Journal of Econometrics 187 (2015) 131–153

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Simulated maximum likelihood estimation for discrete choices using transformed simulated frequencies[☆]

Donghoon Lee[a], Kyungchul Song[b,*]

[a] Federal Reserve Bank of New York, 33 Liberty Street, New York, NY, 10045, United States

[b] Vancouver School of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC, V6T 1Z1, Canada

ARTICLE INFO

Article history:
Received 5 June 2013
Received in revised form
2 June 2014
Accepted 19 December 2014
Available online 12 February 2015

JEL classification:
C12
C24
C52

Keywords:
Simulated MLE
Discrete choice models
Simulation bias
Simulated frequencies
Cube-root asymptotics

ABSTRACT

Many existing methods of simulated likelihood for discrete choice models require additive errors that have normal or extreme value distributions. This paper focuses on a situation where the model does not admit such additive errors so that the popular method of GHK or logit estimation is not applicable. This paper proposes a new method of simulated likelihood that is free from simulation bias for each finite number of simulations, and yet flexible enough to accommodate various model specifications beyond those of additive normal or logit errors. The method begins with the likelihood function involving simulated frequencies and finds a transform of the likelihood function that identifies the true parameter for each finite simulation number. The transform is explicit, containing no unknowns that demand an additional step of estimation. The estimator achieves the efficiency of MLE when the simulation number increases fast enough. This paper presents and discusses results from Monte Carlo simulation studies of the new method.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Discrete choice models have long been used in a wide range of empirical fields of economics. While a discrete choice model typically specifies the data generating process up to a parametric family of distributions, maximum likelihood estimation is infeasible in practice when the explicit evaluation of the likelihood is not possible. Since the seminal work of Lerman and Manski (1981), the approach of simulation-based inference has been increasingly instrumental for overcoming this difficulty, providing the researcher with a wider spectrum of flexibility in modeling. (See Hajivassiliou and Ruud (1994), Stern (1997), Gouriéroux and Monfort (1997), and Train (2003) for a review of the literature and references therein.) More recently, Chernozhukov and Hong (2003) offer a general MCMC based method for M-estimation. Their method can also be used in the simulation-based estimation. Armstrong et al. (2013) analyzed the asymptotic distribution of simulation-based estimators for simulation draws common across individual sample units.

This paper proposes a new approach of simulated likelihood estimation. The first merit of the approach is that the estimator is consistent even with the finite simulation number. As far as we know, our estimator is the only simulated likelihood estimator with such a property. Second, our method can be used for a wide class of models, whenever one can simulate the individual choices. Hence one does not need to assume additive normal or logit errors as often done in simulation-based estimation. Third, the method does not suffer from a log of zero problem even with the finite simulation number, unlike the classical simulated frequency method. Fourth, our estimator achieves the asymptotic efficiency

☆ A previous version was titled, "A Consistent SMLE When the Simulation Number was Finite". We are indebted to Sean Campbell, Joel Horowitz, Eugene Savin, and Chamna Yoon for their valuable comments and inputs. We are also grateful to the seminar participants at Cemmap Seminar at University College London, Columbia University, Econometric Society Meeting, Greater New York Econometrics Colloquium, Lehigh University, University of Iowa, University of Rochester, and New York University for useful comments. We appreciate the valuable comments and suggestions from Editor, Associate Editor, and two referees. We thank Jacob Schwartz for assisting us in the simulation studies of this paper. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System. All errors are ours.

* Correspondence to: Vancouver School of Economics, University of British Columbia, Vancouver, BC, Canada.

E-mail addresses: donghoon.lee1@gmail.com (D. Lee), kysong@mail.ubc.ca (K. Song).

http://dx.doi.org/10.1016/j.jeconom.2014.12.009
0304-4076/© 2015 Elsevier B.V. All rights reserved.

of maximum likelihood estimator when the simulation number grows faster than the square root of the sample size. Finally, our method is easy to use, accompanying almost no computational cost additional to the classical simulated frequency method.

However, our method shares one disadvantage with the simulated frequency method of likelihood estimation or moment estimation (Lerman and Manski (1981) and McFadden (1989)): the sample objective function is discontinuous in the parameter. This may cause a higher computational cost than the method of GHK or mixed logit error modeling. Hence we do not propose our method as a competitor of the latter simulation methods. Rather we propose it as a simulated-likelihood method for an environment where GHK or mixed logit approach is not applicable. Such an environment arises in many structural models of individual decision-making, where unobserved heterogeneity enters nonlinearly in latent processes. (See e.g. Keane and Wolpin (1994, 1997).)

Our method is built on the main finding of this paper that there exists a simple and explicit transform of a simulated likelihood function whose maximization delivers a consistent estimator even with a finite simulation number. The transform is algebraically explicit, depending on no unknowns. Furthermore, the use of the transform does not require any restrictions on the specification of the random utilities, and hence flexibly applies to many discrete choice models that have a nonlinear, nonnormal form of heterogeneity. We call this new method *transformed simulated frequencies (TSF) method*.

In this paper, we formally present conditions for identification and derive the asymptotic theory for the estimator in both the cases of simulation numbers fixed and increasing with the sample size. Our exposition is made through easily verifiable, high-level conditions to emphasize the flexibility of our approach. The conditions require only weak regularity conditions for the stochastic link between the decision variables and the observed covariates.

Here is the summary of the asymptotic properties of the estimators based on the TSF method. When the simulation number is fixed and the sample size n increases, the estimator is consistent at the rate of $\sqrt[3]{n}$, like the maximum score estimator (Manski (1975) and Kim and Pollard (1990)). In the case of an increasing number of simulations, we establish that the estimator is $\sqrt{n}$-consistent and asymptotically normal as the simulation number increases to infinity at a rate faster than $\sqrt{n}$. Under this same condition, the estimator achieves the asymptotic efficiency of MLE.

To illustrate the usefulness of our approach, we performed a Monte Carlo simulation study based on a schooling choice model which involves heterogeneity in discount factor and ability. More specifically, the discount factor is assumed to be correlated with other observed individual characteristics and also an unobserved characteristic. The study considered the simulated MLE based on the classical simulated frequency method, and the simulated method of moments (SMM). Our estimator mostly dominates the classical simulated frequency method. The domination is prominent especially when the simulation number is small and the sample size is large. Also our method performs better than SMM without using an optimal weighting matrix, and performs comparably with SMM using an optimal weighting matrix but with shorter computation time than SMM.

Since the seminal paper by Lerman and Manski (1981), the simulation-based method has been widely used in empirical researches. Many early researches since Lerman and Manski (1981) have focused on developing a new simulation method that overcomes the drawbacks of Lerman and Manski (1981). For example, the method of Stern (1992) and the method of GHK simulator (Geweke (1989), Hajivassiliou (1990) and Keane (1993)) enable one to simulate choice probabilities from a probit model that are smooth in parameters. Hajivassiliou (1990) and Hajivassiliou and McFadden (1998) proposed a different method of simulated likelihood that uses simulated scores to construct simulated moment conditions and proved efficiency of the estimators. Another increasingly popular class of discrete choice models include mixed multinomial logit models (MMLM) (McFadden and Train (2000) and see references therein). The MMLMs offer a flexible way of modeling heterogeneity through random coefficient specifications and yet requires the presence of additive logit errors. Lee (1995) analyzed the bias properties of simulation-based estimators. See Kristensen and Salanié (2013) for a recent contribution in a similar spirit. Fermanian and Salanié (2004) developed nonparametric simulated maximum likelihood estimation and Kristensen and Shin (2012) extended their framework to a wider class of dynamic models. Armstrong et al. (2013) analyzed the asymptotic distribution of simulation-based estimators when the simulation draws are common across the sample units. Chernozhukov and Hong (2003) developed an MCMC approach to various extremum estimation problems. Their approach can also be applied to various simulation-based estimation where the sample objective function is generated by simulations. (See Jun et al. (2011) for a similar MCMC approach for maximum score estimators.) Ackerberg (2009) makes a notable contribution to reduce the simulation burden for various structural estimation environments.

The remainder of this paper is organized as follows. In Section 2, we define the class of discrete choice models, discuss simulated MLE, and offer a preview of our method. In Section 3, we present the main results of this paper which formally establish identification and consistency of the proposed estimator. It is also shown that the estimator is asymptotically normal when the simulation number goes to infinity fast enough. In Section 4, we present and discuss results from Monte Carlo simulation studies. Section 5 concludes. All the technical proofs are relegated to the Appendix.

2. Discrete choice models and TSF

2.1. Methods of simulated likelihoods

Suppose that a binary variable, $D_{ij} \in \{0, 1\}$, of a unit i realizing the jth state, is stochastically linked with an observed covariate vector X_i as follows:

$$D_{ij} = \delta_j(X_i, \eta_i; \theta_0), \tag{2.1}$$

where $X_i = (X_{i1}, \ldots, X_{iJ})^\top$ represents a vector of observed random variables, $\eta_i = (\eta_{i1}, \ldots, \eta_{iJ})^\top$ a vector of unobserved variables, and $\theta_0 \in \Theta \subset \mathbf{R}^d$ the parameter to be estimated. The number J denotes the number of the states and n the number of the cross-sectional units in the data set. The parametric maps $\delta_j(X_i, \eta_i; \theta_0)$ link exogenous variables X_i and η_i to endogenous outcomes D_{ij}, and satisfies that

$$\sum_{j=1}^{J} \delta_j(X_i, \eta_i; \theta) = 1, \quad \text{for all } \theta \in \Theta.$$

We assume that X_i and η_i are independent. For example, in the case of discrete choice model with random utilities, one can take

$$\delta_j(X_i, \eta_i; \theta_0) = 1\left\{\Delta_j(X_i, \eta_i; \theta_0) \geq 0\right\},$$

where

$$\Delta_j(X_i, \eta_i; \theta_0) = u_j(X_i, \eta_i; \theta_0) - \max_{1 \leq k \leq J:\, k \neq j} u_k(X_i, \eta_i; \theta_0),$$

and $u_j(X_i, \eta_i; \theta)$ is a random utility from choosing action j (McFadden (1974)). However, the TSF method does not require that the indicator $\delta_j(X_i, \eta_i; \theta_0)$ be generated through a certain index $\Delta_j(X_i, \eta_i; \theta_0)$.

The conditional choice probability of the ith agent choosing the jth option is defined by

$$p_j(X_i, \theta_0) = P\left\{D_{ij} = 1 | X_i\right\}.$$

Download English Version:

https://daneshyari.com/en/article/5095782

Download Persian Version:

https://daneshyari.com/article/5095782

Daneshyari.com