DRAFT

Identification of the Logit Kernel (or Mixed Logit) Model

Joan Walker, Moshe Ben-Akiva, and Denis Bolduc

July 2003

Abstract

Logit Kernel is a discrete choice model that has both probit-like disturbances as well as an additive i.i.d. extreme value (or Gumbel) disturbance à la multinomial logit. The result is an intuitive, practical, and powerful model that combines the flexibility of probit (and more) with the tractability of logit. For this reason, logit kernel has been deemed the "model of the future" and is becoming extremely popular in the literature. It has been included in popular statistical software packages as well as a recent edition of a widely used econometrics textbook and two texts specializing on discrete choice.

While the basic structure of logit kernel models is well understood, there are important identification issues that are often overlooked. Misunderstanding of these issues can lead to biased estimates as well as a significant loss of fit. This paper presents a general framework for identifying the logit kernel model. Many of the special cases of the logit kernel model are discussed in detail, including heteroscedasticity, error components, nesting structures, random coefficients, auto correlation, and application to panel data. Specification and identification issues related to each special case are identified. Finally the findings are demonstrated with empirical examples using both simulated and real data. The objectives of the paper are to present our specific findings, as well as highlight the broader themes and provide tools for uncovering identification issues pertaining to logit kernel models.

Introduction

The logit kernel model is a straightforward concept: it is a discrete choice model in which the disturbances (of the utilities) consist of both a probit-like portion and an additive i.i.d. Gumbel portion (i.e., a multinomial logit disturbance).

Multinomial logit (MNL) has its well-known blessing of tractability and its equally well-known curse of a rigid error structure leading to the IIA property. The nested logit model relaxes the rigidity of the MNL error structure and has the advantage of retaining a probability function in closed form. Nonetheless, nested logit is still limited and cannot capture many forms of unobserved heterogeneity, including, for example, random taste heterogeneity. The logit kernel model with its probit-like (but even more general) disturbances completely opens up the specification of the disturbances so that almost any desirable error structure can be represented in the model. As with probit, however, this flexibility comes at a cost, namely that the probability functions consist of multi-dimensional integrals that do not have closed form solutions. Standard practice is to estimate such models by replacing the choice probabilities with easy to compute and unbiased simulators. The beauty of the additive i.i.d. Gumbel term is that it leads to a particularly convenient and attractive probability simulator, which is simply the average of a set of logit probabilities. The logit kernel probability simulator has all of the desirable properties of a simulator including being convenient, unbiased, and smooth.

There have been numerous relatively recent applications and investigations into the model, including Bekhor et al. (2002), Bhat (1997 & 1998), Bolduc, Fortin and Fournier (1996), Brownstone, Bunch and Train (2000), Brownstone and Train (1999), Goett, Hudson, and Train (2000), Gönül and Srinivasan (1993), Greene (2000), Mehndiratta and Hansen (1997), Revelt and Train (1998 & 1999), Srinivasan and Mahmassani (2000), and Train (1998) [need to update…]. A very important recent contribution is McFadden and Train's (2000) paper on mixed logit, which both (i) proves that any well-behaved random utility consistent behavior can be represented as closely as desired with a mixed logit specification, and (ii) presents easy to implement specification tests for these models.

While logit kernel has strong computational advantages and allows for rich specifications, empirical applications of the model turn out not to be straightforward, as there are many important issues that arise in identification. While the number of logit kernel applications in the literature is rapidly growing, the identification issue has been largely ignored. The objective of this paper is to lay out a procedure for identifying logit kernel models, and to analyze and develop identification rules for some of the most common forms of the logit kernel model. Just as there are standard identifying practices for a pure multinomial logit model (for example, constraining the scale to 1, or setting one of the alternative-specific constants to zero), a set of guidelines for identifying the logit kernel model must be developed. This is particularly important as the model is poised to become widely applied in the literature and in practice.

Terminology

There are numerous terms floating around the literature that are related to the logit kernel model that is discussed here. McFadden and Train (2000) and others have adopted the term "mixed logit" to reflect that the model is comprised of a mixture of logit models. The term logit kernel is also used for the same model, which refers to the fact that the core of the model (as well as the resulting probability simulator) is a logit formulation. The term "logit kernel probit" has also been used in the special case when the formulation is used to approximate or extend the probit model. There are also numerous terms that are used to describe various error specifications in discrete choice models, including error components, taste variation, random parameters (coefficients), random effects, unobserved heterogeneity, etc. When such models are specified in a form that includes an additive i.i.d. extreme value (or other GEV) term, then they fall within the same broad class of logit kernel (as well as mixed logit) class of models.

We choose to use the term *logit kernel*, because conceptually these models start with a logit model at the core and then are extended by adding a host of different error terms. In addition, the term is descriptive of the form of the likelihood function and the resulting logit kernel simulator.

Organization of the Paper

The paper begins with a discussion of the identification problem in econometrics, which is followed by a discussion of the Logit Kernel model and specific rules of identification. These rules are then used to analyze specific cases of logit kernel, including heteroscedasticity, nesting, and models with panel data. Finally, empirical results are presented that highlight the identification issues.

The Identification Problem

> "We must face up to the fact that we cannot answer all of the questions that we ask."
> – Manski (1995)

All econometric models require identification restrictions in order to be estimated. A model is specified based on some underlying theory or propositions. These models are a function of unknown parameters, and data are used to provide estimates of the unknown parameters. The issue that arises is that there does not exist a unique vector of parameters that solves (optimizes) the equation. In fact, without restrictions being imposed, there are infinite solutions as the objective function at the solution is a plateau (hyperplane). Furthermore, no amount of data can solve the problem. The identification problem is to determine what conclusions can or cannot be drawn from a model, and under what sets of assumptions.

In the context of this paper, the identification problem can be described as determining the set of restrictions to impose in order to obtain a unique vector of parameter estimates. Identifying restrictions can be divided into two types. The first type is the most severe, in which any restrictions that are imposed change the behavior represented by the model, and, therefore, influences its prediction. Estimation of supply and demand equations is a classic example of this type of identification. Manski (1995) provides a good discussion of such identification issues. He points out that strong assumptions must be introduced to solve these problems, and these can lead to widely varying conclusions.

The other type of identification problem is less severe, and is the focus of this paper. In this case, while there are still infinite solutions to the unrestricted model, any one of them is acceptable in that they all represent the same behavior and lead to the same prediction. That is, given the problem being studied, any one of these infinite solutions is adequate. In this case, the identification problem (often called *normalization*) is to impose constraints on a subset of parameters such that there is one unique solution, and to do so in a way such that this solution is a member of the solution set for the unrestricted model. If such restrictions are not imposed to obtain a unique solution, then the standard errors of the estimated parameters cannot be determined, and therefore the statistical tools used for hypothesis testing (for example, t-statistics) and confidence intervals are unavailable. Setting the scale in a discrete choice model, such as logit or probit, is a classic example of this type of identification, and this is closely related to the issues discussed in this paper.

Model Formulation

The Discrete Choice Model

The discrete choice model can be written as follows. For a given individual n, $n = 1,...,N$ where N is the sample size, and an alternative i, $i = 1,...,J_n$ where J_n is the number of alternatives in the choice set C_n of individual n, the model is written as:

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \text{ for } j = 1,...,J_n \\ 0 & \text{otherwise} \end{cases},$$

$$U_{in} = X_{in}\beta + \varepsilon_{in} \ ,$$

where y_{in} indicates the observed choice, and U_{in} is the utility of alternative i as perceived by individual n. X_{in} is a $(1 \times K)$ vector of explanatory variables describing individual n and alternative i, including alternative-specific dummy variables as well as generic and alternative-specific attributes and their interactions with the characteristics of individual n. β is a $(K \times 1)$ vector of coefficients and ε_{in} is a random disturbance. The assumption that the disturbances are i.i.d. Gumbel leads to the tractable, yet restrictive logit model. The assumption that the disturbances are multivariate normal distributed leads to the flexible, but computationally demanding probit model. The logit kernel model presented in this paper is a hybrid between logit and probit and represents an effort to incorporate the advantages of each.

In a more compact vector form, the discrete choice model can be written as follows:

$$y_n = [y_{1n},..., y_{J_n n}]' \ ,$$

$$U_n = X_n\beta + \varepsilon_n \ , \tag{1}$$

where y_n, U_n, and ε_n are $(J_n \times 1)$ vectors and X_n is a $(J_n \times K)$ matrix.

The Logit Kernel Model with Factor Analytic Form

The logit kernel model results from a particular specification of the disturbance, ε_{in}, in which there is a flexible probit-like term (often normal, although not necessarily) and an i.i.d. Gumbel random variate (or, more generally, GEV). The probit-like term captures the interdependencies among the alternatives. We specify these interdependencies using a factor analytic structure. The factor analytic structure was first proposed for probit by McFadden (1984) as a means of reducing the dimensionality of the integral. The specification is used here because it is a flexible specification that includes all known (additive) error structures as special cases, as is shown in Ben-Akiva, Bolduc, and Walker (2002).

Using the factor analytic form, the disturbance vector ε_n is specified as follows:

$$\varepsilon_n = F_n \xi_n + \nu_n \ , \tag{2}$$

where ξ_n is an $(M \times 1)$ vector of M multivariate distributed latent factors, F_n is a $(J_n \times M)$ matrix of the factor loadings that map the factors to the error vector (F_n includes fixed and/or unknown parameters and may also be a function of covariates), and ν_n is a $(J_n \times 1)$ vector of i.i.d. Gumbel random variates. For estimation, it is desirable to specify the factors such that they are independent, and therefore ξ_n is decomposed as follows:

$$\xi_n = T\zeta_n \ , \tag{3}$$

where ζ_n are a set of standard independent factors (often normally distributed), TT' is the covariance matrix of ξ_n, and T is the Cholesky factorization of it. The number of factors, M, can be less than, equal to, or greater than the number of alternatives. To simplify the presentation, we assume that the factors have standard normal distributions; however, they can follow any number of different distributions, such as lognormal, triangular, uniform, etc. (See Kenneth Train's mixed logit software and supporting documents.)

Substituting Equations (2) and (3) into Equation (1), yields:

The Factor Analytic Logit Kernel Specification

$$U_n = X_n\beta + F_n T\zeta_n + \nu_n \ , \tag{4}$$

$$\operatorname{cov}(U_n) = F_n TT' F_n' + (g/\mu^2) I_{J_n} \tag{5}$$

(which we denote as $\Omega_n = \Sigma_n + \Gamma_n$),

where: U_n is a $(J_n \times 1)$ vector of utilities;

X_n is a $(J_n \times K)$ matrix of explanatory variables;

β is a $(K \times 1)$ vector of unknown parameters;

F_n is a $(J_n \times M)$ matrix of factor loadings, including fixed and/or unknown parameters;

T is a $(M \times M)$ lower triangular matrix of unknown parameters, where $TT' = Cov(\xi_n = T\zeta_n)$;

ζ_n is a $(M \times 1)$ vector of i.i.d. random variables with zero mean and unit variance; and

ν_n is a $(J_n \times 1)$ vector of i.i.d. Gumbel random variables with zero location parameter and scale equal to $\mu > 0$. The variance is g/μ^2, where g is the variance of a standard Gumbel ($\pi^2/6$).

The unknown parameters in this model are μ, β, those in F_n, and those in T. X_n are observed, whereas ζ_n and ν_n are unobserved. Note that if $T = 0$ then the model reduces to logit. This formulation assumes cross-sectional data, and will be modified later to address panel data.

It is important to note that we specify the model in level form (i.e., U_{jn}, $j = 1,..., J_n$) rather than in difference form (i.e., $(U_{jn} - U_{J_n n})$, $j = 1,...,(J_n - 1)$). We do this for interpretation purposes, because it enables us to parameterize the covariance structure in ways that capture specific (and conceptual) correlation effects. It also represents the manner in which logit kernel models are specified in all estimation packages that we are aware of. Nonetheless, it is the difference form that is estimable, and there are multiple level structures that can represent any unique difference covariance structure, which is a critical point for identification.

Response Probabilities

The power of the logit kernel model derives from the ease with which complex disturbance structures can be estimated. If the factors ζ_n in Equation (4) are known, the model corresponds to a multinomial logit formulation:

$$\Lambda(i \mid \zeta_n) = \frac{e^{\mu(X_{in}\beta + F_{in}T\zeta_n)}}{\sum_{j \in C_n} e^{\mu(X_{jn}\beta + F_{jn}T\zeta_n)}} \;, \tag{6}$$

where $\Lambda(i \mid \zeta_n)$ is the probability that the choice is i given ζ_n, and F_{jn} is j^{th} row of the matrix F_n, $j = 1,..., J_n$. Since the ζ_n is in fact not known, the unconditional choice probability of interest is:

$$P(i) = \int_{\zeta} \Lambda(i \mid \zeta) n(\zeta, I_M) d\zeta \;, \tag{7}$$

where $n(\zeta, I_M)$ is the joint density function of ζ, which, by construction, is a product of standard univariate normals:

$$n(\zeta, I_M) = \prod_{m=1}^{M} \phi(\zeta_m) \;.$$

$P(i)$ can then be estimated with an unbiased, smooth, tractable simulator, which is computed as:

$$\hat{P}(i) = \frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \Lambda(i \mid \zeta_n^d) \ ,$$

where ζ_n^d denotes draw d from the distribution of ζ, thus enabling the estimation of high dimensional integrals with relative ease.

Identification of Logit Kernel

As described above, all discrete choice models require identification restrictions in order to be estimated. For example, the normalization required to set the scale of a logit model is a special type of identification restriction. Without such a restriction there does not exist a unique set of parameter estimates that maximize the likelihood. Without a unique solution, the standard errors of the estimated parameters cannot be determined, and therefore the statistical tools used for hypothesis testing (for example, t-statistics) and confidence intervals are unavailable. The problem is exacerbated in models that are estimated via simulation (as with the logit kernel), because the simulation can hide identification problems and lead to incorrect conclusions about the model and behavior being studied. The problem we face is to determine a set of restrictions such that a unique solution is obtained.

For the logit kernel model, there are two sets of relevant parameters that need to be considered for identification: the vector β and the unrestricted parameters of the distribution of the disturbance vector ε_n. Note that the nature of the identification problem is that there are infinite sets of restrictions that can be imposed in order to identify any particular model. One manifestation of this is that there is a choice as to whether to impose the restrictions on the disturbance parameters (in ε_n) or the systematic parameters (β). For example, in a multinomial logit model, the scale can be set either by constraining the variance of the disturbance (for example, μ=1 or μ=2) or by constraining one of the systematic parameters (for example, a particular β=1 or β=2). Without loss of generality, the assumption made throughout this paper is that the conventional multinomial logit constraints are imposed on the vector β, for example as described in Ben-Akiva and Lerman (1985), and the scale is constrained through the disturbance. These assumptions allow us to focus solely on the identification of the distribution of the disturbance vector in the logit kernel model. This also best represents the manner in which logit kernel models are typically estimated using available statistical software packages, that is, there is a logit core that is specified as in a conventional multinomial logit specification, and the additional complexity is encased in the disturbance.

A common tool used to establish identification of a particular model is to examine its information matrix (the expected second derivatives matrix of the log likelihood). While this is an invaluable method in checking for identification, it is typically employed after conventional identification rules are imposed on the model (for example, in multinomial logit, after the scale is set and an alternative specific constant is zeroed). This is simply because examining the information matrix requires estimation of the model and is computationally complex to estimate. The focus in this work is to establish identification rules that can be applied to a logit kernel specification before estimation has begun, such as rules on how to normalize a heteroscsedastic logit kernel model or a logit kernel model with a nested specification. For this purpose, it is more straightforward to examine the covariance matrix directly, or, more specifically, the *covariance matrix of utility differences*, which is the approach used in this paper. This approach follows work presented in Bunch (1991) for the probit model, in which he implemented the use of the order and rank conditions to determine identification.

In discussing the normalization of logit kernel, there are two different cases of interest. The first is where the disturbance does not vary based on characteristics of an individual or attributes of the alternatives, that is, it is cases in which $F_n T \zeta_n$ is not a function of X_n. This includes the case of an unrestricted covariance matrix, heteroscedasticity, nesting and cross-nesting, error components, and some panel data structures. The second situation is when the disturbances vary across the population, and the most common example is the case of random parameters on attributes of the alternatives (for example, cost) or characteristics of the decision-maker (for example, gender). This paper deals only with the former (alternative-specific) case. For readers interested in random parameters, see, Walker (2001, 2002).

Logit Kernel with Alternative-Specific Disturbances

Consider the case where the disturbance does not vary based on the characteristics of the person or attributes of the alternatives. We call this the alternative-specific portion of the disturbance structure, since it is not a function of X_n beyond the alternative specific dummy variables in X_n. In this case, the individual subscript n can be dropped from F_n, and the utility equation is then:

Logit Kernel with Alternative-Specific Disturbances $\quad U_n = X_n \beta + FT\zeta_n + \nu_n$

This is a broad class that includes unrestricted covariance structure, heteroscedasticity, nesting, cross-nesting, error components, and some panel data structures.

The procedure involved in identifying the alternative-specific portion of the logit kernel model is described in detail below. Briefly, the steps are (1) to hypothesizing the model of interest, converting the model to differences in utilities (to set the location of the model), (2 and 3) to apply order and rank conditions to the covariance matrix of utility differences to determine the number of estimable parameters in the disturbance, and then (3) to determine a valid normalization such that the resulting choice probabilities of the model do not change. Each of these steps is expanded on below. After the description, the procedure will be applied to special cases.

1. Hypothesize the model of interest.

Comments: This amounts to specifying $FT\zeta_n$, *which is the disturbance structure that is a priori assumed to exist. For example, one could hypothesize an unrestricted covariance matrix or various restricted covariance matrices such as heteroscedasticity or nesting.*

2. Determine the covariance matrix of utility differences of the hypothesized model:

$$\Omega_\Delta = Cov(\Delta U_n) = \Delta FTT'F'\Delta' + \Delta(g/\mu^2)I_J\Delta',$$

where Δ is the linear operator that transforms the J utilities into $(J-1)$ utility differences taken with respect to the J^{th} alternative. Δ is a $(J-1)\times J$ matrix that consists of a $(J-1)\times(J-1)$ identity matrix with a column vector of -1's inserted as the J^{th} column.

Comments: A key point of identification is that while we write (and program) the logit kernel specification in levels form above (i.e., U_{jn}, $j=1,...,J_n$*), in estimation it is only the* ***differences in utility*** *that matter (i.e.,* $(U_{jn}-U_{J_n n})$, $j=1,...,(J_n-1)$*). Taking the differences sets the "location" of the model, which is necessary for all random utility models (see Ben-Akiva and Lerman, 1985, for discussion). The logit kernel model is specified in level form rather than in difference form for interpretation purposes, because it enables the parameterization of the covariance structure in ways that capture specific (and conceptual) correlation effects (for example, nesting and heteroscedasticity). Estimation packages now available for logit kernel also employ a levels specification. Nonetheless, it is the difference form that is estimable, and the levels form is irrelevant to the estimation procedure. Furthermore, there are multiple level structures that can represent any unique difference covariance structure (as was clearly presented in Bunch (1991) and will be shown for nested logit kernel below).*

3. Apply the **Order Condition**, which states that the number of estimable alternative-specific disturbance parameters, S, adheres to:

$$S \le \frac{J(J-1)}{2} - 1.$$

Comments: This upper bound is equal to the number of unique cells in Ω_Δ *(which is symmetric) minus 1 to set the scale of the model (another necessity of random utility models). This is a necessary, but not sufficient, condition of identification. The limit imposed by the Order Condition is a function only of the number of alternatives, and therefore states that:*

with 2 alternatives, no alternative-specific covariance terms can be identified;
with 3 alternatives, up to 2 terms can be identified;
with 4 alternatives, up to 5 terms can be identified;
with 5 alternatives, up to 9 terms can be identified; etc.

It is clearly shown in Bunch (1991) that the number of parameters that can be estimated is often less than suggested by the order condition, depending on the covariance structure postulated. Nonetheless, do not underestimate the usefulness of the Order Condition; it provides for a quick check to avoid major blunders, and there are models that have been published that do not pass this test.

4. Apply the **Rank Condition**, which states that the number of estimable alternative-specific disturbance parameters, S, adheres to:

   $$S = Rank(Jacobian(vecu(\Omega_\Delta))) - 1,$$

   where *vecu* is a function that vectorizes the unique elements of Ω_Δ into a column vector.

   Comments: The rank condition is more restrictive than the order condition. The rank condition is sufficient to ensure identification in that there is a single solution to the normalized model. The order condition simply counts cells, and ignores the internal structure of Ω_Δ. The rank condition, however, counts the number of linearly independent equations available in Ω_Δ (mathematically written above as taking the Rank of the Jacobian) that can be used to estimate the parameters of the error structure. The subtraction of 1 is necessary to set the scale of the model.

5. Apply the **Equality Condition**, which states that any imposed normalization must adhere to:

   $$\Omega_\Delta = \Omega_\Delta^{Normalized},$$

   where Ω_Δ is the covariance matrix of utility differences from any of the solutions in the unrestricted model (they are all the same) and $\Omega_\Delta^{Normalized}$ is from the normalized model.

   Comments: It is necessary to verify that the imposed normalization does not otherwise restrict the model, that is, that the probabilities remain the same as before the restriction is imposed. Keeping the probabilities the same means that the covariance matrix of utility differences also must remain the same, which is stated in the equation. The Equality Condition is sufficient to ensure that the correct model is being identified.

Both the Rank Condition and the Equality Condition must hold for a model to be correctly identified.

In the following sections, the general rules of identification that were established above will be used to address identification of specific instances of the logit kernel model. The objectives are to provide examples for how to apply the identification rules and also to establish rules of identification for common forms of the logit kernel model. In some cases (particularly the first one), the identification conditions will be applied very deliberately. In later cases, steps will be dropped and conclusions drawn.

Identification of Heteroscedastic Logit Kernel

For this first case, each steps of identification will be followed precisely. The cases that follow will be discussed in a more streamlined fashion.

Step 1: The Model

In a heteroscedastic model, the disturbance of each alternative has a different variance. Thus, the model allows for situations in which the systematic portion of the utility better represents the utility of some alternatives more than others. For example, often the carpool alternative in a mode choice model has a higher variance than other alternatives, because it is more difficult to systematically explain why a traveler chooses to carpool.

The heteroscedastic model, assuming a universal choice set ($C_n = C \quad \forall n$), is written as:

$U_n = X_n\beta + T\zeta_n + \nu_n$, ($M = J$ and F_n equals the identity matrix I_J),

$$T = \begin{bmatrix} \sigma_1 & & & \\ 0 & \sigma_2 & & \\ 0 & 0 & \ddots & \\ 0 & 0 & 0 & \sigma_J \end{bmatrix} (J \times J), \quad \zeta_n \quad (J \times 1),$$

and, defining $\sigma_{ii} = (\sigma_i)^2$, the $Cov(U_n)$ is:

$$\Omega = \begin{bmatrix} \sigma_{11} + g/\mu^2 & & & \\ 0 & \sigma_{22} + g/\mu^2 & & \\ 0 & 0 & \ddots & \\ 0 & 0 & 0 & \sigma_{JJ} + g/\mu^2 \end{bmatrix} (J \times J).$$

In scalar notation, the model is $U_{in} = X_{in}\beta + \sigma_i\zeta_{in} + \nu_{in}$, $i \in C$.

Step 2: The Covariance Matrix of Utility Differences

The identification conditions are worked through for a two alternative heteroscedastic model, a three alternative heteroscedastic model, and a four alternative heteroscedastic model, because the three models serve well to highlight various aspects of identification. The conclusions will be generalized to any number of alternatives at the end. The covariance structures for these three models are as follows:

$$J = 2: \ \Omega = \begin{bmatrix} \sigma_{11} + g/\mu^2 & \\ 0 & \sigma_{22} + g/\mu^2 \end{bmatrix},$$

$$J=3:\ \Omega=\begin{bmatrix}\sigma_{11}+g/\mu^2 & & \\ 0 & \sigma_{22}+g/\mu^2 & \\ 0 & 0 & \sigma_{33}+g/\mu^2\end{bmatrix},$$

$$J=4:\ \Omega=\begin{bmatrix}\sigma_{11}+g/\mu^2 & & & \\ 0 & \sigma_{22}+g/\mu^2 & & \\ 0 & 0 & \sigma_{33}+g/\mu^2 & \\ 0 & 0 & 0 & \sigma_{44}+g/\mu^2\end{bmatrix}.$$

The covariance matrices of utility differences are as follows:

$$J=2:\ \Delta_J=\begin{bmatrix}1 & -1\end{bmatrix}, \qquad \Omega_\Delta=\begin{bmatrix}\sigma_{11}+\sigma_{22}+2g\big/\mu^2\end{bmatrix},$$

$$J=3:\ \Delta_J=\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & -1\end{bmatrix}, \quad \Omega_\Delta=\begin{bmatrix}\sigma_{11}+\sigma_{33}+2g\big/\mu^2 & \\ \sigma_{33}+g\big/\mu^2 & \sigma_{22}+\sigma_{33}+2g\big/\mu^2\end{bmatrix},$$

$$J=4:\ \Delta_J=\begin{bmatrix}1 & 0 & 0 & -1\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & -1\end{bmatrix},$$

$$\Omega_\Delta=\begin{bmatrix}\sigma_{11}+\sigma_{44}+2g/\mu^2 & & \\ \sigma_{44}+g/\mu^2 & \sigma_{22}+\sigma_{44}+2g/\mu^2 & \\ \sigma_{44}+g/\mu^2 & \sigma_{44}+g/\mu^2 & \sigma_{33}+\sigma_{44}+2g/\mu^2\end{bmatrix}.$$

Step 3: The Order Condition

Each heteroscedastic model has $J+1$ unknown parameters: J σ_{ii}'s and one μ. The order condition then provides the following information regarding identification:

$J=2$: $unknowns=\{\sigma_{11},\sigma_{22},\mu\}$; $s=0$ → 0 variances are identified

$J=3$: $unknowns=\{\sigma_{11},\sigma_{22},\sigma_{33},\mu\}$; $s=2$ → up to 2 variances are identified

$J=4$: $unknowns=\{\sigma_{11},\sigma_{22},\sigma_{33},\sigma_{44},\mu\}$; $s=5$ → potentially *all* variances are identified

Step 4: The Rank Condition

The first step in applying the Rank Condition is to vectorize the unique elements of Ω_Δ into a column vector (we call this operator vecu):[1]

---

[1] Note that there's no need to continue with identification for the binary heteroscedastic case, since the order condition resolved that none of the error parameters are identified.

$$J=3:\ vecu(\Omega_\Delta)=\begin{bmatrix}\sigma_{11}+\sigma_{33}+2g/\mu^2\\ \sigma_{22}+\sigma_{33}+2g/\mu^2\\ \sigma_{33}+g/\mu^2\end{bmatrix},$$

$$J=4:\ vecu(\Omega_\Delta)=\begin{bmatrix}\sigma_{11}+\sigma_{44}+2g/\mu^2\\ \sigma_{22}+\sigma_{44}+2g/\mu^2\\ \sigma_{33}+\sigma_{44}+2g/\mu^2\\ \sigma_{44}+g/\mu^2\end{bmatrix}.$$

By examination, it is clear that we are short an equation in both cases. This is formally determined by examining the Rank of the Jacobian matrix of vecu(Ω_Δ) with respect to each of the unknown parameters $(\sigma_{11},...,\sigma_{JJ},g/\mu^2)$:

$$J=3:\ \begin{matrix}Jacobian\\ matrix\ \ of\\ vecu(\Omega_\Delta)\end{matrix}=\begin{bmatrix}1&0&1&2\\0&1&1&2\\0&0&1&1\end{bmatrix},\ Rank=3\quad\rightarrow\quad \text{can estimate 2 of the parameters; must normalize } \mu \text{ and one } \sigma_{ii}.$$

$$J=4:\ \begin{matrix}Jacobian\\ matrix\ \ of\\ vecu(\Omega_\Delta)\end{matrix}=\begin{bmatrix}1&0&0&1&2\\0&1&0&1&2\\0&0&1&1&2\\0&0&0&1&1\end{bmatrix},\ Rank=4\quad\rightarrow\quad \text{can estimate 3 of the parameters; must normalize } \mu \text{ and one } \sigma_{ii}.$$

So for both of these cases, the scale term μ as well as one of the σ_{ii}'s must be normalized.

Which σ_{ii} should be fixed? And to what value? This is where the Equality Condition comes into play, and it turns out that the normalizations for logit kernel models are not always arbitrary or intuitive.

Step 5: The Equality Condition

For application of the Equality Condition, we will use the three alternative model as an example. It is useful in the analysis to deal directly with the estimated (i.e., scaled) parameters, so we introduce the notation $\dot{\sigma}_{ii}=(\mu\sigma_i)^2$. Say we impose the normalization that the third heteroscedastic term, $\dot{\sigma}_{33}$, is constrained to some fixed value we denote as $\dot{\sigma}_{ff}^N$, the Equality Condition is then written as:

$$\begin{bmatrix}(\dot{\sigma}_{11}+\dot{\sigma}_{33}+2g)/\mu^2 & \\ (\dot{\sigma}_{33}+g)/\mu^2 & (\dot{\sigma}_{22}+\dot{\sigma}_{33}+2g)/\mu^2\end{bmatrix}=\begin{bmatrix}(\dot{\sigma}_{11}^N+\dot{\sigma}_{ff}^N+2g)/\mu_N^2 & \\ (\dot{\sigma}_{ff}^N+g)/\mu_N^2 & (\dot{\sigma}_{22}^N+\dot{\sigma}_{ff}^N+2g)/\mu_N^2\end{bmatrix}$$

where the matrix on the left represents the theoretical (non-normalized) model and the matrix on the right represents the normalized model. This relationship states that when the normalization is imposed, the remaining parameters in the normalized model will adjust such that the theoretical (or true) covariance matrix of utility differences is recovered. It also provides the following three equations:

$$(\dot{\sigma}_{ff}^N + g)\big/\mu_N^2 = (\dot{\sigma}_{33} + g)\big/\mu^2 \ , \tag{8}$$

$$(\dot{\sigma}_{11}^N + \dot{\sigma}_{ff}^N + 2g)\big/\mu_N^2 = (\dot{\sigma}_{11} + \dot{\sigma}_{33} + 2g)\big/\mu^2 \ \text{, and} \tag{9}$$

$$(\dot{\sigma}_{22}^N + \dot{\sigma}_{ff}^N + 2g)\big/\mu_N^2 = (\dot{\sigma}_{22} + \dot{\sigma}_{33} + 2g)\big/\mu^2 \ . \tag{10}$$

The question is, what values of $\dot{\sigma}_{ff}^N$ guarantee that these relationships hold? To derive the restrictions on $\dot{\sigma}_{ff}^N$, we first use Equations (8) to (10) to develop equations for the unknown parameters of the normalized model $\left(\mu_N^2, \dot{\sigma}_{11}^N, \text{ and } \dot{\sigma}_{22}^N\right)$ as functions of the normalized parameter $\dot{\sigma}_{ff}^N$ and the theoretical parameters $\left(\mu^2, \dot{\sigma}_{11}, \ \dot{\sigma}_{22} \text{ and } \dot{\sigma}_{33}\right)$, which leads to:

$$\mu_N^2 = \mu^2(\dot{\sigma}_{ff}^N + g)\big/(\dot{\sigma}_{33} + g) \ , \tag{11}$$

$$\dot{\sigma}_{11}^N = \left((\dot{\sigma}_{11} + g)\dot{\sigma}_{ff}^N + (\dot{\sigma}_{11} - \dot{\sigma}_{33})g\right)\big/\left(\dot{\sigma}_{33} + g\right) \ \text{, and} \tag{12}$$

$$\dot{\sigma}_{22}^N = \left((\dot{\sigma}_{22} + g)\dot{\sigma}_{ff}^N + (\dot{\sigma}_{22} - \dot{\sigma}_{33})g\right)\big/\left(\dot{\sigma}_{33} + g\right) \ . \tag{13}$$

Additionally, the scales μ^2 and μ_N^2 must be strictly greater than zero, and the variances $\dot{\sigma}_{11}, \ \dot{\sigma}_{22}, \ \dot{\sigma}_{33}, \ \dot{\sigma}_{11}^N, \ \dot{\sigma}_{22}^N$ and $\dot{\sigma}_{ff}^N$ must be greater than or equal to zero, which results in the following set of restrictions:

$$\mu^2(\dot{\sigma}_{ff}^N + g)\big/(\dot{\sigma}_{33} + g) > 0 \ , \tag{14}$$

$$\left((\dot{\sigma}_{11} + g)\dot{\sigma}_{ff}^N + (\dot{\sigma}_{11} - \dot{\sigma}_{33})g\right)\big/\left(\dot{\sigma}_{33} + g\right) \geq 0 \ \text{, and} \tag{15}$$

$$\left((\dot{\sigma}_{22} + g)\dot{\sigma}_{ff}^N + (\dot{\sigma}_{22} - \dot{\sigma}_{33})g\right)\big/\left(\dot{\sigma}_{33} + g\right) \geq 0 \ . \tag{16}$$

Equation (14) will always be satisfied. Equations (15) and (16) are where it gets interesting, because solving for $\dot{\sigma}_{ff}^N$ leads to the following restrictions on the normalization:

$$\dot{\sigma}_{ff}^N \geq \left(\dot{\sigma}_{33} - \dot{\sigma}_{ii}\right) {}^{g}\!\big/_{(g + \dot{\sigma}_{ii})} \quad , \ i = 1,2 \ . \tag{17}$$

($\dot{\sigma}_{33}$ is the heteroscedastic term that is fixed.)

What does this mean? Note that as long as alternative 3 is the minimum variance alternative, the right hand side of Equation (17) is negative, and so the restriction is satisfied for any $\dot{\sigma}_{ff}^N \geq 0$. However, when alternative 3 is not the minimum variance alternative, $\dot{\sigma}_{ff}^N$ must be set "large enough" (and certainly above zero) such that Equation (17) is satisfied. This latter approach to normalization is not particularly practical since the $\dot{\sigma}_{ii}$ are unknown (how large is large enough?), and it has the drawback that MNL is not a case nested within the logit kernel specification. Therefore, the following normalization is recommended:

The preferred normalization for the heteroscedastic logit kernel model is to constrain the heteroscedastic term of the minimum variance alternative to zero.

A method for implementing this normalization is described later in the next section.

Generalizing the Heteroscedastic Case

These results for the heteroscedastic model can be straightforwardly generalized to the following:

Identification

$J=2$ none of the heteroscedastic variances can be identified.

$J \geq 3$ $J-1$ of the heteroscedastic variances can be identified.

Normalization

For $J \geq 3$, a normalization must be imposed on one of the variance terms, denote this as $\dot{\sigma}_{jj} = \dot{\sigma}_{ff}^{N}$ where $\dot{\sigma}_{jj}$ is the true, albeit unknown, variance term that is fixed to the value $\dot{\sigma}_{ff}^{N}$.

This normalization is not arbitrary, and must meet the following restriction:

$$\dot{\sigma}_{ff}^{N} \geq \left(\dot{\sigma}_{jj} - \dot{\sigma}_{ii}\right) g \Big/ (g + \dot{\sigma}_{ii}) \quad , \; i = 1, \ldots, J \; .$$

This restriction shows that the natural tendency to normalize an arbitrary heteroscedastic term to zero (typically the same normalization as the alternative specific constants in the systematic part of the utility) is incorrect. If the alternative does not happen to be the minimum variance alternative, the parameter estimates will be inconsistent, there can be a significant loss of fit (as demonstrated in the empirical section at the end of this paper), and it can lead to the incorrect conclusion that the model is homoscedastic. This is an important issue, which, as far as we can tell, is ignored in the literature. It appears that arbitrary normalizations are being made for models of this form (see, for example Gönül and Srinivasan, 1993, and Greene, 2000, Table 19.15). Therefore, there is a chance that a non-minimum variance was normalized to zero, which would mean that the model is misspecified. It is important to note that it is the addition of the i.i.d. disturbance that causes the identification problem. Therefore, heteroscedastic pure probit models as well as the heteroscedastic extreme value models (see, for example, Bhat, 1995, and Steckel and Vanhonacker, 1988) do not exhibit this property. Walker (2001) shows that an arbitrary normalization is valid for a heteroscedastic probit model.

Ideally, we would like to impose a normalization such that MNL is a special case of the model. Therefore, the best normalization is to fix the minimum variance alternative to zero. However, there is in practice no prior knowledge of the minimum variance alternative. A brute force solution is to estimate J versions of the model, each with a different heteroscedastic term normalized; the model with the best fit is the one with the correct normalization. This is obviously cumbersome as well as time consuming. Alternatively, one can estimate the unidentified model with all J heteroscedastic terms. Although this model is not identified (that is, there are infinite solutions that provide the best fit to the data), it will converge to *one* of these infinite (and correct) solutions and therefore reflect the true covariance structure of the model. (The problem with the results from an unidentified model is that the standard errors are not estimable.) Therefore, the heteroscedastic term with minimum estimated variance in the unidentified model *is* the minimum variance alternative, thus eliminating the need to estimate J different models. Examples of this method are provided in the applications section.

Identification of Logit Kernel Models with Nesting

Nesting and cross-nesting logit kernel is another important special case, and is analogous to nested and cross-nested logit. The nested logit kernel model is specified as follows:

$$U_n = X_n\beta + F_n T \zeta_n + \nu_n \ ,$$

where: ζ_n is $(M \times 1)$, M is the number of nests, and one factor is defined for each nest.

F_n is $(J_n \times M)$, $f_{jm} = \begin{cases} 1 & \text{if alternative } j \text{ is a member of nest } m \\ 0 & \text{otherwise} \end{cases}$

T is $(M \times M)$ diagonal, which contains the standard deviation of each factor.

In a strictly hierarchical nesting structure (analogous to nested logit), the nests do not overlap, and $F_n F_n{}'$ is block diagonal. In a cross-nested structure, the alternatives can belong to more than one group. This section discusses in order the cases of 2 nests, 3 nests, more than 3 nests, cross-nested models, and other extensions.

Models with 2 Hierarchical Nests (2 levels, no cross-nesting)

The summary of identification for a 2 nest structure is that only 1 of the nesting parameters is identified. Furthermore, the normalization of the nesting parameter is arbitrary. This is best shown by example. Take a 5 alternative case (with universal choice set) in which the first 2 alternatives belong to one nest, and the last 3 alternatives belong to a different nest. The model is written as:

$$\begin{aligned} U_{1n} &= \ldots + \sigma_1\zeta_{1n} \qquad\qquad + \nu_{1n} \\ U_{2n} &= \ldots + \sigma_1\zeta_{1n} \qquad\qquad + \nu_{2n} \\ U_{3n} &= \ldots \qquad\qquad + \sigma_2\zeta_{2n} + \nu_{3n} \\ U_{4n} &= \ldots \qquad\qquad + \sigma_2\zeta_{2n} + \nu_{4n} \\ U_{5n} &= \ldots \qquad\qquad + \sigma_2\zeta_{2n} + \nu_{5n} \end{aligned} \quad , \quad \text{where:} \quad F = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}.$$

We denote this specification as 1, 1, 2, 2, 2 (a shorthand notation of the matrix F).

The covariance matrix is:

$$\Omega = \begin{bmatrix} \sigma_{11} + g/\mu^2 & & & & \\ \sigma_{11} & \sigma_{11} + g/\mu^2 & & & \\ 0 & 0 & \sigma_{22} + g/\mu^2 & & \\ 0 & 0 & \sigma_{22} & \sigma_{22} + g/\mu^2 & \\ 0 & 0 & \sigma_{22} & \sigma_{22} & \sigma_{22} + g/\mu^2 \end{bmatrix}.$$

And the covariance matrix of utility differences (with alternative 5 as the base) is as follows:

$$\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \sigma_{22} + 2g/\mu^2 & & & \\ \sigma_{11} + \sigma_{22} + g/\mu^2 & \sigma_{11} + \sigma_{22} + 2g/\mu^2 & & \\ g/\mu^2 & g/\mu^2 & 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix}$$

It can be seen from this matrix that only the sum $(\sigma_{11} + \sigma_{22})$ can be identified. This is verified by the rank condition as follows:

$$vecu(\Omega_\Delta) = \begin{bmatrix} \sigma_{11} + \sigma_{22} + 2g/\mu^2 \\ \sigma_{11} + \sigma_{22} + g/\mu^2 \\ g/\mu^2 \\ 2g/\mu^2 \end{bmatrix} \rightarrow \begin{matrix} \text{Jacobian} \\ \text{matrix} \end{matrix} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix} \rightarrow \text{RANK=2}$$

$\rightarrow$ can estimate 1 of the parameters; must normalize μ and one σ_{ii}.

Furthermore, unlike the heteroscedastic logit kernel model, either one of the variance terms can be normalized to zero (i.e., the normalization is arbitrary). This can be seen intuitively by noticing that only the sum $(\sigma_{11} + \sigma_{22})$ appears in Ω_Δ, and so it is always this sum that is estimated regardless of which term is set to zero. The parameters can also be constrained to be the same $(\sigma_{11} = \sigma_{22})$. This can also be verified via the Equality Condition, as follows.

$$\begin{bmatrix} (\dot{\sigma}_{11} + \dot{\sigma}_{22} + 2g)/\mu^2 & & & \\ (\dot{\sigma}_{11} + \dot{\sigma}_{22} + g)/\mu^2 & (\dot{\sigma}_{11} + \dot{\sigma}_{22} + 2g)/\mu^2 & & \\ g/\mu^2 & g/\mu^2 & 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix}$$

$$=\begin{bmatrix} (\dot{\sigma}_{11}^N+2g)/\mu^2 & & & \\ (\dot{\sigma}_{11}^N+g)/\mu^2 & (\dot{\sigma}_{11}^N+2g)/\mu^2 & & \\ g/\mu^2 & g/\mu^2 & 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix} \text{Normalization 1: } \dot{\sigma}_{22}^N=0$$

$$=\begin{bmatrix} (\dot{\sigma}_{22}^N+2g)/\mu^2 & & & \\ (\dot{\sigma}_{22}^N+g)/\mu^2 & (\dot{\sigma}_{22}^N+2g)/\mu^2 & & \\ g/\mu^2 & g/\mu^2 & 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix} \text{Normalization 2: } \dot{\sigma}_{11}^N=0$$

$$=\begin{bmatrix} (2\dot{\sigma}_{xx}^N+2g)/\mu^2 & & & \\ (2\dot{\sigma}_{xx}^N+g)/\mu^2 & (2\dot{\sigma}_{xx}^N+2g)/\mu^2 & & \\ g/\mu^2 & g/\mu^2 & 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix} \text{Normalization 3: } \dot{\sigma}_{11}^N=\dot{\sigma}_{22}^N=\dot{\sigma}_{xx}^N$$

These equalities will always hold as each sigma term is greater than or equal to zero, and the following is always true:

$$\begin{aligned}(\dot{\sigma}_{11}+\dot{\sigma}_{22}) &= \dot{\sigma}_{11}^N \text{ from normalization 1}\\ &= \dot{\sigma}_{22}^N \text{ from normalization 2}\\ &= 2\dot{\sigma}_{xx}^N \text{ from normalization 3}\end{aligned} \quad .$$

Therefore, while it is not possible to estimate both variance parameters of the 1, 1, 2, 2, 2 structure, the following structures are all identified and result in *identical* covariance structures (i.e., *identical* models):

{ 1, 1, 0, 0, 0 } = { 0, 0, 2, 2, 2 } = { 1, 1, 2, 2, 2 with $\sigma_1=\sigma_2$ } .

These results straightforwardly extend to all hierarchical two nest structures regardless of the number of alternatives (as long as at least one of the nests has 2 or more alternatives).

Relationship with Nested Logit

How is it that only 1 of the nesting parameters is identified in Logit Kernel, whereas 2 nesting parameters (the logsum parameters) are estimable in a 2-nest Nested Logit model? The answer is that the covariance structure shown above for the 2-nest Logit Kernel model is different than that of a Nested Logit model. Nested Logit is a homoscedastic model in that the diagonal elements of the covariance structure are identical to one another. This can be exactly mimicked with Logit Kernel via a more complicated error component structure than what was used above.

$$
\begin{aligned}
U_{1n} &= \ldots + \sigma_1\zeta_{1n} + \sigma_2\zeta_{3n} + \nu_{1n} \\
U_{2n} &= \ldots + \sigma_1\zeta_{1n} + \sigma_2\zeta_{4n} + \nu_{2n} \\
U_{3n} &= \ldots + \sigma_2\zeta_{2n} + \sigma_2\zeta_{5n} + \nu_{3n} \quad , \\
U_{4n} &= \ldots + \sigma_2\zeta_{2n} + \sigma_1\zeta_{6n} + \nu_{4n} \\
U_{5n} &= \ldots + \sigma_2\zeta_{2n} + \sigma_1\zeta_{7n} + \nu_{5n}
\end{aligned}
$$

$$
\text{where: } F = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_1 \end{bmatrix} .
$$

The covariance matrix is:

$$
\Omega = \begin{bmatrix} \sigma_{11} + \sigma_{22} + g/\mu^2 & & & & \\ \sigma_{11} & \sigma_{11} + \sigma_{22} + g/\mu^2 & & & \\ 0 & 0 & \sigma_{11} + \sigma_{22} + g/\mu^2 & & \\ 0 & 0 & \sigma_{22} & \sigma_{11} + \sigma_{22} + g/\mu^2 & \\ 0 & 0 & \sigma_{22} & \sigma_{22} & \sigma_{11} + \sigma_{22} + g/\mu^2 \end{bmatrix} .
$$

The covariance matrix of utility differences is:

$$
\Omega_\Delta = \begin{bmatrix} 2\sigma_{11} + 2\sigma_{22} + 2g/\mu^2 & & & \\ 2\sigma_{11} + \sigma_{22} + g/\mu^2 & 2\sigma_{11} + 2\sigma_{22} + 2g/\mu^2 & & \\ \sigma_{11} + g/\mu^2 & \sigma_{11} + g/\mu^2 & 2\sigma_{11} + 2g/\mu^2 & \\ \sigma_{11} + g/\mu^2 & \sigma_{11} + g/\mu^2 & \sigma_{11} + g/\mu^2 & 2\sigma_{11} + 2g/\mu^2 \end{bmatrix} .
$$

$$
vecu(\Omega_\Delta) = \begin{bmatrix} 2\sigma_{11} + 2\sigma_{22} + 2g/\mu^2 \\ 2\sigma_{11} + \sigma_{22} + g/\mu^2 \\ \sigma_{11} + g/\mu^2 \\ 2\sigma_{11} + 2g/\mu^2 \end{bmatrix} \rightarrow \begin{matrix} \text{Jacobian} \\ \text{matrix} \end{matrix} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \end{bmatrix} \rightarrow \text{RANK=3}
$$

$\rightarrow$ can estimate 2 of the parameters; only need to normalize μ.

So, in this way a Nested Logit model can be exactly mimicked and the correlation within each nest estimated, but at the cost of 5 degrees of integration!

Models with Three Hierarchical Nests (2 levels, no cross-nesting)

The summary of identification for models with 3 or more hierarchical nests is that *all* of the nesting parameters are identified. To show this, we will again look at a 5 alternative model, this time imposing a 3 nest structure (1, 1, 2, 3, 3):

$$\begin{aligned} U_{1n} &= ... + \sigma_1\zeta_{1n} + \nu_{1n} \\ U_{2n} &= ... + \sigma_1\zeta_{1n} + \nu_{2n} \\ U_{3n} &= ... + \sigma_2\zeta_{2n} + \nu_{3n} \\ U_{4n} &= ... + \sigma_3\zeta_{3n} + \nu_{4n} \\ U_{5n} &= ... + \sigma_3\zeta_{3n} + \nu_{5n} \end{aligned} \quad , \quad \text{where: } F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}.$$

The covariance matrix of utility differences is:

$$\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \sigma_{33} + 2g/\mu^2 & & & \\ \sigma_{11} + \sigma_{33} + g/\mu^2 & \sigma_{11} + \sigma_{33} + 2g/\mu^2 & & \\ \sigma_{33} + g/\mu^2 & \sigma_{33} + g/\mu^2 & \sigma_{22} + \sigma_{33} + 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix}.$$

A check of the rank condition verifies that all three variance parameters are identified:

$$vecu(\Omega_\Delta) = \begin{bmatrix} \sigma_{11} + \sigma_{33} + 2g/\mu^2 \\ \sigma_{11} + \sigma_{33} + g/\mu^2 \\ \sigma_{33} + g/\mu^2 \\ \sigma_{22} + \sigma_{33} + 2g/\mu^2 \\ g/\mu^2 \\ 2g/\mu^2 \end{bmatrix} \rightarrow \begin{array}{c}\text{Jacobian} \\ \text{matrix}\end{array} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix} \rightarrow \text{RANK=4}$$

$\rightarrow$ can estimate 3 of the parameters; only need to normalize μ.

It is an interesting result that 1, 1, 0, 2, 2 structure results in both variance parameters being identified (by virtue of having a 3 nest structure) whereas only one parameter of the 1, 1, 2, 2, 2 structure is identified.

Discussion of 2-Nest versus 3-Nest Identification Results

Conceptually, the number of estimable parameters can be thought of in terms of the number of differences and number of covariances that are left in the utility differences. In a two nest structure, only one difference remains and no covariances and therefore one parameter is estimable. Whereas in a three nest structure, there are two differences, plus the covariance between these two differences, and so three parameters are estimable.

Models with Three or More Hierarchical Nests (2 levels, no cross-nesting)

Th 3-nest finding can be extended to any model with 3 or more nests (where 'nests' can have only 1 alternative, as long as at least one nest has 2 or more alternatives) as follows. Without loss of generality, assume that the base alternative is a member of a nest with 2 or more alternatives (as in the example above). Define m_b as the group to which the base alternative belongs, and σ_{bb} as the variance associated with this base. Recall that M is the number of nests. The covariance matrix of utility differences has the following elements:

On the diagonal:

$\sigma_{ii} + \sigma_{bb} + 2g/\mu^2 \quad \forall\, i \notin m_b$,	*M-1* equations,	(18)
$2g/\mu^2$,	*1* equation.	(19)

On the off-diagonal:

$\sigma_{bb} + g/\mu^2$,	*1* equation,	(20)
g/μ^2 ,	irrelevant: a dependent equation,	
$\sigma_{ii} + \sigma_{bb} + g/\mu^2$ for some $i \notin m_b$,	irrelevant: a dependent equation.	

Equations (18) through (20) provide identification for all nesting parameters, and the remaining equations are dependent. In the two-nest case, Equation (20) does not exist, and therefore is an equation short of identification.

Cross-Nested Models

There are no general rules for identification and normalization of cross-nested structures, and one has to check the rank condition on a case-by-case basis. For example, in the five alternative case in which the third alternative belongs to both nests (1, 1, 1-2, 2, 2), the (non-differenced) covariance matrix is:

$$\Omega = \begin{bmatrix} \sigma_{11} + g/\mu^2 & & & & \\ \sigma_{11} & \sigma_{11} + g/\mu^2 & & & \\ \sigma_{11} & \sigma_{11} & \sigma_{11} + \sigma_{22} + g/\mu^2 & & \\ 0 & 0 & \sigma_{22} & \sigma_{22} + g/\mu^2 & \\ 0 & 0 & \sigma_{22} & \sigma_{22} & \sigma_{22} + g/\mu^2 \end{bmatrix}.$$

A check of the order and rank conditions would find that both of the parameters in this cross-nested structure are identified. However, note that the cross-nesting specification can have unintended consequences on the covariance matrix. For example, in the (1, 1, 1-2, 2, 2) specification shown above, the third alternative is forced to have the highest variance (at least in the levels form). There are numerous possible solutions. One is to add a set of heteroscedastic terms, another is to add factors such that all the alternative-specific variances are identical as with the following specification:

$$F = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_2 \end{bmatrix}.$$

The covariance matrix of utility differences for this structure is as follows:

$$\Omega_\Delta = \begin{bmatrix} 2\sigma_{11} + 2\sigma_{22} + 2g/\mu^2 & & & \\ 2\sigma_{11} + \sigma_{22} + g/\mu^2 & 2\sigma_{11} + 2\sigma_{22} + 2g/\mu^2 & & \\ 2\sigma_{11} + g/\mu^2 & 2\sigma_{11} + g/\mu^2 & 2\sigma_{11} + 2g/\mu^2 & \\ \sigma_{11} + g/\mu^2 & \sigma_{11} + g/\mu^2 & \sigma_{11} + g/\mu^2 & 2\sigma_{11} + 2g/\mu^2 \end{bmatrix}.$$

A check of the rank condition verifies that both variance parameters are identified for this specification.

$$vecu(\Omega_\Delta) = \begin{bmatrix} 2\sigma_{11} + 2\sigma_{22} + 2g/\mu^2 \\ 2\sigma_{11} + \sigma_{22} + g/\mu^2 \\ 2\sigma_{11} + g/\mu^2 \\ 2\sigma_{11} + 2g/\mu^2 \\ \sigma_{11} + g/\mu^2 \\ 2\sigma_{11} + 2g/\mu^2 \end{bmatrix} \rightarrow \begin{matrix}\text{Jacobian} \\ \text{matrix}\end{matrix} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \end{bmatrix} \rightarrow \text{RANK=3}$$

$\rightarrow$ can estimate 2 of the parameters, only need to normalize μ.

Extensions to Nested Models

There are various complexities that can be introduced to the nesting structure, including multi-level nests, cross-nested structures with multiple dimensions, and unknown parameters in the loading matrix *(F)*. While we have investigated various special cases of these extended models, we have not yet derived general rules for identification. We recommend that identification be performed automatically on a case-by-case basis by working through the rank and order conditions. More importantly, research should be directed towards establishing more general findings.

Identification of Logit Kernel Models with Panel Data

Another situation for which the Logit Kernel model is being used is for panel data, in which multiple choices $\left(p=1,...,P_n\right)$ are observed for a given individual. Typically, Equation (4) would be modified for a panel data context by making the choice a choice at a particular time period by adding subscripts *p* (or, more often *t*) to denote the time period of the choice and the explanatory variables for that choice. Since the interest here is covariance structure of the utilities, we will modify equation (4) such that there are J_nP_n utilities for each individual as follows:

$$U_n = X_n\beta + F_nT\zeta_n + \nu_n \ ,$$

$$\text{cov}(U_n) = F_nTT'F_n' + (g/\mu^2)I_{J_n}$$

where: U_n is a $(J_nP_n \times 1)$ vector of utilities;

X_n is a $(J_nP_n \times K)$ matrix of explanatory variables;

β is $(K \times 1)$;

F_n is $(J_nP_n \times M)$;

T is $(M \times M)$;

ζ_n is $(M \times 1)$; and

ν_n is $(J_nP_n \times 1)$.

The key in terms of identification is that the covariance matrix to be examined for identification is now of dimension $J_nP_n \times J_nP_n$. Unsurprisingly, this enables potentially many more disturbance parameters to be estimated. This is suggested by the Order Condition alone, which states that the maximum number of (alternative-specific) disturbance parameters may be as high as $JP(JP-1)/2-1$.

We are going to describe identification of two specific ways (of many) that Logit Kernel can be used in the panel data case. The first is an agent effect, in which it is assumed that the disturbances of an alternative are correlated across time periods. Bhat and Gossen (2003) used such a specification in a model of weekend activity type choice. The second is when a factor is included to represent a latent characteristic of the individual, such as environmental friendliness and health consciousness. Toledo (2003) used such a specification in a model of driver following and lane switching behavior, in which the latent factor represented aggressiveness.

In both cases described below, we will use the situation of 3 alternatives ($J=3$) and two time periods ($P=3$).

Agent Effect

The idea of an agent effect is that what is unobserved for one individual in one time period is probably the same as what is unobserved for the same individual in another time period. This is implemented as having person and alternative specific covariances that are repeated in all time periods for any given individual. The model is written as follows:

$$\begin{aligned} U_{\text{int}} = U_{1n1} &= ...+\sigma_1\zeta_{1n}+\nu_{1n} \\ U_{2n1} &= ...+\sigma_2\zeta_{2n}+\nu_{2n} \\ U_{3n1} &= ...+\sigma_3\zeta_{3n}+\nu_{3n} \\ U_{1n2} &= ...+\sigma_1\zeta_{1n}+\nu_{4n} \\ U_{2n2} &= ...+\sigma_2\zeta_{2n}+\nu_{5n} \\ U_{3n2} &= ...+\sigma_3\zeta_{3n}+\nu_{6n} \end{aligned}, \quad F=\begin{bmatrix} 1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&1&0\\0&0&1 \end{bmatrix} \text{ and } T=\begin{bmatrix}\sigma_1&0&0\\0&\sigma_2&0\\0&0&\sigma_3\end{bmatrix}.$$

The covariance matrix is:

$$\Omega=\begin{bmatrix} \sigma_{11}+g/\mu^2 &&&&& \\ 0 & \sigma_{22}+g/\mu^2 &&&& \\ 0 & 0 & \sigma_{33}+g/\mu^2 &&& \\ \sigma_{11} & 0 & 0 & \sigma_{11}+g/\mu^2 && \\ 0 & \sigma_{22} & 0 & 0 & \sigma_{22}+g/\mu^2 & \\ 0 & 0 & \sigma_{33} & 0 & 0 & \sigma_{33}+g/\mu^2 \end{bmatrix},$$

and the covariance matrix of utility differences is:

$$\Omega_\Delta=\begin{bmatrix} \sigma_{11}+\sigma_{33}+2g/\mu^2 &&& \\ \sigma_{33}+g/\mu^2 & \sigma_{22}+\sigma_{33}+2g/\mu^2 && \\ \sigma_{11}+\sigma_{33} & \sigma_{33} & \sigma_{11}+\sigma_{33}+2g/\mu^2 & \\ \sigma_{33} & \sigma_{22}+\sigma_{33} & \sigma_{33}+g/\mu^2 & \sigma_{22}+\sigma_{33}+2g/\mu^2 \end{bmatrix}$$

Application of the Rank Condition leads to:

$$vecu(\Omega_\Delta)=\begin{bmatrix} \sigma_{11}+\sigma_{33}+2g/\mu^2 \\ \sigma_{33}+g/\mu^2 \\ \sigma_{22}+\sigma_{33}+g/\mu^2 \\ \sigma_{11}+\sigma_{33} \\ \sigma_{33} \\ \sigma_{22}+\sigma_{33} \end{bmatrix} \rightarrow \begin{matrix}\text{Jacobian}\\ \text{matrix}\end{matrix} = \begin{bmatrix} 1&0&1&2\\0&0&1&1\\0&1&1&1\\1&0&1&0\\0&0&1&0\\0&1&1&0 \end{bmatrix} \rightarrow \text{RANK=4}.$$

All three sigmas are identified, and only the variance of the Gumbel has to be normalized. Unlike in the cross-sectional heteroscedastic case, the interactions between choices for the same individual (the equations in the lower left quadrant of the variance-covariance matrices) gets us the third parameter.

Furthermore, since only the differences in utilities matter, the case above is identical to one in which:

$$\begin{aligned} U_{\text{int}} = U_{1n1} &= \ldots + \sigma_1\zeta_{1n} + \sigma_3\zeta_{3n} + \nu_{1n} \\ U_{2n1} &= \ldots + \sigma_2\zeta_{2n} + \sigma_3\zeta_{3n} + \nu_{2n} \\ U_{3n1} &= \ldots \qquad\qquad\qquad\quad + \nu_{3n} \\ U_{1n2} &= \ldots + \sigma_1\zeta_{1n} + \sigma_3\zeta_{3n} + \nu_{4n} \\ U_{2n2} &= \ldots + \sigma_2\zeta_{2n} + \sigma_3\zeta_{3n} + \nu_{5n} \\ U_{3n2} &= \ldots \qquad\qquad\qquad\quad + \nu_{6n} \end{aligned}, \quad F = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } T = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix},$$

and also identical to:

$$\begin{aligned} U_{\text{int}} = U_{1n1} &= \ldots + \alpha_1\zeta_{1n} \qquad\qquad\quad + \nu_{1n} \\ U_{2n1} &= \ldots + \alpha_3\zeta_{1n} + \alpha_2\zeta_{2n} + \nu_{2n} \\ U_{3n1} &= \ldots \qquad\qquad\qquad\quad + \nu_{3n} \\ U_{1n2} &= \ldots + \alpha_1\zeta_{1n} \qquad\qquad\quad + \nu_{4n} \\ U_{2n2} &= \ldots + \alpha_3\zeta_{1n} + \alpha_2\zeta_{2n} + \nu_{5n} \\ U_{3n2} &= \ldots \qquad\qquad\qquad\quad + \nu_{6n} \end{aligned}, \quad F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$T = \begin{bmatrix} \alpha_1 & 0 \\ \alpha_3 & \alpha_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\sigma_{11} + \sigma_{33}} & 0 \\ \dfrac{\sigma_{33}}{\sqrt{\sigma_{11} + \sigma_{33}}} & \dfrac{\sqrt{\sigma_{11}\sigma_{22} + \sigma_{11}\sigma_{33} + \sigma_{22}\sigma_{33}}}{\sqrt{\sigma_{11} + \sigma_{33}}} \end{bmatrix}.$$

Only the differences matter, and since all three cases lead to the same variance matrix of utility differences, all three cases are statistically identical. The fact that they each say something different at the level of utility differences is irrelevant, as, statistically, the levels form does not exist. The value of recognizing the equality of the different specifications include not wasting effort by exploring redundant specification and also, more likely than not, one of the possible specifications will be easier to program than another.

Furthermore, note that additional parameters are also estimable beyond the three in the model above. For example, a choice-specific covariance could be added as follows:

$$
\begin{aligned}
U_{\text{int}} = U_{1n1} &= \ldots + \sigma_1\zeta_{1n} + \sigma_4\zeta_{4n} + \nu_{1n} \\
U_{2n1} &= \ldots + \sigma_2\zeta_{2n} + \sigma_4\zeta_{4n} + \nu_{2n} \\
U_{3n1} &= \ldots + \sigma_3\zeta_{3n} \qquad\qquad + \nu_{3n} \\
U_{1n2} &= \ldots + \sigma_1\zeta_{1n} + \sigma_4\zeta_{5n} + \nu_{4n} \\
U_{2n2} &= \ldots + \sigma_2\zeta_{2n} + \sigma_4\zeta_{5n} + \nu_{5n} \\
U_{3n2} &= \ldots + \sigma_3\zeta_{3n} \qquad\qquad + \nu_{6n}
\end{aligned}
, \quad F = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}
$$

$$
T = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 \\ 0 & 0 & 0 & 0 & \sigma_4 \end{bmatrix}.
$$

$$
\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \sigma_{33} + \sigma_{44} + 2g/\mu^2 & & & \\ \sigma_{33} + \sigma_{44} + g/\mu^2 & \sigma_{22} + \sigma_{33} + \sigma_{44} + 2g/\mu^2 & & \\ \sigma_{11} + \sigma_{33} & \sigma_{33} & \sigma_{11} + \sigma_{33} + \sigma_{44} + 2g/\mu^2 & \\ \sigma_{33} & \sigma_{22} + \sigma_{33} & \sigma_{33} + \sigma_{44} + g/\mu^2 & \sigma_{22} + \sigma_{33} + \sigma_{44} + 2g/\mu^2 \end{bmatrix}
$$

$$
vecu(\Omega_\Delta) = \begin{bmatrix} \sigma_{11} + \sigma_{33} + \sigma_{44} + 2g/\mu^2 \\ \sigma_{33} + \sigma_{44} + g/\mu^2 \\ \sigma_{22} + \sigma_{33} + \sigma_{44} + g/\mu^2 \\ \sigma_{11} + \sigma_{33} \\ \sigma_{33} \\ \sigma_{22} + \sigma_{33} \end{bmatrix} \rightarrow \begin{array}{c}\text{Jacobian} \\ \text{matrix}\end{array} = \begin{bmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \rightarrow \text{RANK=5}
$$

Therefore all 4 parameters are identified. However, our experience has been that real data have trouble supporting such an intricate specification, even if the specification is theoretically identifiable. Is it necessary to constrain the parameter associated with ς_4 to be equal to that of ς_5 ? Check the rank condition.

Latent Variable

This is the case in which a latent factor, which is a characteristic of the individual, is included in the model. The idea here is that there is a property of the individual (such as aggressiveness or environmentally friendliness) that is important to the choice behavior, but is unobserved to the analyst. This latent characteristic can be modeled as a factor ς_1. This characteristic enters all utilities for the person (just as another characteristic such as Income would enter the utilities), and the impact of this characteristic on each utility varies as denoted by the different factor loadings f_1, f_2, and f_3 as follows:

$$\begin{aligned} U_{\text{int}} = U_{1n1} &= ... + f_1\zeta_{1n} + \nu_{1n} \\ U_{2n1} &= ... + f_2\zeta_{1n} + \nu_{2n} \\ U_{3n1} &= ... + f_3\zeta_{1n} + \nu_{3n} \\ U_{1n2} &= ... + f_1\zeta_{1n} + \nu_{4n} \\ U_{2n2} &= ... + f_2\zeta_{1n} + \nu_{5n} \\ U_{3n2} &= ... + f_3\zeta_{1n} + \nu_{6n} \end{aligned}, \quad F = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_1 \\ f_2 \\ f_1 \end{bmatrix} \text{ and } T = [1].$$

The covariance matrix is:

$$\Omega = \begin{bmatrix} f_{11} + g/\mu^2 & & & & & \\ f_1 f_2 & f_{22} + g/\mu^2 & & & & \\ f_1 f_3 & f_2 f_3 & f_{33} + g/\mu^2 & & & \\ f_{11} & f_1 f_2 & f_1 f_3 & f_{11} + g/\mu^2 & & \\ f_1 f_2 & f_{22} & f_2 f_3 & f_1 f_2 & f_{22} + g/\mu^2 & \\ f_1 f_3 & f_2 f_3 & f_{33} & f_1 f_3 & f_2 f_3 & f_{33} + g/\mu^2 \end{bmatrix},$$

and the covariance matrix of utility differences is:

$$\Omega_\Delta = \begin{bmatrix} (f_1 - f_3)^2 + 2g/\mu^2 & & & \\ (f_1 - f_3)(f_2 - f_3) + g/\mu^2 & (f_2 - f_3)^2 + 2g/\mu^2 & & \\ (f_1 - f_3)^2 & (f_1 - f_3)(f_2 - f_3) & (f_1 - f_3)^2 + 2g/\mu^2 & \\ (f_1 - f_3)(f_2 - f_3) & (f_2 - f_3)^2 & (f_1 - f_3)(f_2 - f_3) + g/\mu^2 & (f_2 - f_3)^2 + 2g/\mu^2 \end{bmatrix}$$

.

$$vecu(\Omega_\Delta) = \begin{bmatrix} (f_1-f_3)^2+2g/\mu^2 \\ (f_1-f_3)(f_2-f_3)+g/\mu^2 \\ (f_2-f_3)^2+2g/\mu^2 \\ (f_1-f_3)^2 \\ (f_1-f_3)(f_2-f_3) \\ (f_2-f_3)^2 \end{bmatrix} \rightarrow$$

$$\begin{matrix}\text{Jacobian} \\ \text{matrix}\end{matrix} = \begin{bmatrix} 2(f_1-f_3) & 0 & -2(f_1-f_3) & 2 \\ (f_2-f_3) & (f_1-f_3) & -(f_1-f_3)-(f_2-f_3) & 1 \\ 0 & 2(f_2-f_3) & -2(f_2-f_3) & 2 \\ 2(f_1-f_3) & 0 & -2(f_1-f_3) & 0 \\ (f_2-f_3) & (f_1-f_3) & -(f_1-f_3)-(f_2-f_3) & 0 \\ 0 & 2(f_2-f_3) & -2(f_2-f_3) & 0 \end{bmatrix} \rightarrow \text{RANK=3}$$

$\rightarrow$ can estimate only 2 of the parameters, need to normalize one f and μ.

Application of the Equality Condition will show that the normalization is arbitrary, that is, any one of the f s can be normalized. Intuitively, this is because unlike the heteroscedastic case in which the parameters (σ_{11}, σ_{22}, ..., σ_{JJ}) were restricted to be non-negative, the parameters in this agent effect case (f_1, f_2, f_3) are not restricted. In fact, the parameters perform like the parameters of a socio-economic variable X_n such as income, where there is a base alternative (the one with $f_i = 0$) and the other parameters can be either positive or negative to reflect relative preferences given a particular value of X_n.

Estimation Results

In this section, a series of estimation results are presented that empirically verify the identification claims made above for the heteroscedastic and nesting cases. Both synthetic data and real data are used.

Synthetic Data: Heteroscedasticity

The first application concerns the heteroscedastic case, in which the findings were that one heteroscedastic variance must be normalized and the choice of which to normalize is not arbitrary. A synthetic dataset are used that consist of a choice situation among three alternatives. The model specification is as follows.

$$\begin{aligned} U_{1n} &= \alpha_1 + X_{1n}\beta + \sigma_1\zeta_{1n} + \nu_{1n} \quad , \\ U_{2n} &= \alpha_2 + X_{2n}\beta + \sigma_2\zeta_{2n} + \nu_{2n} \; , \\ U_{3n} &= \qquad\; X_{3n}\beta + \sigma_3\zeta_{3n} + \nu_{3n} \; . \end{aligned}$$

The true parameter values used to generate the synthetic data are:

$$\alpha_1 = 1.5,\ \ \alpha_2 = 0.5,\ \ \beta = -1,\ \ \sigma_1 = 3,\ \ \sigma_2 = 2,\ \ \sigma_3 = 1, \text{ and } \mu = 1.$$

The explanatory variable, X, is simulated as a normal variable with a standard deviation of 3, independent across alternatives and observations. The utilities for each observation are generated by drawing a single random draw for each ζ_{jn} from independent standard normal distributions and each ν_{jn} from independent standard Gumbel distributions. The utilities are calculated, and the alternative with the highest utility is then the chosen alternative.

Estimation results using the synthetic data are provided in Table 1, which presents estimation results regarding selecting and setting the base heteroscedastic term. Recall that only $J-1$ heteroscedastic terms are identified, and that it is necessary to either set the minimum variance term to zero, or set any of the other variance terms high enough according to Equation (17), which was derived earlier:

$$\dot{\sigma}_{ff}^{N} \geq \left(\dot{\sigma}_{jj} - \dot{\sigma}_{ii}\right) g \Big/ (g + \dot{\sigma}_{ii}) \quad ,\ i = 1,\ldots,J\ ,$$

where $\dot{\sigma}_{jj}$ is the theoretical (true) variance that is fixed to the value $\dot{\sigma}_{ff}^{N}$.

All of the models in Table 1a are estimated with 10,000 observations and 500 Halton draws. Walker (2001) provides estimation results that verify that 500 Halton draws are sufficient for this model. The first model shows estimation results for an unidentified model; this model is used to determine the minimum variance alternative, and it correctly identifies the third alternative as having minimum variance.[2] Models 2 through 4 show identified models in which the minimum variance alternative is constrained to different values (0, 1, and 2); as expected, the log-likelihoods of these models are basically equivalent and all of these represent correct specifications. Models 5 through 10 show identified models in which the maximum variance alternative is constrained to different values (0, 1, 1.5, 2.25, 3, and 4). Applying Equation (17) (repeated above), the model specification will be correct as long as σ_1 is constrained to a value above 2.2. The empirical results verify this. First, there is a severe loss of fit when the σ_1 is constrained below 2.2. Second, the parameter estimates for the mis-specified models are biased. This can be seen by examining the ratio of the systematic parameters (for example, β / α_1) across models. While the scale shifts for various normalizations (and therefore the parameter estimates also shift), the ratio of systematic parameters should remain constant across normalizations. A cursory examination of the estimation results shows that these ratios begin to drift with successively invalid normalizations. Finally, note that these results indicate a slight loss of fit when the base alternative is constrained to a high value (σ_3 =2 and σ_1 =4), and this is due to the

[2] We were able to calculate t-statistics for the unidentified model here (and elsewhere) for two reasons. First, simulation has the tendency to mask identification issues, and therefore does not always result in a singular Hessian for a finite number of draws. Second, the slight difference between the Gumbel and Normal distributions makes the unidentified model only 'nearly' singular, and not perfectly singular.

issue regarding the slight difference between the Gumbel and normal distributions. It must be emphasized that the normalization in heteroscedastic logit kernel models is not arbitrary.

Table 1: Selecting and Setting the Base Alternative in a Heteroscedastic Model (Synthetic Data with 3 Alternatives)

Parameter	True Value	Unidentified		Identified: Minimum Variance Base						Identified: Maximum Variance Base											
		Est	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
$\alpha 1$	1.5	1.27	*(3.4)*	1.24	*(15.7)*	1.51	*(15.9)*	2.18	*(15.9)*	0.97	*(29.1)*	1.02	*(27.9)*	1.08	*(23.4)*	1.24	*(5.8)*	1.57	*(17.2)*	2.03	*(17.4)*
$\alpha 2$	0.5	0.43	*(2.6)*	0.42	*(8.9)*	0.53	*(9.2)*	0.76	*(9.2)*	0.37	*(11.1)*	0.40	*(11.5)*	0.41	*(10.4)*	0.42	*(2.2)*	0.54	*(6.8)*	0.70	*(7.0)*
β	-1.0	-0.80	*(3.8)*	-0.78	*(14.6)*	-0.94	*(14.1)*	-1.36	*(13.7)*	-0.51	*(55.5)*	-0.57	*(65.0)*	-0.64	*(39.1)*	-0.78	*(16.0)*	-0.98	*(37.1)*	-1.27	*(37.1)*
$\sigma 1$	3.0	2.32	*(2.9)*	2.24	*(9.7)*	2.84	*(10.3)*	4.30	*(11.0)*	*0.00*	---	*1.00*	---	*1.50*	---	*2.25*	---	*3.00*	---	*4.00*	---
$\sigma 2$	2.0	1.27	*(1.9)*	1.21	*(4.7)*	1.69	*(5.9)*	2.80	*(7.7)*	0.06	*(0.1)*	0.03	*(0.3)*	0.50	*(1.8)*	1.22	*(6.6)*	1.82	*(11.7)*	2.58	*(14.5)*
$\sigma 3$	1.0	0.35	*(0.2)*	0.00	---	1.00	---	2.00	---	0.00	*(0.9)*	0.00	*(1.6)*	0.01	*-(0.5)*	0.16	*(0.0)*	1.07	*(4.4)*	1.78	*(7.6)*
(Simul.) Log-Likelihood:		-6837		-6837		-6837		-6838		*-6907*		*-6865*		-6845		-6837		-6837		-6838	
	Model:	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*		*10*	

Empirical Application: Telephone Service

In this section, we apply these methods to residential telephone demand analysis. The model involves a choice among five residential telephone service options for local calling. A household survey was conducted in 1984 for a telephone company and was used to develop a comprehensive model system to predict residential telephone demand (Train, McFadden and Ben-Akiva, 1987). Below we use part of the data to estimate a model that explicitly accounts for inter-dependencies between residential telephone service options. We first describe the data. Then we present estimation results using a variety of error structures.

The Data

Local telephone service typically involves the choice between flat (i.e., a fixed monthly charge for unlimited calls within a specified geographical area) and measured (i.e., a reduced fixed monthly charge for a limited number of calls plus usage charges for additional calls) services. In the current application, five services are involved, two measured and three flat. They can be described as follows:

- *Budget measured* - no fixed monthly charge; usage charges apply to each call made.
- *Standard measured* - a fixed monthly charge covers up to a specified dollar amount (greater that the fixed charge) of local calling, after which usage charges apply to each call made.
- *Local flat* - a greater monthly charge that may depend upon residential location; unlimited free calling within local calling area; usage charges apply to calls made outside local calling area.
- *Extended area flat* - a further increase in the fixed monthly charge to permit unlimited free calling within an extended area.

- *Metro area flat* - the greatest fixed monthly charge that permits unlimited free calling within the entire metropolitan area.

The sample concerns 434 households. The availability of the service options of a given household depends on its geographical location. Details are provided in Table 2. In Table 3, we summarize the service option availabilities over the usable sample.

Table 2: Telephone Data - Availability of Service Options

Service Options	Geographic Location		
	Metropolitan Areas	Perimeter Exchanges Adjacent to Metro Areas	All Other
Budget Measured	Yes	Yes	Yes
Standard Measured	Yes	Yes	Yes
Local Flat	Yes	Yes	Yes
Extended Flat	No	Yes	No
Metro Flat	Yes	Yes	No

Table 3: Telephone Data - Summary Statistics on Availability of Service Options

Service Options	Chosen	Percent	Total Available
Budget Measured	73	0.168	434
Standard Measured	123	0.283	434
Local Flat	178	0.410	434
Extended Flat	3	0.007	13
Metro Flat	57	0.131	280
Total :	434	1.000	1595

Models

The model that we use in the present analysis is intentionally specified to be simple. The explanatory variables used to explain the choice between the five service options are four alternative-specific constants, which correspond to the first four service options, and a generic cost variable (the natural log of the monthly cost of each service options expressed in dollars). We investigated three types of error structures: heteroscedasticity, nested, and cross-nested structures. Unless otherwise noted, all of the logit kernel specifications were estimated using 1000 Halton draws; Walker (2001) provides evidence that this is a sufficient number of draws.

Heteroscedastic

The results for the heteroscedastic case are provided in Table 4. Model 1 is simply the multinomial logit specification. The rest are logit kernel models. Model 2 is the unidentified model. The results of the unidentified model suggest that there is no strong base alternative, and it could be either alternative 1, 2, 4, or 5. Models 3-8 are estimation results for identified heteroscedastic models. Again, to explore the issue of the minimum variance alternatives, 5 identified models were estimated, each one with a different base heteroscedastic term. (Note that this defeats the purpose of estimating the unidentified model, but was done for illustration purposes only.) As indicated by the unidentified models, the identified model estimation results support the conclusion that any of alternatives 1, 2, 4, or 5 could be set as the base. However, constraining σ_3 to zero results in a significant loss of fit as demonstrated by model 5, whereas constraining it to 4.0 brings it in line with the correctly specified model as demonstrated by model 8. Comparing the correctly specified heteroscedastic models with the MNL model, there is an obvious gain in likelihood from incorporating heteroscedasticity, primarily due to capturing the high variance of alternative 3.

Table 4: Telephone Model - Heteroscedastic Specification (1000 Halton Draws)

	MNL		Unidentified Heteroscedastic		Identified Heteroscedastic Model											
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants																
Budget Measured (1)	-2.46	*(8.4)*	-3.28	*(7.7)*	-3.27	*(7.9)*	-3.27	*(7.1)*	-5.03	*(2.4)*	-3.28	*(6.0)*	-3.27	*(7.8)*	-3.91	*(2.2)*
Standard Measured (2)	-1.74	*(6.6)*	-2.53	*(6.4)*	-2.53	*(6.6)*	-2.52	*(6.2)*	-3.85	*(2.2)*	-2.53	*(6.1)*	-2.52	*(6.5)*	-3.02	*(2.4)*
Local Flat (3)	-0.54	*(2.7)*	-1.37	*(3.6)*	-1.37	*(3.8)*	-1.36	*(3.2)*	-1.09	*(2.1)*	-1.37	*(3.6)*	-1.36	*(3.7)*	-1.67	*(3.3)*
Extended Flat (4)	-0.74	*(1.1)*	-1.04	*(1.3)*	-1.04	*(1.3)*	-1.04	*(1.3)*	-1.37	*(1.5)*	-1.04	*(1.4)*	-1.04	*(1.4)*	-1.10	*(1.2)*
Log Cost	-2.03	*(9.6)*	-2.68	*(8.2)*	-2.68	*(8.2)*	-2.67	*(4.9)*	-3.24	*(3.1)*	-2.68	*(6.2)*	-2.67	*(8.2)*	-3.33	*(2.9)*
$\sigma 1$			0.03	*(0.2)*			0.02	*(0.1)*	2.77	*(1.8)*	0.03	*(0.0)*	0.03	*(0.3)*	0.76	*(0.4)*
$\sigma 2$			0.14	*(0.4)*	0.13	*(0.3)*			3.27	*(1.6)*	0.14	*(0.1)*	0.14	*(0.3)*	0.70	*(0.3)*
$\sigma 3$			2.88	*(3.4)*	2.88	*(4.9)*	2.88	*(2.4)*			2.88	*(3.3)*	2.87	*(3.8)*	4.00	----
$\sigma 4$			0.04	*(0.1)*	0.04	*(0.1)*	0.04	*(0.1)*	1.14	*(0.5)*			0.04	*(0.1)*	0.11	*(0.1)*
$\sigma 5$			0.09	*(0.3)*	0.09	*(0.3)*	0.09	*(0.2)*	0.01	*(0.0)*	0.10	*(0.0)*			1.33	*(1.3)*
(Simul.) Log-Likelihood:	-477.56		-471.20		-471.20		-471.20		-476.66		-471.20		-471.20		-471.42	
Model:	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*	

Nested & Cross-Nested Structures

In Table 5, the estimation results of various nested and cross-nested specifications are provided. Table 5a reports results for identified model structures (as can be verified by the rank condition). The best specification is model 3, in which the first two alternatives are nested, the last two alternatives are nested, and the third term has a heteroscedastic term. This provides a significant improvement in fit over the MNL specification shown in the first column of Table 4, and also provides a better fit than the heteroscedastic models in Table 4. The poor fit for many of the nesting and cross-nesting specifications is due to the fact that the variance for alternative 3 is constrained to be in line with the other variances. The heteroscedastic models indicated that it has a much higher variance. Table 5b shows that when an additional variance term for alternative 3 is added to the nested and cross-nested models the fit improves further.[3]

Table 5c provides results for the unidentified model in which the first two alternatives are nested and the last 3 alternatives are nested, and we attempt (incorrectly) to estimate both error parameters. The first model, estimated with 1,000 Halton draws, appears to be identified. However, the second model, estimated using different starting values, shows that this is not the

[3] Therefore, the problem identified earlier with the cross-nested *1, 1, 1-2, 2, 2* structure does not apply to this dataset. In fact, as shown by the models in Table 5c, alternative 3 has an even larger relative variance than the *1, 1, 1-2, 2, 2* structure provides.

case; it has an identical fit, but very different estimates of the error parameters. This is as expected, because only the sum of the variances $(\sigma_1^2 + \sigma_2^2)$ can be identified, which are shown to be identical in all the models (equal to 3.05). The remaining columns show that it can take a very large number of draws to get the telltale sign of an unidentified model, the singular Hessian – in this case, 80,000 Halton draws. (Again, the actual number depends on the specification and the data, and part of the complication is the slight difference between the Gumbel and Normal distributions.) Table 5d shows that the normalization for the 2 nest model is arbitrary. The table presents three normalizations resulting in identical fits where { 1, 1, 0, 0, 0 } = { 0, 0, 2, 2, 2 } = { 1, 1, 2, 2, 2 with $\sigma_1 = \sigma_2$ }.

Table 5: Telephone Model - Nested & Cross-Nested Error Structures

Table a: Identified Nesting & Cross-Nesting Error Structures

	Nested Structures								Cross-Nested Structures			
Specification*:	1, 1, 2, 2, 0		1, 1, 2, 2, 3		1, 1, 2, 3, 3		1, 1, 2, 2, 2 (σ1= σ2)		1, 1, 1-2, 2, 2		1-2, 2-3, 3-4, 4-5, 5-6 (all σ equal)	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants												
Budget Measured (1)	-3.63	*(5.0)*	-3.63	*(5.0)*	-3.79	*(5.4)*	-3.80	*(5.7)*	-3.80	*(5.7)*	-2.83	*(2.4)*
Standard Measured (2)	-2.85	*(4.3)*	-2.85	*(4.3)*	-3.00	*(4.6)*	-3.01	*(4.9)*	-3.00	*(4.9)*	-1.90	*(3.1)*
Local Flat (3)	-1.48	*(3.1)*	-1.48	*(3.1)*	-1.63	*(3.1)*	-1.09	*(3.6)*	-1.09	*(3.5)*	-0.55	*(2.3)*
Extended Flat (4)	-1.52	*(1.5)*	-1.52	*(1.5)*	-1.18	*(1.3)*	-1.19	*(1.4)*	-1.19	*(1.4)*	-0.76	*(1.0)*
Log Cost	-3.05	*(4.5)*	-3.05	*(4.5)*	-3.19	*(5.0)*	-3.25	*(6.1)*	-3.25	*(6.1)*	-2.40	*(2.1)*
σ1	1.32	*(1.1)*	1.32	*(1.1)*	1.55	*(1.5)*	2.16	*(3.0)*	0.01	*(0.8)*	0.65	*(0.6)*
σ2	3.02	*(2.9)*	3.02	*(2.9)*	3.34	*(2.9)*			3.04	*(3.0)*		
σ3			0.00	*(0.0)*	0.01	*(0.1)*						
(Simul.) Log-Likelihood:	-471.26		-471.26		-470.70		-473.04		-473.05		-477.48	

Table b: Nesting / Cross-Nesting plus Heteroscedasticity (0, 0, 1, 0, 0)

	Combined Models					
Specification*:	2, 2, 1-3, 3, 3 ($\sigma2 = \sigma3$)		2, 2, 2-1-3, 3, 3		2-3, 3-4, 4-1-5, 5-6, 6-7 ($\sigma2 \ldots \sigma7$ equal)	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants						
Budget Measured (1)	-3.81	*(5.5)*	-3.80	*(5.3)*	-3.28	*(7.3)*
Standard Measured (2)	-3.02	*(4.7)*	-3.01	*(4.6)*	-2.53	*(6.3)*
Local Flat (3)	-1.64	*(3.1)*	-1.64	*(3.1)*	-1.37	*(3.5)*
Extended Flat (4)	-1.19	*(1.3)*	-1.18	*(1.3)*	-1.04	*(1.3)*
Log Cost	-3.21	*(5.2)*	-3.20	*(5.0)*	-2.68	*(8.0)*
$\sigma1$	3.37	*(2.8)*	3.38	*(2.8)*	2.88	*(3.3)*
$\sigma2$	1.11	*(1.6)*	0.03	*(0.3)*	0.09	*(0.2)*
$\sigma3$			1.55	*(1.6)*		
(Simul.) Log-Likelihood:	-470.64		-470.69		-471.22	

Table c: Unidentified Nested Error Structures

Specification*:	1, 1, 2, 2, 2 (Unidentified - can only estimate ($\sigma1^2+\sigma2^2$))				10000 Halton Draws		80000 Halton Draws	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants								
Budget Measured (1)	-3.80	*(5.7)*	-3.80	*(5.7)*	-3.80	*(5.7)*	-3.80	*n/a*
Standard Measured (2)	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*n/a*
Local Flat (3)	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*n/a*
Extended Flat (4)	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*n/a*
Log Cost	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*n/a*
$\sigma1$	2.65	*(3.1)*	0.78	*(0.5)*	2.55	*(2.5)*	1.93	*n/a*
$\sigma2$	1.51	*(2.2)*	2.95	*(3.3)*	1.67	*(3.8)*	2.36	*n/a*
$(\sigma1^2+\sigma2^2)^{1/2}$	*3.05*		*3.05*		*3.05*		*3.05*	
(Simul.) Log-Likelihood:	-473.02		-472.99		-473.02		-473.02	

Table d: Identical (Identified) Nested Error Structures

Specification*:	1, 1, 0, 0, 0		0, 0, 2, 2, 2		1, 1, 2, 2, 2 ($\sigma_1=\sigma_2$)	
Parameter	*Est*	*T-stat*	*Est*	*T-stat*	*Est*	*T-stat*
Altern. Specific constants						
Budget Measured (1)	-3.80	*(5.7)*	-3.80	*(5.7)*	-3.80	*(5.7)*
Standard Measured (2)	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*(4.9)*
Local Flat (3)	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*(3.6)*
Extended Flat (4)	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*(1.4)*
Log Cost	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*(6.1)*
$\sigma 1$	3.05	*(3.0)*			2.16	*(3.0)*
$\sigma 2$			3.05	*(3.0)*	2.16	---
$(\sigma 1^2+\sigma 2^2)^{1/2}$	*3.05*		*3.05*		*3.05*	
(Simul.) Log-Likelihood:	-473.02		-473.03		-473.04	

* the specification lists the factors (and sigmas) that apply to each of the five alternatives

Other Identification Issues

Near Singularity

Finally, it turns out that the fact that σ_p must be constrained in a logit kernel model is not exactly correct. In a *probit* kernel model (i.e., with an i.i.d. normal term), it is true that σ_p must be constrained. In this case, there is a perfect trade-off between the multivariate normal term and the i.i.d. normal term. However, in the logit kernel model, this perfect trade-off does not exist because of the slight difference between the Gumbel and Normal distributions. Therefore, there will be an optimal combination of the Gumbel and Normal distribution, and this effectively allows another parameter to be estimated. This leads to somewhat surprising results. For example, in a heteroscedastic logit kernel model a variance term can be estimated for *each* of the alternatives, whereas probit, probit kernel, or extreme value logit requires that one of the variances be constrained. The same holds true for an unrestricted covariance structure. Nonetheless, the reality is that without the constraint, the model is nearly singular (i.e., the objective function is very flat at the optimum), as will be demonstrated in the estimation results that follow. Due to the near singularity, it is advisable to impose the additional constraint.

Empirical Identification

While this paper discusses theoretical identification, there is always the practical issue of whether or not the given data can support a given model specification. As Train (2003) has pointed out, "there is a natural limit to how much can be learned about things that cannot be seen". Therefore, one always needs to be careful about asking more from the data than it can provide, and verifying that the parameter estimates are stable with increasing numbers of draws.

Conclusion

This paper presents a procedure for determining identification of the logit kernel model, which consists of the Order Condition, Rank Condition, and Equality Condition. These conditions were described mathematically and used to determine identification of several important logit kernel specifications such as heteroscedasticity, nesting, and panel data. The paper also highlights a variety of issues that can lead to complications when estimating logit kernel models. Some of the findings are that (a) the logit kernel model requires different treatment than that of an analogous probit model (e.g., as with heteroscedasticity), and (b) the imposition of seemingly obvious identification restrictions can be incorrect (e.g., as with heteroscedasticity and nesting). Misspecification can lead to bias of the systematic parameters and incorrect conclusions being drawn from hypothesis tests. The nature of the impact varies dramatically based on the data and specification at hand, and it is difficult to draw generalizations.

This paper just skims the surface of identification and estimation issues related to the logit kernel model. There is much more work to be done in terms of expanding the library of guidelines that we have for identification (such as those presented in this paper for heteroscedasticity and nesting), building robust estimation programs that detect identification errors (through, for example, automatic processing of the information matrix and explicit reporting of identification problems), developing methods for practically addressing complications that arise in identification, and better understanding the consequences of identification errors. This exploration is critical as the model is being included in textbooks, user-friendly estimation software is becoming widely available, and application of the logit kernel model is poised to explode in both the literature and practice.

Finally, while the logit kernel model provides immense flexibility in terms of the richness that can be captured in the disturbance term, the identification issues highlighted in this paper raise computational and performance issues. For example, a large number of draws and multiple model estimation runs are required to verify identification and parameter stability. Therefore, an important area of research is to explore the tradeoffs between logit kernel and analogous model forms (logit kernel, probit, and GEV). Also promising is the idea of a GEV kernel, which allows some of the heteroscedasticity/covariance to be represented in the kernel (which has a closed form solution), thereby reducing the dimensionality of the simulation.

Acknowledgments

Many thanks to Chandra Bhat, David Bunch, Dinesh Gopinath, Tomer Toledo, and Kenneth Train for useful discussions on this topic.

References

Bekhor, S., M.E. Ben-Akiva, and M.S. Ramming (2002) "Adaptation of Logit Kernel to Route Choice Situation", *Transportation Research Record* **1805**, 78-85.

Ben-Akiva, M. and S. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge, MA.

Ben-Akiva, M., D. Bolduc, and J. Walker. Specification, Identification, & Estimation of the Logit Kernel (or Continuous Mixed Logit) Model. Working paper, MIT, 2001. http://web.mit.edu/jwalker/www/home.htm. Accessed April 1, 2002.

Bhat, C.R. (1995) "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice", *Transportation Research B* **29(6)**, 471-483.

Bhat, C.R. (1997) "Accommodating Flexible Substitution Patterns in Multi-dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice", *Transportation Research B* **32(7)**, 455-466.

Bhat, C.R. (1998) "Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling", *Transportation Research A* **32(7)**, 495-507.

Bhat, C.R. and R. Gossen (2003) "Mixed Multinomial Logit Model Analysis of Weekend Recreational Episode Type Choice", Transportation Research Board Annual Meeting.

Bolduc, D., B. Fortin and M.A. Fournier (1996) "The Impact of Incentive Policies to Influence Practice Location of General Practioners: A Multinomial Probit Analysis", *Journal of Labor Economics* **14**, 703-732.

Brownstone, D., D.S. Bunch and K. Train (2000) "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles", *Transportation Research B* **34**, 315-338.

Bunch, D.A. (1991) "Estimability in the Multinomial Probit Model", *Transportation Research B* **25**, 1-12.

Goett, A., K. Hudson and K. Train (2000) "Customers' Choice Among Retail Energy Suppliers: The Willingness-to-Pay for Service Attributes", working paper, AAG Associates and University of California at Berkeley.

Gönül, F. and K. Srinivasan (1993) "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues", *Marketing Science* **12(3)**, 213-229.

Greene, W.H. (2000) *Econometric Analysis Fourth Edition*, Prentice Hall, Upper Saddle River, New Jersey.

Manski, C.F. (1995) *Identification Problems in the Social Sciences*, Harvard University Press.

McFadden, D. (1984) "Econometric Analysis of Qualitative Response Models", *Handbook of Econometrics* **II**, Z. Friliches and M.D. Intriligator, Eds., Elsevier Science Publishers.

McFadden, D. and K. Train (2000) "Mixed MNL Models for Discrete Response", *Journal of Applied Econometrics* **15(5)**, 447-470.

Mehndiratta, R.M. and M. Hansen (1997) "Analysis of Discrete Choice Data with Repeated Observations: Comparison of Three Techniques in Intercity Travel Case", *Transportation Research Record* **1607**, 69-73.

Revelt, D. and K. Train (1998) "Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level", *Review of Economics and Statistics* **80(4)**, 647-657.

Revelt, D. and K. Train (1999) "Customer-Specific Taste Parameters and Mixed Logit", working paper, University of California at Berkeley.

Srinivasan, K.K. and H.S. Mahmassani (2000) "Dynamic Kernel Logit Model for the Analysis of Longitudinal Discrete Choice Data: Properties and Computational Assessment", presented at the International Association of Travel Behavior Research (IATBR) Conference, Gold Coast, Queensland, Australia.

Steckel, J.H. and W.R. Vanhonacker (1988) "A Heterogeneous Conditional Logit Model of Choice", *Journal of Business & Economic Statistics* **6(3)**, 391-398.

Toledo, T. (2003) *Integrating Driver Behavior Modeling*, , Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

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(1)**, 109-123.

Train, K.E. (1998) "Recreational Demand Models with Taste Differences Over People", *Land Economics* **74(2)**, 230-239.

Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press.

Walker, J.L. (2001) *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*, Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Walker, J. (2002) "The Mixed Logit (or Logit Kernel) Model: Dispelling Misconceptions of Identification", *Transportation Research Record* **1805**, 86-98.