



**HAL**  
open science

## Dynamic invariant multinomial probit model: Identification, pretesting and estimation

Roman Liesenfeld, Jean-François Richard

► **To cite this version:**

Roman Liesenfeld, Jean-François Richard. Dynamic invariant multinomial probit model: Identification, pretesting and estimation. *Econometrics*, 2009, 155 (2), pp.117. 10.1016/j.jeconom.2009.09.021 . hal-00628315

**HAL Id: hal-00628315**

**<https://hal.science/hal-00628315>**

Submitted on 2 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Dynamic invariant multinomial probit model: Identification, pretesting and estimation

Roman Liesenfeld, Jean-François Richard

PII: S0304-4076(09)00234-6  
DOI: 10.1016/j.jeconom.2009.09.021  
Reference: ECONOM 3251

To appear in: *Journal of Econometrics*

Received date: 7 September 2007  
Revised date: 7 July 2009  
Accepted date: 24 September 2009

Please cite this article as: Liesenfeld, R., Richard, J.-F., Dynamic invariant multinomial probit model: Identification, pretesting and estimation. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.09.021

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Dynamic Invariant Multinomial Probit Model: Identification, Pretesting and Estimation

Roman Liesenfeld\*

Department of Economics, Christian Albrechts Universität, Kiel, Germany

Jean-François Richard

Department of Economics, University of Pittsburgh, USA

July 7, 2009

## 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.

JEL classification: C35, C15

Keywords: Discrete choice, Efficient Importance sampling, Invariance, Monte-Carlo integration, Panel data, Simulated maximum likelihood;

---

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

# 1 Introduction

In this paper we revisit the Dynamic Multinomial (multiperiod) Probit model (hereafter DMP). 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 the 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 model (hereafter DIMP) 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 specification one can rely on very similar implementations of the 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, 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 misspecified

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

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

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

$$\varepsilon_t = R\varepsilon_{t-1} + \eta_t, \quad \eta_t \sim N_2(0, \Sigma). \quad (2)$$

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 the ease of exposition we initially consider the case when  $R$  is diagonal with diagonal elements  $\rho_1$  and  $\rho_2$  (with  $|\rho_i| < 1$ ). Selecting an appropriate (re)parametrization helps clarifying the issues under consideration. Since the transformation from  $\varepsilon_t$  to  $d'\varepsilon_t$  implies a reduction in dimensionality from 2 to 1 and, therefore, an (implicit) marginalization, we first introduce the auxiliary

non-singular transformation

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

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

$$\varepsilon_t^* = R_* \varepsilon_{t-1}^* + \eta_t^*, \quad \eta_t^* \sim N_2(0, \Sigma_*), \quad (4)$$

with

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

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

$$\Phi_* = Q\Phi Q'. \quad (6)$$

The most relevant parametrization is that which is associated with the factorization of the stationary density of  $\varepsilon_t^*$  into a marginal density for  $e_t$  and a conditional density for  $\varepsilon_{t2}|e_t$ . Whence,  $\Phi_*$  is re-parameterized as

$$\Phi_* = \begin{pmatrix} \Psi & b_2\Psi \\ b_2\Psi & v^2 + b_2^2\Psi \end{pmatrix}, \quad (7)$$

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

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

$$\Psi = \phi_{11} + \phi_{22} - 2\phi_{12}, \quad b_2 = \frac{\phi_{21} - \phi_{22}}{\phi_{11} + \phi_{22} - 2\phi_{12}}, \quad (9)$$

$$v^2 = \frac{\phi_{11}\phi_{22} - \phi_{12}^2}{\phi_{11} + \phi_{22} - 2\phi_{12}}. \quad (10)$$

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

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

The parametrization used for the rest of the discussion consists of  $(\Psi, b_2, v^2, \rho_1, \rho_2)$  together with  $\beta$ . The identification for  $\beta$  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  $v^2$  which represents the variance of the conditional distribution of the utility error term  $\varepsilon_{t2}$  given  $e_t$  is clearly unidentified. We are left discussing identification of  $(\Psi, b_2, \rho_1, \rho_2)$ . Equations (5) to (7) imply that the stationary distribution of  $e_t$  is characterized by the following moments:

$$\text{Var}(e_t) = \Psi, \quad (12)$$

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

with

$$\gamma_s^* = [\rho_1^s + b_2(\rho_1^s - \rho_2^s)]. \quad (14)$$

Identification results for  $\Psi$  are standard. If  $d'U_t$  is observed, then  $\Psi$  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  $\rho_1$  is identified. If  $b_2 = -1$  ( $b_1 = 0$ ), then  $\rho_2$  is identified. Otherwise, the triples

$(\rho_1, \rho_2, b_2)$  and  $(\rho_2, \rho_1, b_1)$  are observationally equivalent, in which case  $(\rho_1, \rho_2, b_2)$  are locally but not globally identified. However, as we shall discuss below, global identification of the  $\rho$ s and  $b$ s obtains from the diagonal elements of  $\text{Cov}(e_{t-s}, e_t)$  when the dimension of  $e_t$  is greater than 1 (except on a subspace of measure zero).

It also follows from Equation (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

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

If neither of these conditions hold, then while it still is the case that  $\text{Cov}(\lambda_t, e_{t-1}) = 0$  by construction, the higher order  $\text{Cov}(\lambda_t, e_{t-s})$  are non zero for  $s > 1$ . In other words,  $\lambda_t$  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  $\gamma_1^*$  is then unrestricted. Note that the first-order covariance associated with  $\gamma_1^*$  in Equation (15) does not suffice to identify the  $\rho$ s and  $b$ s 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 neither of the conditions listed above hold depend upon the nature of the observations. If  $\{d'U_t\}_{t=1}^T$  were observed, then we would be discussing a standard Generalized Least Squares miss-specification and the ML estimate of  $\beta$  would remain consistent but would be inefficient. Since, furthermore, it need not be the case that  $|\gamma_1^*| < 1$ , we might be lead to erroneously conclude that the process is non-stationary. If, as commonly the case for discrete choice applications, we

only observed the signs of the differences  $d'U_t$ , then the Hessian would no longer be block diagonal and all parameter estimates would be inconsistent.

Foremost, we shall see that the  $\rho$ s and  $b$ s are identified except on a zero-measure subspace and that the (EIS-)GHK algorithm applies as discussed below to the more general model. Whence, there is no need to impose a priori the restriction that  $d'\varepsilon_t$  follows an AR(1) process, which eventually can be pretested within the more general model introduced in Equations (1) and (2).

Actually, we might not even restrict  $R$  to be diagonal. Let  $R = (r_{ij})$ . The corresponding elements for  $R_*$  are then given by

$$\begin{pmatrix} r_{11}^* & r_{12}^* \\ r_{21}^* & r_{22}^* \end{pmatrix} = \begin{pmatrix} r_{11} - r_{12} & r_{11} + r_{12} - (r_{22} + r_{21}) \\ r_{21} & r_{22} + r_{21} \end{pmatrix}. \quad (16)$$

The coefficients  $\gamma_s^*$  in Equation (13) then obtain from the recursion

$$\begin{pmatrix} \gamma_s^* \\ c_s \end{pmatrix} = R_* \begin{pmatrix} \gamma_{s-1}^* \\ c_{s-1} \end{pmatrix}, \quad s > 0, \quad (17)$$

initialized by  $\gamma_0^* = 1$  and  $c_0 = b_1$ . It follows that the condition  $r_{12}^* = 0$  is sufficient for  $e_t$  to follow an AR(1) process (the conditions  $b_1 = 0$  or  $b_2 = 0$  are no longer sufficient if  $r_{12}^* \neq 0$ ). Actually, if  $r_{12}^* = 0$ , then it follows from Equations (4) and (16) that

$$e_t = (r_{11} - r_{12}) e_{t-1} + (\eta_{2t} - \eta_{1t}). \quad (18)$$

### 3 Dynamic Invariant Multinomial Probit (DIMP)

#### 3.1 Model

We now discuss the higher dimensional version of the pilot model introduced in Section 2. Let  $U_t' = (u_{t1}, \dots, u_{tJ})$  denote a  $J$ -dimensional vector of utilities at time  $t$ . We assume that  $U_t$  evolves according to the stationary process

$$U_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \mid \varepsilon_{t-1} \sim N_J(R\varepsilon_{t-1}, \Sigma), \quad (19)$$

where  $\mu_t = \mu(x_t; \beta)$  is left unspecified since the focus of our analysis lies in the second order moments of the process and of its subsequent transformations. The stationary covariance matrix of  $\varepsilon_t$  denoted by  $\Phi$  satisfies the identity

$$\Sigma = \Phi - R\Phi R'. \quad (20)$$

The standard approach consists of selecting a baseline alternative, say alternative  $J$ , and of expressing all utilities in difference from  $u_{tJ}$ . As in Section 2, we initially complete that transformation into the following non-singular transformation:

$$U_t^* = \begin{pmatrix} Y_t \\ u_{tJ} \end{pmatrix} = \begin{pmatrix} D \\ i'_J \end{pmatrix} U_t = QU_t, \quad (21)$$

where  $D$  is the  $(J-1) \times J$  matrix

$$D = (I_{J-1} \dot{-} \iota_{J-1}), \quad (22)$$

$\iota_{J-1} = (1, \dots, 1)$  and  $i'_J$  is the unit vector  $(0, \dots, 0, 1)$ . The transformed system

is given by

$$U_t^* = \mu_t^* + \varepsilon_t^*, \quad \varepsilon_t^* | \varepsilon_{t-1}^* \sim N_J(R_* \varepsilon_{t-1}^*, \Sigma_*), \quad (23)$$

with

$$\mu_t^* = Q\mu_t, \quad \varepsilon_t^* = Q\varepsilon_t, \quad \Sigma_* = Q\Sigma Q', \quad R_* = QRQ^{-1}, \quad (24)$$

$$Q^{-1} = \begin{pmatrix} I_{J-1} & \iota_{J-1} \\ 0 & 1 \end{pmatrix}. \quad (25)$$

The stationary covariance matrix of  $\{\varepsilon_t^*\}_{t=1}^T$  consists of the following blocks

$$\Phi_* = \text{Var}(\varepsilon_t^*) = Q\Phi Q', \quad \text{Cov}(\varepsilon_{t-s}^*, \varepsilon_t^*) = (R_*)^s \Phi_*. \quad (26)$$

As in Section 2, we partition  $\Phi_*$  as follows

$$\Phi_* = \begin{pmatrix} \Psi & \Psi \underline{b} \\ \underline{b}' \Psi & \nu^2 + \underline{b}' \Psi \underline{b} \end{pmatrix}. \quad (27)$$

The stationary moments of  $e_t = D\varepsilon_t$  are given by

$$\text{Var}(e_t) = \Psi, \quad (28)$$

$$\text{Cov}(e_{t-s}, e_t) = \begin{pmatrix} I_J & : & 0 \end{pmatrix} (R_*)^s \Phi_* \begin{pmatrix} I_J \\ 0 \end{pmatrix} = \Gamma_s^* \Psi, \quad (29)$$

with

$$\Gamma_s^* = \begin{pmatrix} I_J & : & 0 \end{pmatrix} (R_*)^s \begin{pmatrix} I_J \\ \underline{b}' \end{pmatrix}. \quad (30)$$

In the next two sections we discuss the invariance and the identifications of the parametrization  $(\Psi, \underline{b}, \{\Gamma_s^*\})$ .

### 3.2 Invariance

Suppose we select a different baseline alternative, say  $u_{tj}$  instead of  $u_{tJ}$ . The simplest way to handle this change consists of permuting  $u_{jt}$  and  $u_{Jt}$  before applying the  $Q$  transformation introduced in Equation (21). Whence the error terms associated with the baseline alternative  $u_{tj}$  are given by

$$\varepsilon_t^{*j} = QP_j\varepsilon_t = QP_jQ^{-1}\varepsilon_t^*, \quad (31)$$

where  $P_j$  is the permutation matrix for rows  $J$  and  $j$ . Note that

$$QP_jQ^{-1} = \begin{pmatrix} S_j & 0 \\ i'_j & 1 \end{pmatrix}, \quad (32)$$

where  $S_j$  is the  $(J-1) \times (J-1)$  matrix

$$S_j = \begin{pmatrix} I_{j-1} & -\iota_{j-1} & 0 \\ 0 & -1 & 0 \\ 0 & -\iota_{J-j-1} & I_{J-j-1} \end{pmatrix}, \quad \text{with} \quad S_J = I_{J-1}, \quad (33)$$

$S_j^{-1} = S_j$  and  $-i'_j S_j = i'_j$ . It follows from Equation (27) and (31) that the stationary covariance matrix of  $\varepsilon_t^{*j}$  is given by

$$\Phi_{*j} = \begin{pmatrix} S_j & 0 \\ i'_j & 1 \end{pmatrix} \begin{pmatrix} \Psi & \Psi \underline{b} \\ \underline{b}'\Psi & v^2 + \underline{b}'\Psi \underline{b} \end{pmatrix} \begin{pmatrix} S'_j & i_j \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \Psi_j & \Psi_j \underline{b}_j \\ \underline{b}'_j \Psi_j & v_j^2 + \underline{b}'_j \Psi_j \underline{b}_j \end{pmatrix}, \quad (34)$$

with

$$\Psi_j = S_j \Psi S'_j, \quad \underline{b}_j = S'_j (i_j + \underline{b}), \quad v_j^2 = v^2. \quad (35)$$

Moreover (since  $-i'_j S_j = i'_j$ ),

$$(R_{*j})^s = \begin{pmatrix} S_j & 0 \\ i'_j & 1 \end{pmatrix} (R_{*})^s \begin{pmatrix} S_j & 0 \\ i'_j & 1 \end{pmatrix}. \quad (36)$$

It follows that

$$\Gamma_{js}^* = \begin{pmatrix} I_J & : & 0 \end{pmatrix} (R_{*j})^s \begin{pmatrix} I_J \\ \underline{b}' \end{pmatrix} = S_j \Gamma_s^* S_j. \quad (37)$$

The immediate implication of Equations (35) and (37) is that the parameters  $(\Psi_j, \underline{b}_j, v_j^2, \{\Gamma_{js}^*\}_{s=1}^{T-1})$  associated with baseline alternative  $j$  are in one-to-one correspondence with the parameters  $(\Psi, \underline{b}, v^2, \{\Gamma_s^*\}_{s=1}^{T-1})$  associated with baseline alternative  $J$ . Whence ML estimators of the multinomial probit model introduced in Equation (19) will be invariant w.r.t. the choice of the baseline alternative.

In sharp contrast, the standard dynamic multinomial probit model which assumes a diagonal autocorrelation matrix for the differences relative to a particular alternative is not invariant relative to the choice of that alternative unless all correlations are equal. Specifically, assume that  $R_*$  is diagonal, say

$$R_* = \text{Diag}(\rho_i^*), \quad i = 1, \dots, J. \quad (38)$$

Then it is trivial to verify that  $R_{*j}$  which obtains from equation (36) with  $s = 1$  is not diagonal unless

$$R_* = \rho^* I_J. \quad (39)$$

### 3.3 Identification

Next, we discuss identification of the DIMP model in Equation (19) when observations consist of  $\{x_t, j_t\}_{t=1}^T$ , where  $j_t$  denotes the index of the alternative

selected at time  $t$ . Following our discussion in Section 3.1, we can arbitrarily select a baseline alternative, say alternative  $J$ . Since the transformations associated with permutations of the alternatives are one-to-one we only need to discuss identification of the baseline parameters. The parameters whose identification needs to be discussed are  $(\Psi, \underline{b}, v^2, \beta, R)$ .

Identification of  $\beta$  has been extensively discussed in the literature – see, e.g., Bunch (1991) and Keane (1992) – and need not be considered further here. Moreover, under scale normalization of the mean vector  $\mu(x_t, \beta)$ , the stationary covariance matrix  $\Psi$  in Equation (27) is identified while  $v^2$  is not. Whence, we are left discussing identification of  $(\underline{b}, R)$ .

Actually, the likelihood function depends upon the sequence of auxiliary regression matrices  $\{\Gamma_s^*\}_{s=1}^{T-1}$  introduced in Equation (30) which are clearly over-identified functions of  $\underline{b}$  and  $R$ . (This will not be a problem for ML estimation since optimization will be conducted in terms of  $\underline{b}$  and  $R$  themselves.) Since the relationship between  $(\underline{b}, R)$  and  $\{\Gamma_s^*\}$  is highly non-linear we cannot expect global identification if  $R$  is left unconstrained (beyond stationarity). Local identification might be considered though in the present paper, we choose to focus our attention on the case where  $R$  is diagonal, for which we can establish global identification under simple conditions. Whence let

$$R = \text{Diag}(\rho_i), \quad i = 1, \dots, J, \quad (40)$$

in which case  $\Gamma_s^*$  in Equation (30) is given by

$$\Gamma_s^* = \begin{pmatrix} \rho_1^s & 0 & \cdots & 0 \\ 0 & \rho_2^s & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \rho_{J-1}^s \end{pmatrix} + \begin{pmatrix} \rho_1^s - \rho_J^s \\ \rho_2^s - \rho_J^s \\ \vdots \\ \rho_{J-1}^s - \rho_J^s \end{pmatrix} \cdot (b_1, b_2, \dots, b_{J-1}), \quad (41)$$

where the  $b_j$ s are the individual elements of  $\underline{b}$ . The following special cases are trivial: (i) If  $\rho_i = \rho_J$ , then  $\rho_i$  is identified; and (ii) if  $b_j = 0$ , then  $(\rho_j, b_j)$  are identified. In order to avoid dealing with all possible combinations of these special cases, we first proceed to a reduction of the matrices  $\{\Gamma_s^*\}$  by eliminating all rows  $i$  for which  $\rho_i = \rho_J$  and all columns  $j$  for which  $b_j = 0$ . Moreover, we also permute alternatives in such a way that the remaining  $J_R$  rows and  $J_C$  columns are regrouped in the leading  $J_R \times J_C$  blocks of  $\{\Gamma_s^*\}$ . It follows that all  $b_j$ s for  $j > J_C$  and all  $\rho_i$ s for  $i > \min(J_R, J_C)$  are identified. The following theorem then provides sufficient condition for global identification of  $(\underline{b}, R)$ :

**Theorem.** *If (i)  $J_R > 1$ ; (ii)  $J_C > 0$ ; (iii) either there exist a pair  $\rho_i \neq \rho_{i'}$  ( $i, i' \leq J_R$ ) or a  $b_j \neq -\frac{1}{2}$  ( $j \leq J_C$ ), then  $(\underline{b}, R)$  are globally identified.*

*Proof.* Since we can permute alternatives, we need only to consider the leading  $(2 \times 1)$  block of the reduced  $\{\Gamma_s^*\}$ , say

$$\begin{pmatrix} \rho_1^s + b_1(\rho_1^s - \rho_J^s) \\ b_1(\rho_2^s - \rho_J^s) \end{pmatrix}.$$

As was the case for  $\gamma_s^*$  in Equation (14) we note that

$$\rho_1^s + b_1(\rho_1^s - \rho_J^s) \equiv \rho_J^s - (1 + b_1)(\rho_J^s - \rho_1^s).$$

If, however,  $\rho_1 \neq \rho_2$  or  $\rho_1 = \rho_2$  but  $b_1 \neq -1/2$  (in which case  $-(1 + b_1) \neq -b_1$ ), the off-diagonal element (2, 1) prevents permuting  $\rho_1$  and  $\rho_J$ .  $\square$

### 3.4 Reinterpreting the standard DMP

The standard DMP model (see e.g. Börsch-Supan et al., 1990 or Geweke et al., 1997) requires selecting a baseline alternative, say alternative  $J$ , expressing all utilities in deviation from  $u_{tJ}$  and assuming that the differences  $e_t = D\varepsilon_t$  follow a diagonal AR(1) process. This model is clearly not invariant with respect to the choice of the baseline alternative. Worse, in the absence of additional assumptions on the underlying utility process the DMP models associated with different baseline alternatives are non-nested with one another. This probably explains why, to the best of our knowledge, data based selection of the baseline alternative has not been discussed in the literature.

It turns out that our discussion of invariance and identification enables us to reinterpret the standard DMP model as one which is explicitly nested within a DIMP with diagonal  $R$  matrix. The key to such reinterpretation lies with Equation (41) which indicates that a diagonal  $R$  matrix also produces diagonal  $\Gamma_s^*$  matrices if  $\underline{b}$  is zero. However, in such a case the matrices  $\Gamma_{js}^*$  associated with baseline alternative  $j \neq J$  are not diagonal – confirming the lag of invariance of

the standard DMP model. Instead, following Equation (37) they are given by

$$\Gamma_{js}^* = \begin{pmatrix} \rho_1^s & 0 & \dots & 0 \\ 0 & \rho_2^s & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \rho_{J-1}^s \end{pmatrix} - \begin{pmatrix} \rho_1^s - \rho_j^s \\ \rho_2^s - \rho_j^s \\ \vdots \\ \rho_{J-1}^s - \rho_j^s \end{pmatrix} i_j', \quad (42)$$

which leaves  $\rho_J$  as the sole unidentified autocorrelation coefficient, irrespective of the baseline alternative. A direct comparison between Equations (41) and (42) – see also Equation (35) – indicates that the vector  $\underline{b}_j$  associated with baseline alternative  $j \neq J$  is given by

$$\underline{b}_j = S_j' i_j = -i_j. \quad (43)$$

This reinterpretation of the standard DMP model has fundamental implications for practitioners. Specifically, once the standard DMP model is nested within a DIMP model with diagonal  $R$  matrix, it is no longer advisable to select a priori (arbitrarily) a baseline alternative, say  $J$ , and to impose the assumption that the corresponding  $\Gamma_s^*$  matrices be diagonal. Instead, one can use the DIMP as maintained hypothesis under a parametrization  $(\Psi, \underline{b}, R)$  which is fundamentally invariant with respect to  $J$ . It is then trivial to test whether this DIMP simplifies into a standard DMP for baseline alternative  $j$  denoted by  $\text{DMP}_j$ . The corresponding null hypotheses are given by

$$H_{0j} (\text{“DMP}_j\text{”}) : \underline{b}_j = 0, \quad \text{for } j \neq J, \quad \text{and} \quad H_{0J} (\text{“DMP}_J\text{”}) : \underline{b} = 0, \quad (44)$$

all of which are trivially nested within the DIMP. In view of Equation (35) these

null hypotheses can all be reformulated in terms of  $\underline{b}$

$$H_{0j} : \underline{b} = -i_j, \quad \text{for } j \neq J, \quad \text{and} \quad H_{0J} : \underline{b} = 0. \quad (45)$$

Nesting DMP<sub>*j*</sub>s within a DIMP with diagonal  $R$  matrix removes all ambiguities as well as arbitrariness associated with the conventional interpretation of DMP models. Note, in particular, the complete disconnect between the selection of alternative  $J$  for parametrization purposes (with full invariance with respect to  $J$ ) and the eventual subsequent data based selection of particular DMP<sub>*j*</sub> (likelihood based tests of  $H_{0j}$  are clearly invariant with respect to  $J$ ).

The fact that under the null hypotheses  $H_{0j}$ : DMP<sub>*j*</sub> ( $j = 1, \dots, J$ ) the parameter  $\rho_j$  is not identified poses the problem that the asymptotic distribution of standard tests like the likelihood ratio (LR), Lagrange multiplier (LM), and Wald test are nonstandard, which means that the conventional critical values cannot be used to assess significance (see, e.g., Davies 1977,1987). In order to derive asymptotic optimal tests for such testing problems, Davies (1977, 1987) and Andrews and Ploberger (1994) consider mappings (supremum and average) of the LR, LM, and Wald statistics as functions in the nuisance parameter which is unidentified under the null hypothesis, while Hansen (1996) suggests a useful computational method for simulating appropriate asymptotic critical values. The extension of these methods to our nonlinear simulated Maximum Likelihood (ML) framework seems to be conceptionally straightforward but is beyond the scope of the present paper and is left for future research. A simple but computationally more demanding alternative to those asymptotic methods is to perform bootstrap versions of the LR, LM, and Wald test, where the model under consideration is used to simulate the finite-sample distribution of the corresponding

test statistics under the null hypothesis.

## 4 Estimation

Before we formally discuss simulated ML estimation, it should be noted that since DMP models are nested within DIMP models, the analysis which follows applies identically to both classes of models. The only difference is that a DIMP model includes the  $J$  additional parameters  $(\underline{b}, \rho_J)$ . This justifies our earlier claim of similar ease of computation, and the analysis which follows applies to DMPs as well as DIMP.

### 4.1 GHK and GHK-EIS

Let  $y' = (y_1, \dots, y_M)$  denote a  $M$ -dimensional Normal random vector with mean  $\mu$  and covariance matrix  $V$ :

$$y = \mu + L\eta, \quad \eta \sim N_M(0, I_M), \quad V = LL', \quad (46)$$

where  $L$  denotes the lower triangular Cholesky decomposition of  $V$ . The probability to be computed is that of the event  $D = \{y_i < 0; i = 1, \dots, M\}$ . Let  $\ell'_\tau = (\gamma'_\tau, \delta_\tau)$  denote the  $\tau$ th (lower triangular) row of  $L$  with  $\gamma_\tau \in \mathbb{R}^{\tau-1}$  and  $\delta_\tau > 0$ . The  $\tau$ th row of  $y$  is then given by

$$y_\tau = \mu_\tau + \gamma'_\tau \eta_{(\tau-1)} + \delta_\tau \eta_\tau, \quad (47)$$

with  $\eta'_{(\tau-1)} = (\eta_1, \dots, \eta_{\tau-1})$  and  $\eta_{(0)} = \emptyset$ . It follows that

$$P(D) = \int_{\mathbb{R}^M} \prod_{\tau=1}^M \varphi_{\tau}(\eta_{(\tau)}) d\eta, \quad (48)$$

with

$$\varphi_{\tau}(\eta_{(\tau)}) = \mathbb{I}(\eta_{\tau} < -\frac{1}{\delta_{\tau}}[\mu_{\tau} + \gamma'_{\tau}\eta_{(\tau-1)}]) \cdot \phi(\eta_{\tau}), \quad (49)$$

where  $\mathbb{I}$  denotes the indicator function and  $\phi$  the standardized normal density. Both GHK and GHK-EIS are importance sampling techniques that are described in Liesenfeld and Richard (2009). In short they rely upon a sequential parametric importance sampling density of the form

$$m(\eta; a) = \prod_{\tau=1}^M m_{\tau}(\eta_{\tau} | \eta_{(\tau-1)}; a_{\tau}), \quad (50)$$

with  $a' = (a_1, \dots, a_M) \in A = \times_{\tau=1}^M A_{\tau}$ . The density  $m_{\tau}$  obtains from an auxiliary density kernel  $k_{\tau}(\eta_{(\tau)}; a_{\tau})$  with known functional integral  $\chi_{\tau}(\eta_{(\tau-1)}; a_{\tau})$  in  $\eta_{\tau}$ , i.e.,

$$m_{\tau}(\eta_{\tau} | \eta_{(\tau-1)}; a_{\tau}) = \frac{k_{\tau}(\eta_{(\tau)}; a_{\tau})}{\chi_{\tau}(\eta_{(\tau-1)}; a_{\tau})}, \quad \chi_{\tau}(\eta_{(\tau-1)}; a_{\tau}) = \int_{\mathbb{R}} k_{\tau}(\eta_{(\tau)}; a_{\tau}) d\eta_{\tau}. \quad (51)$$

GHK-EIS relies upon a backward recursive sequence of auxiliary LS problems, selecting values of  $a_{\tau}$  which approximately minimize the MC sampling variances of the ratios  $\varphi_{\tau}\chi_{\tau+1}/k_{\tau}$  for  $\tau = M, \dots, 1$  (with  $\chi_{M+1} \equiv 1$ ). If we momentarily assume that  $\chi_{\tau+1}$  is given by the product of a gaussian kernel in  $\eta_{(\tau)}$  by a Normal cdf of the form  $\Phi(\omega_{\tau})$ , where  $\omega_{\tau}$  is a linear combination of  $\eta_{(\tau)}$ , then the only non-Gaussian kernel in the numerator  $\varphi_{\tau}\chi_{\tau+1}$  is  $\Phi(\omega_{\tau})$  itself. Whence  $k_{\tau}$  obtains by combining the various Gaussian kernels in that numerator with a Gaussian

approximations of  $\Phi(\omega_\tau)$  obtained by the LS approximation

$$\ln \Phi(\omega_\tau) \simeq -\frac{1}{2}(\hat{\alpha}_\tau \omega_\tau^2 + 2\hat{\beta}_\tau \omega_\tau + \hat{\kappa}_\tau), \quad (52)$$

with  $\hat{a}_\tau = (\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\kappa}_\tau)$ .

Analytical integration with respect to  $\eta_\tau$  of the resulting Gaussian kernel  $k_\tau$  over the range associated with the indicator function in Equation (49) produces a (recursive) functional integral which is itself in the form of a Gaussian kernel times a Gaussian cdf in  $\omega_{\tau-1}$ . Note that all Gaussian kernels common to  $\varphi_\tau \chi_{\tau+1}$  and  $k_\tau$  cancel out in the ratios  $\varphi_\tau \chi_{\tau+1}/k_\tau$ . It follows that the GHK-EIS estimate of  $P(D)$  is given by

$$\hat{P}_s(D) = \chi_1(\hat{a}_1) \cdot \frac{1}{S} \sum_{s=1}^S \left[ \prod_{\tau=1}^{M-1} \frac{\Phi(\tilde{\omega}_\tau^{(s)})}{\exp -\frac{1}{2}(\hat{\alpha}_\tau [\tilde{\omega}_\tau^{(s)}]^2 + 2\hat{\beta}_\tau \tilde{\omega}_\tau^{(s)} + \hat{\kappa}_\tau)} \right], \quad (53)$$

where  $\tilde{\omega}_\tau^{(s)}$  is a linear combination of  $\tilde{\eta}_{(\tau)}^{(s)}$ , and  $\{\{\tilde{\eta}_\tau^{(s)}\}_{\tau=1}^M\}_{s=1}^S$  denotes  $S$  i.i.d. trajectories drawn from the EIS sampler  $m(\eta; \hat{a})$  as defined in Equation (51) with  $\hat{a} = \{\hat{a}_\tau\}_{\tau=1}^M$ . As discussed in Richard and Zhang (2007), the values of the EIS parameters  $\hat{a}$  and, therefore, the simulated  $\{\tilde{\eta}_\tau^{(s)}\}_{\tau=1}^M$  trajectories depend upon the values of the model parameters  $(\mu, V)$ . Hence, continuity of the MC likelihood evaluation (53) in  $(\mu, V)$  requires that all trajectories  $\{\{\tilde{\eta}_\tau^{(s)}\}_{\tau=1}^M\}_{s=1}^S$  drawn under different values of  $\hat{a}$  be obtained by a set of Common Random Numbers (CRNs) pre-drawn from a canonical distribution, which does not depend on the parameters  $\hat{a}$ . In the present context, the CRNs consist of  $M \times S$  draws form a uniform distribution on  $[0, 1]$  to be transformed by inversion into truncated Gaussian draws from  $m_\tau(\eta_\tau | \tilde{\eta}_{\tau-1}^{(s)}; \hat{a}_\tau)$ .

As for GHK, it ignores the factor  $\chi_{\tau+1}$  in the construction of the auxiliary

Gaussian kernels  $k_\tau$ , which is equivalent to setting  $\hat{\alpha}_\tau = \hat{\beta}_\tau = \hat{\kappa}_\tau = 0$  in Equations (52) and (53). Whence, for a given simulation sample size  $S$  GHK is intrinsically numerically less efficient than GHK-EIS but computationally faster since it does not require computing the auxiliary LS approximations in Equation (52). See Liesenfeld and Richard (2009) for full details and for numerical examples which indicate that even if we increase the simulation sample size  $S$  for GHK in order to equate computing time, GHK-EIS remains the numerically more efficient procedure. All computations reported below rely upon GHK-EIS.

## 4.2 GHK-EIS likelihood evaluation

Since individuals in a probit panel data set are typically assumed to act independently from one another, we consider here a single individual for whom observations consist of a sequence of  $T$  indicators of selected alternatives  $\{j_1, \dots, j_T; j_t \in (1, \dots, J)\}$  together with a  $T \times K$  matrix of exogenous variables. As discussed in Section 3, the identified parameters are those associated with an (arbitrary) baseline alternative  $J$  and are denoted by  $(\Psi, \underline{b}, R, \beta)$ . The likelihood function is the joint probability of the observed sequence of choices.

The covariance matrix  $V$  of the  $T(J-1)$  vector  $e' = (e'_{j_1}, \dots, e'_{j_T})$  with  $e_{j_t} = S_{j_t} e_t$  consists of the following blocks:

- The variance of  $e_{j_t}$  obtain from Equation (34):

$$\text{Var}(e_{j_t}) = \Psi_{j_t} = S_{j_t} \Psi S'_{j_t} \quad (54)$$

- The covariance between  $e_{j_t}$  and  $e_{j_s}$  obtain from Equation (29):

$$\text{Cov}(e_{j_t}, e_{j_s}) = S_{j_t} \Gamma_{t-s}^* \Psi S'_{j_s}, \quad t > s. \quad (55)$$

Brute force Cholesky decomposition of the joint covariance matrix  $V$  enables us to rewrite the DIMP model in the form given by equation (46), from which GHK-EIS evaluation of the likelihood function proceeds as described in Section 4.1.

As discussed in Section 3.1, D(I)MP models obtain by marginalization with respect to the  $J$ th component of the error term  $\varepsilon_t^*$  in Equation (23). It is precisely this operation which led to the replacement of the correlation matrices  $(R_*)^{t-s}$  by  $\Gamma_{t-s}^*$  in Equation (55) and to the subsequent need for brute force Cholesky decomposition of the joint covariance matrix  $V$ . In contrast, the Cholesky decomposition of the covariance matrix of  $\{\varepsilon_t^*\}$ , transformed as needed to account for  $\{j_t\}$ , can take full advantage of the present AR(1) structure as discussed in the lemma below. Whence, an efficient algorithm would first compute that Cholesky decomposition, and only then marginalizing with respect to the  $T$   $J$ th components of  $\{\varepsilon_t^*\}$ . In order to do so, we complete the rectangular transformation from  $\varepsilon_t^* \in \mathbb{R}^J$  into  $e_{j_t} \in \mathbb{R}^{J-1}$  into the square transformation

$$e_{j_t}^0 = \begin{pmatrix} S_{j_t} & 0 \\ 0 & 1 \end{pmatrix} \varepsilon_t^* = K_{j_t} \varepsilon_t^*, \quad (56)$$

with  $K_{j_t}^{-1} = K_{j_t}$ . The covariance matrix  $V_0$  of the  $TJ$ -dimensional vector  $e_0^0 = (e_{j_1}^0, \dots, e_{j_T}^0)$  consists of the following blocks:

$$\text{Var}(e_{j_t}^0) = K_{j_t} \Phi_* K_{j_t}', \quad (57)$$

$$\text{Cov}(e_{j_t}^0, e_{j_s}^0) = K_{j_t} (R_*)^{t-s} \Phi_* K_{j_s}', \quad t > s. \quad (58)$$

Note that the covariance matrix  $V_0$  preserves the initial AR(1) structure (up to a set of squared transformations), as indicated by the presence of  $(R_*)^{t-s}$  in Equa-

tion (58) - instead of  $\Gamma_{t-s}^*$  in Equation (55). The Cholesky decomposition of  $V_0$  obtains by application of the following lemma (deleting the subscripts 0, \* and  $j$  for the ease of notation).

**Lemma.** *Let the  $TJ$ -dimensional stationary covariance matrix  $V$  be partitioned into  $J$ -dimensional blocks of the form*

$$V_{ts} = K_t(R)^{t-s}\Phi K'_s, \quad t \geq s, \quad (59)$$

with  $K_t^{-1} = K_t$ . Let  $L$  denote the lower triangular Cholesky decomposition of  $V$ , partitioned conformably with  $V$ . The diagonal blocks of  $L$  obtain by the following  $J$ -dimensional Cholesky factorizations:

$$L_{11}L'_{11} = K_1\Phi K'_1, \quad (60)$$

$$L_{tt}L'_{tt} = K_t\Sigma K'_t, \quad \text{with } \Sigma = \Phi - R\Phi R', \quad (61)$$

and its off-diagonal blocks by the products

$$L_{ts} = [K_t(R)^{t-s}K_s] L_{ss}, \quad t > s. \quad (62)$$

*Proof:* See Liesenfeld and Richard (2009, Appendix).

Note that  $K_t$  can only take  $J$  different forms for the D(I)MP, corresponding to each of the  $J$  different alternatives (see Equation 56). Whence the Cholesky factorization of the  $TJ$ -dimensional matrix  $V_0$  requires at most  $J$  different  $J$ -dimensional Cholesky factorizations as given in Equations (60) and (61).

The extension to  $V_0$  only requires two trivial modifications of the GHK-EIS algorithm for  $V$ : (i) Since the baseline utility  $u_{Jt}$  is not observed the corresponding

integral with respect to the last component of  $e_{jt}^0$  ( $t = 1, \dots, T$ ) is un-truncated. It produces a probability  $\Phi(\cdot)$  equal to 1 in Equations (52) and (53) with the corresponding EIS coefficients equal to zero; (ii) The stationary covariance matrix  $\Phi_*$  as given by Equations (26) and (27) includes an additional parameter  $v^2$  which is unidentified and can, therefore, be set equal to any (reasonable) arbitrary value without affecting the likelihood values.

## 5 Monte Carlo simulations

### 5.1 ML-GHK-EIS estimates of the DIMP model

In order to illustrate the identification and ease of estimation of DIMP models we conducted a Monte Carlo experiment for a DIMP panel model with  $J = 3$  alternatives,  $T = 10$  periods and  $N = 500$  individuals. We arbitrarily selected alternative 3 as the baseline (remembering that DIMP models are invariant with respect to the selection of the baseline alternative). Since our focus of attention lies on second order moments we specified the regression function of the utilities, as given in Equation (19) directly in differences as

$$D\mu_{it} = D\mu_i(x_t; \beta) = (\pi_1 + \psi Z_{it1}, \pi_2 + \psi Z_{it2})', \quad (63)$$

with  $\beta' = (\pi_1, \pi_2, \psi)$  and  $Z_{itj} \sim i.i.d.N(0, 1)$  for  $t = 1, \dots, T$ ,  $i = 1, \dots, N$  and  $j \in \{1, 2\}$ . Note that this specification represents a simplified version of that used by Geweke et al. (1997).

The autocorrelation matrix  $R$  in Equation (19) is diagonal with elements  $\rho_1 = 0.8$ ,  $\rho_2 = 0.6$  and  $\rho_3 = 0.3$ . The covariance matrix  $\Sigma = [\sigma_{ik}]$  has diagonal elements  $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$  and off-diagonal elements  $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0.3$ .

The corresponding stationary covariance matrix  $\Phi_*$  as defined in Equation (26) is given by

$$\Phi_* = \begin{pmatrix} 3.087 & & \\ 0.915 & 1.930 & \\ -0.704 & -0.733 & 1.099 \end{pmatrix}. \quad (64)$$

Its partitioning according to Equation (27) produces the following values for  $\Lambda = [\lambda_{ik}]$ , the Cholesky decomposition of  $\Psi$ , and  $\underline{b}$

$$\Lambda = \begin{pmatrix} 1.757 & 0 \\ 0.521 & 1.288 \end{pmatrix} \quad \underline{b} = - \begin{pmatrix} 0.134 \\ 0.316 \end{pmatrix}. \quad (65)$$

Since we are leaving  $D\mu_{it}$  unnormalized,  $\Psi$  is identified only up to a scale factor. Whence, for ease of comparison with the parameter true values we set  $\lambda_{11}$  equal to its true value 1.757. Had we set it equal to one as commonly done, we would have to divide the true values for  $(\pi_1, \pi_2, \psi, \lambda_{11}, \lambda_{12}, \lambda_{22})$  and their estimates by 1.757.

ML-GHK-EIS estimates are computed using for the likelihood evaluations (53) a simulation sample size  $S = 20$  which, as we shall see, suffices to produce high numerical accuracy in likelihood function IS estimates and corresponding ML parameter estimates. Following Richard and Zhang (2007) we conduct two sets of replications. For the first one we keep the simulated data set fixed and replicate ML-GHK-EIS estimation 100 times under different sets of CRNs used for the likelihood evaluations. The MC standard deviations of these replications are direct measures of numerical accuracy (accounting in particular for the simulation based selection of the EIS samplers). Since these replications are conducted for a fixed data set, we also computed Root Mean Squared Errors (RMSEs) relative to the “true” ML estimates (as computed using a simulation sample size  $S = 1,000$ ,

which suffices to produce highly accurate approximations to the infeasible true ML estimates).

Our second set of replications repeats ML-GHK-EIS under 100 different simulated data sets (using a fixed set of CRNs for all 100 replications). The resulting means and standard deviations are numerical estimates of the finite sample (statistical) distribution of the ML estimators. We also produced for comparison asymptotic standard deviations based upon the inverse Hessian (since EIS likelihood estimates are sufficiently smooth to produce numerically well-behaved estimates of the Hessian). The results are reproduced in Table 1. They unequivocally confirm that the DIMP model is identified and statistically well-behaved, including the parameters  $(b_1, b_2, \rho_3)$  which are not identified under the standard DMP (unless, as discussed in Section 3.4, the latter is reinterpreted as nested within the DIMP in which case  $b_1 = b_2 = 0$  a priori though  $\rho_3$  remains unidentified). The finite sample standard deviations are close to their asymptotic counterparts.

The numerical (MC) standard deviations are orders of magnitude smaller than the statistical standard deviations (with ratios ranging from 0.016 to 0.050). Though the results are not reported here, we also reproduced the entire simulation exercise under ML-GHK. Our findings confirm results presented in Liesenfeld and Richard (2009) in that for a common simulation sample size  $S = 20$  ML-GHK estimates are numerically far less accurate than ML-GHK-EIS estimates especially for the DIMP parameters  $(b_1, b_2, \rho_1, \rho_2, \rho_3)$  with ratios of numerical standard deviations ranging from 8 to 95. In that paper, where we provide a more detailed comparison of GHK and GHK-EIS for the estimation of the standard DMP model, we also produce results in which we increase the simulation sample size of the computationally less demanding GHK from 20 to 100 in order to roughly equalize computing time for both procedures. Doing so we found

that GHK-EIS with  $S = 20$  remains numerically more efficient than GHK with  $S = 100$ , though with efficiency gain reductions by a factor ranging between 1.2 and 2.1 relative to the comparison for a common simulation sample size  $S = 20$  (see Liesenfeld and Richard, 2009, Tables 2 and 4).

For each of the 100 ML-GHK-EIS estimates of the DIMP model under different simulated data sets we computed the three Wald statistics testing whether the DIMP simplifies into a corresponding  $DMP_j$  model, which amounts to testing the restrictions  $\underline{b} = -\iota_1$  ( $DMP_1$ ),  $\underline{b} = -\iota_2$  ( $DMP_2$ ), and  $\underline{b} = 0$  ( $DMP_3$ ). Figure 1 provides the histogram of the values for the three Wald statistics. As expected, for all three hypotheses nearly all values of the Wald statistic are significantly larger than the ‘naive’ 1% critical value from a  $\chi^2_{(2)}$ -distribution, which is used as a preliminary benchmark value. Of course, one would need, as mentioned above, to adjust the naive critical value since it ignores the fact that we are ‘scanning’ across a range of values for the nuisance parameter  $\rho_3$  which is unidentified under the null. This suggests that the appropriate critical value is larger than the naive one (see Davis, 1987). However, as shown in Section 5.3 below, the appropriate critical value in the present context remains in the same ballpark as its naive counterpart. Hence, the comparably large values of the Wald-statistics in Figure 1 suggest that such a Wald-type procedure is a useful tool to pretest whether an initial DIMP model simplifies into a DMP model for a particular baseline alternative.

The main conclusions to be drawn from this simulation exercise are that: (i) DIMP models appear to be statistically well-behaved, as expected in view of the identification theorem in Section 3.3; (ii) ML-GHK-EIS can produce numerically accurate ML estimates with as little as  $S = 20$  EIS draws; and (iii) a Wald test used to pretest whether an initial DIMP model can be substituted by a standard

DMP model with a particular reference category appears to have good power.

## 5.2 Estimates of a miss-specified DMP model

In order to analyze the impact of erroneously assuming a DMP<sub>*j*</sub> model, we used the same simulated data sets from the DIMP model as in Section 5.1 and computed the ML-GHK-EIS estimates for a miss-specified DMP<sub>3</sub> model, i.e., the DIMP model with the restriction  $\underline{b} = 0$  and unidentified  $\rho_3$ . The results are reproduced in the left columns of Table 2. They indicate that (statistical) biases of the estimates for the mean parameters  $\pi_{01}$ ,  $\pi_{02}$ , and  $\psi$  are comparably small while, in contrast, the estimates for the covariance parameters  $\lambda_{12}$ ,  $\rho_1$ ,  $\rho_2$  are typically severely biased due to the miss-specification.

Based on the likelihood values of the miss-specified DMP<sub>3</sub> model at the parameter estimates and those of the DIMP model obtained in Section 5.1 we computed the values of the LR statistic testing for the DMP<sub>3</sub> model. The lower right panel of Figure 1 provides the histogram of the LR values. Consistent with the results for the Wald statistic (see lower left panel of Figure 1) most of the LR values are significantly larger than the naive asymptotic critical value. However note that the finite-sample variation of the LR-statistic appears to be much smaller than that of the Wald statistic.

## 5.3 Testing for a DMP model

In order to analyze the finite-sample distribution of the Wald statistic under the null hypothesis of a DMP model, we simulated 100 data sets from a DMP<sub>3</sub> model and computed ML-GHK-EIS estimates for an unrestricted DIMP specification together with the corresponding Wald statistic for the DMP<sub>3</sub> model.

Under the restriction  $\underline{b} = (b_1, b_2)' = 0$ , the DIMP model with  $J = 3$  alternatives and  $j = 3$  as baseline alternative reduces to a  $\text{DMP}_3$  model. This amounts to setting the off-diagonal blocks of the stationary covariance matrix  $\Phi_*$  given by  $\Psi \underline{b}$  and  $\underline{b}' \Psi$  equal to zero (see Equations 26 and 27). This restriction on  $\Phi_*$  is satisfied if the parameters of the autocorrelation and covariance matrix  $(R, \Sigma)$  in the initial specification (19) satisfy the Equations

$$\sigma_{13} = \frac{1 - \rho_1 \rho_3}{1 - \rho_3^2} \sigma_{33} \quad \text{and} \quad \sigma_{23} = \frac{1 - \rho_2 \rho_3}{1 - \rho_3^2} \sigma_{33}. \quad (66)$$

We consider the following values for the  $(R, \Sigma)$  parameters:  $(\rho_1, \rho_2, \rho_3, \sigma_{11}, \sigma_{22}, \sigma_{33}) = (0.8, 0.6, 0.3, 1, 1, 0.2)$ , together with the restrictions given in Equation (66). The implied values for the identified parameters for the Cholesky decomposition of  $\Psi$ , and  $\underline{b}$  are  $(\lambda_{12}, \lambda_{22}, b_1, b_2) = (0.223, 1.137, 0, 0)$ .

The parameter estimates of the unrestricted DIMP model for the simulated data from a  $\text{DMP}_3$  specification are reported in the right columns of Table 2. As expected, the ML-GHK-EIS estimator is statistically well-behaved, except for the parameter  $\rho_3$ , which is unidentified under the data generating  $\text{DMP}_3$  model. Note also that the means of the  $b_1$  and  $b_2$  estimates are not significantly different from zero. The histograms for the Wald statistics testing for the  $\text{DMP}_1$ ,  $\text{DMP}_2$ , and  $\text{DMP}_3$  model are provided in Figure 2. As for the Wald statistic testing the  $\text{DMP}_3$  specification (see lower panel of Figure 1), in 2% of the cases one would reject the null using the naive 1% asymptotic critical value, in 8% for the 5% critical value, and in 10% for the 10% critical value. This clearly suggests that the naive asymptotic  $\chi^2$  distribution delivers a fairly good approximation. Hence, the appropriate critical values which account for the fact that the nuisance parameter  $\rho_3$  is unidentified under the null can be expected to be fairly close to

the naive ones. Finally, we observe that the histograms for the Wald statistic testing for a  $DMP_1$  and  $DMP_2$  model (see upper panels of Figure 2) indicate large deviations from the specification under the corresponding null, as expected.

The main conclusion to be drawn from this simulation experiment are that: (i) the naive asymptotic distribution of the Wald statistic under the null of a DMP model provides a fairly good approximation to the true one, and (ii) the Wald test appears to be able to discriminate between different DMP specifications obtained by using different baseline categories, allowing for data based selection of the latter.

#### 5.4 Predictions under a miss-specified DMP model

The histograms in Figure 1 show that when the true model is a DIMP model, both the Wald- and LR-test reject the miss-specified DMP models at very high rates. However, it is well-known that a miss-specified model can always be rejected given enough data even if the miss-specification leads to only trivial deterioration in fit and predictions. This raises the question of whether a miss-specified DMP model produces significant differences in substantively important predictions relative to a correctly specified DIMP. Note that both models share common specifications for their means with similar estimated mean coefficients in Tables 1 and 2 and differ only in their error correlation structure. It follows, for example, that we should expect DMP miss-specification to have a greater impact on predicted transition probabilities than on predicted marginal selection probabilities.

In order to substantiate these conjectures, we compare marginal selection probabilities and transition probabilities and their responses to changes in the covariates under the DIMP and  $DMP_3$  model estimated in Sections 5.1 and 5.2, setting the parameters at their estimated values in Tables 1 and 2, respectively.

In this context the DIMP parameters serve as true values while the  $DMP_3$  parameters implicitly represent pseudo-true values. Probabilities are accurately estimated by Monte Carlo using 1,000,000 simulated sequences of selected alternatives  $\{j_{i1}, \dots, j_{iT}\}$  for each model.

In Table 3 we compare the partial effects of changes in the covariates  $Z_{itj}$  ( $j = 1, 2$ ) on the marginal selection probabilities

$$\begin{aligned} P(j_{it}|Z_{it1} = 1, Z_{it2} = 0) &- P(j_{it}|Z_{it1} = Z_{it2} = 0), \quad \text{and} \\ P(j_{it}|Z_{it1} = 0, Z_{it2} = 1) &- P(j_{it}|Z_{it1} = Z_{it2} = 0), \quad \text{for } j_{it} = 1, 2, 3. \end{aligned}$$

As expected, percentage differences remain fairly small ranging from -6.82% to 7.34%. For reference, the selection probabilities  $P(j_{it}|Z_{it1} = Z_{it2} = 0)$  for  $j_{it} = 1, 2, 3$  are (0.573, 0.086, 0.342) under the DIMP model and (0.571, 0.084, 0.345) under the  $DMP_3$  model.

In Table 4 we report the transition probabilities evaluated at the means of the covariates

$$P(j_{it}|j_{it-1}, Z_{i.1} = Z_{i.2} = 0), \quad \text{for } (j_{it}, j_{it-1}) \in \{1, 2, 3\} \times \{1, 2, 3\},$$

where  $Z_{i.j} = (Z_{i1j}, \dots, Z_{iTj})$ , and observe larger percentage differences ranging from -16.11% to 29.65%.

Finally, we report in Table 5 the partial effects of a permanent change in the covariates on the transition probabilities

$$\begin{aligned} P(j_{it}|j_{it-1}, Z_{i.1} = 1, Z_{i.2} = 0) &- P(j_{it}|j_{it-1}, Z_{i.1} = Z_{i.2} = 0), \quad \text{and} \\ P(j_{it}|j_{it-1}, Z_{i.1} = 0, Z_{i.2} = 1) &- P(j_{it}|j_{it-1}, Z_{i.1} = Z_{i.2} = 0). \end{aligned}$$

Here we observe the largest percentage changes ranging from -19.35% to 59.65%.

These results fully confirm our conjecture that miss-specified DMP models can significantly distort predicted transition probabilities and their responses to changes in the covariates even when marginal selection probabilities are affected to a much lesser extent. We suspect that we could easily induce larger distortions of the predicted transition probabilities by assigning autocorrelation parameters  $\rho_j$  over the full range from  $-1$  to  $+1$  for the DIMP model instead of the  $[0.3, 0.8]$  interval used here. Therefore, it appears prudent to initially estimate a DIMP model in any substantive empirical application analyzing the dynamics of decision processes since the cost of generalizing a DMP estimation algorithm into a DIMP one essentially consists of a one-time implementation cost associated with modifications of the Cholesky decomposition of the corresponding joint covariance required for GHK(-EIS).

## 6 Conclusion

We have proposed a new specification for the multinomial multiperiod model with autocorrelated errors which is invariant with respect to the selection of the baseline alternative, in sharp contrast with commonly used models. Furthermore, it identifies parameters which are not identified under the standard approach (such as the parameter governing the dynamics of the utility for the reference category). The formal identification of the proposed invariant model has been illustrated by MC experiments. Since our model also nests the standard specifications as special cases, it becomes feasible to test whether it simplifies into a standard model for a particular baseline alternative.

## Acknowledgement

We are grateful to two anonymous referees for their helpful comments which have produced major clarifications on several key issues. Roman Liesenfeld acknowledges research support provided by the Deutsche Forschungsgemeinschaft (DFG) under grant HE 2188/1-1; Jean-François Richard acknowledges the research support provided by the National Science Foundation (NSF) under grant SES-0516642.

## References

- Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present under the alternative. *Econometrica* 62, 1383–1414.
- Bunch, D.S., 1991. Estimability in the multinomial probit model. *Transportation Research B* 25, 1–12.
- Börsch-Supan, A., Hajivassiliou, V., Kotlikoff, L., Morris, J. 1990. Health, children, and elderly living arrangements: a multiperiod multinomial probit model with unobserved heterogeneity and autocorrelated errors. NBER working paper No. 3343.
- Davis, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.
- Davis, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.
- Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, 571–578.
- Geweke, J., Keane, M., Runkle, D. 1997. Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics* 80, 125–165.
- Hajivassiliou, V., 1990. Smooth simulation estimation of panel data LDV models. Mimeo. Yale University.
- Hansen, B., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
- Keane, M., 1992. A note on identification in the multinomial probit model. *Journal of Business & Economics Statistics* 10, 193–200.

- Keane, M., 1994. A computationally practical simulation estimator for panel data. *Econometrica* 62, 95–116.
- Keane, M., 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business and Economic Statistics* 15, 310–327.
- Liesenfeld, R., Richard, J.-F., 2009. Efficient estimation of Probit models with correlated errors. Working paper, University of Kiel.
- Richard, J.-F., Zhang, W., 2007. Efficient high-dimensional importance sampling. *Journal of Econometrics* 141, 1385–1411.

Table 1. ML-EIS-GHK estimates for the DIMP model

Parameter		diff. data sets		fixed data set/diff. CRNs	
		true	ML-GHK-EIS	ML-value	ML-GHK-EIS
$\pi_{01}$	mean	.500	.495	.537	.539
	std. dev.		.063		.0009
	rmse		.063		.0023
	mean asy. s.e.		.049		
$\pi_{02}$	mean	-1.000	-1.008	-.961	-.968
	std. dev.		.083		.0027
	rmse		.084		.0079
	mean asy. s.e.		.066		
$\psi$	mean	1.000	1.003	1.020	1.024
	std. dev.		.036		.0009
	rmse		.036		.0044
	mean asy. s.e.		.030		
$\lambda_{12}$	mean	.521	.523	.550	.552
	std. dev.		.087		.0051
	rmse		.087		.0054
	mean asy. s.e.		.066		
$\lambda_{22}$	mean	1.288	1.288	1.291	1.296
	std. dev.		.073		.0026
	rmse		.073		.0059
	mean asy. s.e.		.056		
$b_1$	mean	-.134	-.139	.0003	-.004
	std. dev.		.095		.0024
	rmse		.095		.0051
	mean asy. s.e.		.068		
$b_2$	mean	-.316	-.313	-.427	-.418
	std. dev.		.099		.0040
	rmse		.099		.0098
	mean asy. s.e.		.072		
$\rho_1$	mean	.800	.802	.775	.775
	std. dev.		.029		.0008
	rmse		.029		.0008
	mean asy. s.e.		.024		

Table 1. Continued

Parameter		diff. data sets		fixed data set/diff. CRNs	
		true	ML-GHK-EIS	ML-value	ML-GHK-EIS
$\rho_2$	mean	.600	.594	.687	.686
	std. dev.		.066		.0019
	rmse		.066		.0021
	mean asy. s.e.		.045		
$\rho_3$	mean	.300	.245	.317	.302
	std. dev.		.160		.0055
	rmse		.169		.0166
	mean asy. s.e.		.110		

*NOTE:* The reported numbers for ML-GHK-EIS are mean, standard deviation, RMSE, and the mean of the asymptotic standard errors obtained for  $S = 20$ . The asymptotic standard errors are obtained from a numerical approximation to the Hessian. For the experiment with a fixed data set and different CRNs, RMSE is computed around the true ML value for that particular data set. The true ML values are the ML-GHK-EIS estimates based on  $S = 1000$ .

Table 2. ML-EIS-GHK estimates for the DIMP/DMP model

Parameter		true model: DIMP / est. model: DMP <sub>3</sub>		true model: DMP <sub>3</sub> / est. model: DIMP	
		ML-GHK- true	ML-GHK- EIS	ML-GHK- true	ML-GHK- EIS
$\pi_{01}$	mean	.500	.469	.500	.515
	std. dev.		.063		.064
	rmse		.071		.066
	mean asy. s.e.		.052		.051
$\pi_{02}$	mean	-1.000	-1.043	-1.000	-.994
	std. dev.		.088		.116
	rmse		.098		.116
	mean asy. s.e.		.073		.070
$\psi$	mean	1.000	1.005	1.000	1.001
	std. dev.		.036		.047
	rmse		.037		.047
	mean asy. s.e.		.031		.032
$\lambda_{12}$	mean	.521	.423	.223	.245
	std. dev.		.094		.079
	rmse		.136		.081
	mean asy. s.e.		.075		.065
$\lambda_{22}$	mean	1.288	1.264	1.137	1.141
	std. dev.		.072		.093
	rmse		.076		.093
	mean asy. s.e.		.060		.058
$b_1$	mean	-.134		.000	-.010
	std. dev.				.410
	rmse				.410
	mean asy. s.e.				.129
$b_2$	mean	-.316		.000	.004
	std. dev.				.247
	rmse				.247
	mean asy. s.e.				.143
$\rho_1$	mean	.800	.726	.800	.810
	std. dev.		.017		.082
	rmse		.075		.082
	mean asy. s.e.		.014		.025

Table 2. Continued

Parameter		true model: DIMP / est. model: DMP <sub>3</sub>		true model: DMP <sub>3</sub> / est. model: DIMP	
		true	ML-GHK- EIS	true	ML-GHK- EIS
$\rho_2$	mean	.600	.537	.600	.604
	std. dev.		.035		.052
	rmse		.072		.052
	mean asy. s.e.		.029		.037
$\rho_3$	mean			.300	.189
	std. dev.				.891
	rmse				.898
	mean asy. s.e.				.346

NOTE: The reported numbers for ML-GHK-EIS are mean, standard deviation, RMSE, and the mean of the asymptotic standard errors obtained for  $S = 20$ . The asymptotic standard errors are obtained from a numerical approximation to the Hessian.

Table 3. Partial effects of a change in  $Z_{itj}$  on the selection probabilities

Alternative	$\Delta Z_{it1} = 1$			$\Delta Z_{it2} = 1$		
	$j_{it} = 1$	$j_{it} = 2$	$j_{it} = 3$	$j_{it} = 1$	$j_{it} = 2$	$j_{it} = 3$
DIMP	.207	-.045	-.162	-.085	.171	-.086
DMP <sub>3</sub>	.206	-.042	-.164	-.079	.172	-.093
Difference in percent	-.51	-6.82	1.24	-6.51	.47	7.34

Table 4. Transition probabilities

Alternative		$j_{it} = 1$	$j_{it} = 2$	$j_{it} = 3$
$j_{it-1} = 1$	DIMP	.781	.039	.180
	DMP <sub>3</sub>	.788	.045	.167
	Difference in percent	.86	16.68	-7.36
$j_{it-1} = 2$	DIMP	.268	.350	.382
	DMP <sub>3</sub>	.347	.332	.321
	Difference in percent	29.65	-5.09	-16.11
$j_{it-1} = 3$	DIMP	.300	.100	.600
	DMP <sub>3</sub>	.267	.089	.644
	Difference in percent	-11.08	-10.63	7.31

Table 5. Partial effects of a permanent change in  $Z_{itj}$  on the transition probabilities

Alternative	$\Delta Z_{it1} = 1$			$\Delta Z_{it2} = 1$			
	$j_{it} = 1$	$j_{it} = 2$	$j_{it} = 3$	$j_{it} = 1$	$j_{it} = 2$	$j_{it} = 3$	
$j_{it-1} = 1$	DIMP	.099	-.019	-.080	-.033	.076	-.043
	DMP <sub>3</sub>	.096	-.022	-.074	-.046	.088	-.042
	Difference in percent	-3.27	14.78	-7.55	40.06	15.37	-3.53
$j_{it-1} = 2$	DIMP	.121	-.048	-.074	-.045	.188	-.142
	DMP <sub>3</sub>	.135	-.059	-.076	-.072	.187	-.115
	Difference in percent	11.32	24.20	3.02	59.65	-28	-19.35
$j_{it-1} = 3$	DIMP	.133	-.026	-.107	-.042	.148	-.106
	DMP <sub>3</sub>	.122	-.021	-.101	-.043	.142	-.100
	Difference in percent	-8.44	18.90	-5.88	1.85	-4.36	-6.79

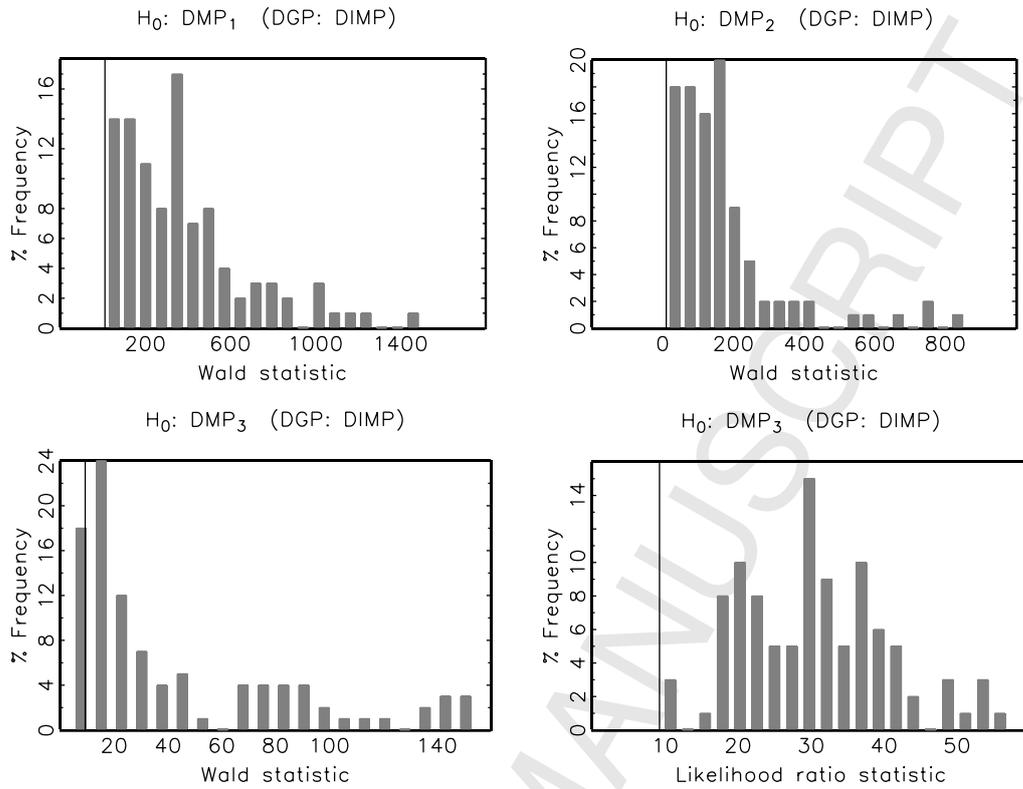


Figure 1. Histogram of the Wald statistic of  $H_0: DMP_j$  for  $j = 1$  (upper left panel),  $j = 2$  (upper right panel),  $j = 3$  (lower left panel) and histogram of the likelihood-ratio statistic for  $H_0: DMP_3$  (lower right panel); The true data generating process is the DIMP model as given by Equations (63) to (65). The vertical line marks the 99%-percentile of a  $\chi^2$ -distribution with 2 degrees of freedom given by 9.21.

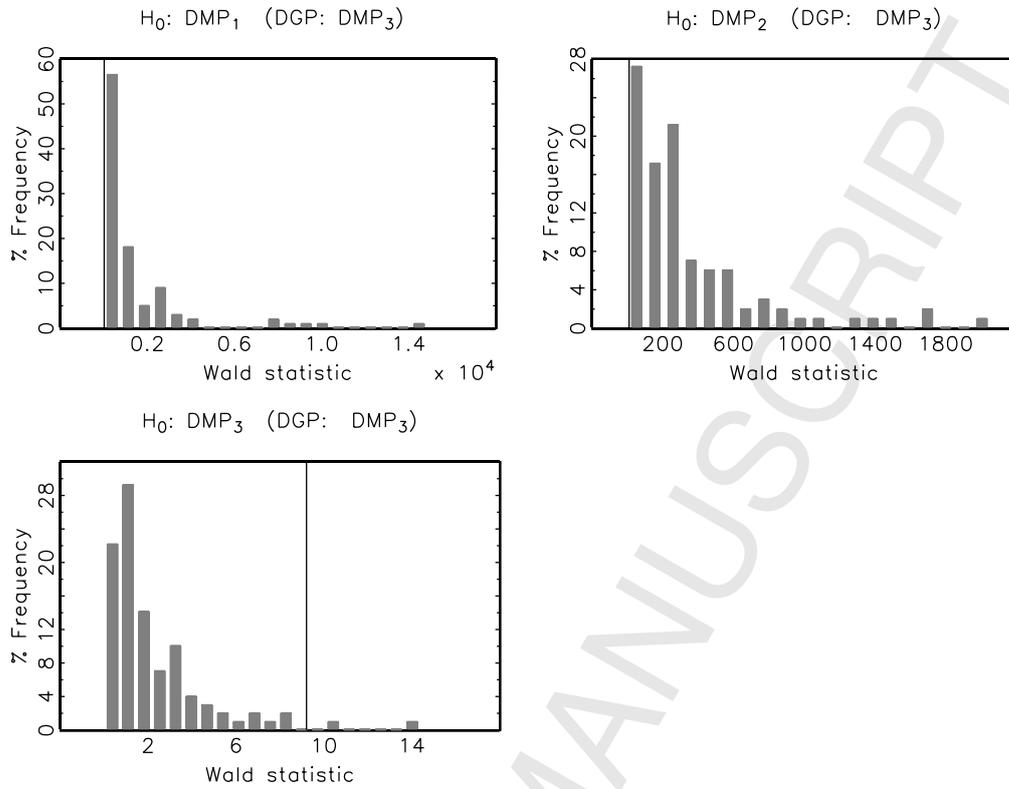


Figure 2. Histogram of the Wald statistic of  $H_0: DMP_j$  for  $j = 1$  (upper left panel),  $j = 2$  (upper right panel),  $j = 3$  (lower left panel); The true data generating process is the  $DMP_3$  model. The vertical line marks the 99%-percentile of a  $\chi^2$ -distribution with 2 degrees of freedom given by 9.21.