Accepted Manuscript

Efficient estimation of Probit models with correlated errors

Roman Liesenfeld, Jean-François Richard

PII: S0304-4076(09)00295-4

DOI: 10.1016/j.jeconom.2009.11.006

Reference: ECONOM 3295

To appear in: *Journal of Econometrics*

Received date: 26 September 2008

Revised date: 22 July 2009

Accepted date: 17 November 2009

Please cite this article as: Liesenfeld, R., Richard, J.-F., Efficient estimation of Probit models with correlated errors. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.11.006

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.

Efficient Estimation of Probit Models with Correlated Errors

Roman Liesenfeld*

Department of Economics, Christian-Albrechts-Universität, Kiel, Germany

Jean-François Richard

Department of Economics, University of Pittsburgh, USA

July 4, 2009

Abstract

Maximum Likelihood (ML) estimation of Probit models with correlated errors typically requires high-dimensional truncated integration. Prominent examples of such models are multinomial Probit models and binomial panel Probit models with serially correlated errors. In this paper we propose to use a generic procedure known as Efficient Importance Sampling (EIS) for the evaluation of likelihood functions for Probit models with correlated errors. Our proposed EIS algorithm covers the standard GHK probability simulator as a special case. We perform a set of Monte-Carlo experiments in order to illustrate the relative performance of both procedures for the estimation of a multinomial multiperiod Probit model. Our results indicate substantial numerical efficiency gains for ML estimates based on GHK-EIS relative to those obtained by using GHK.

JEL classification: C35, C15

Keywords: Discrete choice, Importance sampling, 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 likelihood evaluation of discrete choice Probit models with correlated errors. Prominent examples of such models are the multinomial Probit model introduced by Thurstone (1927) and applied, e.g., by Hausman and Wise (1978) to transit choice problems, the binomial panel Probit model with random effects and serially correlated errors used by Keane (1993) for an analysis of labor supply problems, and the multinomial multiperiod Probit (MMP) model applied by Börsch-Supan et al. (1992) and Keane (1997) in order to study the living arrangements of the elderly and the brand choice in successive purchase occasions, respectively.

The likelihood function of Probit model with correlated errors takes the form of a product of choice probabilities. These obtain as analytically intractable and frequently high-dimensional truncated integrals of multivariate normal distributions. Thus likelihood-based estimation of such models 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, inter alia, to obtain simulated ML estimates of a binomial panel probit with random effects and serially correlated AR(1) errors (see, e.g., Keane, 1993) and to MMP models in order to compute simulated ML estimates as well as estimates based upon the method of simulated moments (MSM) (see, e.g., Börsch-Supan et al., 1992, Keane, 1994, and Geweke et al., 1997, Keane, 1997). In an extensive study of alternative MC-procedures for the evaluation of probit probabilities, Hajivassiliou et al. (1996) find that GHK is the numerically most reliable among the considered alternatives. However, as illus-

trated 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 serial correlation in the innovations is strong. This study also shows that the corresponding MSM estimates based upon GHK surprisingly do not suffer from the same bias problem. However, there are many model specifications such as the mixed discrete/continuous sample selection models à la Heckman (1976) where SML is much easier to implement than MSM. For a detailed discussion of the relative performance of MSM and SML based upon GHK see also Geweke and Keane (2001).

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 MC approximations. The combined GHK-EIS algorithm is well suited to handle the correlation structure in Probit models and, thereby, provides highly accurate likelihood approximations. In order to illustrate this approach we consider the likelihood evaluation of the MMP model as the other probit models of interest mentioned above are all special cases. In particular, GHK-EIS is illustrated through a set of MC experiments where 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, 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 even with as few as 20 draws.

The remainder of this paper is organized as follows. The MMP model is introduced in Section 2. In Section 3 we describe the GHK-EIS procedure and apply it to the MMP model. MC experiments are discussed in Section 4 and conclusions are drawn in Section 5.

2 Multinomial Probit Models

We first consider the case where decisions are independent across individuals as well as over time. Whence we can focus our attention on a single individual choice among $J + 1$ alternatives, omitting individual and time subscripts for the ease of notation. Let $U = (U_1, \dots, U_{J+1})'$ denote a $(J + 1)$ -dimensional vector of random utilities, where U_j denotes the utility of the j th alternative. Alternative k is selected if $U_k > U_j$ for all $j \neq k$.

If one only observes the index of the selected alternative, then the likelihood (choice probability) only depends upon utility differences. The standard approach consists of selecting a baseline alternative, say alternative $J + 1$, and expressing all other utilities in difference from U_{J+1} . Under the standard static multinomial probit model the vector of the J utility differences $Y = (U_1 - U_{J+1}, \dots, U_J - U_{J+1})'$ is assumed to be normally distributed:

$$Y = \mu + \epsilon, \quad \epsilon \sim N_J(0, \Psi), \quad (1)$$

where ϵ denotes a vector of random shocks with covariance matrix $\Psi = (\psi_{kj})$. In most applications μ is assumed to be a linear function of observable exogenous variables X , say $\mu = \mu(X, \beta)$, where β is a corresponding vector of unknown coefficients subject to standard identification restrictions as discussed, e.g., by Bunch (1991) or Keane (1992). Since μ and Ψ are identified only up to a scale factor, it is conventional to normalize them by setting $\text{Var}(\epsilon_1) = \psi_{11} = 1$.

The observation consists of $\{j, X\}$, where j denotes the index of the selected alternative. In order to express the observable choices in terms of utility differences, we introduce the following $J \times J$ non-singular transformation matrix:

$$S_j = \begin{pmatrix} I_{(j-1)} & -\iota_{(j-1)} & 0 \\ 0 & -1 & 0 \\ 0 & -\iota_{(J-j)} & I_{(J-j)} \end{pmatrix}, \quad \text{with} \quad S_1 = \begin{pmatrix} -1 & 0 \\ -\iota_{(J-1)} & I_{(J-1)} \end{pmatrix}, \quad (2)$$

and $S_{J+1} = I_{(J)}$, where $I_{(\ell)}$ denotes a ℓ -dimensional identity matrix and $\iota_{(\ell)} = (1, \dots, 1)'$. Then alternative j is selected iff

$$Y_j = S_j Y < 0, \quad (3)$$

and the corresponding choice probability is given by

$$P(Y_j < 0 | X). \quad (4)$$

Hence, ML estimation requires the evaluation of truncated J -dimensional Gaussian integrals.

A dynamic extension of this model is the Multinomial Multiperiod Probit (MMP) model for repeated choices in each of T periods among $J+1$ alternatives.

Let Y_t denote the corresponding vector of J utility differences w.r.t. the baseline category $J + 1$ in time period t ($t = 1, \dots, T$). Under the MMP model as presented, e.g., by Börsch-Supan et al. (1992) and Geweke et al. (1997) the utility differences are assumed to evolve according to

$$Y_t = \mu_t + \epsilon_t, \quad \epsilon_t | \epsilon_{t-1} \sim N_J(R\epsilon_{t-1}, \Sigma), \quad (5)$$

where R denotes a diagonal matrix with elements $\{\rho_j\}_{j=1}^J$ and $\mu_t = \mu(X_t, \beta)$. If j_t denotes the index of the alternative chosen in period t , the probability for the sequence of T observed choices is

$$P(Y_{j_1} < 0, \dots, Y_{j_T} < 0 | X), \quad \text{where} \quad Y_{j_t} = S_{j_t} Y_t, \quad (6)$$

and $X = (X_1, \dots, X_T)$. Since Y_t is serially correlated, this choice probability requires evaluating a $T \cdot J$ -dimensional truncated Gaussian integral.

3 The GHK-EIS Algorithm

The presentation of the generic GHK-EIS algorithm 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 Sections 3.1 and 3.2 we present the GHK-EIS algorithm and its implementation under streamlined notation ignoring individual and time indices. Its application to the selection probability of the static model (4) and that of the multiperiod model (6) are presented in Section 3.3 and 3.4, respectively.

3.1 GHK-EIS baseline algorithm

The probabilities to be computed are those associated with events of the form $Y < 0$, where $Y' = (Y_1, \dots, Y_M)$ 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

$$Y = \mu + L\eta, \quad \eta \sim N_M(0, I_{(M)}). \quad (7)$$

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

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

$$\ell'_\tau = (\gamma'_\tau, \delta_\tau), \quad (8)$$

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

$$Y_\tau = \mu_\tau + \gamma'_\tau \eta_{(\tau-1)} + \delta_\tau \eta_\tau, \quad (9)$$

with $\eta'_{(\tau-1)} = (\eta_1, \dots, \eta_{\tau-1})$ and $\eta_{(0)} = \emptyset$. The probability to be computed has the form

$$P(D) = \int_{\mathbb{R}^M} \prod_{\tau=1}^M \varphi_\tau(\eta_{(\tau)}) d\eta, \quad (10)$$

with

$$\varphi_\tau(\eta_{(\tau)}) = \mathbb{I}(\eta_\tau < -\frac{1}{\delta_\tau}[\mu_\tau + \gamma'_\tau \eta_{(\tau-1)}]) \cdot \phi(\eta_\tau), \quad (11)$$

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

$$m(\eta; a) = \prod_{\tau=1}^M m_{\tau}(\eta_{\tau} | \eta_{(\tau-1)}, a_{\tau}), \quad (12)$$

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

$$\hat{P}_S(D; a) = \frac{1}{S} \sum_{s=1}^S \omega(\tilde{\eta}^{(s)}; a), \quad \text{where} \quad \omega(\eta; a) = \prod_{\tau=1}^M \frac{\varphi_{\tau}(\eta_{(\tau)})}{m_{\tau}(\eta_{\tau} | \eta_{(\tau-1)}, a_{\tau})}, \quad (13)$$

and $\{\tilde{\eta}^{(s)}\}_{s=1}^S$ denotes S *i.i.d.* simulated trajectories drawn from the IS density m . A trajectory is a sequential draw of η whereby $\tilde{\eta}_{\tau}^{(s)}$ is simulated 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 minimizes the MC sampling variance of $\hat{P}_S(D; a)$. This requires selecting a value $\hat{a}$ which makes the IS sampling density $\prod_{\tau} m_{\tau}$ globally as close as possible to the target function $\prod_{\tau} \varphi_{\tau}$. The principle of the EIS procedure 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 $\varphi_{\tau}(\eta_{(\tau)})$ directly by a density $m_{\tau}(\eta_{\tau} | \eta_{(\tau-1)}, a_{\tau})$, which by definition integrates to one w.r.t. η_{τ} . 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

$$m_{\tau}(\eta_{\tau} | \eta_{(\tau-1)}, a_{\tau}) = \frac{k_{\tau}(\eta_{(\tau)}; a_{\tau})}{\chi_{\tau}(\eta_{(\tau-1)}; a_{\tau})}, \quad \text{with} \quad \chi_{\tau}(\eta_{(\tau-1)}; a_{\tau}) = \int_{\mathbb{R}^1} k_{\tau}(\eta_{(\tau)}; a_{\tau}) d\eta_{\tau}. \quad (14)$$

The integral in Equation (10) is then rewritten as

$$P(D; a) = \chi_1(a_1) \int_{\mathbb{R}^M} \prod_{\tau=1}^M \frac{\varphi_\tau(\eta_{(\tau)}) \cdot \chi_{\tau+1}(\eta_{(\tau)}; a_{\tau+1})}{k_\tau(\eta_{(\tau)}; a_\tau)} \cdot \prod_{\tau=1}^M m_\tau(\eta_\tau | \eta_{(\tau-1)}, a_\tau) d\eta, \quad (15)$$

with $\chi_{M+1}(\cdot) \equiv 1$. Sequential EIS aims at recursively selecting values of a_τ which provide the best match period by period between $\varphi_\tau \chi_{\tau+1}$ and k_τ in order to minimize 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 $\{\hat{a}_\tau; \tau = 1, \dots, M\}$ obtain as solutions of the following backward recursive sequence of fixed point auxiliary Least Squares (LS) problems (for $\tau = M, M-1, \dots, 1$):

$$(\hat{\kappa}_\tau, \hat{a}_\tau) = \arg \min_{\kappa_\tau \in \mathbb{R}^1, a_\tau \in A_\tau} \sum_{s=1}^S \left\{ \ln \left[\varphi_\tau(\tilde{\eta}_{(\tau)}^{(s)}) \cdot \chi_{\tau+1}(\tilde{\eta}_{(\tau)}^{(s)}; \hat{a}_{\tau+1}) \right] - \kappa_\tau - \ln k_\tau(\tilde{\eta}_{(\tau)}^{(s)}; a_\tau) \right\}^2, \quad (16)$$

where $\{\tilde{\eta}^{(s)}\}_{s=1}^S$ denotes S *i.i.d.* trajectories drawn from $m(\eta; \hat{a})$ and κ_τ represents an intercept.

Before we discuss further implementation details it might be useful to point out that Equation (15) follows the very same recursive induction which would produce an exact solution if the integral in Equation (10) was analytically tractable. In particular, this would be the case if φ_τ in Equation (11) was a Gaussian density kernel in $\eta_{(\tau)}$ without an indicator function. Since the family of Gaussian distributions is closed under multiplication it follows by backward induction that if $k_{\tau+1}$ is Gaussian in $\eta_{(\tau+1)}$, then $\chi_{\tau+1}$ and the product $\varphi_\tau \chi_{\tau+1}$ in Equation (15) are Gaussian kernels in $\eta_{(\tau)}$. Setting k_τ equal to that product by selecting a_τ accordingly would indeed result in χ_τ being itself Gaussian in $\eta_{(\tau-1)}$. Equation (15) would then simplify into $P(D; a) = \chi(a_1)$, which is the exact result which

obtains under recursive analytical integration.

Under Equation (11) we can still select a Gaussian kernel for $k_{\tau+1}$ which includes $\varphi_{\tau+1}$ but due to the presence of the indicator function in $\varphi_{\tau+1}$, its integral w.r.t. $\eta_{\tau+1}$ now includes a standardized Gaussian c.d.f. $\Phi(\omega_\tau)$ where, as shown further below, ω_τ denotes an appropriate linear combination of $\eta_{(\tau)}$. Except for that additional term $\Phi(\omega_\tau)$, all other factors in the product $\varphi_\tau \chi_{\tau+1}$ remain Gaussian in $\eta_{(\tau)}$, so that this product can be rewritten as $\varphi_\tau \chi_{\tau+1} = \varphi_\tau \chi_{\tau+1}^* \Phi(\omega_\tau)$, where $\chi_{\tau+1}^*$ denotes a Gaussian kernel in $\eta_{(\tau)}$. Accordingly, we shall define k_τ as $k_\tau = \varphi_\tau \chi_{\tau+1}^* k_\tau^*(\omega_\tau)$, where $k_\tau^*(\omega_\tau)$ denotes a Gaussian kernel approximation of $\Phi(\omega_\tau)$. It follows that k_τ being the product of three Gaussian kernels is itself Gaussian, but is truncated by the indicator function in φ_τ . This particular selection of k_τ implies that all Gaussian factors common to k_τ and $\varphi_\tau \chi_{\tau+1}$ cancel out in the auxiliary EIS regression (16) which reduce to simple LS regressions of $\{\ln \Phi(\tilde{\omega}_\tau^{(s)})\}_{s=1}^S$ on $\{\tilde{\omega}_\tau^{(s)}, [\tilde{\omega}_\tau^{(s)}]^2\}_{s=1}^S$ and an intercept, irrespective of the dimension on $\eta_{(\tau)}$. It follows that GHK-EIS is straightforward to implement using GHK as a template and will remain computationally fast since the additional cost of computing required for EIS is that of running sequences of auxiliary LS regressions.

Note also that as discussed further in Richard and Zhang (2007), the draws used to run the auxiliary regressions in Equation (16) depend themselves upon the $\hat{a}_\tau$'s since the draws need to cover the region of importance in order to guarantee the best global approximation of $\varphi_\tau \chi_{\tau+1}$ by k_τ . This requires iterating over the $\hat{a}_\tau$'s upon the sequence of LS problems until a fixed point solution obtains. As starting values we propose to use the values of the auxiliary parameters a implied by the GHK sampler discussed further below. Furthermore, in order to guarantee fast and smooth fixed-point convergence it is critical that the trajectories

$\{\tilde{\eta}^{(s)}\}_{s=1}^S$ drawn under alternative values of a be all obtained by transformation of a set of Common Random Numbers (CRNs), say $\{\tilde{u}^{(s)}\}_{s=1}^S$, 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 $M \times S$ draws from a uniform distribution on $[0, 1]$ to be transformed by inversion into truncated Gaussian draws from $m_\tau(\eta_\tau | \tilde{\eta}_{(\tau-1)}^{(s)}, a_\tau)$. Furthermore note that $\{a_\tau\}$ is an implicit function of (μ, V) . Therefore, maximal numerical efficiency requires complete reruns of the EIS algorithm for any new value of (μ, V) . See Richard and Zhang (2007) for details. At convergence, the GHK-EIS estimate of $P(D)$ is given by

$$\hat{P}_S^{\text{GHK-EIS}}(D) = \chi_1(\hat{a}_1) \cdot \frac{1}{S} \sum_{s=1}^S \prod_{\tau=1}^M \frac{\varphi_\tau(\tilde{\eta}_{(\tau)}^{(s)}) \cdot \chi_{\tau+1}(\tilde{\eta}_{(\tau)}^{(s)}; \hat{a}_{\tau+1})}{k_\tau(\tilde{\eta}_{(\tau)}^{(s)}; \hat{a}_\tau)}. \quad (17)$$

Before we present the functional forms of the EIS implementation, it is important to note that standard GHK is a special case of the sequential EIS defined by Equations (14) to (17). It obtains by using the individual factors in the probability integral (10) as IS density kernels, i.e.,

$$k_\tau(\eta_{(\tau)}; \cdot) = \varphi_\tau(\eta_{(\tau)}), \quad (18)$$

with an integrating factor

$$\chi_\tau(\eta_{(\tau-1)}; \cdot) = \Phi\left(-\frac{1}{\delta_\tau}[\mu_\tau + \gamma'_\tau \eta_{(\tau-1)}]\right), \quad (19)$$

where Φ denotes the standardized normal c.d.f. The resulting GHK sampling densities are truncated Normals of the form

$$m_\tau(\eta_\tau | \eta_{(\tau-1)}, a_\tau) = \frac{\varphi_\tau(\eta_{(\tau)})}{\Phi\left(-\frac{1}{\delta_\tau}[\mu_\tau + \gamma'_\tau \eta_{(\tau-1)}]\right)}, \quad \tau = 1, \dots, M. \quad (20)$$

Therefore, the GHK estimate of $P(D)$ is given by

$$\hat{P}_S^{\text{GHK}}(D) = \Phi\left(-\frac{\mu_1}{\delta_1}\right) \cdot \frac{1}{S} \sum_{s=1}^S \prod_{\tau=1}^{M-1} \Phi\left(-\frac{1}{\delta_{\tau+1}}[\mu_{\tau+1} + \gamma'_{\tau+1} \tilde{\eta}_{(\tau)}^{(s)}]\right). \quad (21)$$

Equation (18) indicates that, in contrast with GHK-EIS, standard GHK ignores the integrating factor $\chi_{\tau+1}$ in the construction of the IS kernel k_{τ} . According to the EIS-LS regression (16) this particular selection of k_{τ} amounts to regressing the integrating factors $\chi_{\tau+1} = \Phi(\cdot)$ on the intercepts κ_{τ} only. Relatedly, it implies that $\chi_{\tau+1}$ only depends on $\Phi(\cdot)$ and does not include an additional Gaussian kernel (denoted $\chi_{\tau+1}^*$ above). Whence, the IS density kernel of standard GHK provides a perfect fit to φ_{τ} but does not account for the MC variation in $\chi_{\tau+1}$, leading to potential losses of numerical efficiency.

The fact that the GHK algorithm represents a special IS procedure and that other choices of m_{τ} than that given in Equation (20) could potentially lead to procedures which (numerically) dominate the standard GHK was already pointed out by Keane (1994, p. 104) and Vijverberg (1997). Vijverberg pursued this line and experimented with various non-Gaussian m_{τ} -sampling densities such as truncated logit, student- t and transformed Beta densities in combination with the use of antithetic variates. (For the combination of GHK with antithetic variates, see also Hajivassiliou, 2000) However, in contrast to our GHK-EIS procedure, Vijverberg's approach does not provide a systematic way to select m_{τ} in order to maximize the numerical accuracy of the corresponding MC-probability estimation. His results indicate that the substitution of the truncated Gaussian GHK sampler by the proposed alternative sampling densities does not deliver systematic improvements relative to GHK, while the use of antithetic variates leads to some numerical efficiency gains. In preliminary experiments, we also found

that antithetic sampling leads to efficiency gains of GHK probability estimation. However, the improvements of the probability estimates turned out to be substantially smaller than those obtained by GHK-EIS and, more importantly, have only minor effects on the numerical and statistical properties of the corresponding ML-GHK parameter estimation discussed below.

3.2 GHK-EIS implementation

Next, we provide closed form expressions for the GHK-EIS evaluation of $P(D)$ as defined in Equations (14) to (17). The corresponding matrix algebra, which essentially consists of regrouping three Gaussian kernels in $\eta_{(\tau)}$ and integrating out η_{τ} , is straightforward though notationally tedious.

Since the EIS-kernel $k_{\tau+1}(\eta_{(\tau+1)}; \hat{a}_{\tau+1})$ which is used in this application is Gaussian in $\eta_{(\tau+1)}$, its integrating factor $\chi_{\tau+1}(\eta_{(\tau)}; \hat{a}_{\tau+1})$ over the truncated range for $\eta_{\tau+1}$ given $\eta_{(\tau)}$, as implied by the indicator function in Equation (11), takes the form of the product of a Gaussian kernel in $\eta_{(\tau)}$ by a Gaussian c.d.f. in a linear combination of the elements in $\eta_{(\tau)}$. In order to recursively derive its actual expression, let $\chi_{\tau+1}$ be parameterized as

$$\chi_{\tau+1}(\eta_{(\tau)}; \hat{a}_{\tau+1}) = \Phi(\omega_{\tau}) \cdot \chi_{\tau+1}^*(\eta_{(\tau)}; \hat{a}_{\tau+1}), \quad \tau + 1 = M, M - 1, \dots, 1, \quad (22)$$

with

$$\chi_{\tau+1}^*(\eta_{(\tau)}; \hat{a}_{\tau+1}) = \exp -\frac{1}{2}(\eta'_{(\tau)} P_{\tau+1}^* \eta_{(\tau)} - 2\eta'_{(\tau)} q_{\tau+1}^* + r_{\tau+1}^*), \quad (23)$$

$$\omega_{\tau} = c_{\tau+1} - d'_{\tau+1} \eta_{(\tau)}, \quad (24)$$

where $(P_{\tau+1}^*, q_{\tau+1}^*, r_{\tau+1}^*, c_{\tau+1}, d_{\tau+1})$ denote appropriate functions of the EIS auxil-

iary parameter $\hat{a}_{t+1}$, to be obtained by a backward recursion as described below. (As mentioned in relation with Equation (15), the recursion is initialized by setting $\chi_{M+1} \equiv 1$.)

Taking advantage of the fact that the family of Gaussian densities is closed under multiplication, we define k_τ as the product of the Gaussian kernels already included in the product $\varphi_\tau \chi_{\tau+1}$ by an additional Gaussian kernel approximating $\Phi(\omega_\tau)$. Whence, the GHK-EIS density kernel k_τ is defined as the following product of Gaussian density kernels

$$k_\tau(\eta_{(\tau)}; \hat{a}_\tau) = k_\tau^*(\omega_{(\tau)}; \hat{a}_\tau) \cdot \chi_{\tau+1}^*(\eta_{(\tau)}; \hat{a}_{\tau+1}) \cdot \varphi_\tau(\eta_{(\tau)}), \quad (25)$$

where $\ln k_\tau^*$ denotes an EIS quadratic approximation to $\ln \Phi(\omega_\tau)$ of the form

$$\ln \Phi(\omega_\tau) \simeq \ln k_\tau^*(\omega_\tau, \hat{a}_\tau) = -\frac{1}{2}(\hat{\alpha}_\tau \omega_\tau^2 + 2\hat{\beta}_\tau \omega_\tau + \hat{\kappa}_\tau). \quad (26)$$

The EIS auxiliary parameter is defined as $\hat{a}_\tau = (\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\kappa}_\tau)$. Since the last two factors in k_τ as defined in Equation (25) are also part of the product $\varphi_\tau \chi_{\tau+1}$ they cancel out in the auxiliary regression (16) of $\ln(\varphi_\tau \chi_{\tau+1})$ on $\ln k_\tau$. Whence it simplifies into a simple LS regression of simulated values of $\ln \Phi(\omega_\tau)$ on simulated values of ω_τ^2 and ω_τ and a constant according to Equation (26).

The three density kernels in k_τ are defined in Equations (26), (23) and (11), respectively. In order to integrate k_τ w.r.t. η_τ conditionally on $\eta_{(\tau-1)}$ we first combine these three kernels into a single one for $\eta_{(\tau)}$, which is then factorized into a kernel for $\eta_\tau | \eta_{(\tau-1)}$ and one for $\eta_{(\tau-1)}$ by application of standard quadratic form algebra. Combining the three kernels in Equation (25) yields the following

expression for k_τ

$$\begin{aligned} k_\tau(\eta_{(\tau)}; \hat{a}_\tau) &= \mathbb{I}(\eta_\tau < -\frac{1}{\delta_\tau}[\mu_\tau + \gamma'_\tau \eta_{(\tau-1)}]) \\ &\times \exp -\frac{1}{2}\{\eta'_{(\tau)} P_\tau \eta_{(\tau)} - 2\eta'_{(\tau)} q_\tau + r_\tau + \ln(2\pi)\}, \end{aligned} \quad (27)$$

where

$$P_\tau = P_{\tau+1}^* + \hat{\alpha}_\tau d_{\tau+1} d'_{\tau+1} + e_{(\tau)} e'_{(\tau)} \quad (28)$$

$$q_\tau = q_{\tau+1}^* + (\hat{\alpha}_\tau c_{\tau+1} + \hat{\beta}_\tau) d_{\tau+1} \quad (29)$$

$$r_\tau = r_{\tau+1}^* + \hat{\alpha}_\tau c_{\tau+1}^2 + 2\hat{\beta}_\tau c_{\tau+1} + \hat{\kappa}_\tau, \quad (30)$$

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

Next, in order to extract from Equation (27) a Gaussian kernel for $\eta_\tau | \eta_{(\tau-1)}$, we partition P_τ and q_τ conformably with $\eta_{(\tau)} = (\eta'_{(\tau-1)}, \eta_\tau)$ into

$$P_\tau = \begin{pmatrix} P_{00}^\tau & P_{01}^\tau \\ P_{10}^\tau & P_{11}^\tau \end{pmatrix}, \quad q_\tau = \begin{pmatrix} q_0^\tau \\ q_1^\tau \end{pmatrix}. \quad (31)$$

The r.h.s. of Equation (27) is then factorized as:

$$k_\tau(\eta_{(\tau)}; \hat{a}_\tau) = k_\tau^1(\eta_{(\tau)}; \hat{a}_\tau) \cdot k_\tau^2(\eta_{(\tau-1)}; \hat{a}_\tau), \quad (32)$$

with

$$k_\tau^1(\eta_{(\tau)}; \hat{a}_\tau) = \mathbb{I}(\eta_\tau < -\frac{1}{\delta_\tau}[\mu_\tau + \gamma'_\tau \eta_{(\tau-1)}]) \quad (33)$$

$$\times \exp -\frac{1}{2}\{P_{11}^\tau [\eta_\tau - \bar{m}_\tau(\eta_{(\tau-1)})]^2\},$$

$$k_\tau^2(\eta_{(\tau-1)}; \hat{a}_\tau) = \exp -\frac{1}{2}\{\eta'_{(\tau-1)} P_\tau^* \eta_{(\tau-1)} - 2\eta'_{(\tau-1)} q_\tau^* + s_\tau^* + \ln(2\pi)\}, \quad (34)$$

and

$$\bar{m}_\tau(\eta_{(\tau-1)}) = \frac{1}{P_{11}^\tau}(q_1^\tau - P_{10}^\tau \eta_{(\tau-1)}), \quad (35)$$

$$P_\tau^* = P_{00}^\tau - \frac{1}{P_{11}^\tau} P_{01}^\tau P_{10}^\tau, \quad q_\tau^* = q_0^\tau - \frac{1}{P_{11}^\tau} P_{01}^\tau q_1^\tau, \quad s_\tau^* = r_\tau - \frac{1}{P_{11}^\tau} (q_1^\tau)^2. \quad (36)$$

Clearly k_τ^1 provides a truncated Gaussian kernel for the GHK-EIS sampler of $\eta_\tau | \eta_{(\tau-1)}$. Its integrating factor is given by

$$\chi_\tau^1(\eta_{(\tau-1)}; \hat{a}_\tau) = \int_{\mathbb{R}^1} k_\tau^1(\eta_\tau; \hat{a}_\tau) d\eta_\tau = \frac{\sqrt{2\pi}}{\sqrt{P_{11}^\tau}} \Phi(c_\tau - d_\tau' \eta_{(\tau-1)}), \quad (37)$$

with

$$c_\tau = -\sqrt{P_{11}^\tau} \left(\frac{\mu_\tau}{\delta_\tau} + \frac{q_1^\tau}{P_{11}^\tau} \right), \quad d_\tau = \sqrt{P_{11}^\tau} \left(\frac{\gamma_\tau}{\delta_\tau} - \frac{P_{01}^\tau}{P_{11}^\tau} \right). \quad (38)$$

Whence the GHK-EIS sampler for $\eta_\tau | \eta_{(\tau-1)}$ obtains as

$$m_\tau(\eta_\tau | \eta_{(\tau-1)}, \hat{a}_\tau) = \frac{k_\tau^1(\eta_\tau; \hat{a}_\tau)}{\chi_\tau^1(\eta_{(\tau-1)}; \hat{a}_\tau)}, \quad (39)$$

which represents the density of a truncated $N[\bar{m}_\tau(\eta_{(\tau-1)}), 1/P_{11}^\tau]$ -distribution.

The overall integrating factor of k_τ is given by

$$\chi_\tau(\eta_{(\tau-1)}; \hat{a}_\tau) = \chi_\tau^1(\eta_{(\tau-1)}; \hat{a}_\tau) \cdot k_\tau^2(\eta_{(\tau-1)}; \hat{a}_\tau), \quad (40)$$

and is of the form assumed in Equations (22) to (24) with (P_τ^*, q_τ^*) given by Equation (36) and

$$r_\tau^* = s_\tau^* + \ln P_{11}^\tau. \quad (41)$$

Note that for $\tau = 1$ with $\eta_{(0)} = \emptyset$ and $\eta_1 | \eta_{(0)} = \eta_1$ the function k_τ in Equation (27) is a truncated Gaussian kernel for η_1 . Whence, the factorization step defined by Equations (32) to (34) can be skipped and all subsequent terms with subscript

(0) can be deleted.

All in all, Equations (28)-(30) and (36), (38) and (41) fully characterize the GHK-EIS recursion whereby the coefficients $(P_{\tau+1}^*, q_{\tau+1}^*, r_{\tau+1}^*)$ are combined with the period τ EIS regression coefficients $(\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\kappa}_\tau)$ in order to produce back-recursively the coefficients (P_τ, q_τ) characterizing the GHK-EIS sampling densities as well as the coefficients $(P_\tau^*, q_\tau^*, r_\tau^*, c_\tau, d_\tau)$ needed for the EIS step $\tau - 1$.

Based on the functional forms provided above the computation of the GHK-EIS estimate of the probability $P(D)$ requires the following simple steps:

- (i) Simulate S independent trajectories $\{\{\tilde{\eta}_\tau^{(s)}\}_{\tau=1}^{M-1}\}_{s=1}^S$ from the GHK-sampling densities given in Equation (20).
- (ii) Run the back-recursive sequence of the EIS-approximations of $\varphi_\tau \chi_{\tau+1}$ by k_τ for $\tau = M, M-1, \dots, 1$. Specifically:
 - (ii.1) For $\tau = M$, set $\chi_{M+1} \equiv 1$ and select as GHK-EIS density kernel

$$k_M(\eta_{(M)}; \hat{a}_M) = \varphi_M(\eta_{(M)}),$$

generating immediately a perfect fit to $\varphi_M \chi_{M+1}$ which explains why we do not have to draw η_M in step (i) and (iii). (The corresponding EIS parameters in Equation (26) are $\hat{\alpha}_M = \hat{\beta}_M = \hat{\kappa}_M = 0$.) Integrating k_M w.r.t. η_M yields

$$\chi_M(\eta_{(M-1)}; a_M) = \Phi(\omega_{M-1}), \quad \omega_{M-1} = c_M - d'_M \eta_{(M-1)}, \quad (42)$$

with $c_M = -\mu_M/\delta_M$ and $d'_M = \gamma'_M/\delta_M$. It follows from Equations (22) and (23) that $P_M^* = 0, q_M^* = 0, r_M^* = 0$.

(ii.2) For $1 < \tau < M$, use $(c_{\tau+1}, d_{\tau+1})$ to construct the simulated values of ω_τ , i.e., $\tilde{\omega}_\tau^{(s)} = c_{\tau+1} - d'_{\tau+1} \tilde{\eta}_{(\tau)}^{(s)}$ and to run the EIS regression:

$$-2 \ln \Phi(\tilde{\omega}_\tau^{(s)}) = \alpha_\tau [\tilde{\omega}_\tau^{(s)}]^2 + 2\beta_\tau \tilde{\omega}_\tau^{(s)} + \kappa_\tau + \zeta_\tau^{(s)}, \quad s = 1, \dots, S, \quad (43)$$

where $\zeta_\tau^{(s)}$ denotes the implicit regression error term. Use the LS estimates $(\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\kappa}_\tau)$ to compute the auxiliary parameters (P_τ, q_τ, r_τ) according to Equations (28)–(30). The partitioning of (P_τ, q_τ) in Equation (31) together with Equation (35) provides the GHK-EIS sampler (39).

Next, compute $(P_\tau^*, q_\tau^*, r_\tau^*)$ according to Equations (36) and (41), and (c_τ, d_τ) according to Equation (38). The integrating constant χ_τ to be transferred back into the $(\tau - 1)$ -EIS step is then given by Equation (40).

(ii.3) For $\tau = 1$, proceed as above to obtain the auxiliary parameters (P_1, q_1, r_1) .

The corresponding GHK-EIS sampler is $m_1(\eta_1; \hat{a}_1) \sim \mathbb{I}(\eta_1 < -\mu_1/\delta_1) \cdot N[q_1/P_1, 1/P_1]$ and the integration of k_1 w.r.t. η_1 yields

$$\chi_1(\hat{a}_1) = \Phi\left(-\left[\frac{\mu_1}{\delta_1} + \frac{q_1}{p_1}\right]\sqrt{P_1}\right) \cdot \sqrt{\frac{1}{P_1}} \cdot \exp\left(-\frac{1}{2}\left(r_1 - \frac{q_1^2}{P_1}\right)\right). \quad (44)$$

(iii) Simulate S independent trajectories $\{\{\tilde{\eta}_\tau^{(s)}\}_{\tau=1}^{M-1}\}_{s=1}^S$ from the GHK-EIS sampling densities $\{m_\tau(\eta_\tau | \eta_{(\tau-1)}, \hat{a}_\tau)\}_{\tau=1}^{M-1}$ obtained in step (ii) either to repeat step (ii) or – at convergence – to compute the GHK-EIS estimate of the probability $P(D)$ as given by Equation (17), which simplifies into:

$$\hat{P}_S^{\text{GHK-EIS}}(D) = \chi_1(\hat{a}_1) \cdot \frac{1}{S} \sum_{s=1}^S \prod_{\tau=1}^{M-1} \frac{\Phi(\tilde{\omega}_\tau^{(s)})}{\exp\left(-\frac{1}{2}(\hat{\alpha}_\tau [\tilde{\omega}_\tau^{(s)}]^2 + 2\hat{\beta}_\tau \tilde{\omega}_\tau^{(s)} + \hat{\kappa}_\tau)\right)}, \quad (45)$$

where $\tilde{\omega}_\tau^{(s)} = c_{\tau+1} - d'_{\tau+1} \tilde{\eta}_{(\tau)}^{(s)}$.

Convergence of this iterative construction of the EIS sampling density can be checked by monitoring the values of the auxiliary EIS parameters $\{\hat{a}_\tau\}$ across successive iterations and using a stopping rule based, e.g., on an appropriate relative change threshold. In the applications described below, the probability integrands in Equation (10) turn out to be well-behaved functions in $\eta_{(\tau)}$ so that convergence obtains in two or three iterations, depending upon the degree of correlation among the errors, under a relative change threshold of the order of 10^{-3} . In particular, for modest correlation among the errors in Equation (5) (with AR(1) parameters ρ_j and cross-correlations in Σ of the order of 0.5) convergence typically obtains in two iterations, while for higher correlations (of the order of 0.8) an additional iteration may be required.

As noted in Section 3.1, GHK-EIS covers GHK as a special case with $\hat{a}_\tau = \hat{\beta}_\tau = 0$ leading to the GHK-sampling densities (20) and the GHK estimate (21). 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 $(Y_1, \dots, Y_\tau) < 0$, but neglects the correlated information $(Y_{\tau+1}, \dots, Y_M) < 0$. This implies that draws from m_τ ignore potentially critical information, which would allow to adjust the region of importance for η_τ conditionally on $\eta_{(\tau-1)}$, leading to potential efficiency losses of the MC-GHK estimate for the probability $P(D)$ (see also Stern, 1997). In contrast, the GHK-EIS auxiliary parameters $\hat{a}_\tau = (\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\kappa}_\tau)'$ are constructed backward recursively in such a way that they account for that information. In particular, $\hat{a}_\tau$ accounts for the one-period-ahead integrating factor $\chi_{\tau+1}$ which conveys information on the constraint $Y_{\tau+1} < 0$ and – since it depends by recursion on $\chi_{\tau+2}, \dots, \chi_M$ – also on $(Y_{\tau+2}, \dots, Y_M) < 0$. Accordingly, the GHK density m_τ can be interpreted as a filtering density in-

corporating the information about the constraints on Y only up to element τ . In contrast, GHK-EIS produces sequential sampling densities for η_τ , which are conditional on the entire set of constraints on Y .

In the following sections we illustrate the application of the GHK-EIS probability simulator to ML estimation of multinomial probit models. However, it should be mentioned that GHK-EIS could also be used to implement the Method of Simulated Moments (MSM) estimator analogously to the MSM implementation of Keane (1994) based on the standard GHK simulator. This is an important avenue for future research.

3.3 GHK-EIS application for the static model

The application of GHK-EIS to the selection probability of the static multinomial Probit model given in Equation (4) is straightforward. In particular, this probability, which represents the likelihood contribution of a particular observation, is an integral of the form given in Equations (10) and (11) with $M = J$. Let j_i denote the index of the alternative chosen by observation i . According to Equations (1)–(3), Equation (7) is rewritten as

$$Y_{j_i} = S_{j_i} Y_i = S_{j_i} \mu_i + L_{j_i} \eta_i, \quad \eta_i \sim N_J(0, I_{(J)}), \quad (46)$$

where L_{j_i} denotes the Cholesky decomposition of the the covariance matrix $\text{Cov}(S_{j_i} \epsilon_i) = S_{j_i} \Psi S_{j_i}'$. Note that since there are only $J + 1$ alternatives, we have at most $J + 1$ Cholesky decompositions to compute.

3.4 GHK-EIS application for the multiperiod model

Under autocorrelation in the MMP model the likelihood function for a particular individual given by Equation (6) 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 (7) to (11) with $Y = (Y'_{j1}, \dots, Y'_{jT})'$. The vector μ in Equation (7) then denotes $(\mu'_{j1}, \dots, \mu'_{jT})'$, where $\mu_{j_t} = S_{j_t} \mu_t$, and the lower triangular matrix L is the Cholesky decomposition of the joint covariance matrix of $(\epsilon_{j1}, \dots, \epsilon_{jT})$, where $\epsilon_{j_t} = S_{j_t} \epsilon_t$. This joint covariance matrix contains the following blocks:

$$\text{Var}(\epsilon_{j_t}) = S_{j_t} \Omega S'_{j_t}, \quad \text{Cov}(\epsilon_{j_t}, \epsilon_{j_s}) = S_{j_t} R^{t-s} \Omega S'_{j_s}, \quad t > s, \quad (47)$$

where Ω represents the stationary covariance matrix of the shocks ϵ_t . According to Equation (5) it satisfies $\Omega = R\Omega R + \Sigma$. The Cholesky decomposition of the joint covariance matrix of $(\epsilon_{j1}, \dots, \epsilon_{jT})$ can be computed either by brute force or, more efficiently, by application of lemma A1 in the Appendix. The latter exploits the particular structure of the joint covariance matrix and is based on individual Cholesky decomposition of matrices of the form $S_{j_t} \Omega S'_{j_t}$.

The main advantage of this one-shot procedure (also used to implement the GHK for ML estimation of a MMP model, 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 (26) depend upon only three coefficients irrespectively of the size $J \cdot T$.

Nevertheless, if $J \cdot T$ were significantly larger, there is an alternative to the one-

shot procedure based on the Cholesky decomposition of a single $J \cdot T$ -dimensional covariance matrix which could be considered at the cost of additional programming. It 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 (9) to (40) as the $J + (\tau - 1)$ -dimensional vector $\eta'_{(\tau-1)} = (\epsilon_{-1}', \eta_1, \dots, \eta_{\tau-1})$, where ϵ_{-1} denotes the vector of innovations $\epsilon_{j_{t-1}}$ associated with the alternative selected in period $t - 1$ and $\eta_1, \dots, \eta_{\tau-1}$ 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 (14) 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.

4 Monte Carlo Results for the MMP model

In order to analyze the sampling distribution and numerical accuracy of the ML estimator based upon GHK and GHK-EIS for the MMP model given by Equations (5) and (6), we use the same design as Geweke et al. (1997). They consider a three alternative ($J + 1 = 3$) probit model with $T = 10$ periods and $N = 500$ individuals. In particular, they assume the following data generating process (DGP) for the utility differences of individual i :

$$Y_{it} = \mu_{it} + \epsilon_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N, \quad (48)$$

with

$$\mu_{it} = (\pi_{01} + \pi_{11}X_{it} + \psi Z_{it1}, \pi_{02} + \pi_{12}X_{it} + \psi Z_{it2})' \quad (49)$$

$$\epsilon_{it} = R\epsilon_{it-1} + v_{it} \quad (50)$$

$$v_{it} \sim N_2 \left(0, (1 - \rho_1)^2 \cdot \begin{bmatrix} 1 & \omega_{12} \\ \omega_{12} & \omega_{12}^2 + \omega_{22}^2 \end{bmatrix} \right), \quad (51)$$

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

$$X_{it} = \phi\zeta_i + \sqrt{1 - \phi^2}\omega_{it}, \quad Z_{itj} = \phi\tau_{ij} + \sqrt{1 - \phi^2}\xi_{itj}, \quad (52)$$

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

We use this DGP to construct sampling distribution of the ML-GHK and ML-GHK-EIS estimator. Richard and Zhang (2007) advocate distinguishing between MC *numerical* standard deviations (obtained for a single data set under different sets of CRNs) and *statistical* standard deviations (obtained for different data sets under a single set of CRNs). However, in order to make our results directly comparable to those presented by Geweke et al. (1997) we ran our MC simulations with a different set of CRNs for each simulated data set. These produce *compound* standard deviations of the corresponding ML estimator which account jointly for numerical and statistical variations. This being said, it will be the case for all the results reported below in Tables 1 to 4 that numerical variation is always dominated by statistical variation. Whence the compound standard deviations we report under GHK and GHK-EIS all are very close approximations to the actual statistical standard deviations of the corresponding ML estimators. In

a second experiment we then focus our attention on the numerical properties of ML-GHK and ML-GHK-EIS estimates as MC approximations for the unfeasible exact ML estimate. We do so by repeating the corresponding ML estimation 50 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 parameter 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 1–3 where we ran, as mentioned above, two experiments for each set, one based upon 50 simulated data sets (each with its own set of CRNs), the other on 50 different sets of CRNs for the first simulated data set. For the first experiment we report the (compound) means,

standard deviations and RMSEs around the true parameter values (see column three and four of Tables 1–3). 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 iterations¹. For the second experiment we report the (numerical) means, standard deviations and RMSEs around the “true” ML estimates (see column six and seven of Tables 1–3). The latter are obtained by an ML-GHK-EIS estimate based on simulation sample size of $S = 1000$ ². For $S = 20$, one GHK-EIS likelihood evaluation takes 5 s and a GHK evaluation 1 s on an Intel Core 2 CPU notebook with 2GHz for a code written in GAUSS. This 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 $\sqrt{5}$. Figure 1 plots the computing time of GHK and GHK-EIS for one likelihood evaluation of the MMP model against the dimension of the probability integral $M = J \cdot T$ for different simulation sample sizes S . Obviously, the computing time for GHK as well as GHK-EIS is almost linear in M and S , while GHK is between 4 and 6 times faster than GHK-EIS.

Our results for the compound 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

¹For all ML-estimations on our MC-study we use the BFGS optimizer with the true parameter values as starting values.

²In order to verify that the “true” ML 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.

and ρ_2 under parameter set 1 and 3 (see Tables 1 and 3) and for all covariance parameters under set 2 (see Table 2).

Next, the results reported under different data sets indicate that the compound means, standard deviations and RMSEs for the ML-GHK-EIS estimates of the π -coefficients are nearly the same as those for their ML-GHK counterparts for all three data structure. This is not the case for the compound means (and RMSEs) of the ML estimates of the covariance parameters. While their ML-GHK estimates suffer from significant biases their ML-GHK-EIS estimates are virtually unbiased even under simulation sample sizes as low as $S = 20$.

As for numerical accuracy, the results obtained for the repeated parameter estimates under different sets of CRNs and for a fixed data set 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 7 (ω_{12}) and 16 times (ρ_1) smaller than their GHK counterpart under the first parameter set (see Table 1). Furthermore, the mean GHK-EIS estimates are very close to the true ML values under all three data structures and for all parameters. GHK, on the other hand, while producing estimates close to the true ML values for the mean parameters, exhibits relatively large numerical biases for the covariance parameters. Thus, the 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 exact ML estimate).

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 2 plots the GHK and EIS-GHK MC estimates of the sectional log-likelihood functions for the mean parameter ψ and the covariance parameter

ρ_2 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_2 are obtained by setting the remaining parameters equal to their true ML value as 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.

As shown so far, GHK-EIS provides significant improvements over GHK given the frequently used simulation sample size $S = 20$. However, as mentioned above, the likelihood evaluation using GHK-EIS is about five times slower than that based upon GHK. In order to analyze which sample size S and computing time one needs to achieve with ML-GHK the same numerical and statistical accuracy as ML-GHK-EIS with $S = 20$, we increased S for ML-GHK from 20 to 100, 500 and 1280, respectively, and repeated the MC experiments for the second parameter set. The results are summarized in Table 4 and indicate that for ML-GHK a simulation sample size of at least $S = 500$ is needed to obtain the same level of numerical accuracy as ML-GHK-EIS with $S = 20$ (see Table 2). Furthermore, for $S = 500$ the statistical biases of ML-GHK found for the covariance parameter disappear and its statistical accuracy is about the same as that of GHK-EIS with $S = 20$. Since GHK based on $S = 500$ is about five times slower than GHK-EIS with $S = 20$ (see Figure 1), the GHK procedure needs for this parameter set significantly more computing time to achieve the same statistical and numerical accuracy as the GHK-EIS algorithm.

5 Conclusion

We have proposed to combine the GHK probability simulator with Efficient Importance Sampling (EIS) in order to compute choice probabilities for standard multinomial Probit models as well as for 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 for multinomial Probit models as illustrated by the MC results we have reported for the MMP. We have shown that GHK-EIS can lead to significant numerical efficiency gains relative to GHK, even under comparable computing times for likelihood evaluation and ML estimation. Hence, GHK-EIS adds a powerful tool to the simulation arsenal, and depending on the context it can lead to substantial improvements over other methods.

Acknowledgement

We are grateful to three 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 research support provided by the National Science Foundation (NSF) under grant SES-0516642.

Appendix: Efficient Cholesky decomposition for $\text{Cov}(\epsilon_{j_1}, \dots, \epsilon_{j_T})$

According to Equation (47), the $J \cdot T$ -dimensional stationary covariance matrix V of $(\epsilon_{j_1}, \dots, \epsilon_{j_T})$ is partitioned into J -dimensional quadratic blocks of the form

$$V_{ts} = \text{Cov}(\epsilon_{j_t}, \epsilon_{j_s}') = S_{j_t} R^{t-s} \Omega S_{j_s}', \quad t \geq s, \quad (\text{A-1})$$

with $S_j = S_j^{-1}$ (note the S_{j_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 $t \geq s$.

Lemma A1. *The diagonal blocks of L are given by the following J -dimensional Cholesky decompositions*

$$L_{11} L_{11}' = S_{j_1} \Omega S_{j_1}' \quad (\text{A-2})$$

$$L_{tt} L_{tt}' = S_{j_t} \Sigma S_{j_t}', \quad \text{with } \Sigma = \Omega - R \Omega R', \quad t > 1, \quad (\text{A-3})$$

and the off-diagonal blocks by the products

$$L_{ts} = (S_{j_t} R^{t-s} S_{j_s}) L_{ss}, \quad t \geq s. \quad (\text{A-4})$$

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

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

$$L_{t1}L'_{11} = S_{jt}R^{t-1}\Omega S'_{j1} = (S_{jt}R^{t-1}S_{j1})S_{j1}\Omega S'_{j1} = (S_{jt}R^{t-1}S_{j1})L_{11}L'_{11}. \quad (\text{A-5})$$

For $s > 1$, we have

$$L_{t1}L'_{s1} + \sum_{j=2}^{s-1} L_{tj}L'_{sj} + L_{ts}L'_{ss} = S_{jt}R^{t-s}\Omega S'_{js}, \quad (\text{A-6})$$

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

$$\begin{aligned} L_{ts}L'_{ss} &= S_{jt} \left[-R^{t-s}\Omega R'^{s-1} - \sum_{j=2}^{s-1} R^{t-j}(\Omega - R\Omega R')R'^{s-j} \right. \\ &\quad \left. + R^{t-s}\Omega \right] S'_{js} \end{aligned} \quad (\text{A-7})$$

$$= S_{jt}R^{t-s}S_{js} [S_{js}(\Omega - R\Omega R')S'_{js}], \quad (\text{A-8})$$

which, together with (A-3), completes the proof. $\square$

Note that the proof critically relies on the fact that S_j is square non-singular with $S_j^{-1} = S_j$.

References

- 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. 1992. Health, children, and elderly living arrangements: a multiperiod multinomial probit model with unobserved heterogeneity and autocorrelated errors. In Wise, D.A., *Topics in the Economics of Aging*. University of Chicago Press, Chicago, 79–108.
- 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, 3463–3568.
- 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., 2000. Some practical issues in maximum simulated likelihood. In Mariano, R., Schuermann, T., Weeks, M., *Simulation-based Inference in Econometrics: methods applications*. Cambridge University Press, Cambridge, UK, 71–99.
- 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.

- Hausman, J.A., Wise D.A., 1978. A conditional Probit model for qualitative choice: discrete decisions recognizing interdependence and heterogenous preferences. *Econometrica* 46, 403–426.
- Heckman, J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 120–137.
- Keane, M., 1992. A note on identification in the multinomial probit model. *Journal of Business & Economics Statistics* 10, 193–200.
- Keane, M., 1993. Simulation Estimation for Panel Data Models with Limited Dependent Variables. In Maddala, G.S., Rao, C.R., Vinod, H.D., *The Handbook of Statistics* 11, North Holland publisher, 545–571.
- 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.
- Richard, J.-F., Zhang, W., 2007. Efficient high-dimensional importance sampling. *Journal of Econometrics* 141, 1385–1411.
- Stern, S., 1997. Simulation-based estimation. *Journal of Economic Literature* 35, 2006–2039.
- Thurstone, L., 1927. A law of comparative judgement. *Psychological Review* 34, 273–286.
- Vijverberg, W.P.M., 1997. Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics* 76, 281–307.

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

Parameter	true	diff. data sets		fixed data set/diff. CRNs		
		GHK	GHK EIS	true ML	GHK	GHK EIS
π_{01}	.500	.501 (.031) [.031]	.501 (.031) [.031]	.548	.551 (.0048) [.0057]	.548 (.0004) [.0008]
π_{11}	1.000	.994 (.036) [.036]	.996 (.036) [.036]	1.030	1.031 (.0040) [.0042]	1.031 (.0004) [.0011]
π_{02}	-1.200	-1.187 (.062) [.063]	-1.189 (.061) [.061]	-1.197	-1.205 (.0155) [.0178]	-1.199 (.0019) [.0032]
π_{12}	1.000	.990 (.056) [.056]	.992 (.054) [.054]	1.058	1.063 (.0105) [.0116]	1.060 (.0013) [.0022]
ψ	1.000	.994 (.025) [.025]	.998 (.025) [.025]	1.008	1.008 (.0051) [.0052]	1.009 (.0006) [.0011]
ω_{12}	.500	.519 (.071) [.073]	.504 (.064) [.064]	.511	.533 (.0304) [.0376]	.512 (.0042) [.0042]
ω_{22}	.866	.878 (.061) [.062]	.863 (.063) [.063]	.849	.874 (.0210) [.0320]	.850 (.0031) [.0032]
ρ_1	.500	.452 (.029) [.056]	.498 (.030) [.030]	.518	.472 (.0098) [.0474]	.518 (.0006) [.0006]
ρ_2	.500	.411 (.050) [.102]	.493 (.049) [.049]	.475	.379 (.0394) [.1031]	.477 (.0039) [.0044]

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 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 and ML-GHK for the Multiperiod Multinomial Probit: Parameter Set 2.

Parameter	true	diff. data sets		fixed data set/diff. CRNs		
		GHK	GHK EIS	true ML	GHK	GHK EIS
π_{01}	.500	.496 (.041) [.041]	.502 (.043) [.043]	.540	.541 (.0099) [.0101]	.542 (.0009) [.0027]
π_{11}	1.000	.987 (.039) [.041]	0.999 (.038) [.038]	1.017	1.009 (.0082) [.0115]	1.020 (.0008) [.0030]
π_{02}	-1.200	-1.163 (.073) [.081]	-1.185 (.071) [.072]	-1.136	-1.114 (.0295) [.0366]	-1.143 (.0029) [.0079]
π_{12}	1.000	.979 (.055) [.059]	0.994 (.051) [.051]	1.047	1.037 (.0172) [.0199]	1.051 (.0025) [.0049]
ψ	1.000	.986 (.031) [.034]	1.000 (.030) [.030]	1.005	0.992 (.0099) [.0161]	1.007 (.0009) [.0025]
ω_{12}	.500	.544 (.081) [.091]	.511 (.069) [.069]	.431	.490 (.0342) [.0684]	.443 (.0081) [.0143]
ω_{22}	.866	.916 (.061) [.078]	.871 (.057) [.057]	.749	.832 (.0430) [.0933]	.765 (.0083) [.0183]
ρ_1	.800	.746 (.018) [.057]	.798 (.014) [.014]	.798	.751 (.0095) [.0482]	.796 (.0008) [.0017]
ρ_2	.800	.702 (.041) [.106]	.791 (.026) [.027]	.842	.757 (.0220) [.0879]	.834 (.0028) [.0083]

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

Parameter	true	diff. data sets		fixed data set/diff. CRNs		
		GHK	GHK EIS	true ML	GHK	GHK EIS
π_{01}	.500	.500 (.033) [.033]	.501 (.034) [.034]	.500	.500 (.0041) [.0042]	.501 (.0004) [.0008]
π_{11}	1.000	.996 (.035) [.035]	1.000 (.035) [.035]	.938	0.936 (.0044) [.0046]	0.940 (.0004) [.0025]
π_{02}	-1.200	-1.197 (.082) [.082]	-1.204 (.084) [.084]	-1.102	-1.112 (.0192) [.0216]	-1.115 (.0025) [.0134]
π_{12}	1.000	.986 (.052) [.054]	.988 (.053) [.053]	.934	0.939 (.0127) [.0135]	0.937 (.0016) [.0035]
ψ	1.000	1.002 (.042) [.042]	1.009 (.042) [.042]	.936	0.934 (.0038) [.0042]	0.941 (.0008) [.0053]
ω_{12}	.800	.798 (.059) [.059]	.785 (.062) [.063]	.694	.714 (.0251) [.0322]	.689 (.0042) [.0062]
ω_{22}	.600	.599 (.046) [.046]	.599 (.049) [.049]	.572	.588 (.0161) [.0223]	.579 (.0022) [.0066]
ρ_1	.500	.464 (.028) [.045]	.494 (.029) [.029]	.509	.478 (.0087) [.0321]	.507 (.0007) [.0018]
ρ_2	.500	.459 (.040) [.057]	.497 (.043) [.043]	.549	.504 (.0159) [.0482]	.549 (.0022) [.0022]

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 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 4. ML-GHK for the Multiperiod Multinomial Probit
for Alternative Simulation Sample Sizes: Parameter Set 2.

Param.	diff. data sets				fixed data set/diff. CRNs			
	true	S			true ML	S		
		100	500	1280		100	500	1280
π_{01}	.500	.498 (.043) [.043]	.499 (.043) [.043]	.500 (.043) [.043]	.540	.540 (.0058) [.0058]	.540 (.0043) [.0043]	.539 (.0027) [.0028]
π_{11}	1.000	.993 (.037) [.037]	.996 (.038) [.038]	.997 (.038) [.038]	1.017	1.014 (.0063) [.0073]	1.015 (.0032) [.0039]	1.017 (.0020) [.0022]
π_{02}	-1.200	-1.173 (.076) [.080]	-1.180 (.074) [.076]	-1.181 (.073) [.075]	-1.136	-1.123 (.0179) [.0222]	-1.132 (.0098) [.0104]	-1.134 (.0056) [.0061]
π_{12}	1.000	.986 (.052) [.053]	.989 (.053) [.053]	.991 (.052) [.052]	1.047	1.041 (.0112) [.0126]	1.044 (.0072) [.0078]	1.046 (.0035) [.0035]
ψ	1.000	.993 (.031) [.031]	.997 (.031) [.031]	.997 (.031) [.031]	1.005	.998 (.0063) [.0093]	1.002 (.0032) [.0038]	1.003 (.0021) [.0025]
ω_{12}	.500	.521 (.074) [.076]	.507 (.070) [.070]	.509 (.071) [.071]	.431	.447 (.0219) [.0272]	.434 (.0156) [.0158]	.435 (.0091) [.0100]
ω_{22}	.866	.892 (.070) [.074]	.872 (.063) [.063]	.869 (.063) [.062]	.749	.774 (.0280) [.0374]	.756 (.0205) [.0216]	.754 (.0115) [.0128]
ρ_1	.800	.786 (.014) [.020]	.796 (.014) [.015]	.798 (.013) [.013]	.798	.784 (.0053) [.0146]	.796 (.0028) [.0037]	.797 (.0015) [.0018]
ρ_2	.800	.769 (.032) [.044]	.789 (.027) [.028]	.793 (.028) [.028]	.842	.821 (.0107) [.0239]	.838 (.0066) [.0078]	.839 (.0041) [.0050]

NOTE: The reported numbers for ML-GHK are mean, standard deviation (in parentheses) and RMSE (in brackets) obtained for $S = 100, 500, 1280$. 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 $S = 1000$.

Figure 1. Computing time in seconds for one likelihood evaluation of the multiperiod multinomial Probit for $N = 500$ individuals using GHK (left panel) and GHK-EIS (right panel) for different values of $M = J \cdot T \in \{4, 8, 16, 32, 64\}$ (dimension of the probability integral) and different simulation sample sizes S . The times are obtained on an Intel Core 2 CPU notebook with 2 GHz with GAUSS 6.0. The variation in M is obtained by a variation in T while fixing $J = 2$.

Figure 2. Sectional log-likelihood functions for the 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 true ML values (see Table 2). The vertical lines indicate the range of the parameter values which maximize the individual simulated sectional log-likelihood functions.