Journal of Econometrics 155 (2010) 117–127

Contents lists available at ScienceDirect

Journal of Econometrics

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

The dynamic invariant multinomial probit model: Identification, pretesting and estimation

Roman Liesenfeld [a,*], Jean-François Richard [b]

[a] Department of Economics, Christian Albrechts Universität, Kiel, Germany
[b] Department of Economics, University of Pittsburgh, USA

ARTICLE INFO

Article history:
Received 7 September 2007
Received in revised form
7 July 2009
Accepted 24 September 2009
Available online 2 October 2009

JEL classification:
C35
C15

Keywords:
Discrete choice
Efficient Importance Sampling
Invariance
Monte Carlo integration
Panel data
Simulated maximum likelihood

ABSTRACT

We present a new specification for the multinomial multiperiod probit model with autocorrelated errors. In sharp contrast with commonly used specifications, ours is invariant with respect to the choice of a baseline alternative for utility differencing. It also nests these standard models as special cases, allowing for data-based selection of the baseline alternatives for the latter. Likelihood evaluation is achieved under an Efficient Importance Sampling (EIS) version of the standard GHK algorithm. Several simulation experiments highlight identification, estimation and pretesting within the new class of multinomial multiperiod probit models.

1. Introduction

In this paper we revisit the Dynamic Multinomial (multiperiod) Probit (hereafter DMP) model. DMP models offer a flexible and operational framework for analyzing correlated sequences of discrete choices such as living arrangement decisions for elderlies (Börsch-Supan et al., 1990) or brand choices in successive purchases (Keane, 1997).

The standard DMP specification commonly used in the literature initially expresses all utilities in differences from that of a baseline alternative which is selected a priori among all available alternatives. It then assumes that the error terms associated with these differences follow a stationary diagonal AR(1) process. One common interpretation of that approach amounts to treating the selected baseline utility as non-random; see, e.g., Börsch-Supan et al. (1990) or Geweke et al. (1997). However, as we shall discuss below, the standard DMP model suffers from a major drawback in that it is not invariant with respect to the choice of the baseline alternative. Specifically, DMP models derived under different baseline alternatives are non-nested and their respective parameterizations are not one-to-one transformations of one another. It follows that results (estimations or test statistics) derived under different baseline alternatives are mutually incompatible and, therefore, not easily comparable.

In the present paper we propose a dynamic version of the multinomial probit model which is specified in terms of utilities prior to differencing. It still relies upon an arbitrary baseline alternative in order to construct the likelihood function. However, parameters associated with different selections of baseline alternative will be in one-to-one correspondence with one another. Whence, our specification will be invariant with respect to that selection.

In addition, our Dynamic Invariant Multinomial Probit (hereafter DIMP) model offers the critical advantage that it actually nests all DMP versions thereof, corresponding to different baseline categories. Whence it becomes trivial to test whether an initial DIMP model simplifies into a DMP model for a particular baseline alternative (whose selection is now data based instead of arbitrary).

Last but not least, the Monte Carlo (MC) evaluation of the likelihood function of the DIMP model is not more demanding than that of the standard DMP model. For the likelihood evaluation of both specifications one can rely on very similar implementations of the

* Corresponding address: Institut für Statistik und Ökonometrie, Christian-Albrechts-Universität zu Kiel, Olshausenstraße 40-60, D-24118 Kiel, Germany. Tel.: +49 (0)431 8803810; fax: +49 (0)431 8807605.
E-mail address: liesenfeld@stat-econ.uni-kiel.de (R. Liesenfeld).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.09.021

GHK probability simulator as developed by Geweke (1991), Hajivassiliou (1990) and Keane (1994). Actually, in the present paper we shall rely upon a numerically more Efficient Importance Sampling version of the GHK algorithm (hereafter GHK-EIS) as developed in the companion paper by Liesenfeld and Richard (2009).

Invariance, nesting of DMPs, and similar ease of computation offer strong incentives for the adoption of our proposed DIMP specification by practitioners. In particular, it allows for pretesting of whether a DIMP model can be subsequently simplified into a standard DMP model under data-based selection of a baseline alternative.

The remainder of the paper is organized as follows. In Section 2, we use a simple bivariate example in order to introduce some of the key features of the DIMP model under a simplified notation. The general DIMP specification is introduced in Section 3.1, followed by a discussion of its invariance (Section 3.2), identification (Section 3.3) and nesting properties (Section 3.4). Estimation is presented in Section 4 with a brief description of GHK-EIS (Section 4.1) followed by its application to likelihood evaluation (Section 4.2). MC experiments are presented in Section 5: first a correctly specified DIMP (Section 5.1), next a mis-specified DMP (Section 5.2) and finally a sample-based pretesting of a correctly specified DMP (Section 5.3). Section 6 concludes.

2. Introductory example

Consider the case where there are only two categories with utilities given by

$$U_t = \begin{pmatrix} u_{t1} \\ u_{t2} \end{pmatrix} = \mu\,(x_t;\,\beta) + \varepsilon_t, \tag{1}$$

where $\mu(\cdot)$ is momentarily left unspecified and ε_t follows a stationary AR(1) process

$$\varepsilon_t = R\varepsilon_{t-1} + \eta_t, \quad \eta_t \sim \mathrm{N}_2(0,\,\Sigma). \tag{2}$$

Assume that we only observe the difference $Y_t = d'U_t$ with $d' = (1, -1)$, or later only its sign. The following related three questions are central to our paper:

(i) Which parameters remain identified?
(ii) Under what conditions would $d'\varepsilon_t$ itself follow a stationary AR(1) process?
(iii) What would be the consequences of incorrectly assuming that $d'\varepsilon_t$ follows an AR(1) process?

For ease of exposition we initially consider the case when R is diagonal with diagonal elements ρ_1 and ρ_2 (with $|\rho_i| < 1$). Selecting an appropriate (re)parameterization helps to clarify the issues under consideration. Since the transformation from ε_t to $d'\varepsilon_t$ implies a reduction in dimensionality from 2 to 1 and, therefore, an (implicit) marginalization, we first introduce the auxiliary nonsingular transformation

$$\varepsilon_t^* = \begin{pmatrix} e_t \\ \varepsilon_{t2} \end{pmatrix} = Q\,\varepsilon_t, \quad Q = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \tag{3}$$

with $e_t = d'\varepsilon_t = \varepsilon_{t1} - \varepsilon_{t2}$. Note that ε_t^* follows the stationary AR(1) process

$$\varepsilon_t^* = R_*\varepsilon_{t-1}^* + \eta_t^*, \quad \eta_t^* \sim \mathrm{N}_2\,(0,\,\Sigma_*)\,, \tag{4}$$

with

$$R_* = QRQ^{-1} \quad \text{and} \quad \Sigma_* = Q\,\Sigma Q'. \tag{5}$$

Let $\Phi = (\phi_{ij})$ denote the stationary covariance matrix of ε_t and Φ_* that of ε_t^*:

$$\Phi_* = Q\Phi Q'. \tag{6}$$

The most relevant parameterization is that which is associated with the factorization of the stationary density of ε_t^* into a marginal density for e_t and a conditional density for $\varepsilon_{t2}|e_t$. Whence, Φ_* is reparameterized as

$$\Phi_* = \begin{pmatrix} \Psi & b_2\Psi \\ b_2\Psi & \upsilon^2 + b_2^2\Psi \end{pmatrix}, \tag{7}$$

with $\Psi > 0$, $\upsilon^2 > 0$ and $b_2 \in \mathbb{R}$. For ease of reference, the relationships between the successive parameterizations just introduced are given by

$$\phi_{11} = \frac{\sigma_1^2}{1-\rho_1^2}, \qquad \phi_{12} = \frac{\sigma_{12}}{1-\rho_1\rho_2}, \qquad \phi_{22} = \frac{\sigma_2^2}{1-\rho_2^2}, \tag{8}$$

$$\Psi = \phi_{11} + \phi_{22} - 2\phi_{12}, \qquad b_2 = \frac{\phi_{21} - \phi_{22}}{\phi_{11} + \phi_{22} - 2\phi_{12}}, \tag{9}$$

$$\upsilon^2 = \frac{\phi_{11}\phi_{22} - \phi_{12}^2}{\phi_{11} + \phi_{22} - 2\phi_{12}}. \tag{10}$$

For obvious reasons of symmetry we shall also consider the (stationary) regression coefficient of ε_{t1}, on $(\varepsilon_{t2} - \varepsilon_{t1})$, which is given by

$$b_1 = \frac{\phi_{21} - \phi_{11}}{\phi_{11} + \phi_{22} - 2\phi_{12}} = -\,(1 + b_2)\,. \tag{11}$$

The parameterization used for the rest of the discussion consists of $(\Psi, b_2, \upsilon^2, \rho_1, \rho_2)$ together with β. The identification for β is standard and has to be achieved by means of restrictions on the difference $\mu_1(x_t, \beta) - \mu_2(x_t, \beta)$, while υ^2, which represents the variance of the conditional distribution of the utility error term ε_{t2} given e_t, is clearly unidentified. We are left discussing the identification of $(\Psi, b_2, \rho_1, \rho_2)$. Eqs. (5) to (7) imply that the stationary distribution of e_t is characterized by the following moments:

$$\mathrm{Var}\,(e_t) = \Psi, \tag{12}$$

$$\mathrm{Cov}\,(e_{t-s}, e_t) = \begin{pmatrix} 1 & 0 \end{pmatrix}(R_*)^s\Phi_*\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \gamma_s^*\Psi, \tag{13}$$

with

$$\gamma_s^* = \left[\rho_1^s + b_2\left(\rho_1^s - \rho_2^s\right)\right]. \tag{14}$$

Identification results for Ψ are standard. If $d'U_t$ is observed, then Ψ is identified. If only the sign of $d'U_t$ is observed then it is identified only up to a constant and this indeterminacy is typically resolved by setting $\Psi = 1$. If $b_2 = 0$, then ρ_1 is identified. If $b_2 = -1 (b_1 = 0)$, then ρ_2 is identified. Otherwise, the triples (ρ_1, ρ_2, b_2) and (ρ_2, ρ_1, b_1) are observationally equivalent, in which case (ρ_1, ρ_2, b_2) are locally but not globally identified. However, as we shall discuss below, global identification of the ρs and bs is obtained from the diagonal elements of $\mathrm{Cov}(e_{t-s}, e_t)$ when the dimension of e_t is greater than 1 (except on a subspace of measure zero).

It also follows from Eq. (14) that if any of the following three conditions hold: (i) $b_1 = 0$; (ii) $b_2 = 0$; (iii) $\rho_1 = \rho_2$, then $\gamma_s^* = (\gamma_1^*)^s$ and e_t follows the stationary AR(1) process

$$e_t = \gamma_1^* e_{t-1} + \lambda_t. \tag{15}$$

If none of these conditions holds, then while it still is the case that $\mathrm{Cov}(\lambda_t, e_{t-1}) = 0$ by construction, the higher-order $\mathrm{Cov}(\lambda_t, e_{t-s})$ are non-zero for $s > 1$. In other words, λ_t is no longer an innovation relative to $\{e_\tau\}_{\tau=1}^{t-2}$ and e_t no longer follows an AR(1) process. Furthermore, even though $|\rho_i| < 1$, it does not even follow that $|\gamma_1^*| < 1$ since $b_2 \in \mathbb{R}$. Actually γ_1^* is then unrestricted. Note that the first-order covariance associated with γ_1^* in Eq. (15) does not suffice to identify the ρs and bs by itself. As we shall formally demonstrate in Section 3.3 below, identification of these parameters requires taking into consideration the higher-order covariances.

The consequences of erroneously assuming that e_t follows an AR(1) process when none of the conditions listed above holds