



The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models

Chandra R. Bhat *

The University of Texas at Austin, Department of Civil, Architectural and Environmental Engineering, 1 University Station, C1761, Austin, TX 78712-0278, United States

ARTICLE INFO

Article history:

Received 3 August 2010

Received in revised form 18 April 2011

Accepted 19 April 2011

Keywords:

Multinomial probit

Mixed models

Composite marginal likelihood

Discrete choice models

Spatial econometrics

Panel data

ABSTRACT

The likelihood functions of multinomial probit (MNP)-based choice models entail the evaluation of analytically-intractable integrals. As a result, such models are usually estimated using maximum simulated likelihood (MSL) techniques. Unfortunately, for many practical situations, the computational cost to ensure good asymptotic MSL estimator properties can be prohibitive and practically infeasible as the number of dimensions of integration rises. In this paper, we introduce a maximum approximate composite marginal likelihood (MACML) estimation approach for MNP models that can be applied using simple optimization software for likelihood estimation. It also represents a conceptually and pedagogically simpler procedure relative to simulation techniques, and has the advantage of substantial computational time efficiency relative to the MSL approach. The paper provides a “blueprint” for the MACML estimation for a wide variety of MNP models.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The “workhorse” multinomial logit model, used extensively in practice for econometric discrete choice analysis, was introduced by [Luce and Suppes \(1965\)](#) and [McFadden \(1974\)](#), and has a very simple and elegant structure. However, it is also saddled with the familiar independence from irrelevant alternatives (IIA) property – that is, the ratio of the choice probabilities of two alternatives is independent of the characteristics of other alternatives in the choice set. This has led to several extensions of the MNL model through the relaxation of the independent and identically distributed (IID) error distribution (across alternatives) assumption. Two common model forms of non-IID error distribution include the generalized extreme-value (GEV) class of models proposed by [McFadden \(1978\)](#) and the multinomial probit (MNP) model that allow relatively flexible error covariance structures (up to certain limits of identifiability; see [Train, 2009](#), Chapter 5). Both of these non-IID kernel structures (or even the IID versions of the GEV and the MNP models, which lead to the MNL and the independent MNP models) can further be combined with (a) continuous mixing factor error structures to accommodate unobserved taste variation across choice occasions through random coefficients (see [Bhat and Sardesai, 2006](#)), (b) individual-specific (time-stable or time-dissipating) error terms that generate error dependencies across the choice occasions of the same decision-maker in panel or repeated choice contexts (see [Li et al., 2010](#)), (c) spatial/social dependency-based error structure models that generate error dependencies across choice occasions of different decision-makers (see [Franzese and Hays, 2008](#)), or (d) some combinations of these model forms. While many different continuous error distributions can be used to accommodate these additional structures, it is most common to adopt a normal distribution for these. For instance, when introducing random coefficients, it is typical to use the multivariate normal distribution for the mixing coefficients, almost to the point that

* Tel.: +1 512 471 4535; fax: +1 512 475 8744.

E-mail address: bhat@mail.utexas.edu

the terms mixed logit or mixed GEV or mixed probit are oftentimes used synonymously with normal mixing (see Fiebig et al., 2010; Dube et al., 2002).¹ Similarly, when introducing panel effects or spatial/social interdependency, the use of normal error structures is ubiquitous, except for some recent developments using general copula forms by Bhat and Sener (2009) and Bhat et al. (2010a).

In the context of the normal error distributions just discussed, the use of a GEV kernel structure leads to a mixing of the normal distribution with a GEV kernel, while the use of an MNP kernel leads once again to an MNP model. Both structures have been widely used in the past, with the choice between a GEV kernel or an MNP kernel really being a matter of “which is easier to use in a given situation” (Ruud, 2007). In recent years, the mixing of the normal with the GEV kernel has been the model form of choice in the economics and transportation fields, mainly due to the relative ease with which the probability expressions in this structure can be simulated (see Bhat et al. (2008) and Train (2009) for detailed discussions). On the other hand, the use of an MNP kernel has not seen as much use in recent years, because the simulation estimation is generally more difficult. In any case, while there have been several approaches proposed to simulate these models with a GEV or an MNP kernel, most of these involve pseudo-Monte Carlo or quasi-Monte Carlo simulations in combination with a quasi-Newton optimization routine in a maximum simulated likelihood (MSL) inference approach (see Bhat, 2001, 2003). In such an inference approach, consistency, efficiency, and asymptotic normality of the estimator is critically predicated on the condition that the number of simulation draws rises faster than the square root of the number of individuals in the estimation sample. This effectively implies that the desirable asymptotic properties of the MSL estimator are obtained at the expense of computational cost. Unfortunately, for many practical situations, the computational cost to ensure good asymptotic estimator properties can be prohibitive and literally infeasible (in the context of the computation resources available and the time available for estimation) as the number of dimensions of integration increases. This is particularly so because the accuracy of simulation techniques is known to degrade rapidly at medium-to-high dimensions, and the simulation noise increases substantially. This leads to convergence problems during estimation, unless a very high number of simulation draws is used. Besides, an issue generally ignored in simulation-based approaches is the accuracy (or lack thereof) of the covariance matrix of the estimator, which is critical for good inference even if the asymptotic properties of the estimator are well established. Specifically, the hessian (or second derivatives) needed with the MSL approach to estimate the asymptotic covariance matrix of the estimator is itself estimated on a highly nonlinear and non-smooth second derivatives surface of the log-simulated likelihood function. This is also usually undertaken numerically because the complex analytic nature of the second derivatives makes them difficult to code. The net result is that Monte Carlo simulation with even three to four decimal places of accuracy in the probabilities embedded in the log-likelihood function can work poorly (see Bhat et al., 2010b), suggesting a critical need to evaluate the likelihood function at a very high level of accuracy and precision. This further increases computational cost. Craig (2008) also alludes to this problem when he states that “(...) the randomness that is inherent in such methods [referring to the Genz-Bretz algorithm (Genz and Bretz, 1999), but applicable in general to MSL methods] is sometimes more than a minor nuisance.”

In this paper, we propose a new methodology (which is labeled as the maximum approximate composite marginal likelihood or MACML inference approach) that allows the estimation of models with both GEV and MNP kernels using simple, computationally very efficient, and simulation-free estimation methods. In the MACML inference approach, models with the MNP kernel, when combined with additional normal random components, are much easier to estimate because of the conjugate addition property of the normal distribution (which puts the structure resulting from the addition of normal components to the MNP kernel back into an MNP form). On the other hand, the MACML estimation of models obtained by superimposing normal error components over a GEV kernel requires a normal scale mixture representation for the extreme value error terms, and adds an additional layer of computational effort (see Bhat, 2011). Given that the use of a GEV kernel or an MNP kernel is simply a matter of convenience, we will henceforth focus in this paper on the MNP kernel within the MACML inference approach.

The proposed MACML estimation of the resulting MNP-based models involves only univariate and bivariate cumulative normal distribution function evaluations, regardless of the number of alternatives or the number of choice occasions per individual or the nature of social/spatial dependence structures. This implies substantial computational efficiency relative to the MSL inference approach for models that can be estimated using the MSL approach, and also allows the estimation of several model structures that are literally infeasible to estimate using traditional MSL approaches (with the available computational resources and time available for estimation). For instance, consider the case of dealing with panel data or repeated unordered choice data from each individual in a sample, with individual-specific normally distributed random coefficients. In such a case, the full likelihood function contribution of an individual entails taking the product of the individual's choices across choice occasions conditional on the random coefficients, and then unconditioning out the random coefficients by integration over the multivariate normal domain. The multivariate integration is over the real line with respect to each random coefficient, with the dimensionality being equal to the number of random coefficients. As the number of random coefficients increases, evaluating the likelihood using simulation techniques becomes computationally expensive. Another example where the full likelihood becomes near impractical to work with is when there are individual-specific normal random

¹ To be sure, there have been models with non-normal mixing distributions too, such as the log-normal distribution, the triangular distribution, and the Rayleigh distribution (see, Bhat et al., 2008 for a review). However, it has been well known that using non-normal distributions can lead to convergence/computational problems, and it is not uncommon to see researchers consider non-normal distributions only to eventually revert to the use of a normal distribution (see, for example, Bartels et al., 2006 and Small et al., 2005).

Download English Version:

<https://daneshyari.com/en/article/1132253>

Download Persian Version:

<https://daneshyari.com/article/1132253>

[Daneshyari.com](https://daneshyari.com)