistically different, are qualitatively very similar. The parameter estimates of the three models are of the same sign, generally of comparable statistical significance, and, with few exceptions, of comparable magnitude. Train (2003) suggested that the results of a conditional logit can often be used as a general approximation of models that relax IIA. He further suggested that which model researchers should use depends on the goals of their research. When researchers are more concerned with knowing the individuals' average preferences, violating IIA may not be much of an issue and the relatively simple conditional logit model should suffice. IIA becomes a serious issue when researchers attempt to forecast the substitution patterns among the alternatives, e.g., if researchers need to forecast how much the demand for alternative A would change due to changes in its characteristics or the characteristics of other choices.

The more thoroughly a researcher specifies a conditional logit model, the more likely that it will serve as a good approximation, regardless of the intended use. One way to improve a conditional logit model is by incorporating more individual characteristics into the model. This would let the model capture some effects of taste variations that a mixed logit or a multinomial probit model usually captures. Note that the IIA assumption, now perceived as a restrictive assumption, was originally perceived as a natural outcome of a well-specified conditional logit model that captures all sources of correlation over alternatives. That is, a well specified conditional logit model would yield results where the residuals are independent and identical (Train (2003)). With computer technology continuing to advance, perhaps one day, researchers will not need to consider the problem of trading off between realism and computational cost.

Table 1

Discrete Choice Models and Probability Distribution Assumptions

Model Type	**Utility Function**	**Assumption on the random components**
Conditional Logit (CL)	$U_{ij} = \alpha' Z_{ij} + \varepsilon_{ij}$	ε_{ij} are iid & follow extreme value distribution function.
Nested Logit (NL)	$U_{ij} = \alpha' Z_{ij} + \varepsilon_{ij}$ Since $Z_{ij} = A_{ij} + B_{ij} \rightarrow$ $U_{ij} = \alpha' A_{ij} + \alpha' B_{ij} + \varepsilon_{ij}$ A are variables that vary across nests; B vary within each nest.	ε_{ij} are correlated among choices within the same nests, but are iid among choices across nests. ε_{ij} follow the generalized extreme value distribution function.
Mixed Logit (ML)	$U_{ij} = \alpha_i' Z_{ij} + \varepsilon_{ij}$ Since α varies across individuals $\rightarrow \alpha = a + \xi_i$, where a = mean of α. $\rightarrow U_{ij} = a' Z_{ij} + \xi_i' Z_{ij} + \varepsilon_{ij}.$	ε_{ij} are iid & follow extreme value distribution. $\xi_i' Z_{ij}$, however, are correlated & can follow different kinds of distributions.
Multinomial Probit (MNP)	$U_{ij} = \alpha' Z_{ij} + \varepsilon_{ij}$ Like mixed logit, this model also allows α to vary across individuals $\rightarrow U_{ij} = a' Z_{ij} + \xi_i' Z_{ij} + \varepsilon_{ij.}$ $= a' Z_{ij} + \eta_{ij}$	ε_{ij} are correlated & follow normal distribution. In this second case, η_{ij} are correlated & follow joint normal distribution. If the $\xi_i' Z_{ij}$ are correlated and follow normal distribution, iid ε_{ij} still give rise to a MNP model.
Heteroscedastic Extreme Value (HEV)	$U_{ij} = \alpha' Z_{ij} + \varepsilon_{ij}$	ε_{ij} are independent but not identical (not *homoscedastic*) and follow the heteroscedastic extreme value distribution.

Source: Summarized from Dahlberg and Eklöf (2003), SAS Institute (2002), and Train (2003).

Table 2

Regression Results:

Comparison among Conditional Logit, Nested Logit, and Mixed Logit Models

Explanatory Variables	Conditional Logit	Nested Logit-1 (Warm vs Cold)	Nested Logit-2 (Coastal vs Non-Coastal)	Mixed Logit
Age*Non-Origin	-0.0922 ***	-0.0948 ***	-0.0937 ***	-0.15 ***
	(-36.94)	(-36.04)	(-36.06)	(-13.24)
Education*Distance	0.0119 ***	0.0123 ***	0.0121 ***	0.0136 ***
	(12.18)	(11.96)	(12.09)	(10.44)
South Region Dummy	-0.1260	-0.0895	-0.116	-0.1464
	(-1.25)	(-0.75)	(-1.13)	(-0.99)
West Region Dummy	-0.2948 **	-0.2845 **	-0.2704 **	-0.487 **
	(-2.40)	(-2.17)	(-2.17)	(-2.57)
Midwest Region Dummy	0.1709 **	0.1379	0.1968 **	0.2287 *
	(2.01)	(1.61)	(2.27)	(1.79)
Distance	-0.2386 ***	-0.2431 ***	-0.2417 ***	-0.2956 ***
	(-18.5)	(-17.91)	(-18.29)	(-10.54)
Employment Growth	0.1362 ***	0.1362 ***	0.1357 ***	0.1483 ***
	(3.40)	(4.83)	(3.36)	(2.76)
State's Employment Share	0.1240 ***	0.1546 ***	0.1279 ***	0.1494 ***
	(10.47)	(14.87)	(10.67)	(10.53)
Adjacent Dummy	0.8967 ***	0.9830 ***	0.9332 ***	0.94 ***
	(9.86)	(10.39)	(10.04)	(8.40)
January Temperature	0.0175 ***	0.0139 **	0.018 ***	0.0249 ***
	(3.98)	(2.33)	(4.03)	(4.35)
λ_k Cold States	-------	0.9279 *	-------	-------
		(1.41)		
λ_k Warm States	-------	0.8669 ***	-------	-------
		(2.60)		
λ_k Coastal/Non-Coastal	-------	-------	0.9382 ***	-------
			(2.90)	
Log Likelihood	-6,025	-6,019	-6,021	-5,837
Covergence Time	1:23	35:20	31:07	9:52:29
Number of Sample	11,433	11,433	11,433	11,433

Note: * = significantly different from zero at 10% level; ** = significantly different from zero at 5% level, and *** = significantly different from zero at 1% level.

T-statistics are in paranthesis. The statistical tests for H0: λ_k .$\geq$ 1; and H1: $\lambda_k < 1$.

Table 3

Likelihood Ratio Tests

	Nested Logit	Mixed Logit
Log Likelihood	-6,019	-5,837
Likelihood Ratio Test Statistics	12	376
Degree of Freedom	2	10
Critical Value (5% level)	5.99	20.48
Decision	Reject Ho	Reject Ho

Note: Ho = that the nested logit or the mixed logit produce the same results as the conditional logit model (for a more complete description of this test, see Green (2000) and Econometrics Laboratory (2000)).

Figure 1

Nesting Specification of the First Nested Logit Model

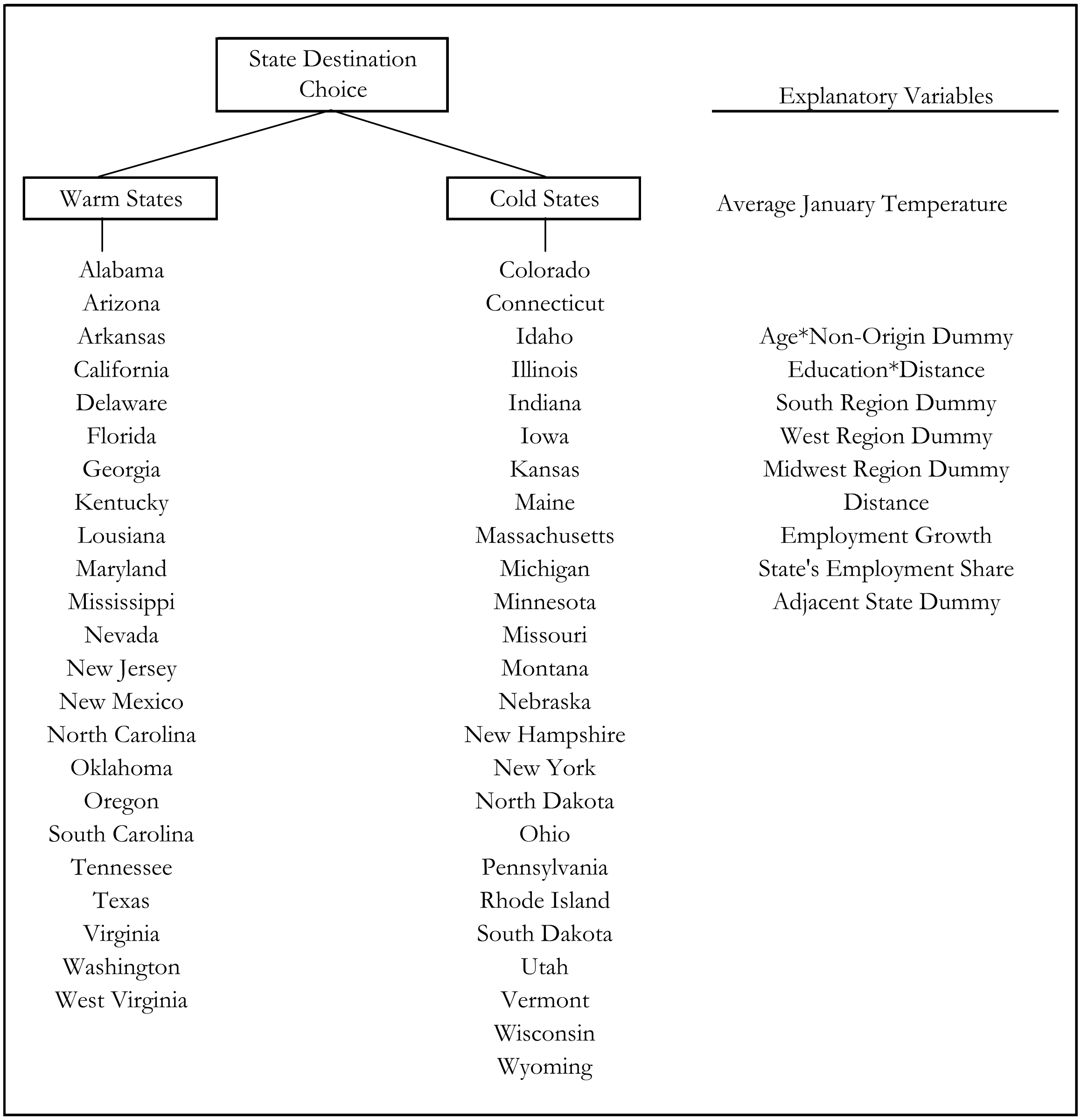

Figure 2
Nesting Specification of the Second Nested Logit Model

State Destination Choice

Coastal States | **Non-Coastal States**

Explanatory Variables (nest level): --

Coastal States
- Alabama
- California
- Connecticut
- Delaware
- Florida
- Georgia
- Louisiana
- Maine
- Maryland
- Massachusetts
- Mississippi
- New Jersey
- New York
- North Carolina
- Oregon
- Rhode Island
- South Carolina
- Texas
- Virginia
- Washington

Non-Coastal States
- Arizona
- Arkansas
- Colorado
- Idaho
- Illinois
- Indiana
- Iowa
- Kansas
- Kentucky
- Michigan
- Minnesota
- Missouri
- Montana
- Nebraska
- Nevada
- New Hampshire
- New Mexico
- North Dakota
- Ohio
- Oklahoma
- Pennsylvania
- South Dakota
- Tennessee
- Utah
- Vermont
- West Virginia
- Wisconsin
- Wyoming

Explanatory Variables
- Average January Temperature
- Age*Non-Origin Dummy
- Education*Distance
- South Region Dummy
- West Region Dummy
- Midwest Region Dummy
- Distance
- Employment Growth
- State's Employment Share
- Adjacent State Dummy

Appendix 1
Explanatory Variables

Name of Variable	Explanation
Distance	Distance, in thousands of miles, between the center of the origin and destination states.
Employment Share	Average of state's total employment divided by the U.S. total employment, 1995-1999 (*REIS*, U.S. Bureau of Economic Analysis).
Employment Growth	State's average annual employment growth, 1995-1999 (*REIS*, U.S. Bureau of Economic Analysis).
January Temperature	Area-Weighted State Average January Temperature, 1971-2000 (National Climatic Data Center).
Adjacent Dummy	A dummy taking a value of 1 if a destination state is adjacent to the origin state, and 0 otherwise.
Dummy of Non-Origin	A dummy taking a value of 1 if destination is not the same as the origin, and 0 otherwise.
Age*Non-Origin	An interaction term between Age and the dummy of non-origin.
Education*Distance	An interaction term between educational attainment and distance.
Dummies of Region	Binary dummy variables to distinguish the four U.S. regions. Three dummies are included in the model, with Northeast as the reference region.

Appendix 2
Characteristics of the Individuals in the Sample

Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Age	42.22	12.43	20	93
Years of Education	10.32	2.66	1	16

REFERENCES

Bhat, Chandra R. 1985. "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice." *Transportation Research B* 29: 471-483.

Cushing, Brian and Jacques Poot. 2004. "Crossing Boundaries and Borders: Regional Science Advances in Migration Modelling." *Papers in Regional Science* 83: 317-338.

Dahlberg, Matz and Matias Eklöf. 2003. "Relaxing the IIA Assumption in Locational Choice Models: A Comparison between Conditional Logit Model, Mixed Logit, and Multinomial Probit Models," *Working Paper 2003: 9, Department of Economics Uppsala University*.

Davies, Paul, Michael J. Greenwood, and Haizheng Li. 2001. "A Conditional Logit Approach to U.S. State-to-State Migration." *Journal of Regional Science* 41: 337–360.

Econometrics Laboratory. 2000. *Qualitative Choice Analysis Short Course*. California: University of California, Berkeley (available at http://elsa.berkeley.edu/eml/qca1_archive.shtml).

Frey William H., K. Liaw, Y. Xie, and M. Carlson. 1996. "Interstate Migration of the U.S. Poverty Population: Immigration 'Pushes' and Welfare Magnet 'Pulls'." *Population and Environment* 17: 491–536

Greene, William H. 2003. *Econometric Analysis, Forth Edition*. New Jersey, USA: Prentice-Hall.

Hensher, David A. 1986. "Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model." *Review of Economics and Statistics* 68: 657-667.

Herriges, J. and C. Kling. 1996. "Testing the Consistency of Nested Logit Models with Utility Maximization." *Economics Letters* 50: 33-39.

Kling, C. and J. Herriges. 1995. "An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization." *American Journal of Agricultural Economics* 77: 875-884.

Knapp, Thomas A., Nancy E. White, and David E. Clark. 2001. "A Nested Logit Approach to Household Mobility." *Journal of Regional Science* 41: 1-22.

Lee, Bosang. 1999. "Calling Patterns and Usage of Residential Toll Service under Self Selecting Tariffs." *Journal of Regulatory Economics* 16: 45-81.

Manski, C. and D. McFadden, eds. 1981. *Structural Analysis of Discrete Choice Data with Econometric Application*. Cambridge, USA : MIT Press.

Mueller, Charles F. 1985. *The Economics of Labor Migration, A Behavioral Analysis*. New York: Academic Press.

Newbold K. B. 1997. "Primary, Return and Onward Migration in the U.S. and Canada: Is There a Difference?" *Papers in Regional Science* 76: 175-198.

Ortuzar, J. and L.G.Willumsen. 1994. *Modelling Transport, Second Edition*. New York: Wiley.

SAS Institute. 2002. "The MDC Procedure," *SAS Online Documentation 9.1.3.* (available at https://bearhaus.baylor.edu/SASdoc/docMainpage.jsp).

Train, K. 1986. *Qualitative Choice Analysis*. Cambridge, MA: MIT Press.

Train, K. 2003. *Discrete Choice Method with Simulation*. Cambridge, MA: Cambridge University Press.

Train, K., D. McFadden, and M. Ben-Akiva. 1987. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." *Rand Journal of Economics* 18: 109-123.

White, M. and Z. Liang. 1998. "The Effect of Immigration on the Internal Migration of the Native-born Population, 1981–1990." *Population Research and Policy Review* 17: 141–166.