Liesenfeld, Roman; Richard, Jean-François

Working Paper

The Multinomial Multiperiod Probit Model: Identification and Efficient Estimation

Economics Working Paper, No. 2007-26

Provided in Cooperation with:
Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Liesenfeld, Roman; Richard, Jean-François (2007) : The Multinomial Multiperiod Probit Model: Identification and Efficient Estimation, Economics Working Paper, No. 2007-26

This Version is available at:
http://hdl.handle.net/10419/22042

The Multinomial Multiperiod Probit Model: Identification and Efficient Estimation

by Roman Liesenfeld and Jean-Francois Richard

Christian-Albrechts-Universität Kiel

Department of Economics

Economics Working Paper

No 2007-26

The Multinomial Multiperiod Probit Model: Identification and Efficient Estimation

Roman Liesenfeld*
Department of Economics, Christian Albrechts Universität, Kiel, Germany

Jean-François Richard
Department of Economics, University of Pittsburgh, USA

September 5, 2007

Abstract

In this paper we discuss parameter identification and likelihood evaluation for multinomial multiperiod Probit models. It is shown in particular that the standard autoregressive specification used in the literature can be interpreted as a latent common factor model. However, this specification is not invariant with respect to the selection of the baseline category. Hence, we propose an alternative specification which is invariant with respect to such a selection and identifies coefficients characterizing the stationary covariance matrix which are not identified in the standard approach. For likelihood evaluation requiring high-dimensional truncated integration we propose to use a generic procedure known as Efficient Importance Sampling (EIS). A special case of our proposed EIS algorithm is the standard GHK probability simulator. To illustrate the relative performance of both procedures we perform a set Monte-Carlo experiments. Our results indicate substantial numerical efficiency gains of the ML estimates based on GHK-EIS relative to ML estimates obtained by using GHK.

JEL classification: C35, C15
Keywords: Discrete choice, Importance sampling, Monte-Carlo integration, Panel data, Parameter identification, 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 multinomial multiperiod Probit (MMP) model and discuss formal parameter identification and likelihood evaluation. The MMP model represents a flexible framework to analyze repeated discrete choices such as, e.g., the living arrangement of the elderly (Börsch-Supan et al., 1990) or the brand choice in successive purchase occasion (McCullock and Rossi, 1994).

The standard dynamic specification commonly used in the literature assumes that the innovations to the utility differences w.r.t. the utility of the baseline decision follow a diagonal AR process, implicity treating the utility of the baseline decision as non-random – see, e.g., Börsch-Supan et al. (1990), McCullock and Rossi (1994), Geweke et al. (1997). However, such a specification is not invariant w.r.t. the choice of the baseline decision. This implies that parameter estimates obtained under different baseline alternatives are not one-to-one transformations of one another and, thus, not directly comparable. Here we propose a dynamic specification of the MMP model which is invariant w.r.t. the chosen baseline alternative. Moreover, it identifies parameters of the stationary covariance matrix which are not identified under the standard specification. These formal identification results will be illustrated by MC experiments.

The main obstacle to the practical implementation of the MMP is the difficulty in computing the choice probabilities involving high-dimensional truncated integration of a multivariate normal distribution. Thus likelihood-based estimation of the MMP model typically relies upon Monte Carlo (MC) integration (see Geweke and Keane, 2001). The most popular MC technique used for the computation of Gaussian choice probabilities is the GHK procedure developed by Geweke (1991), Hajivassiliou (1990), and Keane (1994). It has been applied to the

MMP model to obtain simulated ML estimates as well as estimates based on the method of simulated moments (see, e.g., Börsch-Supan et al., 1990, Keane, 1994, and Geweke et al., 1997). In an extensive study of alternative MC-procedures for the evaluation of probabilities, Hajivassiliou et al. (1996) find that GHK is the numerically most reliable among the considered alternatives. However, as illustrated by the MC study of Geweke et al. (1997), parameter estimates for the MMP model obtained by ML under GHK likelihood evaluation with the frequently used simulation sample size of 20 draws can be significantly biased, especially when the serial correlation in the innovations is strong.

As we shall argue further below the GHK procedure relies on importance sampling densities which ignore critical information relative to the underlying correlation structure of the model under consideration, leading to potentially significant numerical efficiency losses. In order to incorporate such information, we propose here to combine GHK with the Efficient Importance Sampling (EIS) methodology developed by Richard and Zhang (2007). EIS represents a powerful and generic high-dimensional integration technique, which is based on simple Least-Square approximations designed to maximize the numerical efficiency of the probability MC approximations. As such the GHK-EIS is well suited to handle the correlation structure in the MMP model and, thereby, provides highly accurate likelihood approximations. This approach is illustrated through a set of MC experiments. We compare the sampling distribution and the numerical accuracy of the ML estimator for the MMP model using GHK-EIS with those based on standard GHK. Our most important result is that under a common simulation sample size for both procedures, GHK-EIS leads to substantial numerical efficiency gains relative to GHK. Furthermore, the large biases of the ML estimators for the MMP model under GHK become negligible under the GHK-EIS

with only 20 draws.

The remainder of this paper is organized as follows. In the next section we discuss formal identification of the MMP model and propose a specification of the MMP which is invariant w.r.t. the selection of the baseline category. In Section 3 we describe the GHK-EIS procedure in the present context. The results of the MC experiments are discussed in Section 4 and conclusions are drawn in Section 5.

2 Parametrization and Identification

Identification of multinomial Probit models has been extensively discussed in the literature - see, e.g., Bunch (1991), Keane (1992), and Train (2003). The static one-period model is well understood and is discussed below mainly for the purpose of introducing notation. The dynamic multi-period specification is revisited in greater details. We shall argue that the standard autoregressive model, as discussed, e.g., by Börsch-Supan et al. (1990) and Geweke et al. (1997) can be reinterpreted as a latent common factor model and, relatedly, is not invariant with respect to the selection of the baseline category. We shall propose an alternative specification which is invariant w.r.t. such selection and identifies coefficients which are not identified in the standard model.

2.1 Static Multinomial Probit

Let $U = (U_1, ..., U_{J+1})'$ denote a $(J+1)$-dimensional vector of normally distributed random utilities

$$U \sim \mathrm{N}_{J+1}(\mu, \Sigma), \tag{1}$$

where U_j $(j = 1, \ldots, J+1)$ denotes the utility of the jth alternative. Alternative k is chosen if $U_k > U_j$ for all $j \neq k$. In most applications μ would be a linear function of observable exogenous variables but we shall treat it here as an unconstrained vector of unknown coefficients, focussing our attention on the (partial) identification of Σ . Actually, identification (whether formal or qualitative) of the coefficients of the exogenous variables has been extensively discussed elsewhere - see, e.g., Bunch (1991) and Keane (1992). Keane in particular shows that exclusive restrictions on the exogenous coefficients can significantly contribute to the qualitative identification of covariance parameters in Σ.

Observations consist solely of the indices of the selected alternatives, therefore, only depending upon utility differences. The standard approach consists of selecting a baseline alternative, say alternative $J+1$ and expressing all other utilities in differences from U_{J+1}. This amounts to introducing the non-singular transformation of variables

$$U^* = \begin{pmatrix} Y_J \\ U_{J+1} \end{pmatrix} = \begin{pmatrix} \Delta_J \\ e'_{(J+1)} \end{pmatrix} U = Q_J U, \tag{2}$$

where Δ_J is the $J \times (J+1)$ matrix

$$\Delta_J = (I_{(J)} \vdots -\iota_{(J)}), \tag{3}$$

$\iota'_{(J)} = (1, \ldots, 1)$ and $e'_{(J+1)}$ is the unit vector (0,...,0,1). Let partition the covariance matrix of U^* conformably with (Y'_J, U_{J+1}), say

$$\mathrm{Var}(U^*) = Q_J \Sigma Q'_J = \begin{pmatrix} \Psi & \Psi b \\ b'\Psi & \upsilon^2 + b'\Psi b \end{pmatrix}. \tag{4}$$

It follows that

$$Y_J \sim \mathrm{N}_J(\Delta_J\mu, \Psi) \qquad \text{and} \qquad U_{J+1}|Y_J \sim \mathrm{N}_1(a + b'Y, \upsilon^2), \tag{5}$$

with $a = (e'_{J+1} - b'\Delta_J)\mu$. Under the baseline category $J+1$ the parameters $\theta_1 = (\Delta_J\mu, \Psi)$ are identified up to a proportionality constant but since only the utility differences Y_J (or, as below, one-to-one transformations of Y_J) are relevant for the decisions, the parameters $\theta_2 = (a, b, \upsilon^2)$ are unidentified. Next consider what happens when alternative $j \neq J+1$ is selected as baseline alternative. Since $U_i - U_j = (U_i - U_{J+1}) - (U_j - U_{J+1})$, $U_{J+1} - U_j = -(U_j - U_{J+1})$, and $U_j = (U_j - U_{J+1}) + U_{J+1}$, we introduce the following non singular transformation

$$U_j^* = \begin{pmatrix} Y_j \\ U_j \end{pmatrix} = \begin{pmatrix} P_j & 0 \\ e'_{(j)} & 1 \end{pmatrix} U^*, \tag{6}$$

where P_j denotes the $J \times J$ non-singular matrix

$$P_j = \begin{pmatrix} I_{(j-1)} & -\iota_{(j-1)} & 0 \\ 0 & -1 & 0 \\ 0 & -\iota_{(J-j)} & I_{(J-j)} \end{pmatrix}, \tag{7}$$

with $P_j^{-1} = P_j$. It immediately follows that

$$Y_j \sim \mathrm{N}_J(P_j\Delta_J\mu \ , \ P_j\Psi P'_j) \qquad \text{and} \qquad U_j|Y_j \sim \mathrm{N}_1(a + b'_jY_j \ , \ \upsilon^2), \tag{8}$$

with $b'_j = -e'_{(j)} + b'P_j$. Note that Equation (8) remains valid for $j = J+1$ with $P_{J+1} = I_{(J)}$ and $b_{J+1} = b$. The parameters $(P_j\Delta_J\mu, P_j\Psi P'_j)$ and (a, b'_j, υ^2) are trivial one-to-one transformations of θ_1 and θ_2, respectively. Whence, an

approach which consists of leaving θ_2 unspecified is invariant with respect to the selection of the baseline alternative since the likelihood function depends solely on θ_1, irrespective of the reference category.

2.2 Dynamic Multiperiod Multinomial Probit

In order to explore the implicit restrictions underlying the conventional approach, as presented, e.g., by Börsch-Supan et al. (1990) and Geweke et al. (1997), we start by assuming first-order autocorrelation for the shocks to the individual utilities. In particular, let $U_t = (U_{1t}, \ldots, U_{J+1t})'$ denote the vector of utilities in time period t which evolve according to

$$U_t = \mu_t + \epsilon_t, \qquad \epsilon_t|\epsilon_{t-1} \sim \mathrm{N}_{J+1}(R\epsilon_{t-1}, \Sigma), \qquad t = 1, \ldots, T. \tag{9}$$

We apply the same baseline transformation as in the static case - see Equations (2) and (3). Thus we obtain

$$U_t^* = Q_J U_t = \mu_t^* + \epsilon_t^*, \qquad \epsilon_t^*|\epsilon_{t-1}^* \sim \mathrm{N}_{J+1}(R_*\epsilon_{t-1}^*, \Sigma_*), \tag{10}$$

with

$$\mu_t^* = Q_J\mu_t, \qquad \epsilon_t^* = Q_J\epsilon_t, \qquad \Sigma_* = Q_J\Sigma Q_J', \qquad R_* = Q_J R Q_J^{-1}, \tag{11}$$

where Q_J is defined in Equation (2) and

$$Q_J^{-1} = \begin{pmatrix} I_{(J)} & \iota_{(J)} \\ 0 & 1 \end{pmatrix}. \tag{12}$$

The first step consists in deriving the joint distribution of $\{\epsilon_t^*\}_{t=1}^T$, For small T it is generally assumed that ϵ_1^* is drawn from the stationary distribution of ϵ_t^*. Let Φ_* denote the corresponding stationary covariance matrix. It satisfies the identity

$$\Sigma_* = \Phi_* - R_*\Phi_* R_*'. \tag{13}$$

The stationary covariance between ϵ_s^* and ϵ_t^* is given by

$$\text{Cov}(\epsilon_t^*, \epsilon_s^{*\prime}) = R_*^{t-s}\Phi_*, \qquad s \leq t. \tag{14}$$

As in Equation (4), the non-observability of the baseline utility calls for partitioning Φ_* into

$$\Phi_* = \begin{pmatrix} \Psi & \Psi b \\ b'\Psi & \upsilon^2 + b'\Psi b \end{pmatrix}. \tag{15}$$

Next, let j_t represents the particular choice observed in period t and let $\epsilon_{j_t}^*$ denote the transformation of ϵ_t^* associated with the observation j_t. Following Equation (6), the J-dimensional vector $\epsilon_{j_t}^*$ is given by

$$\epsilon_{j_t}^* = (P_{j_t} \vdots 0)\, \epsilon_t^*. \tag{16}$$

As in the static case the probability that alternative j_t is chosen in period t depends only on $\epsilon_{j_t}^*$. The stationary distribution of $\{\epsilon_{j_t}^*\}_{t=1}^T$ is characterized by a covariance matrix containing the following blocks:

$$\text{Var}(\epsilon_{j_t}^*) = P_{j_t}\Psi P_{j_t}', \qquad \text{Cov}(\epsilon_{j_t}^*, \epsilon_{j_s}^{*\prime}) = P_{j_t}R_{t-s}^*\Psi P_{j_s}', \tag{17}$$

where

$$R^*_{t-s} = (I_{(J)} \vdots 0)\ R^{t-s}_* \begin{pmatrix} I_{(J)} \\ b' \end{pmatrix}. \tag{18}$$

According to these expressions we conclude the following: (i) As in the static case $\Delta_J \mu$ and Ψ are identified up to a proportionality factor. (Note that here Ψ denotes the stationary covariance matrix of $\Delta_J \epsilon_t$.) (ii) Identification of (R, b) requires that the transformation of (R, b) into $\{R^*_i\}_{i=1}^{T-1}$ is injective. Since there are at most $J(J+1)$ distinct elements in (R, b), the moments in (17) imply over-identification restrictions on $\{R^i_*\}_{i=1}^{T-1}$ and ML estimation as discussed further below has to account for these implied restrictions. But we have also to account for the possibility that (R, b) might be under-identified.

2.2.1 Standard Multinomial Multiperiod Probit

In the present paper we restrict our attention to two particular MMP specifications of R. The first one is the one commonly discussed in the literature (see, e.g., Börsch-Supan et al., 1990 and Geweke et al., 1997) whereby the differences $\Delta_J \epsilon_t$ are assumed to follow a diagonal AR(1) process. This implies that the first J rows of R_* are of the form

$$(I_{(J)} \vdots 0)\ R_* = (\Gamma \vdots 0), \tag{19}$$

where Γ denotes a diagonal matrix with elements $\{\rho_j\}_{j=1}^J$. It immediately follows that the ρ_js are identified but that, as for the static case, b is not. It is instructive to examine more closely the implications of this specification in terms of the initial R matrix in Equation (9). One verifies that Equation (19) requires that R be of

the following form

$$R = \begin{pmatrix} \Gamma & -\Gamma \cdot \iota_{(J)} \\ 0 & 0 \end{pmatrix} + \iota_{(J+1)} \cdot r', \tag{20}$$

with $r \in \mathbb{R}^{J+1}$ unrestricted (except for stationarity constraints).

This MMP specification calls for three qualifications. First, it is obviously not invariant with respect to the baseline alternative except in the special case where $R = \rho I_{(J+1)}$. Next, the implied form of the matrix R given by Equation (20) suggests that $r'\epsilon_{t-1}$ can be interpreted as a latent common factor to all $J+1$ components of ϵ_t. It only differs from ϵ_{J+1t} by an innovation and is eliminated by differencing w.r.t. U_{J+1t}. Finally, R includes additional restrictions which implies that after differencing it simplifies into an diagonal AR(1) in the differences. Under such interpretation it would be natural to model U_{J+1t} exclusively in terms of exogenous variables which are constant across alternatives. Moreover, interpreting U_{J+1t} as a latent common factor rather than a baseline alternative leads to considering that there are only J actual alternatives. An alternative interpretation to this specification suggested by Geweke et al. (1997) is that the baseline utility U_{J+1t} is non random with $\epsilon_{J+1t} = 0$ for all t (and $r' = 0$).

2.2.2 Invariant Multinomial Multiperiod Probit

Short of the above justifications, one might prefer dynamic MMP specifications which are invariant with respect to the choice of the baseline alternative as for the static case. One such specification which we discuss next is that where R is diagonal with diagonal elements $\{\rho_j\}_{j=1}^{J+1}$. In such a case the elements of R^*_{t-s} in

Equation (18) are given by

$$[R^*_{t-s}]_{i,i} = \rho_i^{t-s} + (\rho_i^{t-s} - \rho_{J+1}^{t-s})b_i, \qquad i = 1, \dots J \tag{21}$$

$$[R^*_{t-s}]_{i,j} = (\rho_i^{t-s} - \rho_{J+1}^{t-s})b_j, \qquad i \neq j = 1, \dots J. \tag{22}$$

Invariance obtains as the result of the following theorem together with the fact that the stationary covariance matrix in Equation (17) can be rewritten as $\mathrm{Cov}(\epsilon^*_{j_t}, \epsilon^{*\prime}_{j_s}) = P_{j_t} R^*_{t-s} P_{j_t} P_{j_t} \Psi P'_{j_s}$, where P_j is defined in Equation (7).

Theorem 1. *The matrix $P_j R^*_{t-s} P_j$ has the same analytical form as R^*_{t-s} up to a permutation between the pairs (ρ_j, b_j) and (ρ_{J+1}, b_{J+1}) with $b_{J+1} = -(1 + \sum_{i=1}^J b_i)$.*

Proof. The proof is similar for all values of $t - s \geq 1$ and is given here only for $t - s = 1$. In view of Equation (7), we have the following five combinations to consider:

(i) $[P_j R^*_1 P_j]_{j,j} = e'_{(j)}\, R^*_1\, \iota_{(J)} = \rho_j + (\rho_j - \rho_{J+1}) \sum_{i=1}^J b_i$
$= \rho_{J+1} + (\rho_{J+1} - \rho_j) b_{J+1}$

(ii) $[P_j R^*_1 P_j]_{j,k} = -e'_{(j)}\, R^*_1\, e_{(k)} = (\rho_{J+1} - \rho_j) b_k,$

(iii) $[P_j R^*_1 P_j]_{k,j} = (e'_{(j)} - e'_{(k)})\, R^*_1\, \iota_{(j)} = \rho_j + (\rho_j - \rho_{J+1}) \sum_{i=1}^J b_i - \rho_k$
$-(\rho_k - \rho_{J+1}) \sum_{i=1}^J b_i = (\rho_k - \rho_j) b_{J+1}$

(iv) $[P_j R^*_1 P_j]_{k,l} = (e'_{(k)} - e'_{(j)})\, R^*_1\, e_{(l)} = (\rho_k - \rho_{J+1}) b_l - (\rho_j - \rho_{J+1}) b_l$
$= (\rho_k - \rho_j) b_l$

(v) $[P_j R^*_1 P_j]_{k,k} = (e'_{(k)} - e'_{(j)})\, R^*_1\, e_{(k)} = \rho_k + (\rho_k - \rho_j) b_k.$ $\square$

Next, we discuss the identification of $(\rho_1, \ldots, \rho_{J+1})$ and b. We first note that for $J+1=2$, R^*_{t-s} reduces to a scalar and, furthermore, that

$$\rho_1^{t-s} + (\rho_1^{t-s} - \rho_2^{t-s})b_1 \equiv \rho_2^{t-s} + (\rho_2^{t-s} - \rho_1^{t-s})b_2, \tag{23}$$

with $b_2 = -(1+b_1)$. Whence, (ρ_1, ρ_2) and b_1 are identified only up to a permutation between (ρ_1, b_1) and (ρ_2, b_2). For $J+1>2$, identification follows from the following theorem.

Theorem 2. *$(\rho_1, \ldots, \rho_{J+1})$ and b are identified as long as there are at least two pairs of distinct ρ_js.*

Proof. In view of theorem 1, we only need to consider the pairs (ρ_j, ρ_{J+1}). Consider first the case where $\rho_1 \neq \rho_{J+1}$ and $\rho_j = \rho_{J+1}$ for $j > 1$. Then, except for its first row, R^*_{t-s} is diagonal (see Equations 21 and 22). It follows that $\rho_2, ..., \rho_J$ are identified. But we can still permute ρ_1 and ρ_{J+1} (as above for the case $J+1=2$) with $b_{J+1} = -(1+b_1)$ and changing the sign of all other b_js.

Suppose next two (or more) ρ_i are different form ρ_{J+1} and different from each other, say ρ_1 and ρ_2. The leading 2×2 block of R^*_{t-s} is given by

$$[R^*_{t-s}]_{1:2,1:2} = \begin{pmatrix} \rho_1^{t-s} + (\rho_1^{t-s} - \rho_{J+1}^{t-s})b_1 & (\rho_1^{t-s} - \rho_{J+1}^{t-s})b_2 \\ (\rho_2^{t-s} - \rho_{J+1}^{t-s})b_1 & \rho_2^{t-s} + (\rho_2^{t-s} - \rho_{J+1}^{t-s})b_2 \end{pmatrix}, \tag{24}$$

for $t-s = 1, 2, \ldots$. As for the case where $J+1=2$ in Equation (23), we only need to consider permutations between (ρ_i, b_i) for $i=1$ and/or 2 and (ρ_{J+1}, b_{J+1}). But any such permutation is excluded by the off-diagonal elements of the leading block. □

Theorem 2 implies that the invariant specification of the MMP with diagonal R identifies coefficients of the stationary covariance matrix which are not identified in the standard specification used, e.g., by Börsch-Supan et al. (1990) and Geweke et al. (1997). These identification results for $J+1 \geq 2$ will be illustrated in section 4.3 below. Finally, we note that the standard specification in Equation (20) and the invariant specification with diagonal R are non-nested within one another.

3 GHK and GHK-EIS Algorithm

The presentation of the generic GHK and GHK-EIS is fairly straightforward as it relies upon standard Gaussian algebra. Moreover, GHK turns out to be a special case of the GHK-EIS so that only the latter needs to be presented in full. In section 3.1 we present the GHK-EIS algorithm under streamlined notation ignoring individual and time indices. Its application to the static model and to the multiperiod models introduced above are presented in Section 3.2 and 3.3, respectively.

3.1 GHK-EIS baseline algorithm

The probabilities to be computed are those associated with events of the form $y < 0$, where $y' = (y_1, \ldots, y_M)$ denotes a M-dimensional multivariate normal latent random vector with mean μ and covariance matrix V. Let L denote the lower triangular Cholesky decomposition of V so that $V = LL'$. It follows that y is given by

$$y = \mu + L\eta, \qquad \eta \sim \mathrm{N}_M(0, I_{(M)}). \tag{25}$$

We aim at computing efficiently the probability that $y \in D$, where $D = \{y; y_\tau < 0, \tau = 1, \ldots, M\}$.

Let ℓ'_τ denote the τth (lower triangular) row of L, partitioned as

$$\ell'_\tau = (\gamma'_\tau, \delta_\tau), \tag{26}$$

with $\gamma_\tau \in \mathbb{R}^{\tau-1}$ and $\delta_\tau > 0$. The τth component of y is given by

$$y_\tau = \mu_\tau + \gamma'_\tau \eta_{(\tau-1)} + \delta_\tau \eta_\tau, \tag{27}$$

with $\eta'_{(\tau-1)} = (\eta_1, \ldots, \eta_{\tau-1})$ and $\eta_{(0)} = \emptyset$. The probability to be computed is given by

$$P(D) = \int_{\mathbb{R}^M} \prod_{\tau=1}^{M} \varphi_\tau(\eta_{(\tau)}) d\eta, \tag{28}$$

with

$$\varphi_\tau(\eta_{(\tau)}) = \mathbb{I}(\eta_\tau < -\frac{1}{\delta_\tau}[\mu_\tau + \gamma'_\tau \eta_{(\tau-1)}])\, \phi(\eta_\tau), \tag{29}$$

where $\mathbb{I}$ denotes the indicator function and ϕ the standardized normal density function. Both GHK and GHK-EIS are MC Importance Sampling (IS) techniques which aim at constructing auxiliary parametric sequential samplers of the form

$$m(\eta; a) = \prod_{\tau=1}^{M} m_\tau(\eta_\tau | \eta_{(\tau-1)}, a_\tau), \tag{30}$$

with $a' = (a_1, \ldots, a_M) \in A = \times_{\tau=1}^{M} A_\tau$. The corresponding IS estimate of $P(D)$ is then given by

$$\hat{P}_S(D; a) = \frac{1}{S} \sum_{s=1}^{S} \omega(\tilde{\eta}^{(s)}; a), \quad \text{where} \quad \omega(\eta; a) = \prod_{\tau=1}^{M} \frac{\varphi_\tau(\eta_{(\tau)})}{m_\tau(\eta_\tau | \eta_{(\tau-1)}, a_\tau)} \tag{31}$$

and $\{\tilde{\eta}^{(s)}; s = 1, \ldots, S\}$ denotes S *i.i.d.* simulated trajectories drawn from m. A trajectory is a sequential draw of η whereby $\tilde{\eta}_\tau^{(s)}$ is drawn from $m_\tau(\eta_\tau|\tilde{\eta}_{(\tau-1)}^{(s)}, a_\tau)$. For a preassigned class of auxiliary samplers $M = \{m(\eta; a); a \in A\}$ whose selection is discussed below, the objective of EIS is that of selecting $\hat{a} \in A$ which (approximately) minimizes the MC sampling variance of $\hat{P}_S(D; a)$. The EIS algorithm is briefly presented next in order to establish notation. See Richard and Zhang (2007) for details.

Note that the integral of $\varphi_\tau(\eta_{(\tau)})$ with respect to η_τ is a function of $\eta_{(\tau-1)}$. Whence, we cannot approximate it directly by a proper density $m_\tau(\eta_\tau|\eta_{(\tau-1)}, a_\tau)$ which integrates to one w.r.t. η_τ by definition. Instead we shall approximate $\varphi_\tau(\eta_{(\tau)})$ as a function of $\eta_{(\tau)}$ by a density kernel $k_\tau(\eta_{(\tau)}; a_\tau)$ with known functional integral $\chi_\tau(\eta_{(\tau-1)}; a_\tau)$ in η_τ. The relationship between χ_τ and m_τ is given by

$$m_\tau(\eta_\tau|\eta_{(\tau-1)}, a_\tau) = \frac{k_\tau(\eta_{(\tau)}; a_\tau)}{\chi_\tau(\eta_{(\tau-1)}; a_\tau)}, \text{ with } \chi_\tau(\eta_{(\tau-1)}; a_\tau) = \int_{\mathbb{R}} k_\tau(\eta_{(\tau)}; a_\tau) d\eta_\tau. \quad (32)$$

The integral in Equation (28) is then rewritten as

$$P(D; a) = \chi_1(a_1) \int_{\mathbb{R}^M} \prod_{\tau=1}^{M} \frac{\varphi_\tau(\eta_{(\tau)}) \cdot \chi_{\tau+1}(\eta_{(\tau)}; a_{\tau+1})}{k_\tau(\eta_{(\tau)}; a_\tau)} \cdot \prod_{\tau=1}^{M} m_\tau(\eta_\tau|\eta_{(\tau-1)}, a_\tau) d\eta, \quad (33)$$

with $\chi_{M+1}(\cdot) \equiv 1$. EIS aims at selecting values of a_τ which minimizes the MC sampling variances of the ratios $\varphi_\tau \chi_{\tau+1}/k_\tau$. As described in greater details in Richard and Zhang (2007), near optimal values $\{\hat{a}_\tau; \tau = 1, \ldots, M\}$ obtain as solutions of the following backward recursive sequence of fixed point auxiliary

Least Squares (LS) problems (for $\tau = M, M-1, \ldots, 1$):

$$(\hat{\kappa}_\tau, \hat{a}_\tau) = \arg\min_{\kappa_\tau, a_\tau} \sum_{s=1}^{S} \left\{ \ln\left[\varphi_\tau\big(\tilde{\eta}_{(\tau)}^{(s)}\big) \cdot \chi_{\tau+1}\big(\tilde{\eta}_{(\tau)}^{(s)}; \hat{a}_{\tau+1}\big)\right] - \kappa_\tau - \ln k_\tau\big(\tilde{\eta}_{(\tau)}^{(s)}; a_\tau\big) \right\}^2, \tag{34}$$

where $\{\tilde{\eta}^{(s)}, s = 1, \ldots, S\}$ denotes *i.i.d.* trajectories drawn from $m(\eta; \hat{a})$ – whence the need for fixed point iterations on these auxiliary LS problems. As starting values we propose to use the values of the auxiliary parameters a implied by the GHK sampler discussed further below. Note that if k_τ is a kernel of a (truncated) Gaussian density, as it is the case below, the LS problems in Equation (34) are linear under their natural parametrization in the sense of Lehmann (1986, Section 2.7). In order to guarantee fast and smooth fixed-point convergence it is critical that all trajectories $\{\tilde{\eta}^{(s)}\}$ be obtained by a transformation of a set of Common Random Numbers (CRNs) $\{\tilde{u}^{(s)}\}$ pre-drawn from a canonical distribution, i.e. one that does not depend on the parameters a. In the present context, the CRNs consists of draws from a uniform distribution on $[0, 1]$ to be transformed into truncated Gaussian draws from $m_\tau(\eta_\tau | \tilde{\eta}_{(\tau-1)}^{(s)}, a_\tau)$ by inversion.

The following theorem provides closed form recursive expressions for GHK-EIS evaluation of $P(D)$ as defined in Equations (32) to (34).

Theorem 3. *If*

(i) $\chi_{\tau+1}(\eta_{(\tau)}, a_{\tau+1})$ *is of the form*

$$\chi_{\tau+1}(\eta_{(\tau)}; a_{\tau+1}) = \Phi(c_{\tau+1} - d'_{\tau+1}\eta_{(\tau)}) \cdot \chi^*_{\tau+1}(\eta_{(\tau)}), \tag{35}$$

where Φ *denotes the standardized normal c.d.f. and* $\chi^*_{\tau+1}$ *the Gaussian den-*

sity kernel

$$\chi^*_{\tau+1}(\eta_{(\tau)}) = \exp -\frac{1}{2}(\eta'_{(\tau)} P^*_{\tau+1}\eta_{(\tau)} - 2\eta'_{(\tau)} q^*_{\tau+1} + r^*_{\tau+1}); \tag{36}$$

(ii) $k_\tau(\eta_{(\tau)}; a_\tau)$ *is defined as the following product of Gaussian density kernels*

$$k_\tau(\eta_{(\tau)}; a_\tau) = \varphi_\tau(\eta_{(\tau)}) \cdot \chi^*_{\tau+1}(\eta_{(\tau)}) \cdot k^*_\tau(\eta_{(\tau)}), \tag{37}$$

where $\ln k^*_\tau$ *denotes an EIS quadratic approximation to* $\ln \Phi$ *of the form*

$$-2 \ln \Phi(\omega_\tau) \doteq \hat{\alpha}_\tau \omega^2_\tau + 2\hat{\beta}_\tau \omega_\tau + \hat{\kappa}_\tau, \tag{38}$$

where $\omega_\tau = c_{\tau+1} - d'_{\tau+1}\eta_{(\tau)}$*;*

Then $\chi_\tau(\eta_{(\tau-1)}, a_\tau)$ *has the same analytical form as* $\chi_{\tau+1}(\eta_{(\tau)}, a_{\tau+1})$ *with coefficients* $(c_\tau, d_\tau, P^*_\tau, q^*_\tau, r^*_\tau)$ *obtained as described in the proof which follows.*

Proof. The proof follows from a sequence of standard algebraic operations on Gaussian kernels.

Step 1: Recombine the three kernels in Equation (37) into a single one of the form

$$-2 \ln k_\tau(\eta_{(\tau)}; a_\tau) \doteq \eta'_{(\tau)} P_\tau \eta_{(\tau)} - 2\eta'_{(\tau)} q_\tau + r_\tau + \ln(2\pi) \tag{39}$$

with the symbol $\doteq$ momentarily accounting for the omission of the indicator

function in Equation (29) and

$$P_\tau = P^*_{\tau+1} + \hat{\alpha}_\tau d_{\tau+1} d'_{\tau+1} + e_{(\tau)} e'_{(\tau)} \tag{40}$$

$$q_\tau = q^*_{\tau+1} + (\hat{\alpha}_\tau c_{\tau+1} + \hat{\beta}_\tau) d_{\tau+1} \tag{41}$$

$$r_\tau = r^*_{\tau+1} + \hat{\alpha}_\tau c^2_{\tau+1} + 2\hat{\beta}_\tau c_{\tau+1} + \hat{\kappa}_\tau, \tag{42}$$

with $e'_{(\tau)} = (0, \ldots, 0, 1)$

Step 2 (for $\tau > 1$): Partition P_τ and q_τ conformably with $\eta_{(\tau)} = (\eta'_{(\tau-1)}, \eta_\tau)$ into

$$P_\tau = \begin{pmatrix} P^\tau_{00} & P^\tau_{01} \\ P^\tau_{10} & P^\tau_{11} \end{pmatrix}, \qquad q_\tau = \begin{pmatrix} q^\tau_0 \\ q^\tau_1 \end{pmatrix}. \tag{43}$$

Next, factorize k_τ into the product of a Gaussian kernel for $\eta_\tau | \eta_{(\tau-1)}$ and one for $\eta_{(\tau-1)}$, say

$$\begin{aligned} -2 \ln k_\tau(\eta_{(\tau)}; a_\tau) \doteq P^\tau_{11} [\eta_\tau - \bar{m}_\tau(\eta_{(\tau-1)})]^2 + \eta'_{(\tau-1)} P^*_\tau \eta_{(\tau-1)} \\ -2\eta'_{(\tau-1)} q^*_\tau + s^*_\tau + \ln(2\pi), \end{aligned} \tag{44}$$

with

$$\bar{m}_\tau(\eta_{(\tau-1)}) = \frac{1}{P^\tau_{11}}(q^\tau_1 - P^\tau_{10}\eta_{(\tau-1)}), \qquad P^*_\tau = P^\tau_{00} - \frac{1}{P^\tau_{11}} P^\tau_{01} P^\tau_{10} \tag{45}$$

$$q^*_\tau = q^\tau_0 - \frac{1}{P^\tau_{11}} P^\tau_{01} q^\tau_1, \qquad s^*_\tau = r_\tau - \frac{1}{P^\tau_{11}} (q^\tau_1)^2. \tag{46}$$

Step 3: Integrate the η_τ Gaussian kernel over the support associated with

$\varphi_\tau(\eta_{(\tau)})$ as defined by Equation (29) – accounting for the the indicator

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}^1}\mathbb{I}(\eta_\tau < -\frac{1}{\delta_\tau}[\mu_\tau+\gamma_\tau'\eta_{(\tau-1)}])\cdot\exp-\frac{1}{2}P_{11}^\tau[\eta_\tau-\bar{m}_\tau(\eta_{(\tau-1)})]^2 d\eta_\tau \\ = \frac{1}{\sqrt{P_{11}^\tau}}\Phi(c_\tau - d_\tau'\eta_{(\tau-1)}),\end{aligned} \tag{47}$$

with

$$c_\tau = -\sqrt{P_{11}^\tau}\left(\frac{\mu_\tau}{\delta_\tau}+\frac{q_1^\tau}{P_{11}^\tau}\right), \qquad d_\tau = \sqrt{P_{11}^\tau}\left(\frac{\gamma_\tau}{\delta_\tau}-\frac{P_{01}^\tau}{P_{11}^\tau}\right). \tag{48}$$

Finally, the log of the multiplicative constant in Equation (47) is combined with s_τ^* in Equation (46) so that r_τ^* is given by

$$r_\tau^* = s_\tau^* + \ln P_{11}^\tau. \tag{49}$$

Note that for $\tau = 1$ with $\eta_1|\eta_{(0)} = \eta_1$, we skip step 2 and delete all subsequent terms with a subscript 0. □

The simplicity of the GHK-EIS auxiliary regression follows from the fact that the first two factors in k_τ as defined in Equation (37) are also included in the product $\varphi_\tau\chi_{\tau+1}$, where $\chi_{\tau+1}$ was defined in Equation (35). Whence, these two factors cancel out in the auxiliary regression of $\ln(\varphi_\tau\chi_{\tau+1})$ on $\ln k_\tau$ which simplifies into a trivial bivariate OLS regression of $\ln\Phi(\omega_\tau)$ on ω_τ^2 and ω_τ and a constant as defined in Equation (38), with $\omega_\tau = c_{\tau+1} - d_{\tau+1}'\eta_{(\tau)}$. Additional implementation details for the static and multiperiod models are discussed in the next two subsections.

Note also that theorem 3 covers standard GHK as special case with $\hat{\alpha}_\tau = \hat{\beta}_\tau = \hat{\kappa}_\tau = 0$ such that the GHK sampling densities have the form

$$m_\tau(\eta_\tau|\eta_{(\tau-1)}, a_\tau) = \frac{\mathbb{I}(\eta_\tau < -\frac{1}{\delta_\tau}[\mu_\tau+\gamma_\tau'\eta_{(\tau-1)}])\,\phi(\eta_\tau)}{\Phi(-\frac{1}{\delta_\tau}[\mu_\tau+\gamma_\tau'\eta_{(\tau-1)}])}, \qquad \tau = 1,...,M. \tag{50}$$

It trivially follows that GHK is numerically less efficient than GHK-EIS. Note in particular that the GHK density m_τ incorporates the constraints that $(y_1, ..., y_\tau) < 0$, but neglects the correlated information $(y_{\tau+1}, ..., y_M) < 0$. This implies that draws form m_τ ignore potentially critical information, which would allow to adjust the region of importance for η_τ, leading to potential efficiency losses of the MC-GHK estimate for the probability $P(D)$ (see also Stern, 1997). Accordingly, the GHK density m_τ can be interpreted as a filtering density incorporating the constraints on y only up to element τ. In contrast, EIS-GHK produces by its back-recursive transfer of the integrating constants χ_t – implemented by the back-recursive LS-problems (34) – sequential sampling densities for η_τ, which are conditional on the entire set of constraints on y.

3.2 GHK-EIS implementation for the static model

The application of GHK-EIS to the static model introduced in section 2.1 is straightforward. Under the assumption that observations are independent of one another the likelihood function for a particular observation is an integral of the form given in Equations (28) and (29) with $M = J$. Let j_i denote the index of the alternative chosen by observation i. According to Equation (8), Equation (25) is rewritten as

$$Y_{j_i} = P_{j_i}\Delta_J\mu_i + L_{j_i}\eta_i, \qquad \eta_i \sim \mathrm{N}_J(0, I_{(J)}), \tag{51}$$

where L_{j_i} denotes the Cholesky decomposition of the the covariance matrix $P_{j_i}\Psi P'_{j_i}$. Note that since there are only $J+1$ alternatives, we have at most $J+1$ Cholesky decompositions to compute.

3.3 GHK-EIS implementation for the multiperiod models

Under autocorrelation in the MMP model discussed in section 2.2 , the likelihood function for a particular individual has to properly account for time dependence across T successive observations. For moderate time dimensions, the simplest way to evaluate the likelihood for an individual amounts to express it as a single $M = J \cdot T$ dimensional integral of the form given by Equations (25) to (29) with $y = (Y'_{j_1}, ..., Y'_{j_T})'$, where $Y_{j_t} = P_{j_t} \Delta_J U_t$. The lower triangular matrix L in Equation (25) then denotes the Cholesky decomposition of the joint covariance matrix of $(\epsilon^*_{j_1}, ..., \epsilon^*_{j_T})$ as defined in Equations (17) and (18). The main advantage of this one-shot procedure (also used to implement the GHK for a multiperiod multinomial Probit, e.g., by Geweke et al., 1997) lies in its relative ease of programming since, beyond the construction of the larger $J \cdot T$-dimensional covariance matrix, it relies upon the same GHK-EIS steps as the static model. Note in particular that the EIS auxiliary regressions in Equation (38) depend upon only three coefficients irrespectively of the size $J \cdot T$.

Nevertheless, if $J \cdot T$ were significantly larger, there are two alternatives to the brute force Cholesky decomposition of a single $J \cdot T$-dimensional covariance matrix which could be considered at the cost of additional programming. The first alternative would consist of applying the baseline GHK-EIS procedure one-period at the time to the J-dimensional integrals with appropriate back-transfer of the integrating factor $\chi(\cdot)$ in order to account for autocorrelation. In a nutshell, this would require redefining $\eta_{(\tau-1)}$ in Equations (27) to (47) as the $t + (\tau - 1)$-dimensional vector $\eta'_{(\tau-1)} = ({\epsilon^*_{-1}}', \eta_1, ..., \eta_{\tau-1})$, where ϵ^*_{-1} denotes the vector of innovations $\epsilon^*_{j_{t-1}}$ associated with the alternative selected in period $t-1$ and $\eta_1, ..., \eta_{\tau-1}$ represents the first $\tau-1$ standardized innovations of period t associated with the choice j_t. The integration factor $\chi_1(a_{1t})$ in Equation (31) would then

depend on ϵ^*_{-1} and would have to be transferred back into the period $t-1$ integral. This would imply that, except for period T for which $\chi_{M+1}(\cdot)$ remains set to one, all other period integrals include an initial carry-over term of the form $\chi_{1t+1}(\epsilon^*_{j_t}; a_{1t+1})$. The principle of such a sequence of J-dimensional integrals is conceptually straightforward but tedious to implement.

A second alternative consists of constructing the $(J+1)\cdot T$ dimensional covariance matrix of $(\epsilon^*_t, ..., \epsilon^*_T)$ instead of that of $(\epsilon^*_{j_1}, ..., \epsilon^*_{j_T})$. While doing so increases the dimension of the relevant covariance matrix by T, it also replaces the rectangular transformation in Equation (16) by the square transformation

$$\epsilon^\diamond_{j_t} = \begin{pmatrix} P_{j_t} & 0 \\ 0 & 1 \end{pmatrix} \epsilon^*_t = Q_{j_t}\epsilon^*_t, \tag{52}$$

with $Q^{-1}_{j_t} = Q_{j_t}$. The Cholesky decomposition of the joint covariance matrix of $(\epsilon^\diamond_{j_1}, ..., \epsilon^\diamond_{j_T})$ can be efficiently computed by application of lemma A1 in the Appendix and is based upon individual Cholesky decomposition of (at most $J+1$) matrices of the form $Q_{j_t}\Phi_* Q'_{j_t}$. Note that the T additional integrals with respect to the $(J+1)$-th element of $\epsilon^\diamond_{j_t}$ are un-truncated and produce a probability $\Phi(\cdot)$ equal to one in Equation (35). Theorem 3 still applies with $(\hat{\alpha}_{J+1}, \hat{\beta}_{J+1}, \hat{\kappa}_{J+1})$ all set equal to zero in Equation (38).

4 Monte Carlo Results

4.1 Simulated Choice Probabilities for a Static Multinomial Probit

In order to evaluate the relative numerical accuracy of GHK-EIS and standard GHK for the static multinomial Probit, we consider the four simple examples used by Stern (1992) and Börsch-Supan and Hajivassiliou (1993). In these studies, choice probabilities according to Equation (5) with $J+1=5$ categories are computed for different parameter values of $\Delta_J\mu$ and Ψ. The parameter values are given by

$$\text{Example 1:} \quad \Delta_J\mu = (-1, -0.75, -0.5, -0.2), \quad \Psi = \begin{pmatrix} 1 & & & \\ 0.2 & 1 & & \\ 0.3 & 0.4 & 1 & \\ 0.1 & 0.3 & 0.5 & 1 \end{pmatrix};$$

$$\text{Example 2:} \quad \Delta_J\mu = (0, 0, 0, 0), \quad \Psi = \begin{pmatrix} 1 & & & \\ 0.2 & 1 & & \\ 0.2 & 0.4 & 1 & \\ 0.2 & 0.4 & 0.6 & 1 \end{pmatrix};$$

$$\text{Example 3:} \quad \Delta_J\mu = (1, 1, 1, 1), \quad \Psi = \begin{pmatrix} 1 & & & \\ 0.9 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 0.95 & 1 \end{pmatrix};$$

Example 4: $\Delta_J\mu = (1.5, 0.75, 0.5, 0.75), \quad \Psi = \begin{pmatrix} 1 & & & \\ 0.5 & 1 & & \\ 0.2 & 0.5 & 1 & \\ 0.1 & 0.2 & 0.5 & 1 \end{pmatrix}.$

Table 1 summarizes the results for the GHK-EIS and GHK MC approximations of the choice probabilities. The results which are reported are sample means, standard deviations and root mean squared errors (RMSE) based upon 1,000 independent replications of both algorithms. Each individual estimate is based upon a simulation sample size $S = 100$. The number of EIS (fixed point) iterations is set equal to three. One GHK-EIS probability evaluation requires 0.0060 s on a Intel Core 2 CPU notebook with 2 GHz for a code written in GAUSS and a GHK evaluation takes 0.0017 s. The true probability values are computed using iterated applications of product Gauss formulas (see Atkinson, 1978).

Our results for the standard GHK given in Table 1 are essentially the same as those reported by Börsch-Supan and Hajivassiliou (1993). Furthermore, we note that in all four cases the MC standard deviations of GHK-EIS are smaller than their GHK counterparts indicating that GHK-EIS is, as expected, numerically more accurate than GHK. We also notice that in the examples 1, 2, and 4 the improvement of GHK-EIS relative to GHK is substantially larger than in example 3. In fact, while in example 3 the GHK-EIS standard deviation is only 1.6 times smaller than the GHK counterpart, the GHK-EIS standard deviations in the remaining cases are between 19 (example 4) and 70 times (example 1) smaller. An obvious explanation for the comparably small efficiency gain of GHK-EIS relative to GHK in example 3 is found in the fact that in this case only the first

and second element and the third and fourth element of $\Delta_J U$ are correlated. Accordingly, the integrating factor χ_3 to be transferred into the approximation problem in $\eta_{(2)}$ (see Equation, 34) does not depend on $\eta_{(2)}$. Hence, the GHK-EIS and the GHK sampling density for η_2 are equivalent (in addition to that for the last element, η_4, which obtains by construction for all GHK-EIS applications). Finally, we note that while GHK-EIS requires about three times the computing time of GHK, the payoff is very substantial, at least for the non pathological examples 1, 2, and 4, as GHK would require between 360 and 4,900 times as many draws as GHK-EIS to reach the same accuracy.

4.2 Standard Multinomial Multiperiod Probit

In order to analyze the sampling distribution and numerical accuracy of the ML estimator based upon GHK and GHK-EIS for the MMP model, we use the same design of as Geweke et al. (1997). They consider a three alternative ($J+1=3$) probit model with $T=10$ periods and $N=500$ individuals based on the non-invariant normalization rule discussed in section 2.2.1. In particular, they use the following data generating process (DGP) for the utility differences of individual i:

$$\Delta_J U_{it} = \Delta_J \mu_{it} + \Delta_J \epsilon_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N \tag{53}$$

with

$$\Delta_J \mu_{it} = (\pi_{01} + \pi_{11} X_{it} + \psi Z_{it1} \ , \ \pi_{02} + \pi_{12} X_{it} + \psi Z_{it2})' \tag{54}$$

$$\Delta_J \epsilon_{it} = \Gamma \cdot (\Delta_J \epsilon_{it-1}) + v_{it} \tag{55}$$

$$v_{it} \sim \mathrm{N}_2 \left(0 \ , \ (1-\rho_1)^2 \cdot \begin{bmatrix} 1 & \omega_{12} \\ \omega_{12} & \omega_{12}^2 + \omega_{22}^2 \end{bmatrix} \right), \tag{56}$$

where Γ is a diagonal matrix with elements (ρ_1, ρ_2). The regressors X_{it} and Z_{itj} $(j = 1, 2)$ are constructed as follows:

$$X_{it} = \phi\zeta_i + \sqrt{1-\phi^2}\omega_{it}, \qquad Z_{itj} = \phi\tau_{ij} + \sqrt{1-\phi^2}\xi_{itj}, \tag{57}$$

with $|\phi| < 1$ and ζ_i, ω_{it}, τ_{ij} and ξ_{itj} being $i.i.d.$ standard normal random variables which are independent among each other.

We use this DGP to construct 20 artificial data sets to obtain the sampling distribution of the ML-GHK and ML-GHK-EIS estimator. In order to make our results directly comparable to those of Geweke et al. (1997), we estimated the MMP for each simulated data set under a different set of CRNs. The resulting sampling distribution compounds the statistical and numerical variation of the simulation based estimators. As discussed in Richard and Zhang (2007), the analysis of the conventional statistical properties of the estimators would actually require to obtain estimates for the different data set under a fixed set of CRNs. However, since in the present case the numerical variation of the estimates is dominated by the statistical variation, the compound sampling distribution of the estimators provides a very close approximation to their statistical distribution. In a second experiment we focus our attention on the numerical properties of ML-GHK and ML-GHK-EIS estimates as MC approximations for the unfeasible exact ML estimate, by repeating the estimation 20 times under different CRNs for the first of the simulated data sets.

In our MC study, we consider three out of the 12 different sets of parameter

values used by Geweke et al. (1997). The three sets considered here are given by

$$(\rho_1, \rho_2, \omega_{12}, \omega_{22}, \phi^2) = \begin{cases} (0.5, 0.5, 0.5, 0.866, 0), & \text{(set 1)} \\ (0.8, 0.8, 0.5, 0.866, 0), & \text{(set 2)} \\ (0.5, 0.5, 0.8, 0.6, 0.8), & \text{(set 3)} \end{cases},$$

with the mean parameters fixed at

$$(\pi_{10}, \pi_{11}, \pi_{02}, \pi_{12}, \psi) = (0.5, 1, -1.2, 1, 1).$$

The first set of parameters values implies low serial and cross correlation of the innovations and no serial correlation in the regressors. The second set with increased serial correlation of the innovations represents a worse case scenario for ML-GHK relative to a Bayesian Gibbs procedure. Finally, the last set, in which the correlations are low, high and high, respectively, represents the best case scenario for ML-GHK. Results for these three scenarios are found in tables 1, 4, and 9, respectively, in Geweke et al. (1997).

The results of our MC experiments based on these three different sets of parameter values are summarized in Tables 2–4 where we ran two experiments for each set, one based upon 20 simulated data sets, the other on 20 different sets of CRNs for the first simulated data set. For the first experiment we report the mean, standard deviation and RMSE around the true parameter values (see column three and four of Tables 2–4). The GHK as well as the GHK-EIS results are based on a simulation sample size of $S = 20$, and for EIS we use three fixed point iterations. For the second experiment we report the mean, standard deviation and RMSE around the pseudo-true values (see column five and six of Tables 2–4). The latter are obtained by an ML-GHK-EIS estimate based

on simulation sample size of $S = 1000$[1]. For $S = 20$, one GHK-EIS likelihood evaluation takes 5 s and a GHK evaluation 1 s for a code written in GAUSS, which implies that GHK-EIS is computationally more efficient than GHK as soon as the resulting efficiency gain measured by the ratio of the respective MC standard deviations exceeds $\sqrt{5}$.

Our results for the statistical distribution of the ML-GHK estimator under different data sets are essentially the same as those reported by Geweke et al. (1997). They indicate that the biases of the estimates for the mean parameters $(\pi_{10}, \pi_{11}, \pi_{02}, \pi_{12})$ are typically very small, while, in contrast, the ML-GHK estimates for the covariance parameters $(\rho_1, \rho_2, \omega_{12}, \omega_{22})$ are often severely biased. In fact, the t-statistic constructed for the difference between the true parameter value and the mean point estimates indicate highly significant biases for ρ_1 and ρ_2 under parameter set 1 and 3 (see Table 2 and 4) and for all covariance parameters under set 2 (see Table 3).

Next, the results obtained for GHK-EIS under different data sets indicate that for the mean parameters the mean point estimates, standard deviations and RMSEs are nearly the same as their GHK counterparts for all three data structure. However, the mean of the GHK-EIS estimates for all covariance parameters are very close to the data generating values with biases which are not statistically significant. Thus, in contrast to the standard GHK, a simulation sample size of $S = 20$ seems to be sufficient for GHK-EIS to produce nearly unbiased parameter estimates for the standard MMP model. As illustrated by Geweke et al. (1997), a much larger size than $S = 20$ is typically necessary in order to eliminate the biases of ML-GHK for the covariance parameters. For example, under the sec-

[1]In order to verify that the pseudo-true values obtained by GHK-EIS with $S = 1000$ are close to those obtained from GHK, we also computed the ML-GHK estimates with $S = 5000$. The results, not reported here, show that both procedures lead indeed to values which are essentially identical.

ond parameter set, a simulation size of at least $S = 1280$ is needed to reduce the biases of GHK to the same level as those of GHK-EIS with $S = 20$. In fact, for $\rho_1, \rho_2, \omega_{12}$, and ω_{22} Geweke et al. (1997, Table 16) report RMSEs for GHK with $S = 1280$ of 0.072, 0.079, 0.013, and 0.029 while those for GHK-EIS with $S = 20$ are according to Table 2 given by 0.066, 0.058, 0.014, and 0.021, respectively.

The results obtained for the repeated parameter estimates under different sets of CRNs indicate substantial numerical efficiency gains of ML-GHK-EIS relative to the ML-GHK for all three data structures. For example, the (numerical) standard deviations for GHK-EIS are between 8 (ω_{12}) and 18 times (ρ_1) smaller than their GHK counterpart under the first parameter set (see Table 2). Furthermore, the mean GHK-EIS estimates are very close to the pseudo-true ML values under all three data structures and for all parameters. GHK, on the other hand, while producing estimates close to the pseudo-true values for the mean parameters, exhibits relatively large numerical biases for the covariance parameters. Thus, the significant statistical biases of the ML-GHK estimates (as estimates for the parameters) found for the covariance parameters are largely driven by numerical biases of the ML-GHK estimates (as MC estimates of the unfeasible true ML estimate). This is consistent with Geweke et al.'s result showing that the statistical biases of ML-GHK disappear if the simulation size for GHK is (substantially) increased, leading to a reduction of the numerical biases.

In order to illustrate how the numerical accuracy of the probability estimates of GHK and GHK-EIS affects that of the corresponding ML parameter estimates, Figure 1 plots the GHK and EIS-GHK MC estimates of the sectional log-likelihood functions for the mean parameter ψ and the covariance parameter ρ_1 obtained under 20 different sets of CRNs and a fixed data set. The data are generated under parameter set 2 and the sectional functions for ψ and ρ_1 are

obtained by setting the remaining parameters equal to their pseudo-true value given in Table 2. Note that the GHK MC estimates of the sectional log-likelihood function exhibit a substantially larger variation than their GHK-EIS counterparts leading to a much broader range of parameter values maximizing the single GHK MC estimates of the sectional log-likelihood. Moreover, notice that the GHK estimates of the log-likelihood appear to be significantly downward biased.

4.3 Invariant Multinomial Multiperiod Probit

In order to illustrate the results on formal identification discussed in Section 2.2.2, we fitted the invariant MMP model with $J + 1 = 3$ alternatives to simulated samples of size $N = 500$ and $T = 10$.

In particular, we consider the following specification for the utility differences (w.r.t. to the utility of the third alternative as the baseline utility):

$$\Delta_J U_{it} = \Delta_J \mu + \Delta_J \epsilon_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N \tag{58}$$

with

$$\begin{aligned} \Delta_J \mu &= (\pi_{01} + \psi Z_{it1}\ ,\ \pi_{02} + \psi Z_{it2})' && (59) \\ \epsilon_{it} &= R\epsilon_{it-1} + v_{it} && (60) \\ v_{it} &\sim \mathrm{N}_3\left(0, \Sigma\right), \quad \Sigma = [\sigma_{ik}] && (61) \end{aligned}$$

where R is a diagonal matrix with elements (ρ_1, ρ_2, ρ_3). The regressor Z_{itj} is constructed according to Equation (57) with $\phi = 0$. From the specification of ϵ_{it} given in Equations (60) and (61), we obtain the stationary distribution of

$\epsilon_{it}^* = Q_J \epsilon_{it} = ((\Delta_J \epsilon_{it})', \epsilon_{it3})'$, which is parameterized according to Section 2.2 as

$$\epsilon_{it}^* \sim \mathrm{N}_3 \left(0, \begin{bmatrix} \Psi & \Psi b \\ b'\Psi & \upsilon^2 + b'\Psi b \end{bmatrix} \right), \quad \text{with} \quad \Psi = \begin{pmatrix} l_{11}^2 & l_{11} l_{12} \\ l_{11} l_{12} & l_{12}^2 + l_{22}^2 \end{pmatrix} \tag{62}$$

and $b = (b_1, b_2)'$.

We consider the following values for the original parameters in Equations (58)–(61):

$$(\pi_{01}, \pi_{02}, \psi, \sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{13}, \sigma_{23}, \rho_1, \rho_2, \rho_3) \\ = (0.5, -1, 1, 1, 1, 1, 0.3, 0.3, 0.3, 0.8, 0.6, 0.3).$$

The implied values for the identified parameters of the stationary distribution for ϵ_{it}^* are given by

$$(l_{11}, l_{12}, l_{22}, b_1, b_2) = (1.757, 0.521, 1.288, -0.134, -0.316),$$

up to a scaling factor for Ψ. For complete identification, we fix the square root of the first diagonal element of Ψ given by l_{11} to its true value. Alternatively, one could set l_{11} equal to one, which amounts to dividing the parameter true values and their estimates for $(\pi_{01}, \pi_{02}, \psi, l_{11}, l_{12}, l_{22})$ by 1.757.

As above, we estimated this invariant MMP specification for 20 artificial data sets by ML-GHK-EIS and ML-GHK both with a simulation sample size of $S = 20$, and repeated the estimation for the first data set 20 times under different CRNs. The results are summarized in Table 5. As for the previous MC experiments, we report the mean of the point estimates, standard deviation and RMSE across the 20 different data sets as well as across the 20 different sets of CRNs. Additionally,

Table 5 contains the mean of the asymptotic standard errors across the simulated data sets.

The MC results for different data sets are fully in line with our earlier results in Section 2.2.2, confirming that the invariant MMP specification is formally identified. In particular, the standard deviations and RMSEs of the ML-GHK-EIS estimates indicate that the three additional parameters of the stationary covariance matrix (b_1, b_2, ρ_3) which are not identified under the standard MMP model can be estimated with a reasonable precision, even though their estimates appear to exhibit, as expected, a somewhat larger variation than those for the remaining parameters. In fact, the standard deviations for b_1, b_2, and ρ_3 are 0.097, 0.098 and 0.140, while those of the remaining parameters are all below 0.074. Furthermore, we note that the mean of the ML-GHK-EIS asymptotic standard errors are in fairly close agreement with the corresponding standard deviation of the GHK-EIS estimate for all parameters of the invariant model, indicating that the Hessian of the log-likelihood is well behaved.

Comparing the ML-GHK and the ML-GHK-EIS estimates under different data set reveals the same feature as that observed under the standard MMP: the GHK-EIS produces point estimates with biases which are not statistically significant, while the ML-GHK estimates are reasonably close to their true values only for the mean parameters, but are significantly biased for the covariance parameters, except for ρ_3. Furthermore, we note that for most of the covariance parameters the mean of the asymptotic standard errors under GHK are substantially smaller than the RMSE which could lead to strongly biased test results in practical applications. For example, the RMSE for ρ_2 is 0.147 while the mean asymptotic standard errors is only 0.047. In contrast, under GHK-EIS the RMSE and the mean asymptotic standard errors are much closer to each other. Also

in line with the results for the standard MMP, GHK-EIS leads to a substantial increase of numerical precision relative to GHK, with significantly smaller (numerical) standard deviations and RMSEs obtained for a fixed data set and under different CRNs. Once again, the large numerical biases of the ML-GHK estimates relative to the pseudo-true ML estimates – in particular for the covariance parameters – are in close accordance with the significant statistical biases of the ML-GHK estimates relative to the data generating parameter values.

Note finally that the GHK numerical standard deviation is larger than the statistical standard deviation of the ML-GHK estimates for some of the covariance parameters (b_1, b_2, ρ_2, ρ_3). An inspection of the individual estimation results obtained under different CRNs reveals that the comparably large numerical standard deviations for those parameters are mainly driven by single 'outliers' producing parameters estimates which are very far from the average estimate.

5 Conclusion

We have proposed to combine the GHK probability simulator with Efficient Importance Sampling (EIS) in order to obtain simulated ML estimates of multinomial multiperiod probit (MMP) models. The proposed GHK-EIS procedure uses simple linear Least-Squares approximations designed to maximize the numerical accuracy of Monte Carlo (MC) estimates for Gaussian probabilities of rectangular domains within a parametric class of importance sampling densities. The implementation of GHK-EIS is straightforward and allows for numerically very accurate and reliable ML estimates of MMP models as illustrated by the MC results we have reported. In particular, GHK-EIS significantly reduces the biases of ML estimates obtained under GHK with the commonly used simulation

sample size of 20 draws.

We have also proposed a MMP specification which is invariant w.r.t. the selection of the baseline category and identifies parameters which are not identified under the standard approach (such as the parameters governing the dynamics of the utility for the reference category). The formal identification of the proposed invariant MMP specification has been illustrated by MC experiments.

Acknowledgement

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.

Appendix 1: Efficient Cholesky decomposition for $(\epsilon^{\diamond}_{j_1}, ..., \epsilon^{\diamond}_{j_T})$

According to Equations (14) and (52), the $(J+1)\cdot T$-dimensional stationary covariance matrix V of $(\epsilon^{\diamond}_{j_1}, ..., \epsilon^{\diamond}_{j_T})$ is partitioned into $(J+1)$ dimensional blocks of the form

$$V_{ts} = \mathrm{Cov}(\epsilon^{\diamond}_t, \epsilon^{\diamond\prime}_s) = Q_t R^{t-s} \Phi Q'_s, \qquad s \leq t, \tag{A-1}$$

with $Q_t = Q_t^{-1}$ (the subscripts $*$ and j are deleted for the ease of notation; note the Q_t can only take one of $J+1$ different forms, corresponding to each of the alternatives). Let L denote the lower triangular Cholesky decomposition of V. L is partitioned conformably with V into blocks L_{ts} for $s \leq t$.

Lemma A1. *The diagonal blocks of L are given by the following $(J+1)$-dimensional Cholesky decompositions*

$$L_{11}L'_{11} = Q_1 \Phi Q'_1 \tag{A-2}$$

$$L_{tt}L'_{tt} = Q_t \Sigma Q'_t, \quad \textit{with} \quad \Sigma = \Phi - R\Phi R', \quad t > 1, \tag{A-3}$$

and the off-diagonal blocks by the products

$$L_{ts} = (Q_t R^{t-s} Q_s) L_{ss}, \qquad s \leq t. \tag{A-4}$$

Proof. The proof follows by recursion over the sequence $(((t,s), t = s, ..., T), s = 1, ..., T)$. Equation (A-2) trivially follows from the (block) lower-triangular form

of L. Then for $s = 1$ and $t > 1$ we have

$$L_{t1}L'_{11} = Q_t R^{t-1}\Phi Q'_1 = (Q_t R^{t-1}Q_1)Q_1\Phi Q'_1 = (Q_t R^{t-1}Q_1)L_{11}L'_{11}. \qquad \text{(A-5)}$$

For $s > 1$, we have

$$L_{t1}L'_{s1} + \sum_{j=2}^{s-1} L_{tj}L'_{sj} + L_{ts}L'_{ss} = Q_t R^{t-s}\Phi Q'_s, \qquad \text{(A-6)}$$

(under the usual summation convention that for $s = 2$ the middle summation is omitted). Whence

$$\begin{aligned} L_{ts}L'_{ss} &= Q_t\big[- R^{t-s}\Phi R'^{s-1} - \sum_{j=2}^{s-1} R^{t-j}(\Phi - R\Phi R')R'^{s-j} \qquad \text{(A-7)} \\ &\qquad\qquad +R^{t-s}\Phi\big]Q'_s \\ &= Q_t R^{t-s}Q_s\big[Q_s(\Phi - R\Phi R')Q'_s\big], \qquad \text{(A-8)} \end{aligned}$$

which, together with (A-3), completes the proof. □

Note that the proof critically relies on the fact that Q_t is square non-singular with $Q_t^{-1} = Q_t$. It does not generalize to the rectangular transformation (16) which is why this efficient Cholesky decomposition requires extending $\epsilon^*_{j_t}$ into $\epsilon^\diamond_{j_t}$.

References

Atkinson, K., 1978. An Introduction to Numerical Analysis. John Wiley & Sons, New York.

Bunch, D.S., 1991. Estimability in the multinomial probit model. Transportation Research B 25, 1–12.

Börsch-Supan, A., Hajivassiliou, V., 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. Journal of Econometrics 58, 347–368.

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.

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., 2001. Computationally intensive methods for integration in econometrics. In Heckman, J., Leamer, E., Handbook of Econometrics 5, Chapter 56. Elsevier.

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.

Hajivassiliou, V., McFadden, D., Ruud, P. 1996. Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. Journal of Econometrics 72, 85–134.

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.

Lehmann, E.L., 1986. Testing Statistical Hypotheses. John Wiley & Sons.

McCulloch R., Rossi, P., 1994. An exact likelihood analysis of the multinomial probit model. Journal of Econometrics 64, 207–240.

Richard, J.-F., Zhang, W., 2007. Efficient high-dimensional importance sampling. Forthcoming in: Journal of Econometrics.

Stern, S., 1992. A method for smoothing simulated moments of discrete probabilities in multinomial probit models. Econom trica 60, 943–952.

Stern, S., 1997. Simulation-based estimation. Journal of Economic Literature 35, 2006–2039.

Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press.

Table 1. Simulated Choice Probabilities for the Static Multinomial Model

		true	mean	std. dev.	rmse
Example 1	GHK	.02401	.02396	.00070	.00070
	GHK-EIS		.02401	.00001	.00001
Example 2	GHK	.14989	.14956	.00448	.00449
	GHK-EIS		.14984	.00018	.00019
Example 3	GHK	.64718	.64713	.00867	.00867
	GHK-EIS		.64638	.00529	.00536
Example 4	GHK	.49557	.49457	.01356	.01360
	GHK-EIS		.49537	.00071	.00074

NOTE: Reported statistics are obtained from 1,000 independent replications of the MC estimation of the probabilities. The GHK and GHK-EIS MC-estimates are based upon a simulation sample size of $S = 100$. The true value is calculated using the ISML subroutine DQAND with a relative accuracy of at least $1e-6$.

Table 2. ML-EIS-GHK and ML-GHK for the Standard Multiperiod Multinomial Probit: Parameter Set 1.

		diff. data sets		fixed data set/diff. CRNs		
Parameter	true	GHK	GHK EIS	pseudo true	GHK	GHK EIS
π_{01}	.500	.512	.512	.548	.551	.548
		(.027)	(.027)		(.0050)	(.0005)
		[.030]	[.029]		[.0060]	[.0008]
π_{11}	1.000	.998	1.000	1.030	1.031	1.031
		(.035)	(.036)		(.0051)	(.0005)
		[.035]	[.036]		[.0053]	[.0011]
π_{02}	−1.200	−1.177	−1.179	−1.197	−1.208	−1.199
		(.058)	(.057)		(.0170)	(.0016)
		[.062]	[.061]		[.0206]	[.0027]
π_{12}	1.000	.997	.998	1.058	1.064	1.060
		(.056)	(.056)		(.0134)	(.0015)
		[.056]	[.056]		[.0145]	[.0022]
ψ	1.000	.991	.994	1.008	1.009	1.009
		(.024)	(.024)		(.0054)	(.0005)
		[.025]	[.024]		[.0054]	[.0010]
ω_{12}	.500	.523	.506	.511	.532	.511
		(.056)	(.051)		(.0351)	(.0046)
		[.060]	[.052]		[.0410]	[.0046]
ω_{22}	.866	.878	.860	.849	.872	.849
		(.056)	(.063)		(.0256)	(.0029)
		[.057]	[.063]		[.0344]	[.0029]
ρ_1	.500	.459	.504	.518	.472	.518
		(.028)	(.031)		(.0106)	(.0006)
		[.050]	[.031]		[.0472]	[.0006]
ρ_2	.500	.413	.495	.475	.388	.477
		(.047)	(.053)		(.0381)	(.0035)
		[.099]	[.053]		[.0945]	[.0041]

NOTE: The reported numbers for ML-GHK and ML-GHK-EIS are mean, standard deviation (in parentheses) and RMSE (in brackets) obtained for $S = 20$. For the experiment with different data sets (fixed data set and different CRNs) RMSE is computed around the true (pseudo-true) value. The pseudo-true values are the ML-GHK-EIS estimates based on $S = 1000$.

Table 3. ML-EIS-GHK and ML-GHK for the Standard Multiperiod Multinomial Probit: Parameter Set 2.

		diff. data sets		fixed data set/diff. CRNs		
Parameter	true	GHK	GHK EIS	pseudo true	GHK	GHK EIS
π_{01}	.500	.504	.510	.540	.540	.542
		(.047)	(.048)		(.0104)	(.0012)
		[.047]	[.049]		[.0104]	[.0028]
π_{11}	1.000	.995	1.005	1.017	1.005	1.020
		(.042)	(.041)		(.0079)	(.0009)
		[.043]	[.041]		[.0144]	[.0032]
π_{02}	−1.200	−1.149	−1.171	−1.136	−1.111	−1.143
		(.070)	(.061)		(.0330)	(.0035)
		[.087]	[.068]		[.0416]	[.0076]
π_{12}	1.000	.995	1.006	1.047	1.031	1.051
		(.049)	(.048)		(.0155)	(.0033)
		[.050]	[.048]		[.0223]	[.0056]
ψ	1.000	.985	.998	1.005	0.989	1.007
		(.033)	(.030)		(.0098)	(.0010)
		[.036]	[.030]		[.0183]	[.0024]
ω_{12}	.500	.556	.510	.431	.483	.445
		(.082)	(.065)		(.0330)	(.0104)
		[.100]	[.066]		[.0613]	[.0171]
ω_{22}	.866	.908	.859	.749	.817	.766
		(.057)	(.058)		(.0463)	(.0107)
		[.071]	[.058]		[.0822]	[.0202]
ρ_1	.800	.750	.799	.798	.754	.796
		(.019)	(.014)		(.0089)	(.0009)
		[.053]	[.014]		[.0448]	[.0019]
ρ_2	.800	.712	.799	.842	.768	.834
		(.033)	(.021)		(.0210)	(.0035)
		[.094]	[.021]		[.0772]	[.0091]

NOTE: The reported numbers for ML-GHK and ML-GHK-EIS are mean, standard deviation (in parentheses) and RMSE (in brackets) obtained for $S = 20$. For the experiment with different data sets (fixed data set and different CRNs) RMSE is computed around the true (pseudo-true) value. The pseudo-true values are the ML-GHK-EIS estimates based on $S = 1000$.

Table 4. ML-EIS-GHK and ML-GHK for the Standard Multiperiod Multinomial Probit: Parameter Set 3.

		diff. data sets		fixed data set/diff. CRNs		
Parameter	true	GHK	GHK EIS	pseudo true	GHK	GHK EIS
π_{01}	.500	.503	.505	.500	.501	.501
		(.033)	(.033)		(.0044)	(.0005)
		[.033]	[.034]		[.0045]	[.0009]
π_{11}	1.000	.995	.999	.938	0.937	0.940
		(.029)	(.030)		(.0037)	(.0005)
		[.029]	[.030]		[.0037]	[.0025]
π_{02}	−1.200	−1.204	−1.210	−1.102	−1.115	−1.115
		(.089)	(.099)		(.0189)	(.0028)
		[.089]	[.100]		[.0232]	[.0136]
π_{12}	1.000	.983	.984	.934	0.941	0.937
		(.043)	(.045)		(.0122)	(.0019)
		[.046]	[.048]		[.0138]	[.0037]
ψ	1.000	1.008	1.016	.936	0.934	0.941
		(.044)	(.044)		(.0042)	(.0009)
		[.045]	[.047]		[.0045]	[.0054]
ω_{12}	.800	.790	.780	.694	.712	.689
		(.064)	(.071)		(.0233)	(.0051)
		[.064]	[.074]		[.0296]	[.0070]
ω_{22}	.600	.607	.608	.572	.590	.579
		(.049)	(.058)		(.0170)	(.0024)
		[.050]	[.059]		[.0242]	[.0068]
ρ_1	.500	.457	.489	.509	.479	.507
		(.025)	(.027)		(.0072)	(.0006)
		[.049]	[.029]		[.0307]	[.0018]
ρ_2	.500	.453	.489	.549	.507	.549
		(.038)	(.042)		(.0149)	(.0024)
		[.060]	[.043]		[.0451]	[.0024]

NOTE: The reported numbers for ML-GHK and ML-GHK-EIS are mean, standard deviation (in parentheses) and RMSE (in brackets) obtained for $S = 20$. For the experiment with different data sets (fixed data set and different CRNs) RMSE is computed around the true (pseudo-true) value. The pseudo-true values are the ML-GHK-EIS estimates based on $S = 1000$.

Table 5. ML-EIS-GHK and ML-GHK for the Invariant Multiperiod Multinomial Probit.

Para-meter			diff. data sets		fixed data set/diff. CRNs		
		true	GHK	GHK EIS	pseudo true	GHK	GHK EIS
π_{01}	mean	.500	.501	.515	.537	.517	.539
	std. dev.		.061	.060		.014	.001
	rmse		.061	.062		.024	.003
	mean asy. s.e.		.049	.050			
π_{02}	mean	−1.000	−.984	−1.011	−.961	−.963	−.969
	std. dev.		.064	.063		.026	.002
	rmse		.066	.064		.026	.009
	mean asy. s.e.		.069	.067			
ψ	mean	1.000	.995	1.009	1.020	1.009	1.024
	std. dev.		.021	.019		.007	.0008
	rmse		.021	.021		.013	.005
	mean asy. s.e.		.031	.031			
l_{12}	mean	.521	.476	.521	.550	.484	.552
	std. dev.		.093	.074		.039	.005
	rmse		.103	.074		.077	.005
	mean asy. s.e.		.070	.064			
l_{22}	mean	1.288	1.260	1.296	1.291	1.273	1.298
	std. dev.		.066	.064		.024	.002
	rmse		.072	.064		.030	.007
	mean asy. s.e.		.057	.054			
b_1	mean	−.134	−.266	−.151	.000	−.145	−.004
	std. dev.		.122	.097		.139	.002
	rmse		.180	.098		.201	.005
	mean asy. s.e.		.078	.075			
b_2	mean	−.316	−.275	−.312	−.427	−.352	−.416
	std. dev.		.098	.098		.119	.003
	rmse		.107	.098		.141	.011
	mean asy. s.e.		.065	.075			
ρ_1	mean	.800	.821	.803	.775	.782	.775
	std. dev.		.042	.030		.038	.0006
	rmse		.047	.030		.039	.0006
	mean asy. s.e.		.030	.026			

Table 5. Continued.

Parameter		true	diff. data sets GHK	diff. data sets GHK EIS	fixed data set/diff. CRNs pseudo true	fixed data set/diff. CRNs GHK	fixed data set/diff. CRNs GHK EIS
ρ_2	mean	.600	.489	.600	.687	.529	.685
	std. dev.		.096	.064		.193	.002
	rmse		.147	.064		.251	.003
	mean asy. s.e.		.047	.045			
ρ_3	mean	.300	.298	.286	.317	.275	.302
	std. dev.		.160	.140		.190	.004
	rmse		.160	.140		.194	.016
	mean asy. s.e.		.088	.110			

NOTE: The reported numbers for ML-GHK and 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 different data sets (fixed data set and different CRNs) RMSE is computed around the true (pseudo-true) value. The pseudo-true values are the ML-GHK-EIS estimates based on $S = 1000$.

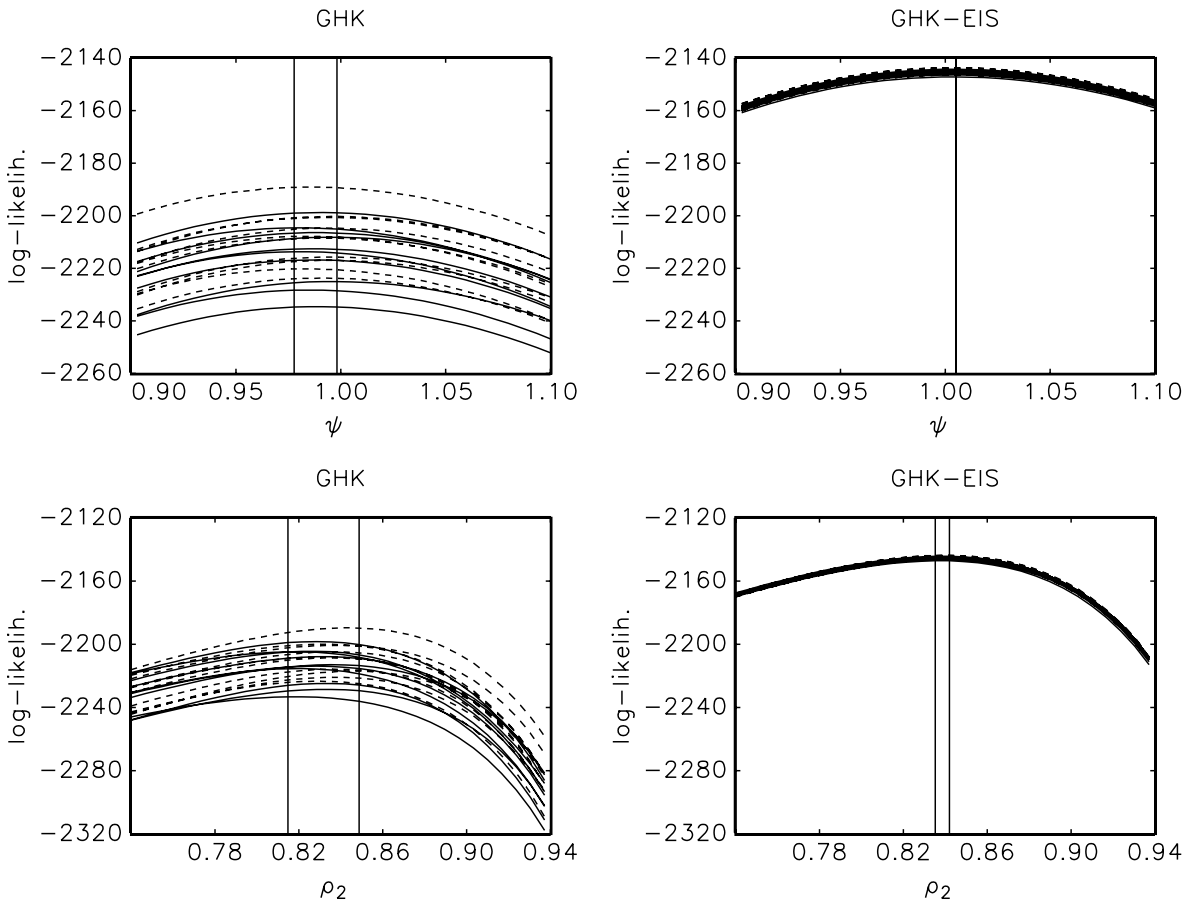

Figure 1. Sectional log-likelihood functions for the standard multiperiod multinomial Probit for parameter ψ (upper panels) and ρ_2 (lower panels). The sectional log-likelihood functions are constructed for a fixed data set (generated under parameter set 2) using GHK (left panels) and GHK-EIS (right panels) under 20 different sets of CRNs. The remaining parameters are set to their pseudo true values (see Table 3). The vertical lines indicate the range of the parameter values which maximize the individual simulated sectional log-likelihood functions.