Working Paper n° 09-02

Estimation of Multivariate Probit Models by Exact Maximum Likelihood

Jacques Huguenin, Florian Pelgrin and Alberto Holly

UNIL | Université de Lausanne

IEMS – Institute of Health Economics and Management

www.hec.unil.ch/iems

Estimation of Multivariate Probit Models by Exact Maximum Likelihood

Jacques Huguenin[1], Florian Pelgrin[2] and Alberto Holly[2]

[1] Institute of Health Economics and Management (IEMS), University of Lausanne
[2] Faculty of Business and Economics (HEC) and Institute of Health Economics and Management (IEMS), Switzerland

Working Paper n° 09-02

February 2009

Institute of Health Economics and Management (IEMS)
UNIL Dorigny
Extranef
1015 Lausanne
Switzerland
Phone +41 (0)21 692 33 20
Fax +41 (0)21 692 36 55

www.hec.unil.ch/iems

Estimation of Multivariate Probit Models by Exact Maximum Likelihood*

Jacques Huguenin

Institute of Health Economics and Management (IEMS)

University of Lausanne, Switzerland

Florian Pelgrin

Faculty of Business and Economics (HEC)

and Institute of Health Economics and Management (IEMS)

University of Lausanne, Switzerland

Alberto Holly

Faculty of Business and Economics (HEC)

and Institute of Health Economics and Management (IEMS)

University of Lausanne, Switzerland

July 11, 2008

Abstract

In this paper, we develop a new numerical method to estimate a multivariate probit model. To this end, we derive a new decomposition of normal multivariate integrals that has two appealing properties. First, the decomposition may be written as the sum of normal multivariate integrals, in which the highest dimension of the integrands is reduced relative to the initial problem. Second, the domains of integration are bounded and delimited by the correlation coefficients. Application of a Gauss-Legendre quadrature rule to the exact likelihood function of lower dimension allows for a major reduction of computing time while simultaneously obtaining consistent and efficient estimates for both the slope and the scale parameters. A Monte Carlo study shows that the finite sample and asymptotic properties of our method compare extremely favorably to the maximum simulated likelihood estimator in terms of both bias and root mean squared error.

JEL Classification: C1 and C3.

Keywords: Multivariate Probit Model, Simulated and Full Information Maximum Likelihood, Multivariate Normal Distribution, Simulations.

*We are grateful to participants at the international workshop on computational and financial econometrics (Geneva, 2007) and at the Swiss society for economics and statistics (Lausanne, 2008). We thank for very helpful comments on preliminary drafts of this paper Nicolas Chopin and Alain Monfort. All errors are our own. Corresponding authors: Florian Pelgrin, fpelgrin@unil.ch, University of Lausanne—Extranef, Quartier Dorigny-Unil, 1015 Lausanne, Switzerland. Alberto Holly, aholly@unil.ch, University of Lausanne—Extranef, Quartier Dorigny-Unil, 1015 Lausanne, Switzerland.

1 Introduction

The multivariate probit is an appealing model of choice behavior because it allows a flexible correlation structure for the unobservable variables. However, until now, applications have been limited because they require high dimensional numerical- or simulation-based integration, and integration (or simulation) of the multivariate normal density over subsets of a Euclidean space is computationally burdensome. More generally, estimation of limited dependent variables models is often hampered by computational complexity. In particular, there is a widespread consensus in the literature that the use of numerical integration or quadrature techniques is possible in principle but it is too time-consuming and/or imprecise to consider in practice, except for the case of low dimensional problems.

Since the seminal work of Ashford and Sowden (1970) on multivariate probit models, numerous attempts have been made to circumvent the curse of dimensionality in evaluating the multivariate probabilities involved. These have been evaluated using either deterministic or Monte Carlo integration, with the latter generally being preferred as the dimension of the problem increases. On one hand, quadrature methods have been used in models that assume special structures of the correlation matrix and for which closed-form expressions are available. Here, the multiple integration problems are typically greatly reduced (Ashford and Sowden, 1970; Sickles and Taubman, 1986; Bock and Gibbons, 1996).[1] In addition, numerical integration has been extensively studied in the evaluation of multivariate normal probabilities with a given mean vector and variance-covariance or correlation matrix (Tong, 1990; Kotz, Balakrishnan, and Johnson, 2000).[2]

On the other hand, various simulation-based methods, as suggested by Lerman and Manski (1981), McFadden (1989), and Pakes and Pollard (1989), have been developed for discrete choice models: the maximum simulated likelihood, the method of simulated moments, and the method of simulated scores.[3] Briefly speaking, these simulation-based methods are a combination of a probability simulator, which determines the multivariate cumulative (normal) density function, and standard estimation procedures through the optimization of the objective function.[4]

In this paper we propose a new procedure, which relies on a numerical integration of an exact dimensional reduction formula of the maximum likelihood (henceforth exact maximum likelihood) function, or the score vector.[5] The key idea is to provide an exact decomposition of the cumulative density function of the M-variate (standardized) normal vector encountered in the likelihood

[1]Quadrature methods have been used by, among others, Heckman (1981) and Butler and Moffitt (1982) in the analysis of discrete choices panel data. Trapezoidal integration, which is a member of the Newton-Cotes formulae, has been implemented by Heckman and Willis (1975). An alternative to quadrature methods is to implement the Laplace approximation (De Bruijn, 1981; Tierney and Kadane, 1986). To the best of our knowledge, this method has not yet been implemented in multivariate probit models.

[2]See also Miwa, Hayter, and Kuriki (2003), and Craig (2008).

[3]For a review, see Gourieroux and Monfort (1996), Stern (1997) and Train (2003).

[4]Other approaches include full Bayesian estimation (Chib and Greenberg, 1998; Lawrence et al., 2004) and GMM estimators (Bertschek and Lechner, 1998).

[5]The results presented in this paper are based on the work of Huguenin (2004) who, using a different approach, generalized and extended the results originally derived in Lazard Holly and Holly (2003).

function of a multivariate probit model. In doing so, we obtain a sum of multivariate integrals, in which the highest dimension of the integrands is $M-1$. It is worth noting that some of our results are in the spirit of the reduction formula for normal multivariate integrals proposed by Plackett (1954). Despite the importance of the results derived by Plackett (1954), they have not received the attention that they deserve. One of the advantages of our suggested procedure is that the domains of integration are bounded and delimited by the correlation coefficients and are thus of Lebesgue measure less than one. We therefore obtain, as a first step, an exact decomposition of the (log-) likelihood function or the corresponding score vector. In a second step, we use a Gauss-Legendre quadrature over bounded intervals for each term of the sum (the approximation afforded by the Gauss-Legendre quadrature rule may be made arbitrarily precise by increasing the number of nodes).[6] Hence, the features of the exact decomposition that we suggest allow for obtaining consistent and efficient estimates of the slope and variance-covariance parameters. Moreover, our method is time-efficient, meaning that the computing time is by no means comparable to that of the maximum simulated likelihood estimator, especially as the dimension of the multivariate probit model increases.

In the sequel, the maximum simulated likelihood method (McFadden, 1989; Pakes and Pollard, 1989) is our benchmark and we compare our results with those of this approach. As is well known, the consistency and asymptotic normality of the maximum simulated likelihood estimator require, among others features, that the number of draws, R, grows without bound faster than the square root of the number of observations. This, in turn, means that consistency is often obtained at the cost of increasing the computing time, which may make such methods less attractive even with the new developments in computer sciences. Computational cost is important *per se* since simulation-based methods are also designed to reduce the computation time. If this were not an issue, one could simulate while letting R be an arbitrarily large number. In addition, the consistency and efficiency of the variance-covariance or correlation parameters are generally ignored in the simulation-based literature. However, even if the accuracy of the simulator is sufficient for locating a relative maximum, it may not be sufficiently accurate for the calculation of the Hessian and the parameter variance estimates (Breslaw, 1994; Lee, 1997). The higher nonlinearity characteristic of the first order conditions with respect to the variance-covariance or correlation parameters makes the consistency of the maximum simulated likelihood even more critical. In contrast, the new approach based on numerical integration proposed in this paper dramatically reduces the computation time while it simultaneously increases the numerical accuracy of both the slope and the variance-covariance parameters, thereby yielding an efficient estimation of the parameters of interest.

In addition to addressing this main objective, two additional points are worth noting in this paper. First, we show that our method easily extends to the panel probit model as long as the time dimension is not too large (as it is commonly assumed in the literature).[7] More generally, our method applies to almost any estimation of limited dependent variables models that are based on the multivariate normal distribution. In particular, multivariate probit models may be interpreted as the reduced form of a simultaneous equations model with latent and observable variables. Sec-

[6] As we will see later on, the computing time increases less than linearly with the number of nodes.

[7] Using the same specification as in Butler and Moffitt (1992), we are not constrained by the time dimension.

ond, we note that, up to a certain order, our method allows for the estimation of flexible correlation matrices and to test such restrictions. This is not often the case in standard econometrics packages (e.g., Stata).[8] It is hoped that our approach to the reduction problem as well as the availability of the associated software will allow researchers to apply our results in a large number of situations.[9]

The rest of the paper proceeds as follows. In Section 2, we propose two new decomposition formulae for normal multivariate integrals and discuss their main implications. An example is provided in the case of a trivariate normal vector. In section 3, we describe the multivariate probit model and its main assumptions. We then apply our decomposition in order to derive the exact maximum likelihood function or the exact score vector, and the numerical evaluation procedure (Gauss-Legendre quadrature) is detailed. In Section 4, we provide some Monte Carlo simulations and comment on the merits of our approach with respect to the maximum simulated likelihood estimator. Some concluding remarks are presented in Section 5. Proofs are provided in the appendix.

2 New Decomposition Formulae for Normal Multivariate Integrals

In this section, we describe our new decomposition formulae for normal multivariate integrals (say of dimension M). In doing so we proceed in two steps. First, we decompose the multivariate normal cumulative distribution function into a sum of multiple integrals in a unique way. In such integrals, the domains of integration are bounded and delimited by the correlation coefficients and are thus of Lebesgue measure less than one. The strength of this decomposition lies in the transformation of the integration domain. We then show in a second step how to preserve the appeal of this decomposition while reducing the highest dimension of the integrands to $M-1$ and the number of normal integrals in the sum. Both dimensional reduction formulae are in the spirit of Plackett (1954) in a sense to be clarified in Subsection 2.1.

Throughout this paper, we use the following notation. $\Phi_{M,Z}\left(z \mid \mu, \Psi\right) \equiv \Phi_{M,Z}\left(\mu, \Psi\right)$ and $\varphi_{M,Z}\left(z \mid \mu, \Psi\right) \equiv \varphi_{M,Z}\left(\mu, \Psi\right)$ denote, respectively, the cumulative density function (cdf) and probability density function (pdf) of an M-variate normal vector Z of a mean vector μ and a variance-covariance matrix. The cdf and pdf of an M-variate normal vector Z evaluated at w are denoted, respectively, as: $\Phi_{M,Z}\left(w \mid \mu, \Psi\right)$ and $\varphi_{M,Z}\left(w \mid \mu, \Psi\right)$. In the case of an M-variate *standardized* normal vector Z of mean 0_M and correlation matrix Ω evaluated at w, the cdf and pdf are denoted, respectively, as: $\Phi_{M,Z}\left(w \mid 0, \Omega\right) \equiv \Phi_{M,z}\left(w, \Omega\right)$ and $\varphi_{M,Z}\left(w \mid 0, \Omega\right) \equiv \varphi_{M,Z}\left(w, \Omega\right)$.

Before going through all of the details, we consider the simple example of a bivariate probit model. This model requires calculating the bivariate integral of a bivariate normal density function (or the cumulative density function of a bivariate normal vector). Without loss of generality, we consider the *standardized* bivariate normal density:

$$\varphi_{2,z}\left(0, \Omega\right) = \left(2\pi\right)^{-1} \left|\Omega\right|^{-1/2} \exp\left(-\frac{1}{2} z' \Omega^{-1} z\right)$$

[8] See Cappellari and Jenkins (2003).

[9] The Gauss codes are available upon request. See Huguenin (2008).

where $z' = (z_1, z_2)$ and Ω is a 2×2 correlation matrix whose main diagonal elements equal 1 and the symmetric off-diagonal element, $\omega_{12} = \omega_{21}$, belongs to $(-1, 1)$. The cumulative density function is the bivariate integral:

$$\begin{aligned} \Phi_{2,z}(w, \Omega) &= \Pr\left[\bigcap_{m=1}^{2} (z_m \leq w_m)\right] \\ &= \int_{-\infty}^{w_2} \int_{-\infty}^{w_1} \varphi_{2,z}(0, \Omega)\, dz_1 dz_2. \end{aligned}$$

Using our new decomposition of normal multivariate integrals, we show that the bivariate cumulative density function may be rewritten as follows:

$$\begin{aligned} \Phi_{2,z}(w, \Omega) &\equiv \Phi_{2,z}(w \mid \omega_{12}) \\ &= \Phi(w_1)\Phi(w_2) + \frac{1}{2\pi}\int_0^{\omega_{12}} \exp\left(-\frac{1}{2}\frac{w_1^2 + w_2^2 - 2\lambda_{12} w_1 w_2}{1 - \lambda_{12}^2}\right)\frac{d\lambda_{12}}{\sqrt{1 - \lambda_{12}^2}}. \end{aligned}$$

This decomposition corresponds exactly to the dimensional reduction formula of Plackett (Eq. 7 p. 353) and the expression derived in Lazard Holly and Holly (2003) for this particular case. More generally, the appeal of this decomposition rests on the following points. First, this transformation is *exact*, and no approximation has been made at this stage. Second, the highest integral to compute is of dimension 1. Third, the domain of integration changes from a rectangular semi-infinite domain $((-\infty, w_2] \times (-\infty, w_1])$ to a bounded domain $([0, \omega_{12}])$.[10] Finally, the integration of the second left-hand side term is with respect to the unrestricted coefficient of the correlation matrix. This decomposition may be generalized to any M-variate integral of an M-variate normal vector so that any cumulative density function of an M-variate normal vector may be decomposed into a sum of multiple finite range integrals of lower dimension.

2.1 Decomposition Formulae

Before proceeding, we introduce some further notation. Let $\widetilde{\Omega}$ denote the vector of unique elements of the normalized matrix Ω (where $\omega_{ij} = \rho_{ij}$ is either the correlation coefficient or the covariance between two random variables i and j), and n denote its dimension, $n = M(M-1)/2$. We set:

$$\begin{aligned} \overline{\Omega} &= \bigotimes_{i=1}^{M-1} \bigotimes_{j=i+1}^{M} [0, \omega_{ij}] \\ d\widetilde{\Omega} &= \prod_{i=1}^{M-1} \prod_{j=i+1}^{M} d\omega_{ij} \end{aligned}$$

[10] If w is the zero vector, the previous expression simplifies to the well known formula:

$$\Phi_{2,z}(0 \mid 0, \Omega) = \frac{1}{4} + \frac{\arcsin \omega_{12}}{2\pi}$$

where $\Phi(z_1) = \frac{1}{2}$ and $\int_0^{\omega_{12}} \varphi_{2,z}(0, \lambda_{12})\, d\lambda_{12} = \frac{\arcsin \omega_{12}}{2\pi}$.

where $\widetilde{\Omega}_{pq}$ is defined as the set of correlation parameters resulting from the qth combination of $\widetilde{\Omega}$, which contains p zero elements:

$$\begin{aligned}
\widetilde{\Omega}_{01} &= \widetilde{\Omega} \\
\widetilde{\Omega}_{11} &= \widetilde{\Omega}|_{\omega_{12}=0}, \widetilde{\Omega}_{12} = \widetilde{\Omega}|_{\omega_{13}=0}, \ldots, \widetilde{\Omega}_{1n} = \widetilde{\Omega}|_{\omega_{1n}=0} \\
& \ldots \\
\widetilde{\Omega}_{n1} &= \varnothing,
\end{aligned}$$

and $\overline{\Omega}_{pq}$ is the constrained set defined from the correlation parameters of Ω when considering only the $(n-p)$ non-zero elements of $\widetilde{\Omega}_{pq}$. In this context, $d\widetilde{\Omega}_{pq}$ defines the $(M-p)$ non-zero elements of $\widetilde{\Omega}_{pq}$ over which we integrate. Finally, $I_k\left(\widetilde{\Omega}_{pq}\right)$ is the number of indices being equal to k, which are present in the i and j indices of the elements ω_{ij} in $\widetilde{\Omega}_{pq}$.

For example, in a trivariate probit model, the $\widetilde{\Omega}_{pq}$'s correspond to the set of non-zero off-diagonal elements of the following matrices:

$$\begin{aligned}
\Omega_{01} &= \begin{pmatrix} 1 & \omega_{12} & \omega_{13} \\ \omega_{12} & 1 & \omega_{23} \\ \omega_{13} & \omega_{23} & 1 \end{pmatrix}, \Omega_{31} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
\Omega_{11} &= \begin{pmatrix} 1 & 0 & \omega_{13} \\ 0 & 1 & \omega_{23} \\ \omega_{13} & \omega_{23} & 1 \end{pmatrix}, \Omega_{12} = \begin{pmatrix} 1 & \omega_{12} & 0 \\ \omega_{12} & 1 & \omega_{23} \\ 0 & \omega_{23} & 1 \end{pmatrix}, \Omega_{13} = \begin{pmatrix} 1 & \omega_{12} & \omega_{13} \\ \omega_{12} & 1 & 0 \\ \omega_{13} & 0 & 1 \end{pmatrix} \\
\Omega_{21} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \omega_{23} \\ 0 & \omega_{23} & 1 \end{pmatrix}, \Omega_{22} = \begin{pmatrix} 1 & 0 & \omega_{13} \\ 0 & 1 & 0 \\ \omega_{13} & 0 & 1 \end{pmatrix}, \Omega_{23} = \begin{pmatrix} 1 & \omega_{12} & 0 \\ \omega_{12} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

It turns out that the $\overline{\Omega}_{ij}$, say $\overline{\Omega}_{11}$, is defined by $[0, \omega_{13}] \otimes [0, \omega_{23}]$, and $d\widetilde{\Omega}_{11} = d\omega_{13} d\omega_{23}$.

We now turn to the main propositions of our paper. To this end, we first need the following two lemmas, which are well known and are stated here for completeness and notational purposes. The first one establishes the factorization of an M-variate normal cdf. The second lemma rewrites the partial differential matrix equation proposed by Plackett (1954), which allows for the dimensional reduction formula of normal multivariate integrals.

Lemma 1 *The cumulative density function $\Phi_{M,Z}(z, \Omega)$ may be weakly factorized as:*

$$\begin{aligned}
\Phi_{M,Z}(z, \Omega) &= \int_{\mathcal{Z}_1} \varphi_{M_1,Z_1}(t_1, \Omega_{11}) \times \\
& \Phi_{M_2,Z_2}\left(\Delta_{22}^{-1}\left(z_2 - \Omega_{21}\Omega_{11}^{-1}t_1\right), \Delta_{22}^{-1}\left(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}\right)\Delta_{22}^{-1}\right) \prod_{k=1}^{M_1} dt_{1k}
\end{aligned}$$

where $Z = (Z_1, Z_2)$ is a partitioning of the M-variate normal vector Z, $\dim(Z_1) = M_1$, $\dim(Z_2) = M_2$, $\mathcal{Z} = \mathcal{Z}_1 \otimes \mathcal{Z}_2$ is the corresponding partitioning of the integration domain, $t_1 = (t_{11}, \cdots, t_{1M_1})'$, Δ_{22} is a diagonal matrix containing the square roots of the diagonal elements of $\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$, *and Ω is partitioned with respect to Z_1:*

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix}.$$

If Z_1 and Z_2 are independent, the cumulative distribution function $\Phi_{M,Z}(z,\Omega)$ may be strictly factorized in the M_1-dimensional cumulative distribution function $\Phi_{M_1,Z_1}(z_1,\Omega_{11})$ and the M_2-dimensional cumulative distribution function $\Phi_{M_2,Z_2}(z_2,\Omega_{22})$ such that

$$\Phi_{M,Z}(z,\Omega) = \Phi_{M_1,Z_1}(z_1,\Omega_{11})\,\Phi_{M_2,Z_2}(z_2,\Omega_{22}). \qquad \square$$

Applying Lemma 1, it is quite straightforward to show that if $M_1 = 2$, the factorization of the cumulative distribution function with respect to Z_i and Z_j is given by:

$$\begin{aligned} \Phi_{2,Z}(z,\Omega) = & \int_{-\infty}^{z_i}\int_{-\infty}^{z_j} \varphi_{2,Z_1}(t_{ij},\Omega_{ij}) \times \\ & \Phi_{M-2,Z_2}\left(\Delta_{-ij}^{-1}\left(z_{-ij}-\Gamma_{ij}\Omega_{ij}^{-1}t_{ij}\right), \Delta_{-ij}^{-1}\left(\Omega_{-ij}-\Gamma_{ij}\Omega_{ij}^{-1}\Gamma_{ij}'\right)\Delta_{-ij}^{-1}\right) dt_i dt_j \end{aligned}$$

where $t_{ij} = (t_i, t_j)'$, z_{-ij} is the vector z without the elements z_i and z_j, Ω_{ij} is the matrix with the ith and jth rows and columns of Ω, Γ_{ij} is the covariance matrix between z_{ij} and z_{-ij}, and Δ_{-ij} is a diagonal matrix whose the kth element is $\sqrt{\frac{1-\omega_{ij}^2-\omega_{ki}^2-\omega_{kj}^2+2\omega_{ij}\omega_{ki}\omega_{kj}}{1-\omega_{ij}^2}}$, with ω_{ki} and ω_{kj} being the elements of the kth row of Γ_{ij}.

Following Plackett (1954), we now define the partial differential matrix equation that states the relationship between the second order partial derivatives of the cdf (respectively pdf) of an M-variate normal vector with respect to $z = (z_1, ..., z_M)'$ and the first order partial derivative of this cdf (respectively pdf) with respect to the correlation or covariance matrix.[11]

Lemma 2 *Assume that Z is an M-variate (standardized) normal vector. Then we have:*

$$\begin{aligned} \frac{\partial^2}{\partial z \partial z'}\varphi_{M,Z}(z,\Omega) &= \frac{\partial}{\partial \Omega}\varphi_{M,Z}(z,\Omega) \\ \frac{\partial^2}{\partial z \partial z'}\Phi_{M,Z}(z,\Omega) &= \frac{\partial}{\partial \Omega}\Phi_{M,Z}(z,\Omega) \end{aligned}$$

where $z = (z_1, ..., z_M)'$.

$\square$

Proof: See Appendix 1.

To the best of our knowledge, the results of Lemma 2 and its implications for the numerical or stochastic evaluation of normal multivariate integrals have been largely ignored in the literature, except for a few applications. For instance, Hausman and Wise (1978, footnote 17, p. 417) use them to estimate a conditional probit model. Breslaw (1994) develops a low variance simulator to

[11] This result has been known for some time when $M = 2$ and has been established for all M by Plackett (Eq. 3 p. 352).

approximate multivariate normal probability integrals, which uses the derivation formulae of Plackett, the line integral approach and the GHK simulator.

Taking only Lemma 2, we are now in a position to prove one of our main results of this paper—an exact and unique decomposition of the normal cumulative distribution function, $\Phi_{M,Z}(z,\Omega)$, into a finite-countable sum of multiple integrals with bounded domains (Proposition 1).

Proposition 1 *Assume that each matrix Ω_{pq} is positive definite, the cumulative distribution function $\Phi_{M,Z}(z,\Omega)$ may be decomposed as follows:*

$$\Phi_{M,Z}(z,\Omega)=\prod_{k=1}^{M}\Phi(z_k)+\sum_{p=0}^{n-1}\sum_{q=1}^{\binom{n}{p}}\int_{\widetilde{\Omega}_{pq}}\xi_{Mp}\left(z,\widetilde{\Lambda}_{pq}\right)d\widetilde{\Lambda}_{pq} \quad (1)$$

with:

$$\xi_{Mp}\left(z,\widetilde{\Omega}_{pq}\right)=\frac{\partial^{n-p}}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\widetilde{\omega}_{ij}}\Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right)$$

where $\partial\widetilde{\omega}_{ij}=\partial\omega_{ij}$ if $\omega_{ij}\neq 0$ and $\partial\widetilde{\omega}_{ij}=1$ otherwise, and $n=M(M-1)/2$.
If $\Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right)$ is weakly factorizable (Lemma 1), then $\xi_{Mp}\left(z,\widetilde{\Omega}_{pq}\right)$ is the derivative of the density function $\varphi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right)$ with respect to the elements of z such that:

$$\xi_{Mp}\left(z,\widetilde{\Omega}_{pq}\right)=\frac{\partial^{M(M-2)-2p}}{\prod_{k=1}^{M}\partial z_k^{I_k\left(\widetilde{\Omega}_{pq}\right)-1}}\varphi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right).$$

Otherwise, if $\Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right)$ is strictly factorizable as follows:

$$\Phi_{M-L}\left(z_{-\{k_1,\ldots,k_L\}},\widetilde{\Omega}_{-\{k_1,\ldots,k_L\},pq}\right)\prod_{l=1}^{L}\Phi\left(z_{k_l}\right)$$

then $\xi_{Mp}\left(z,\widetilde{\Omega}_{pq}\right)$ is the product of the L independent univariate cumulative distribution functions and the derivative of the density function $\varphi_{M-L,Z}\left(z_{-\{k_1,\ldots,k_L\}},\widetilde{\Omega}_{-\{k_1,\ldots,k_L\},pq}\right)$ such that:

$$\xi_{Mp}\left(z,\widetilde{\Omega}_{pq}\right)=\left[\prod_{l=1}^{L}\Phi\left(z_{k_l}\right)\right]\frac{\partial^{\left[\sum_{k\neq k_l}I_k\left(\widetilde{\Lambda}_{pq}\right)\right]-2p}}{\prod_{k\neq k_l}\partial z_k^{I_k\left(\widetilde{\Lambda}_{pq}\right)-1}}\varphi_{M-L}\left(z_{-\{k_1,\ldots,k_L\}},\widetilde{\Omega}_{-\{k_1,\ldots,k_L\},pq}\right).$$

□

Proof: See Appendix 2.

While allowing for a bounded integration domain, the two limitations of Proposition 1 are that (i) the number of elements in the decomposition is large and (ii) the highest dimension of the multiple integrals is larger than M, except for $M=2$ and $M=3$. Indeed, it is straightforward to show that the total number of elements is 2^n. Therefore, in a multivariate probit model of order 5, one has to compute 1,024 elements! Moreover, the highest dimension of the multiple integrals equals n, the number of unique off-diagonal elements of Ω. In the situation where M is greater than or equal to four, the highest dimension of the integral in the decomposition (Eq. 1) is higher than M. From a practical view, we may consider multivariate probit models only up to order 3, as otherwise this decomposition is not useful from the viewpoint of computing time.

[Insert Table 1 around here]

However, using both Lemma 1 and Lemma 2, we can reduce both the number of elements being summed up and the highest dimension of the multivariate integrals involved. The corresponding decomposition is given in Proposition 2.

Proposition 2 *The normal cumulative distribution function $\Phi_M(z,\Omega)$ may be decomposed into the sum of two terms. The first one is the product of M univariate normal cumulative distribution functions and the second is the sum of multiple integrals with bounded domains, with the range of each integral being of the magnitude of one of the elements of Ω. Specifically, $\Phi_{M,Z}$ may be written as:*

$$\Phi_{M,Z}(z,\Omega) = \prod_{k=1}^{M} \Phi(z_k) + \sum_r \int_{\overline{\Omega}_r} \psi_{Mr}\left(z, \widetilde{\Lambda}_r(\Omega)\right) d\widetilde{\Lambda}_r \tag{2}$$

where each matrix $\widetilde{\Lambda}_r(\Omega)$ is assumed to be positive definite and $\psi_{Mr}\left(z,\widetilde{\Lambda}_r(\Omega)\right)$ is given by some linear combination of the $\xi_{Mp}\left(z,\widetilde{\Lambda}_r(\Omega)\right)$ defined in Proposition 1. $\widetilde{\Lambda}_r(\Omega)$ is the corresponding matrix $\widetilde{\Lambda}_{pq}$ filled with some non-zero elements ω_{ij}, and $\overline{\Omega}_r$ is a bounded integration domain.
The highest dimension of the multivariate integrals in the sum is $(M-1)$. $\square$

Proof: See Appendix 3.

Several points are worth noticing. First, as in Proposition 1, the decomposition is exact. However, it is no more unique as is stated in the proof of Proposition 2 and illustrated in the example below. Second, the highest dimension of the multiple integrals is $M-1$ so that a bivariate (trivariate) probit model is analyzed as a univariate (bivariate) integration problem. Moreover, since the integration domains are bounded and of Lebesgue measure less than one, the numerical integration is tremendously simplified. Third, the number of elements is also reduced considerably (Table 1). Finally, the number of elements to evaluate may be further reduced if we impose restrictions on the variance-covariance matrix or the correlation matrix, as is often done in the literature.[12]

[12]In order to solve the dimensionality problem of multivariate probit models, some of the literature assumes special structures of the correlation matrix for which closed-form expressions for the probabilities are available and for which the multiple integration problem is greatly reduced (Ashford and Sowden, 1970; Sickles and Taubman, 1986). The correlation matrix of the multivariate probit model is generally defined in terms of a multi-factor structure (Ochi and Prentice, 1984; Bock and Gibbons , 1996) and the multivariate normal probabilities are evaluated using a Gauss-Hermite quadrature method. But, except under these simplifying assumptions on the correlation (covariance) matrix, the likelihood function remains difficult to evaluate and the curse of dimensionality is still present. In contrast, here we do not need to impose any restrictions on the correlation matrix and numerical integration is simplified by making use of a Gauss-Legendre quadrature method.

Using Lemma 2, Plackett (1954) also obtains a dimension reduction formula of any normal multivariate integral (Eq. 7 p. 353). It consists of expressing a multivariate normal probability as the sum of an easily computed reference probability and a probability correction that may be found as the sum of one-dimensional integrals (after recursion).[13] For completeness, the corresponding decomposition is given in Lemma 3.

Lemma 3 *Assume that the reference matrix is the identity matrix. Then we have:*

$$\Phi_{M,Z}(z,\Omega) = \prod_{k=1}^{M} \Phi(z_k) + \int_0^1 \sum_{i<j} \omega_{ij}\varphi_{2,Z_{ij}}(z_{ij}, \Sigma_{ij}(s))\, \Phi_{M-2,Z_{-ij}}\left(\widetilde{z}_{-ij}(s), \widetilde{\Sigma}_{-ij}(s)\right) ds \tag{3}$$

where $\Sigma(s) = s\Omega + (1-s) I_M$, $\widetilde{z}_{-ij}(s) = \Delta_{-ij}^{-1}(s)\left(z_{-ij} - \Gamma_{ij}(s)\Sigma_{ij}^{-1}(s)\Gamma'_{ij}(s)\right)$, *and*
$\widetilde{\Sigma}_{-ij}(s) = \Delta_{-ij}^{-1}(s)\left(\Sigma_{-ij}(s) - \Gamma_{ij}(s)\Sigma_{ij}^{-1}(s)\Gamma'_{ij}(s)\right)\Delta_{-ij}^{-1}(s)$. $\square$

Proof: See Appendix 4.

Both decompositions (Eq. 2 and 3) may be interpreted as the sum of a target probability $\Phi_{M,Z}(z,\Omega)$ into a reference probability (the reference matrix being the identity matrix), $\prod_{k=1}^{M}\Phi(z_k)$, and a probability correction term—the second term on the right-hand side. Moreover, as already noted, the two decompositions are the same in the case of a bivariate normal integral. However, for $M > 2$, they substantially differ from an estimation point of view. The Plackett-based decomposition is particularly useful when one evaluates multivariate normal probabilities with a given variance-covariance (correlation) matrix (Gassmann, 2003). If one needs to estimate the variance-covariance matrix as in any multivariate probit models or multivariate discrete choice models, then the effectiveness of the recursive formula is greatly reduced relative to our new decomposition formula (Eq. 2). In particular, the Plackett-based log-likelihood function becomes much more nonlinear that the one based on Proposition 2.[14] This, in turn, implies that the optimization procedure (or the determination of the zeroes of the score vector) leads to a higher computing time and generally a loss of precision. All in all, the computational complexity is considerably reduced when using Proposition 2. In the sequel, we only report Monte-Carlo results using our new decomposition (Eq. 2).[15]

2.2 Illustrative Example for $M = 3$

To conclude this section, we illustrate the application of Propositions 1 and 2 in the case of a trivariate cumulative density function. Applying Proposition 1 leads to the following unique decom-

[13]Gassmann (2003) tests the numerical properties of Plackett's method with respect to other numerical procedures (Gassmann, Deák, and Szántai, 2002). Monte-Carlo results show that the recursive Plackett-based method competes very favorably with the most accurate numerical methods and that it may be recommended when the problem's dimension does not exceed 10. A partial implementation of these formulae is also done by Drezner (1994) to calculate trivariate normal probabilities.

[14]See proof of Lemma 3.

[15]We compared the computing time and the accuracy of estimates with both methods. Our results clearly support the use of Proposition 2. Results are not reported in Section 4 but are available upon request.

position:

$$
\begin{aligned}
\Phi_3(z,\Omega) &= \Phi(z_1)\Phi(z_2)\Phi(z_3) \\
&+\Phi(z_3)\int_0^{\omega_{12}}\varphi_2(z_1,z_2,\lambda_{12})\,d\lambda_{12} \\
&+\Phi(z_2)\int_0^{\omega_{13}}\varphi_2(z_1,z_3,\lambda_{13})\,d\lambda_{13} \\
&+\Phi(z_1)\int_0^{\omega_{23}}\varphi_2(z_2,z_3,\lambda_{23})\,d\lambda_{23} \\
&+\int_0^{\omega_{12}}\int_0^{\omega_{13}}\frac{\partial\varphi_3(z,\lambda_{12},\lambda_{13},0)}{\partial z_1}d\lambda_{12}d\lambda_{13} \\
&+\int_0^{\omega_{12}}\int_0^{\omega_{23}}\frac{\partial\varphi_3(z,\lambda_{12},0,\lambda_{23})}{\partial z_2}d\lambda_{12}d\lambda_{23} \\
&+\int_0^{\omega_{13}}\int_0^{\omega_{23}}\frac{\partial\varphi_3(z,0,\lambda_{13},\lambda_{23})}{\partial z_3}d\lambda_{13}d\lambda_{23} \\
&+\int_0^{\omega_{12}}\int_0^{\omega_{13}}\int_0^{\omega_{23}}\frac{\partial^3\varphi_3(z,\Lambda)}{\partial z_1\partial z_2\partial z_3}d\lambda_{12}d\lambda_{13}d\lambda_{23}.
\end{aligned}
$$

Using now Proposition 2, the last integral may be decomposed in a non-unique way as follows:

$$
\begin{aligned}
&\int_0^{\omega_{12}}\int_0^{\omega_{13}}\int_0^{\omega_{23}}\frac{\partial^3\varphi_3(z,\Lambda)}{\partial z_1\partial z_2\partial z_3}d\lambda_{12}d\lambda_{13}d\lambda_{23} \\
&\qquad= \int_0^{\omega_{13}}\int_0^{\omega_{23}}\frac{\partial\varphi_3(z,\omega_{12},\lambda_{13},\lambda_{23})}{\partial z_3}d\lambda_{13}d\lambda_{23} \\
&\qquad\quad -\int_0^{\omega_{13}}\int_0^{\omega_{23}}\frac{\partial\varphi_3(z,0,\lambda_{13},\lambda_{23})}{\partial z_3}d\lambda_{13}d\lambda_{23} \\
&\qquad= \int_0^{\omega_{12}}\int_0^{\omega_{23}}\frac{\partial\varphi_3(z,\lambda_{12},\omega_{13},\lambda_{23})}{\partial z_2}d\lambda_{12}d\lambda_{23} \\
&\qquad\quad -\int_0^{\omega_{12}}\int_0^{\omega_{23}}\frac{\partial\varphi_3(z,\lambda_{12},0,\lambda_{23})}{\partial z_2}d\lambda_{12}d\lambda_{23} \\
&\qquad= \int_0^{\omega_{12}}\int_0^{\omega_{13}}\frac{\partial\varphi_3(z,\lambda_{12},\lambda_{13},\omega_{23})}{\partial z_1}d\lambda_{12}d\lambda_{13} \\
&\qquad\quad -\int_0^{\omega_{12}}\int_0^{\omega_{13}}\frac{\partial\varphi_3(z,\lambda_{12},\lambda_{13},0)}{\partial z_1}d\lambda_{12}d\lambda_{13}.
\end{aligned}
$$

Therefore, the cumulative distribution function may be written as:

$$
\begin{aligned}
\Phi_3(z,\Omega) &= \Phi(z_1)\Phi(z_2)\Phi(z_3) \\
&+\Phi(z_3)\Psi_2(z_1,z_2,\omega_{12}) \\
&+\Phi(z_2)\Psi_2(z_1,z_3,\omega_{13}) \\
&+\Phi(z_1)\Psi_2(z_2,z_3,\omega_{23}) \\
&+\Psi_3(z_1,z_2,z_3,\omega_{12},\omega_{13},0) \\
&+\Psi_3(z_2,z_3,z_1,\omega_{23},\omega_{12},0) \\
&+\Psi_3(z_3,z_1,z_2,\omega_{13},\omega_{23},\omega_{12}).
\end{aligned}
$$

Again this is only one of the three possible specifications due to the non-unique decomposition of the three-dimension integral exposed above.

3 Application to the Multivariate Probit Model

In this section, we first present for notational purpose the multivariate probit model and the underlying assumptions. We then derive the exact likelihood function based on Proposition 2 and explain how the numerical evaluation may be carried out. Finally, some estimation issues are discussed.

3.1 The Model

The multivariate probit model, for observation i and equation m, is:

$$\begin{aligned} y^*_{im} &= x'_{im}\beta_m + \epsilon_{im} \\ y_{im} &= \mathbb{I}\left(y^*_{im} > \tau_m\right) \end{aligned} \tag{4}$$

where $i = 1, \cdots, N$, $m = 1, \cdots, M$, y_{im} equals 1 if $y^*_{im} > \tau_m$ and 0 otherwise, x_{im} is an $k_m \times 1$ vector of covariates, $\beta_m \in \mathbb{R}^{k_m}$ is the vector of parameters, τ_m is the cut-off point or threshold of the m^{th} response variable, and ϵ_{im} is the error term. Without loss of generality, we assume that $\tau_m = 0$, for all m.

The data consist of N observations on $(y_i, x_i)_{i=1,\cdots,N}$ where $y_i = (y_{i1}, \cdots, y_{iM})'$ denotes the collection of responses on all M variables, and $x_i = \text{diag}\left(x'_{i1}, \cdots, x'_{iM}\right)$ is a $M \times K$ matrix, where $K = \sum_{m=1}^{M} k_m$.[16] After stacking all observations, we denote $y = (y'_1, \cdots, y'_N)'$, $X = (x_1; \cdots; x_N) \in \mathcal{M}_{NM\times K}$, and $\beta = (\beta'_1, \cdots, \beta'_M)'$.
The vector of disturbances $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{iM})'$ is M-variate normally distributed with an $M \times M$ symmetric positive definite covariance matrix Ω_i:

$$\Omega_i = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1M}\sigma_1\sigma_M \\ \rho_{12}\sigma_1\sigma_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{M-1,M}\sigma_{M-1}\sigma_M \\ \rho_{1M}\sigma_1\sigma_M & \cdots & \rho_{M-1,M}\sigma_{M-1}\sigma_M & \sigma_M^2 \end{pmatrix}.$$

where all ρ_{kl} represent the correlation coefficients and thus belong to $(-1; 1)$. Note that no specific form is imposed on the variance-covariance matrix. We further assume that ϵ_i are i.i.d., so that $\Omega_i = \Omega$, for all i, and the variance-covariance matrix of $\epsilon = (\epsilon'_1, \cdots, \epsilon'_N)' \in \mathbb{R}^{NM}$ is defined by:

$$V(\epsilon) \equiv \Sigma = I_N \otimes \Omega$$

where I_N is the identity matrix of order N.
The data on x_{im}, $m = 1, \cdots, M$, are assumed throughout to be strictly exogenous, which implies that $\text{cov}(x'_{im}, \epsilon_{js}) = 0$ across all observations i and j and all response variables m and s. This assumption rules out, for example, the presence of y_{is} $(s \neq m)$ in Eq. (4).

Given these assumptions, a convenient representation of the multivariate probit model is in terms of

[16] For the sake of simplicity and to avoid supplementary notation, we assume that there is no common regressor(s). Our results also apply, however, in this context, if using the same notation, $x_{im} = s_m z_i$, with z_i being a row vector of all exogenous regressors and s_m being a selection matrix. See Section 4.

latent variables y_{im}^*. Let $y_i^* = (y_{i1}^*, \cdots, y_{iM}^*)'$ denote the M-vector of latent variables for observation i. Then y_i^* has the following multivariate normal distribution:

$$y_i^* \sim \mathcal{N}_M(x_i\beta, \Omega)$$

where $\mathcal{N}_M(.,)$ stands for the M-variate normal distribution. It turns out that $y^* = (y_1^{*\prime}, \cdots, y_N^{*\prime})'$ is NM-variate normally distributed with mean being the $NM \times 1$ vector, $X\beta$, and variance-covariance matrix being the $NM \times NM$ positive definite matrix, Σ.

As in the univariate case, there is an identification problem associated with the variance-covariance matrix of ϵ—the slope and covariance parameters are not likelihood identified. Since ordinal data are invariant under monotonic transformations of y_i^* (or y^*), the variance-covariance matrix Ω (and thus Σ) may be estimated only up to scaling constants. Therefore, for identifiability purposes, we have to restrict the variance-covariance matrix Ω to be a correlation matrix and to redefine the vector β. Using $C = \text{diag}(\sigma_{11}^{-1}, \cdots, \sigma_{MM}^{-1})$, an identifiable (symmetric) positive semi-definite correlation matrix $\widetilde{\Omega}$ is defined by $\widetilde{\Omega} = C\Omega C'$ in which the main diagonal elements equal 1 and off-diagonal elements equal ρ_{kl} $(= \rho_{lk})$. The correlation matrix is thus defined by $\widetilde{\Sigma} = I_n \otimes \widetilde{\Omega}$. A set of identified slope parameters is thus given by $\widetilde{\beta}_m = \sigma_{mm}^{-1}\beta_m$, $m = 1, \cdots, M$. The parameters of the identified model thus consist of the K parameters of $\widetilde{\beta} = \left(\widetilde{\beta}_1', \cdots, \widetilde{\beta}_M'\right)'$ and the $M(M-1)/2$ parameters $\rho = (\rho_{12}, \cdots, \rho_{1M}, \cdots, \rho_{M-1,M})'$. This, in turn, means that the re-normalized latent vector $z^* = (I_N \otimes C)(y^* - X\beta)$ follows an M-variate *standardized* normal distribution with a symmetrical positive semi-definite correlation matrix $\widetilde{\Sigma}$.

As a final remark, note that we may obtain a panel probit model by assuming that the parameter vectors are identical across response equations, which implies that the set of explanatory variables is common across m. In this respect, by a simple change of notation, $t = m$, Eq. (4) may be interpreted as a panel probit model. In that case, normalization of all diagonal elements of the variance-covariance matrix, Ω, is unnecessary, precisely because the slope vector is invariant across m. Only one main diagonal element of Ω, say σ_{11}^2, is normalized to one for identification purpose.

3.2 The Maximum Likelihood Function

Partially recycling notation, let β denote the vector of identified parameters, $c_i = x_i\beta$, and Ω denote the correlation matrix (e.g., after imposing the identifying restrictions). Under the usual regularity conditions, which we assume to hold throughout the paper, the likelihood function is the joint density for observed outcomes[17]:

$$L(y \mid X; \beta, \Omega) = \prod_{i=1}^{N} L_i(y_i \mid x_i; \beta, \Omega) \tag{5}$$

where the likelihood of observation i, L_i, is given by Lemma 4. The following result is well known in the literature and is reported here as a practical matter (see Greene, 2002).

[17] For a discussion of the existence and uniqueness of the maximum likelihood estimator in the case of a multivariate probit model, see Lesaffre and Kaufmann (1992).

Lemma 4 *The likelihood of observation i is the cumulative density function, evaluated at the vector $W_i c_i$, of an M-variate standardized normal vector with correlation matrix $W_i \Omega W_i$,*

$$L_i\left(y_i \mid x_i; \beta, \Omega\right) = \Phi_{M,\mathcal{E}_i}\left(W_i c_i, W_i \Omega W_i\right) \tag{6}$$

where W_i is a diagonal matrix whose main diagonal elements equal $w_{im} = 2y_{im} - 1$ and depend on the sign of y^*_{im}. $\square$

Proof: See Appendix 5.

The full-information maximum likelihood estimates are obtained by maximizing the log-likelihood

$$\ln L\left(y \mid X; \beta, \Omega\right) = \sum_{i=1}^{N} \ln \Phi_{M,\mathcal{E}_i}\left(W_i c_i, W_i \Omega W_i\right)$$

with respect to β and Ω. Using the dimensional reduction formula of normal multivariate integrals (Proposition 2), this may be rewritten as follows.

Corollary 1 *The exact log-likelihood function is given by:*

$$\ln L\left(y \mid X; \beta, \Omega\right) = \sum_{i=1}^{N} \ln \left[\prod_{m=1}^{M} \Phi\left(w_{im} c_{im}\right) + \sum_{r} \int_{\overline{\Omega}_r} \psi_{Mr}\left(W_i c_i, \widetilde{\Lambda}_r\left(W_i \Omega W_i\right)\right) d\widetilde{\Lambda}_r\right]. \tag{7}$$

$\square$

Neglecting the second term of the sum in the previous expression, one may interpret, $\ln\left[\prod_{m=1}^{M} \Phi\left(w_{im} c_{im}\right)\right]$, as being the exact likelihood of y_i in a multivariate probit, in which we assume that the error terms are independent across m. The correlation matrix is thus the identity matrix of order M (normalization assumption). Consequently, if the model is correctly specified (with such an assumption), the second right-hand side term of the likelihood of observation i may cancel each other out and thus the probability correction term tends toward zero. In contrast, if the identity correlation matrix is far from the target matrix, the reference probability will underestimate or overestimate the target, placing greater emphasis on the probability correction term.

3.3 Numerical Computation and Convergence

Using Corollary 1, we can directly maximize the log-likelihood function with respect to the parameters of interest. Alternatively, we can first derive the first-order conditions with respect to the slope and correlation parameters and then apply our proposed method to obtain the exact score vector with the two methods being equivalent.[18] In both methods, we need to evaluate numerically the finite-range multiple integrals in Proposition 2. To do so, we use a Gauss-Legendre quadrature rule (Golub and Welsch, 1969; Davis and Rabinowitz, 1981; Press et al., 1992) over bounded intervals

[18]They may differ numerically, especially when the objective function is not sufficiently smooth. This may happen when there is a large number of values near one or zero. Note also that the approximation error of the Gauss-Legendre quadrature rule may be different.

for each term of the sum in the exact first-order conditions or the exact log-likelihood function.[19] Consequently, the exact maximum likelihood estimator is defined as follows.

Definition 1 *Let* $\Upsilon(m, r; c_i, \Omega)$ *denote the m-node Gauss-Legendre quadrature rule of* $\int_{\overline{\Omega}_r} \psi_{Mr}\left(W_i c_i, \widetilde{\Lambda}_r\left(W_i \Omega W_i\right)\right)$. *The exact maximum likelihood estimator of* $\theta = (\beta', \Omega)'$ *is defined by:*

$$\hat{\theta}_{EML} = \underset{\theta \in \Theta}{argmax} \sum_{i=1}^{N} \ln \left[\prod_{m=1}^{M} \Phi\left(w_{im} c_{im}\right) + \sum_{r} \Upsilon(m, r; c_i, \Omega) \right]. \tag{8}$$

□

Under the regularity conditions of Lesaffre and Kaufman (1992), the exact maximum likelihood estimator exists and is unique. Moreover, $\hat{\theta}_{EML}$ converges to the true value, θ_0, and it is asymptotically normally distributed.

There is generally a nonvanishing error of the Gauss-Legendre quadrature, which depends on the number of nodes, m, and the integrand. In particular, more nodes do not necessarily reduce the approximation error of the quadrature as a high order of approximation does not necessarily imply high accuracy, unless the integrand is smooth (Press et al., 1992; Stroud and Secrest, 1996). In this case, accuracy may be evaluated by specifying error criteria as the absolute or relative difference between an m-point and a q-point Gauss-Legendre quadrature. On the other hand, given that the integrand is an analytic function in the interior of the parameter space of the multivariate probit model, one can derive an upper bound on the error estimates for the Gauss-Legendre quadrature. Indeed, using the results of Chawla, Jain (1968a, b) and Kambo (1970), error estimates are bounded and the Gauss-Legendre quadrature converges to the true unknown integral.

3.4 Discussion

There are three remaining issues that concern, respectively, the positive definiteness of each Ω_{pq}, the singularity of the correlation matrix, and the number of elements in the sum to compute (e.g., the dimension problem). On one hand, one may still impose restrictions so that the variance-covariance matrices are positive definite.[20] On the other hand, if the correlation matrix is singular (e.g., there is equicorrelation), some linear transformation of the Z-vector exists such that the transformed correlation matrix is non-singular.[21] Therefore, the dimension of the normal cumulative density function is reduced by the number of equicorrelated error terms, r, but it may still be expressed as in Proposition 2. Moreover, although the reduction of the dimension of the normal cumulative density

[19] The standard m-point Gauss-Legendre quadrature rule over a bounded arbitrary interval, (a, b), is given by:

$$\int_a^b f(x)\,dx = \frac{b-a}{2} \sum_{i=1}^{m} w_i f(y_i) + R_m$$

where $y_i = \frac{b-a}{2} x_i + \frac{b+a}{2}$, the nodes x_i are zeros of the Legendre polynomial $P_m(x)$, all w_i represent the corresponding weights, $w_i = \frac{2}{\left(1-x_i^2\right)\left(P'_m(x_i)\right)^2}$, and R_m is the error term, $R_m = Q_m f^{(2m)}(\xi) = \frac{(b-a)^{2m+1}(m!)^4}{(2m+1)(2m!)^3} f^{(2m)}(\xi)$ with $\xi \in (a, b)$.

[20] For an example, see the technical report.

[21] Assume that the rank of Ω equals $r < M$. For some $r \times M$ matrix G, $\widetilde{Z} = GZ$ has a non-singular distribution and the correlation matrix $G \Omega G'$ is non-singular.

function leads to a (slightly) faster numerical evaluation, the decomposition in terms of correlation parameters presents two main advantages: (i) it is continuously differentiable and (ii) the singular multivariate normal cumulative density function may be exactly decomposed as the limiting case of a regular probability.[22] Finally, it is worth emphasizing that the curse of dimensionality may still be an issue. The decomposition in Proposition 2 may be applied in theory to any M. However, from a practical view, this leads to the evaluation of more and more elements in the sum (Proposition 2), which in turn increases the computation time. So far, we have estimated multivariate probit models up to $M = 6$ (empirical applications in health, labor or education economics generally do not exceed $M = 4$), and the computing time of our procedure outperforms the one of the maximum simulated likelihood, while preserving the consistency and efficiency properties. We further discuss this issue in Section 4.

4 Simulations

In this section, we report some Monte Carlo simulations.[23] Sample data for our experiments are alternatively generated by a bivariate, a trivariate, and a quadrivariate probit model. Results of our exact maximum likelihood estimator are compared with those of the maximum simulated likelihood estimator (henceforth MSL).

To investigate the small and large sample properties, each model is estimated using $N = 1,000$ and $N = 10,000$ observations.[24] All results reported below are based on 1,000 simulation repetitions, except for the quadrivariate probit model where 500 repetitions were run due to the computation time. For each estimator, we report (1) the mean bias and the Root Mean Squared Error (RMSE) of the parameters of interest (e.g. the slope and scale parameters as well as their standard deviation) and (2) the average computation time.[25]

Before presenting our main Monte Carlo results, one issue deserves some comment: the number of draws (respectively nodes) *per* simulation for the MSL (respectively the EML) estimator. On one hand, as is well known, an unbiased simulator of the likelihood function is neither necessary nor sufficient for consistent maximum simulated log-likelihood estimation since the latter estimator is obtained as a non-linear function (through optimization) of the simulator. Consequently, while unbiased simulation of the likelihood is generally straightforward, unbiased simulation of the log-likelihood is generally infeasible.[26] In that respect, a critical issue is the selection of the number of draws to obtain a negligible level of asymptotic efficiency loss due to simulations. In particular,

[22] The technical report provides an example in the case of a bivariate probit model.

[23] All experiments were performed on a personal computer with a Pentium P4/2.8GHz processor and 1 G of memory. The software package was Gauss 6.0 with library Maxlik 5.0. The simulated maximum likelihood estimator was also computed using Stata 9.2. (Cappellari and Jenkins, 2003) and Limdep 8.0—results were comparable with our Gauss codes (Huguenin, 2008) both in terms of efficiency and computing time.

[24] We also conducted Monte Carlo simulations in which the number of observations equals 500. Results are qualitatively similar to those reported here. Results are available upon request.

[25] We also calculated the median bias and the Median Absolute Deviation (MAD). Results are not reported here but are available upon request.

[26] For further details, see Hajivassiliou (2000) and Train (2003).

the MSL estimator is consistent, efficient, and asymptotically equivalent to the ML estimator if the number of draws, R, and the number of observations, N, goes to infinity and R grows without bound faster than $\sqrt{N}$ (i.e., $\sqrt{N}/R \to 0$).[27] In contrast, for any (finite) R, the MSL estimator is inconsistent, whereas, if R increases more slowly than $\sqrt{N}$, the MSL estimator is consistent but not asymptotically normal.[28] All in all, this means that the number of draws may increase, and thus the computing time may become great before consistent and efficient estimates may be obtained.[29] To partially circumvent this issue, we use the standard rule of thumb, $R = \sqrt{N}$, as our benchmark case, especially in large samples.[30] Since the consistency and efficiency of parameters in small samples may require a larger number of draws, we generally report evidence by setting $R = 100$. The case $R = 10$ is also displayed in order to compare the computing time and the statistical properties of the MSL estimator with those obtained with a larger number of draws. In all cases, we implement the GHK smooth recursive conditional simulator due to Geweke (1991), Hajivassiliou and McFadden (1990), and Keane (1990, 1994).[31] On the other hand, we test the sensitivity of our EML estimator by using a different number of nodes in the Gauss-Legendre quadrature rule. Results are reported here for 6 nodes.[32]

4.1 Bivariate Probit Models

We consider the following structural bivariate probit model:

$$y_{ik} = \mathbb{1}\left(y_{ik}^* > 0\right) \text{ for } k = 1, \cdots, 2$$

where:

$$y_{ik}^* = x_{ik}'\beta_k + u_{ik}$$

with $x_{ik} = (1, x_{ik,0}, x_{ik,k})'$, and:

$$\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} \sim \mathcal{N}(0, \Sigma), \ \Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}.$$

The structural parameter vector is thus given by $\theta = (\beta_1', \beta_2', \gamma_{21}, \alpha_{21}, \sigma_{12})'$, where $\beta_1 = (\beta_{1c}, \beta_{10}, \beta_{11})'$ and $\beta_2 = (\beta_{2c}, \beta_{20}, \beta_{22})'$.

[27] The MSL estimator is actually consistent under the weaker condition that R, $N \to \infty$.

[28] For further details, see Lee (1999).

[29] Börsch-Supan and Hajivassiliou (1993) show that the bias caused by finite R is small for moderately sized R as long as better smoothed probability simulators are used. In that respect, Hajivassiliou (2000) proposes using a diagnostic test to assess the simulation bias and thus choose a lower bound for the number of replications. In contrast, Gourieroux and Monfort (1991) propose a bias-corrected estimator. Lee (1995, 1997) also considers bias correction in panel choice regression and Markov models. The usefulness of the bias-reduction technique is, however, case dependent and thus may not always hold.

[30] We also use the test proposed by Hajivasilliou (2000). However, given the number of replications, it was too time-consuming. Some experiments suggested that the number of draws inferred by this test was larger than the one suggested by the standard rule of thumb in the literature. Results are available upon request.

[31] Briefly speaking, the GHK simulator switches back and forth between computing univariate, truncated normal probabilities conditional on previously drawn truncated normal random variables.

[32] Other results are available upon request.

Table 2 reports the simulation results. Several points are worth discussing. First, as $N = 1,000$, the results are roughly comparable for both methods. With the exception of the correlation parameter, the mean bias for each parameter of interest is almost the same. This result also holds for the standard deviation estimate of each slope parameter. In contrast, the mean bias and the RMSE of the correlation parameter, σ_{12}, and its standard deviation are generally higher than those of the slope parameters. In particular, the EML and MSL estimators of σ_{12} have similar properties as long as the number of draws is significantly larger than the one implied by the MSL rule of thumb, $R = \sqrt{N}$. To gain further intuition, we summarize this finding graphically. Figure 1 shows the density function of σ_{12} for the EML and MSL estimators (with 5, 10, 50, 100, and 200 draws).

[Insert Figures 1 and 2 around here]

As would be expected, the MSL estimator of σ_{12} has a significant bias and a large RMSE until R grows faster than $\sqrt{N}$. Moreover, while a fairly high level of accuracy is obtained as $R = 100$, the gain becomes marginal when the number of draws is further augmented.[33] The same pattern is observed for the standard deviation estimate of the correlation parameter (Figure 2).[34] One possible explanation of this behavior is that slope parameters enter less nonlinearly in the maximum likelihood function or the score vector. Consequently, the corresponding simulator bias is smaller than the one for the correlation parameter. On the other hand, while the accuracy of the probability simulator might be sufficient for locating a relative maximum, it may not be sufficiently accurate for calculating the Hessian matrix (using finite differences or score-based methods) and thus the parameter variance estimates.[35]

Second, the mean bias and the RMSE of the slope parameters are further reduced when the number of observations increases ($N = 10,000$). Regarding the correlation parameter (and its standard deviation), we observe that the EML mean bias (RMSE) substantially decreases whereas the MSL estimator is significantly downward biased and less efficient when the number of draws is either too small (10 draws) or is fixed with the rule of thumb (100 draws). This suggests that small-sample bias reduction might be achieved at the cost of increasing the number of replications, and thus the computing time.[36] Third, the average computing time (in CPU time in minutes per replication) provides evidence that our procedure competes extremely favorably with respect to the MSL estimator. Indeed, the computing time of the EML method increases less than proportionally to the number of observations, whereas it is proportional to N (and the number of draws) in the MSL estimator.[37]

[33] Breslaw (1994) provides evidence that beyond a certain level, the increase of computing time necessary to achieve a desired level of accuracy may become unacceptable.

[34] Lee (1997) examines the class of panel discrete choice regressions and Markov models and reports that the MSL method performs better for models with moderate serial correlation than for models with high serial correlations. McCulloch (1997) and Jank and Booth (2003) outline that the variance-covariance matrix of a simple logit-normal model is poorly estimated by the MSL estimator.

[35] The accuracy of the probability simulator may be improved by using alternative sampling methods (Sándor and András, 2004).

[36] Increasing the number of draws to 200 still leads to a substantial downward mean bias and yields a larger RMSE than the EML estimator.

[37] Our simulation results also show that the computing time is roughly proportional to the number of nodes.

[Insert Table 2 around here]

In a companion paper, we explore the robustness of our results using a simultaneous bivariate probit equations model in which we subsequently introduce the endogenous variable y_{i1}, the latent variable y_{i1}^*, and both variables in the specification of the second equation. All in all, the results available in the technical report show that the EML estimator displays better finite sample statistical properties than the MSL estimator for the slope and scale parameters as well as for the standard deviation estimates.

4.2 Trivariate and Quadrivariate Probit Models

We now discuss the simulation results for both a trivariate and a quadrivariate probit model. The general specification is given by:

$$y_{ik} = \mathbb{1}\left(y_{ik}^* > 0\right) \text{ for } k = 1, \cdots, 4$$

where:

$$y_{ik}^* = x_{ik}\beta_k + u_{ik}$$

with $x_{ik} = (1, x_{ik,0}, x_{ik,k})$, $u_i \sim \mathcal{N}(0, \Sigma)$. The parameter vectors are respectively defined by $(\beta_1', \beta_2', \beta_3', \sigma_{12}, \sigma_{13}, \sigma_{23})'$ and $(\beta_1', \beta_2', \beta_3', \beta_4', \sigma_{12}, \sigma_{13}, \sigma_{14}, \sigma_{23}, \sigma_{24}, \sigma_{34})'$, with $\beta_k = (\beta_{kc}, \beta_{k0}, \beta_{kk})'$.

Since the interpretation is quite similar, we only report the results of the trivariate (respectively quadrivariate) probit model when $N = 1,000$ (resp. $N = 10,000$). Table 6 provides evidence for the trivariate probit model when the number of draws is either 10 or 35. As in the benchmark bivariate probit model, the two estimators are roughly similar in terms of mean bias and RMSE for the slope parameters. In contrast, we notice that results differ markedly according to the second-order parameters and the corresponding estimated standard deviations. More specifically, the EML estimator has better finite sample bias properties than the MSL estimator for σ_{12}, σ_{13}, and σ_{23} and their standard deviation estimates. In addition, the RMSE of these parameters tends to favor our estimator relative to the MSL estimator. Increasing the number of draws only slightly improves the consistency and efficiency properties at the expense of the computing time. Unreported simulations (with $R = 100$) provide evidence that the two estimators have comparable RMSE but our estimator still outperforms the MSL estimator in terms of mean bias. On the other hand, the number of nodes contributes marginally to reduce the mean bias while still allowing for a computing time far less than that of the MSL estimator. For instance, the computing time of the EML estimator with 12 nodes is similar to the one of the simulated maximum likelihood estimator, with $R = 10$.

[Insert Table 3 around here]

We now turn to the simulation results of a quadrivariate probit model with $N = 10,000$. Table 7 confirms that the EML estimator shares the same large sample properties as the MSL estimator (with $R = 100$) for the slope parameters whereas it generally outperforms it for the scale parameters in terms of both mean bias and RMSE. At the same time, while the mean computing time per simulation repetition is around four minutes with our method, it is nearly three hours with the maximum simulated likelihood method! In that respect, our method is not only accurate but also extremely time-efficient.

[Insert Table 4 around here]

To sum up, Monte Carlo results show that the finite and large sample properties of our method are very competitive in terms of both bias and RMSE with respect to the MSL method. These properties were obtained with a few nodes in the Gauss-Legendre quadrature rule. In particular, augmenting the number of nodes marginally improves the results.[38] Moreover the computing time of our method is fast and by no means comparable to the one of the simulated maximum likelihood, especially as the dimension of the multivariate probit model increases. Finally, our method outperforms the stated benchmark in the presence of endogenous regressors. Such situations are often encountered in applied economics and thus deserve particular attention.

5 Conclusion

The estimation of the multivariate probit models has historically been dealt with mainly by useful and increasingly complex limited-information techniques. Although its importance hardly needs to be stressed, estimation by full-information approaches is seriously hindered by the obstacle of the numerical evaluation of the multivariate normal cumulative distribution function. Several estimation procedures have recently been developed using techniques based on simulations.

In contrast to limited-information or simulation-based approaches, this paper proposes a full-information estimation procedure based on exact analytical maximum likelihood. It offers the possibility of overcoming the difficulty of the numerical evaluation of the multiple integrals involved by using a new decomposition of the multivariate normal cumulative distribution function that yields lesser dimension small-range finite multiple integrals. The practicality of the approach has been widely tested with Monte Carlo simulations. The results strongly support using this approach, as it has been demonstrated to be both highly accurate and computationally time-efficient, especially with respect to the maximum simulated likelihood method.

In companion papers, we show how to extend our methodology to the general class of simultaneous equation models with latent variables and how to improve the numerical procedure of Butler and Moffitt (1992) for panel data models.

[38] It is worth recalling that we use the standard Gauss-Legendre quadrature rule. However, numerical improvements have been proposed in the literature. The sophistication of the Gauss-Legendre quadrature is in the strategic selection of locations at which the function is to be evaluated. For instance, Babolian et al. (2005) proposes estimating numerical values of nodes and weights so that the absolute error of the Gauss-Legendre quadrature is less than a preassigned tolerance. We leave this issue for future research.

References

[1] Ashford, J. R. and Sowden, R. R. (1970), Multi-Variate Probit Analysis , *Biometrics*, vol. 26, pp. 535–546.

[2] Babolian, E., MasjedJamei, M. and Eslahchi, M.R. (2005), On Numerical Improvement of Gauss-Legendre Quadrature Rules, *Applied Mathematics and Computation*, vol. 160, pp. 779–789.

[3] Bertschek, I and Lechner, M. (1998), Convenient Estimators for the Panel Probit Model , *Journal of Econometrics*, vol. 87(2), pp. 329–372.

[4] Bock, R. D. and Gibbons, R. D. (1996), High-Dimensional Multivariate Probit Analysis , *Biometrics*, vol. 52, pp. 1183–1194.

[5] Börsch-Supan, A. and Hajivassiliou, V.A. (1993), Smooth Unbiased Multivariate Probability Simulators for Maximum Likelhood Estimation of Limited Dependent Variable Models, *Journal of Econometrics*, vol. 58(3), pp. 347–368.

[6] Butler, J.S. and Moffitt, R. (1982), A Computationally Efficient Quadrature Procedure for the One-factor Multinomial Probit Model, *Econometrica*, vol. 50, pp. 761–764.

[7] Breslaw, J.A. (1994), Evaluation of Multivariate Normal Probability Integrals using a Low Variance Simulator, *The Review of Economics and Statistics*, vol. 76(4), pp. 673–682.

[8] Cappellari, L. and Jenkins, S. P. (2003), Multivariate Probit Regression Using Simulated Maximum Likelihood, *Stata Journal*, vol. 3, pp. 278–294.

[9] Chawla, M.M. and Jain, M.K. (1968a), Error Estimates for Gauss Quadrature Formulas for Analytic Functions, *Mathematics of Computation*, vol. 22(101), pp. 82–90.

[10] — (1968b), Asymptotic Error Estimates for the Gauss Quadrature Formula, *Mathematics of Computation*, vol. 22(101), pp. 91–97.

[11] Chib, S. and Greenberg, E. (1998), Analysis of Multivariate Probit Models, *Biometrika*, vol. 85(2), pp. 347–361.

[12] Craig, P. (2008), A New Reconstruction of Multivariate Orthant Probabilities, *Journal of the Royal Statistical Society Series B*, vol. 70, forthcoming.

[13] Davis, P.J. and Rabinowitz, P. (1984), *Methods of Numerical Integration*, Academic Press New York, 2nd edition.

[14] De Bruijn, N.G. (1981), *Asymptotic Methods in Analysis*, New York, Dower Publications, third edition.

[15] Drezner, Z. (1994), Computing the Trivariate Normal Integral, *Mathematics of Computation*, vol. 62, pp. 289–294.

[16] Gassmann, H.I. (2003), Multivariate Normal Probabilities: Implementing an Old Idea of Plackett's, *Journal of Computational and Graphical Statistics*, vol. 12(3), pp. 731–752.

[17] Gassmann, H.I., Deák, I. and T. Szántai, T. (2002), Computing Multivariate Normal Probabilities: A New Look, *Journal of Computational and Graphical Statistics*, vol. 11, pp. 920–949.

[18] 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*, pp. 571–578.

[19] Golub, G.H. and Welsch, J. (1967), Calculation of Gauss Quadrature Rules, *Mathematics of Computation*, vol. 23(106), pp. 221–230.

[20] Gourieroux, C. and Monfort, A. (1996), *Simulation-Based Econometric Methods*, CORE Lectures, Oxford University Press, Oxford.

[21] — (1991), Simulation Based Inference in Models with Heterogeneity, *Annales d'Economie et de Statistique*, vol. 20/21, pp. 69–107.

[22] Greene, W. H. (2002), *Econometric Analysis*, 5th ed., Prentice Hall, New Jersey.

[23] Hajivassiliou, V. (2000), Some Practical Issues in Maximum Simulated Likelihood, in *Simulation-based Inference in Econometrics*, edited by R. Mariano, T. Schuerman and M. Weeks, Cambridge University Press.

[24] Hajivassiliou, V. and McFadden, D. (1990), The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crises, Cowles Foundation Discussion Paper 967, Yale University.

[25] Hausman, J.A., and Wise, D.A. (1978), A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences, *Econometrica*, vol. 46(2), pp. 403–426.

[26] Heckman, J. (1981), Statistical Models for Discrete Panel Data, in *Structural Analysis of Discrete Data with Econometric Applications*, edited by Manski, C.F. and McFadden, D., Cambridge, MA: MIT Press, pp. 114–178.

[27] Heckman, J. and Willis, R.J. (1975), Estimation of a Stochastic Model of Reproduction: An Econometric Approach, in *Household Production and Consumption*, edited by Terleckyj, Cambridge University Press.

[28] Huguenin, J. (2004), Multivariate Normal Distribution and Simultaneous Equation Probit Analysis, PhD dissertation, University of Lausanne.

[29] —. (2008), A Gauss Library for the Exact Maximum Likelihood Estimator: An Application to Simultaneous Equations Probit Analysis, Working Paper, University of Lausanne, forthcoming.

[30] Jank, W. and Booth, J. (2003), Efficiency of Monte Carlo EM and Simulated Maximum Likelihood in Two-Stage Hierarchical Models, *Journal of Computational and Graphical Statistics*, vol. 12, pp. 214-229.

[31] Kambo, N.S. (1970), Error of teh Newton-Cotes and Gauss-Legendre Quadrature Formulas, *Mathematics of Computation*, vol. 24(110), pp. 261–269.

[32] Keane, M. (1990), Four Essays in Empirical Macro and Labor Economics, PhD dissertation, Brown University.

[33] Keane, M. (1994), A Computationally Practical Simulation Estimator for Panel Data, *Econometrica*, vol. 62(1), pp. 95–116.

[34] Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000) *Continous Multivariate Distributions, Volume 1: Models and Applications* (second edition), Wiley Series in Probability and Statistics.

[35] Lawrence, E., Bingham, D., Liu, C. and Nair, V.N. (2004), Bayesian Inference for Ordinal Data Using Multivariate Probit Models, Working Paper.

[36] Lazard Holly, H. and Holly, A. (2003), Computation of the Probability that a d–dimensional Normal Variable Belongs to a Polyhedral Cone with Arbitrary Vertex, *Mimeo*, University of Lausanne.

[37] Lee, L-F. (1995), Asymptotic Bias in Maximum Likelihood Estimation of Discret Choice Models, *Econometric Theory*, vol. 11, pp. 437-83.

[38] —. (1997), Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models: Some Monte Carlo Results, *Journal of Econometrics*, vol. 82, pp. 1-35.

[39] —. (1999), Statistical Inference with Simulated Likelihood Functions, *Econometric Theory*, vol. 15, pp. 337–360.

[40] Lerman, S. and Manski, C. (1981), On the Use of Simulated Frequencies to Approximate Choice Probabilities, in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MIT Press.

[41] Lesaffre, E. and Kaufmann, H. (1992), Existence and Uniqueness of the Maximum Likelihood Estimator for a Multivariate Probit Model, *Journal of the American Statistical Association*, vol. 87(419), pp. 805–811.

[42] McCulloch, C.E. (1997), Maximum Likelihood Algorithms for Generalized Linear Mixed Models, *Journal of the American Statistical Association*, vol. 92, pp. 162-70.

[43] McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, vol. 57, pp. 995–1026.

[44] Miwa, T., Hayter, A.J. and Kuriki, S. (2003), The Evaluation of General non-centred Orthant Probabilities, *Journal of the Royal Statistical Society Series B*, vol. 65, pp. 223-34.

[45] Ochi, Y. and Prentice, R.L. (1984), Likelihood Inference in a Correlated Probit Regression Model, *Biometrika*, vol. 71, pp. 531-43.

[46] Packes,A., and Pollard, D. (1989), Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, vol. 57, pp. 1027-57.

[47] Plackett, R. L. (1954), A Reduction Formula for Normal Multivariate Integrals , *Biometrika*, vol. 41, pp. 351–360.

[48] Press, W.H., Teukolsky, S.A., Vetterling, W.T., and B.P. Flannery (1992), *Numerical Recipes*, 2nd Edition, Cambridge University Press.

[49] Sándor, Z. and András, P. (2004), Alternative Sampling Methods for Estimating Multivariate Normal Probabilities, *Journal of Econometrics*, vol. 120, pp. 207–234.

[50] Sickles, R.C. and Taubman, P. (1986), An Analysis of the Health and Retirement Status of the Elderly, *Econometrica*, vol. 6, pp. 1339–1356.

[51] Stern, S. (1997), Simulation-Based Estimation, *Journal of Economic Literature*, vol. 35(4), pp. 2006–2039.

[52] Stroud, A.H., and Secrest, D. (1996), *Gaussian Quadrature Formulas*, Englewood Cliffs, NJ, Prentice Hall.

[53] Tierney, L. and Kadane, J.B. (1986), Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the American Statistical Association*, vol. 81, pp. 82–86.

[54] Tong, Y.L. (1990), *The Multivariate Normal Distribution*, New-York: Springer Verlag.

[55] Train, K.E. (2003), *Discrete Choice Models with Simulations*, Cambridge University Press.

Table 1: Number of elements according to Proposition 2

Dimension of the distribution	**1**	**2**	**3**	**4**	**5**	**6**
Highest dimension of the multiple integrals	0	1	2	3	4	5
Number of summed multiple integrals	1	2	7	26	111	472

Table 2: Bivariate probit Model (benchmark)

True values		EML			MSL (10 draws)			MSL (100 draws)		
		Estimates	Bias	RMSE	Estimates	Bias	RMSE	Estimates	Bias	RMSE
$N = 1,000$										
β_{1c}	1	1.0221	0.0221	0.1146	1.0238	0.0238	0.1140	1.0224	0.0224	0.1144
β_{10}	2	2.0329	0.0329	0.1488	2.0357	0.0357	0.1490	2.0333	0.0333	0.1487
β_{11}	-1	-1.0262	-0.0262	0.1124	-1.0273	-0.0273	0.1135	-1.0261	-0.0261	0.1123
β_{2c}	-0.5	-0.5122	-0.0122	0.0764	-0.5111	-0.0111	0.0781	-0.5120	-0.0120	0.0765
β_{20}	1	1.0158	0.0158	0.0826	1.0158	0.0158	0.0833	1.0152	0.0152	0.0827
β_{22}	-1	-1.0147	-0.0147	0.0803	-1.0144	-0.0144	0.0810	-1.0141	-0.0141	0.0800
σ_{12}	0.5	0.5255	0.0255	0.1655	0.3677	-0.1323	0.1936	0.5127	0.0127	0.1597
$\sigma_{\beta_{1c}}$	0.1062	0.1068	0.0006	0.0107	0.1075	0.0012	0.0108	0.1070	0.0008	0.0107
$\sigma_{\beta_{10}}$	0.1474	0.1487	0.0013	0.0197	0.1496	0.0022	0.0197	0.1489	0.0015	0.0197
$\sigma_{\beta_{11}}$	0.1122	0.1126	0.0003	0.0106	0.1132	0.0010	0.0106	0.1127	0.0004	0.0106
$\sigma_{\beta_{2c}}$	0.0792	0.0795	0.0003	0.0055	0.0798	0.0006	0.0056	0.0795	0.0004	0.0055
$\sigma_{\beta_{20}}$	0.0791	0.0789	-0.0002	0.0100	0.0791	0.0001	0.0101	0.0790	-0.0001	0.0101
$\sigma_{\beta_{22}}$	0.0727	0.0724	-0.0003	0.0099	0.0727	0.0000	0.0101	0.0725	-0.0002	0.0100
$\sigma_{\sigma_{12}}$	0.1599	0.1389	-0.0210	0.0333	0.1318	-0.0281	0.0380	0.1383	-0.0216	0.0336
Computing time			0.01			0.07			0.74	
$N = 10,000$										
β_{1c}	1	0.9996	-0.0004	0.0338	1.0008	0.0008	0.0339	0.9998	-0.0002	0.0339
β_{10}	2	2.0068	0.0068	0.0473	2.0092	0.0092	0.0481	2.0072	0.0072	0.0475
β_{11}	-1	-1.0010	-0.0010	0.0355	-1.0019	-0.0019	0.0360	-1.0011	-0.0011	0.0356
β_{2c}	-0.5	-0.5014	-0.0014	0.0258	-0.5011	-0.0011	0.0261	-0.5012	-0.0012	0.0257
β_{20}	1	1.0011	0.0011	0.0272	1.0005	0.0005	0.0271	1.0008	0.0008	0.0273
β_{21}	-1	-1.0002	-0.0002	0.0245	-0.9995	0.0005	0.0244	-0.9998	0.0002	0.0244
σ_{12}	0.5	0.5021	0.0021	0.0468	0.3407	-0.1593	0.1647	0.4811	-0.0189	0.0505
$\sigma_{\beta_{1c}}$	0.0325	0.0326	0.0001	0.0010	0.0327	0.0002	0.0010	0.0326	0.0001	0.0010
$\sigma_{\beta_{10}}$	0.0458	0.0460	0.0002	0.0020	0.0461	0.0003	0.0020	0.0460	0.0002	0.0020
$\sigma_{\beta_{11}}$	0.0354	0.0353	0.0000	0.0011	0.0355	0.0001	0.0011	0.0354	0.0000	0.0011
$\sigma_{\beta_{2c}}$	0.0251	0.0250	-0.0001	0.0005	0.0251	0.0000	0.0005	0.0250	0.0000	0.0005
$\sigma_{\beta_{20}}$	0.0242	0.0241	-0.0001	0.0009	0.0242	0.0000	0.0009	0.0241	-0.0001	0.0009
$\sigma_{\beta_{21}}$	0.0221	0.0220	-0.0001	0.0009	0.0220	0.0000	0.0009	0.0220	-0.0001	0.0009
$\sigma_{\sigma_{12}}$	0.0477	0.0473	-0.0004	0.0026	0.0421	-0.0056	0.0099	0.0466	-0.0011	0.0027
Computing time			0.11			0.94			8.68	

Note: The number of simulations is 1,000.

Table 3: Trivariate probit model

		EML			MSL (10 draws)			MSL (35 draws)		
True values		Estimates	Bias	RMSE	Estimates	Bias	RMSE	Estimates	Bias	RMSE
$N = 1,000$										
β_{1c}	1	1.0069	0.0069	0.0877	1.0077	0.0077	0.0881	1.0071	0.0071	0.0880
β_{10}	1	1.0196	0.0196	0.0718	1.0199	0.0199	0.0718	1.0199	0.0199	0.0720
β_{11}	-1	-1.0190	-0.0190	0.0962	-1.0196	-0.0196	0.0968	-1.0193	-0.0193	0.0963
β_{2c}	-0.5	-0.5056	-0.0056	0.0770	-0.5058	-0.0058	0.0771	-0.5057	-0.0057	0.0771
β_{20}	1	1.0300	0.0300	0.0966	1.0311	0.0311	0.0963	1.0304	0.0304	0.0966
β_{22}	-1	-1.0250	-0.0250	0.0851	-1.0255	-0.0255	0.0845	-1.0251	-0.0251	0.0849
β_{3c}	-1	-1.0148	-0.0148	0.0773	-1.0152	-0.0152	0.0773	-1.0145	-0.0145	0.0777
β_{30}	0.5	0.5069	0.0069	0.0367	0.5069	0.0069	0.0370	0.5068	0.0068	0.0369
β_{33}	2	2.0353	0.0353	0.1224	2.0362	0.0362	0.1233	2.0347	0.0347	0.1232
σ_{12}	0.2	0.2079	0.0079	0.1503	0.1223	-0.0777	0.1490	0.1915	-0.0085	0.1502
σ_{13}	-0.3	-0.2923	0.0077	0.0844	-0.1892	0.1108	0.1331	-0.2672	0.0328	0.0905
σ_{23}	0.1	0.0929	-0.0071	0.1218	0.0427	-0.0573	0.1002	0.0774	-0.0226	0.1160
$\sigma_{\beta_{1c}}$	0.0760	0.0764	0.0004	0.0060	0.0767	0.0007	0.0061	0.0766	0.0005	0.0060
$\sigma_{\beta_{10}}$	0.0632	0.0635	0.0003	0.0061	0.0637	0.0005	0.0061	0.0636	0.0004	0.0062
$\sigma_{\beta_{11}}$	0.0845	0.0851	0.0006	0.0055	0.0854	0.0009	0.0056	0.0852	0.0007	0.0055
$\sigma_{\beta_{2c}}$	0.0799	0.0799	0.0000	0.0047	0.0802	0.0003	0.0047	0.0800	0.0001	0.0047
$\sigma_{\beta_{20}}$	0.0787	0.0801	0.0014	0.0098	0.0805	0.0017	0.0099	0.0802	0.0015	0.0098
$\sigma_{\beta_{22}}$	0.0695	0.0701	0.0005	0.0080	0.0703	0.0008	0.0081	0.0701	0.0006	0.0080
$\sigma_{\beta_{3c}}$	0.0711	0.0718	0.0007	0.0037	0.0720	0.0009	0.0038	0.0719	0.0008	0.0037
$\sigma_{\beta_{30}}$	0.0351	0.0353	0.0002	0.0026	0.0354	0.0002	0.0027	0.0353	0.0002	0.0026
$\sigma_{\beta_{33}}$	0.1185	0.1198	0.0012	0.0063	0.1201	0.0016	0.0065	0.1199	0.0013	0.0064
$\sigma_{\sigma_{12}}$	0.1416	0.1356	-0.0061	0.0148	0.1154	-0.0262	0.0294	0.1295	-0.0121	0.0186
$\sigma_{\sigma_{13}}$	0.0927	0.0913	-0.0014	0.0060	0.0760	-0.0167	0.0174	0.0888	-0.0038	0.0075
$\sigma_{\sigma_{23}}$	0.1194	0.1154	-0.0040	0.0088	0.0863	-0.0331	0.0340	0.1089	-0.0105	0.0135
Computing time			0.04			0.17			1.15	

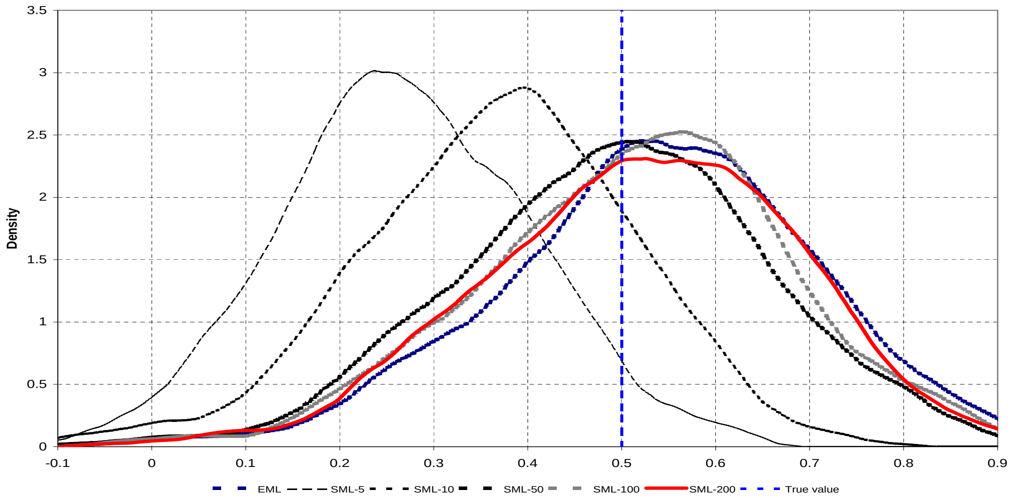
Note: The number of simulations is 1,000.

Table 4: Quadrivariate Probit

		EML			MSL (10 draws)			MSL (100 draws)		
True values		Estimates	Bias	RMSE	Estimates	Bias	RMSE	Estimates	Bias	RMSE
$N = 10,000$										
β_{1c}	1	1.0019	0.0019	0.0214	1.0021	0.0021	0.0214	1.0020	0.0020	0.0214
β_{10}	1	1.0007	0.0007	0.0187	1.0009	0.0009	0.0187	1.0007	0.0007	0.0187
β_{11}	-1	-0.9969	0.0031	0.0274	-0.9972	0.0028	0.0274	-0.9969	0.0031	0.0275
β_{2c}	-0.5	-0.4987	0.0013	0.0263	-0.4991	0.0009	0.0256	-0.4991	0.0009	0.0258
β_{20}	1	1.0039	0.0039	0.0295	1.0050	0.0050	0.0291	1.0045	0.0045	0.0293
β_{22}	-1	-1.0023	-0.0023	0.0262	-1.0035	-0.0035	0.0258	-1.0028	-0.0028	0.0260
β_{3c}	-1	-0.9990	0.0010	0.0192	-0.9993	0.0007	0.0195	-0.9988	0.0012	0.0193
β_{30}	0.5	0.5004	0.0004	0.0120	0.5005	0.0005	0.0121	0.5003	0.0003	0.0120
β_{33}	2	2.0038	0.0038	0.0342	2.0057	0.0057	0.0346	2.0035	0.0035	0.0339
β_{4c}	0.5	0.4996	-0.0004	0.0193	0.4990	-0.0010	0.0194	0.4992	-0.0008	0.0192
β_{40}	1	1.0000	0.0000	0.0165	0.9978	-0.0022	0.0166	0.9990	-0.0010	0.0168
β_{44}	-0.5	-0.5001	-0.0001	0.0223	-0.4982	0.0018	0.0226	-0.4991	0.0009	0.0226
σ_{12}	0.25	0.2444	-0.0056	0.0459	0.1469	-0.1031	0.1096	0.2324	-0.0176	0.0462
σ_{13}	-0.05	-0.0568	-0.0068	0.0292	-0.0321	0.0179	0.0290	-0.0517	-0.0017	0.0271
σ_{14}	0.1	0.0978	-0.0022	0.0299	0.0687	-0.0313	0.0403	0.0945	-0.0055	0.0271
σ_{23}	0.1	0.1029	0.0029	0.0390	0.0385	-0.0615	0.0685	0.0914	-0.0086	0.0341
σ_{24}	-0.3	-0.2956	0.0044	0.0436	-0.1528	0.1472	0.1507	-0.2756	0.0244	0.0486
σ_{34}	0.5	0.5035	0.0035	0.0245	0.3830	-0.1170	0.1188	0.4927	-0.0073	0.0224
$\sigma_{\beta_{1c}}$	0.0248	0.0246	-0.0001	0.0006	0.0247	-0.0001	0.0006	0.0247	-0.0001	0.0006
$\sigma_{\beta_{10}}$	0.0196	0.0196	0.0000	0.0006	0.0196	0.0000	0.0006	0.0196	0.0000	0.0006
$\sigma_{\beta_{11}}$	0.0273	0.0272	-0.0001	0.0007	0.0273	-0.0001	0.0007	0.0272	-0.0001	0.0007
$\sigma_{\beta_{2c}}$	0.0248	0.0248	0.0001	0.0006	0.0250	0.0002	0.0006	0.0249	0.0001	0.0006
$\sigma_{\beta_{20}}$	0.0240	0.0240	0.0000	0.0011	0.0241	0.0002	0.0011	0.0240	0.0001	0.0011
$\sigma_{\beta_{22}}$	0.0220	0.0221	0.0000	0.0011	0.0222	0.0002	0.0011	0.0221	0.0001	0.0010
$\sigma_{\beta_{3c}}$	0.0225	0.0226	0.0000	0.0003	0.0226	0.0001	0.0003	0.0226	0.0000	0.0003
$\sigma_{\beta_{30}}$	0.0107	0.0107	0.0000	0.0002	0.0107	0.0000	0.0002	0.0107	0.0000	0.0002
$\sigma_{\beta_{33}}$	0.0373	0.0372	0.0000	0.0006	0.0375	0.0003	0.0007	0.0373	0.0000	0.0006
$\sigma_{\beta_{4c}}$	0.0214	0.0213	-0.0001	0.0004	0.0215	0.0001	0.0003	0.0213	0.0000	0.0003
$\sigma_{\beta_{40}}$	0.0186	0.0185	-0.0001	0.0005	0.0186	0.0000	0.0005	0.0185	-0.0001	0.0005
$\sigma_{\beta_{44}}$	0.0197	0.0196	-0.0001	0.0005	0.0199	0.0002	0.0004	0.0197	0.0000	0.0004
$\sigma_{\sigma_{12}}$	0.0419	0.0423	0.0004	0.0030	0.0346	-0.0073	0.0074	0.0412	-0.0007	0.0015
$\sigma_{\sigma_{13}}$	0.0303	0.0302	-0.0002	0.0009	0.0245	-0.0059	0.0059	0.0298	-0.0005	0.0007
$\sigma_{\sigma_{14}}$	0.0289	0.0288	0.0000	0.0011	0.0261	-0.0028	0.0028	0.0288	0.0000	0.0005
$\sigma_{\sigma_{23}}$	0.0358	0.0358	0.0000	0.0011	0.0267	-0.0091	0.0091	0.0350	-0.0008	0.0011
$\sigma_{\sigma_{24}}$	0.0390	0.0391	0.0001	0.0014	0.0308	-0.0082	0.0083	0.0387	-0.0004	0.0015
$\sigma_{\sigma_{34}}$	0.0249	0.0247	-0.0002	0.0013	0.0232	-0.0017	0.0018	0.0247	-0.0002	0.0006
Computing time			3.35			8.81			159.17	

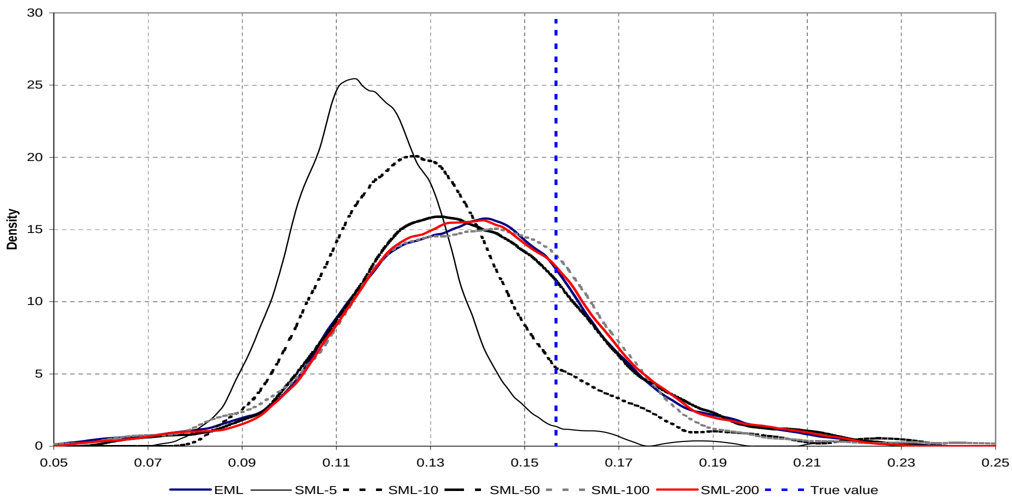
Note: The number of simulations is 500.

Figure 1: Density function of σ_{12} in the benchmark bivariate probit model



Note: The number of observations (resp. simulations) is equal to 1,000 (resp. 1,000).

Figure 2: Density function of $\sigma_{\sigma_{12}}$ in the benchmark bivariate probit model



Appendix 1: Proof of Lemma 2

To establish Lemma 2, we write $\varphi_{M,Z}$ as the transform of its characteristic function:

$$\varphi_{M,Z}(z,\Omega) = (2\pi)^{-M} \int_{\mathcal{Z}} \exp\left(-it'z - \frac{1}{2}t'\Omega t\right) \otimes dt$$

where $\otimes dt = dt_1 \times dt_2 \times \cdots \times dt_M$.

The first-order derivative of $\varphi_{M,Z}$ with respect to $z = (z_1, \cdots, z_M)'$ is given by:

$$\frac{\partial}{\partial z}\varphi_{M,Z}(z,\Omega) = -(2\pi)^{-M} \int_{\mathcal{Z}} it\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) \otimes dt.$$

We next differentiate this expression with respect to z':

$$\begin{aligned}
\frac{\partial^2}{\partial z \partial z'}\varphi_{M,Z}(z,\Omega) &= (2\pi)^{-M} \int_{\mathcal{Z}} i^2 tt'\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) \otimes dt \\
&= -(2\pi)^{-M} \int_{\mathcal{Z}} tt'\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) \otimes dt.
\end{aligned}$$

At the same time, we have:

$$\begin{aligned}
\frac{\partial}{\partial \Omega}\varphi_{M,Z}(z,\Omega) &= (2\pi)^{-M} \int_{\mathcal{Z}} \frac{\partial}{\partial \Omega}\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) \otimes dt \\
&= -(2\pi)^{-M} \int_{\mathcal{Z}} tt'\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) \otimes dt
\end{aligned}$$

since (using the differential operator and the symmetry property of the correlation matrix):

$$\begin{aligned}
d\left[\exp\left(-it'z - \frac{1}{2}t'\Omega t\right)\right] &= \exp\left(-it'z - \frac{1}{2}t'\Omega t\right) d\left(-it'z - \frac{1}{2}t'\Omega t\right) \\
&= -\frac{1}{2}\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) d\left(t'\Omega t\right) \\
&= -\frac{1}{2}\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) d\left(Tr\left(\Omega tt'\right)\right) \\
&= -\frac{1}{2}\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) Tr\left(d\left(\Omega tt'\right)\right) \\
&= -\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) Tr(tt'd\Omega)
\end{aligned}$$

and thus:

$$\frac{\partial}{\partial \Omega}\exp\left(-it'z - \frac{1}{2}t'\Omega t\right) = -tt'\exp\left(-it'z - \frac{1}{2}t'\Omega t\right).$$

This shows the first part of Lemma 2. Finally:

$$\begin{aligned}
\frac{\partial}{\partial \Omega}\Phi_{M,Z}(z,\Omega) &= \int \frac{\partial}{\partial \Omega}\varphi_{M,Z}(z,\Omega)dz = \int \frac{\partial^2}{\partial z \partial z'}\varphi_{M,Z}(z,\Omega)dz \\
&= \frac{\partial^2}{\partial z \partial z'}\Phi_{M,Z}(z,\Omega).
\end{aligned}$$

□

Appendix 2: Proof of Proposition 1

Integrating over the range determined by the n elements of $\widetilde{\Omega}$, the derivative of $\Phi_{M,Z}(z,\Lambda)$ with respect to all elements of $\widetilde{\Lambda}$ yields:

$$\int_{\Omega} \frac{\partial^n}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\lambda_{ij}} \Phi_{M,Z}(z,\Lambda)\, d\widetilde{\Lambda} = \sum_{p=0}^{n}\sum_{q=1}^{\binom{n}{p}} (-1)^p\, \Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right).$$

This may be rewritten as follows:

$$\int_{\Omega} \frac{\partial^n}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\lambda_{ij}} \Phi_{M,Z}(z,\Lambda)\, d\widetilde{\Lambda} = \Phi_{M,Z}(z,\Omega) + \sum_{p=1}^{n}\sum_{q=1}^{\binom{n}{p}} (-1)^p\, \Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right)$$

or, similarly, for any $\Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right)$ with $p = 1,\ldots,n-1$:

$$\int_{\widetilde{\Omega}_{pq}} \frac{\partial^{n-p}}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\widetilde{\lambda}_{ij}} \Phi_{M,Z}\left(z,\widetilde{\Lambda}_{pq}\right) d\widetilde{\Lambda}_{pq} = \Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right) + \sum_{r=p+1}^{n}\sum_{s=1}^{\binom{n}{r}} (-1)^{r-p}\, \Phi_{M,Z}\left(z,\widetilde{\Omega}_{rs}\right).$$

Summing up both sides of these two equations for all $p = 1,\ldots,n-1$ yields, as all $\Phi_{M,Z}\left(z,\widetilde{\Omega}_{pq}\right)$ but for $p = 0$ and $p = n$ cancel each other out on the right-hand side:

$$\int_{\Omega} \frac{\partial^n}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\lambda_{ij}} \Phi_{M,Z}(z,\Lambda)\, d\widetilde{\Lambda} + \sum_{p=1}^{n-1}\sum_{q=1}^{\binom{n}{p}} \int_{\widetilde{\Omega}_{pq}} \frac{\partial^{n-p}}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\widetilde{\lambda}_{ij}} \Phi_{M,Z}\left(z,\widetilde{\Lambda}_{pq}\right) d\widetilde{\Lambda}_{pq} =$$
$$\Phi_{M,Z}(z,\Omega) - \Phi_{M,Z}(z,0).$$

Since $\Phi_{M,Z}(z,0)$ is strictly factorizable, we can write:

$$\begin{aligned}\Phi_{M,Z}(z,\Omega) &= \prod_{k=1}^{M}\Phi(z_k) + \int_{\Omega} \frac{\partial^n}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\lambda_{ij}} \Phi_{M,Z}(z,\Lambda)\, d\widetilde{\Lambda} \\ &\quad + \sum_{p=1}^{n-1}\sum_{q=1}^{\binom{n}{p}} \int_{\widetilde{\Omega}_{pq}} \frac{\partial^{n-p}}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\widetilde{\lambda}_{ij}} \Phi_{M,Z}\left(z,\widetilde{\Lambda}_{pq}\right) d\widetilde{\Lambda}_{pq}.\end{aligned}$$

Given that:

$$\begin{aligned}\frac{\partial^n}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\lambda_{ij}} \Phi_{M,Z}(z,\Lambda) &= \frac{\partial^{M(M-1)}}{\prod_{k=1}^{M}\partial z_k^{M-1}} \Phi_{M,Z}(z,\Lambda) \\ &= \frac{\partial^{M(M-2)}}{\prod_{k=1}^{M}\partial z_k^{M-2}} \varphi_{M,Z}(z,\Lambda)\end{aligned}$$

the second term of the right-hand side expression of the decomposition is given by:

$$\int_{\Omega} \frac{\partial^n}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\lambda_{ij}} \Phi_{M,Z}(z,\Lambda)\, d\widetilde{\Lambda} = \int_{\Omega} \frac{\partial^{M(M-2)}}{\prod_{k=1}^{M}\partial z_k^{M-2}} \varphi_{M,Z}(z,\Lambda)\, d\widetilde{\Lambda}.$$

Similarly, factorizing the cumulative distribution function, whenever at least one $I_k\left(\widetilde{\Lambda}_{pq}\right)$ equals zero, leads to two interesting cases:

- If $\prod_{k=1}^{M} I_k\left(\widetilde{\Lambda}_{pq}\right) \neq 0$,

$$\frac{\partial^{n-p}}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\widetilde{\lambda}_{ij}}\Phi_{M,Z}\left(z,\widetilde{\Lambda}_{pq}\right) = \frac{\partial^{M(M-2)-2p}}{\prod_{k=1}^{M}\partial z_k^{I_k\left(\widetilde{\Lambda}_{pq}\right)-1}}\varphi_{M,Z}\left(z,\widetilde{\Lambda}_{pq}\right)$$

- If $I_k\left(\widetilde{\Lambda}_{pq}\right) = 0$,

$$\begin{aligned}\frac{\partial^{n-p}}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\widetilde{\lambda}_{ij}}\Phi_{M,Z}\left(z,\widetilde{\Lambda}_{pq}\right) &= \Phi\left(z_k\right)\frac{\partial^{M(M-1)-2p}}{\prod_{l=1}^{M}\partial z_l^{I_l\left(\widetilde{\Lambda}_{pq}\right)}}\Phi_{M-1,Z_{-k}}\left(z_{-k},\widetilde{\Lambda}_{-k,pq}\right)\\ &= \Phi\left(z_k\right)\frac{\partial^{(M-1)^2-2p}}{\prod_{\substack{l=1\\ l\neq k}}^{M}\partial z_l^{I_l\left(\widetilde{\Lambda}_{pq}\right)-1}}\varphi_{M-1,Z_{-k}}\left(z_{-k},\widetilde{\Lambda}_{-k,pq}\right)\end{aligned}$$

and so on if any other $I_k\left(\widetilde{\Lambda}_{pq}\right) = 0$, until all but one element of $\widetilde{\Lambda}$ are zero (or $p = n-1$). More specifically, if $\lambda_{ij} = 0$, $\forall\{i,j\} \neq \{k,l\}$ and $\lambda_{kl} \neq 0$, then $I_j\left(\widetilde{\Lambda}_{pq}\right) = 0$, $\forall j \neq \{k,l\}$ and $I_k\left(\widetilde{\Lambda}_{pq}\right) \neq 0$,$I_l\left(\widetilde{\Lambda}_{pq}\right) \neq 0$. Therefore, we obtain the following expression:

$$\begin{aligned}\frac{\partial^{n-p}}{\prod_{i=1}^{M-1}\prod_{j=i+1}^{M}\partial\widetilde{\lambda}_{ij}}\Phi_{M,Z}\left(z,\widetilde{\Lambda}_{pq}\right) &= \left[\prod_{\substack{j=1\\ j\neq\{k,l\}}}^{M}\Phi\left(z_j\right)\right]\frac{\partial^2}{\partial z_k\partial z_l}\Phi_{2,Z_{kl}}\left(z_k,z_l,\lambda_{kl}\right)\\ &= \varphi_{2,Z_{kl}}\left(z_k,z_l,\lambda_{kl}\right)\prod_{\substack{j=1\\ j\neq\{k,l\}}}^{M}\Phi\left(z_j\right)\end{aligned}$$

which completes the proof. The provision merely ensures that the elements in the sum are well defined. □

Appendix 3: Proof of Proposition 2

Starting from Proposition 1, we have:

$$\Phi_{M,Z}(z,\Omega) = \prod_{k=1}^{M} \Phi(z_k) + \sum_{p=0}^{n-1} \sum_{q=1}^{\binom{n}{p}} \int_{\widetilde{\Omega}_{pq}} \xi_{Mp}\left(z, \widetilde{\Lambda}_{pq}\right) d\widetilde{\Lambda}_{pq}.$$

We must show that the second right-hand term may be simplified. For notational simplicity, we consider the non-factorizable case:

$$\xi_{Mp}\left(z, \widetilde{\Lambda}_{pq}\right) = \frac{\partial^{M(M-2)-2p}}{\prod_{k=1}^{M} \partial z_k^{I_k\left(\widetilde{\Lambda}_{pq}\right)-1}} \varphi_{M,Z}\left(z, \widetilde{\Lambda}_{pq}\right)$$

where $\sum_{k=1}^{M} I_k\left(\widetilde{\Lambda}_{pq}\right)$ is always an even number.
If there is any pair $\{i,j\}$ with $i \neq j$ such that $I_i\left(\widetilde{\Lambda}_{pq}\right) > 1$ and $I_j\left(\widetilde{\Lambda}_{pq}\right) > 1$, then the previous expression may be written:

$$\begin{aligned}
\xi_{Mp}\left(z, \widetilde{\Lambda}_{pq}\right) &= \frac{\partial^{M(M-2)-2p}}{\partial z_i^{I_i\left(\widetilde{\Lambda}_{pq}\right)-2} \partial z_j^{I_j\left(\widetilde{\Lambda}_{pq}\right)-2} \prod_{\substack{k=1 \\ k\neq\{i,j\}}}^{M} \partial z_k^{I_k\left(\widetilde{\Lambda}_{pq}\right)-1}} \left[\frac{\partial^2}{\partial z_i \partial z_j} \varphi_{M,Z}\left(z, \widetilde{\Lambda}_{pq}\right)\right] \\
&= \frac{\partial^{M(M-2)-2p}}{\partial z_i^{I_i\left(\widetilde{\Lambda}_{pq}\right)-2} \partial z_j^{I_j\left(\widetilde{\Lambda}_{pq}\right)-2} \prod_{\substack{k=1 \\ k\neq\{i,j\}}}^{M} \partial z_k^{I_k\left(\widetilde{\Lambda}_{pq}\right)-1}} \left[\frac{\partial}{\partial \lambda_{ij}} \varphi_{M,Z}\left(z, \widetilde{\Lambda}_{pq}\right)\right].
\end{aligned}$$

The partial integration with respect to λ_{ij} is now given by:

$$\int_0^{\omega_{ij}} \frac{\partial}{\partial \lambda_{ij}} \varphi_{M,Z}\left(z, \widetilde{\Lambda}_{pq}\right) d\lambda_{ij} = \varphi_{M,Z}\left(z, \widetilde{\Lambda}_{pq}|_{\lambda_{ij}=\omega_{ij}}\right) - \varphi_{M,Z}\left(z, \widetilde{\Lambda}_{pq}|_{\lambda_{ij}=0}\right).$$

Thus the $(n-p)$-dimensional multiple integral is reduced to two $(n-p-1)$-dimensional multiple integrals.

The same procedure may be repeated until there is no such pair left, which is the case if and only if there is only one element ∂z_k, albeit elevated to some power, in the denominator of the derivative.

This procedure may, of course, be applied to the same extent to the factorizable case. Hence, there are three possible cases:

- If $p = 0, 1, \ldots, M$, then $\xi_{Mp}\left(z, \widetilde{\Lambda}_{pq}\right)$ may always be reduced to some multiple integrals of dimension $(M-1)$ or less.
- If $p = M-1, \ldots, n-3$, then only some of the $\xi_{Mp}\left(z, \widetilde{\Lambda}_{pq}\right)$ may be reduced to some lesser-dimensional multiple integrals.
- If $p = n-2$ or $p = n-1$, then $\xi_{Mp}\left(z, \widetilde{\Lambda}_{pq}\right)$ cannot be reduced.

The highest-dimensional resulting non-reducible multiple integrals are those with $\widetilde{\Lambda}_{M-1,q}$ containing the integrating variables $\{\lambda_{12}, \ldots, \lambda_{1M}\}, \{\lambda_{12}, \lambda_{23}, \ldots, \lambda_{2M}\}, \ldots, \{\lambda_{1M}, \ldots, \lambda_{M-1,M}\}$, and thus there are M such $(M-1)$-dimensional multiple integrals. □

Appendix 4: Proof of Lemma 3

Using the definition of the multivariate normal cumulative distribution function, and setting $\Sigma(s) = s\Omega + (1-s) I_M$ for any positive definite correlation matrix Ω and with $s \in [0,1]$, we have:

$$\Phi_{M,Z}(z, \Sigma(s)) = \int_{\mathcal{Z}} \varphi_{M,Z}(t, \Sigma(s)) \prod_{i=1}^{n} dt_i$$

where $\Sigma(s)$, with elements $\sigma_{ij}(s)$, is a positive definite correlation matrix.

Then, integrating the differential element $d\Phi_{M,Z}(z,\Sigma)$ along the line $I_M - \Omega$ reverts to integrating it over the range of s:

$$\int_0^1 \frac{\partial \Phi_{M,Z}(z, \Sigma(s))}{\partial s} ds = \Phi_{M,Z}(z, \Omega) - \Phi_{M,Z}(z, I_M)$$

with $\Phi_{M,Z}(z, I_M)$ being strictly factorizable, according to Lemma 1, into $\Phi_{M,Z}(z, I_M) = \prod_{k=1}^{M} \Phi(z_k)$.

Furthermore, the integrand of the left-hand side may be written:

$$\frac{\partial \Phi_{M,Z}(z, \Sigma(s))}{\partial s} = \sum_{i<j} \frac{\partial \sigma_{ij}(s)}{\partial s} \frac{\partial \Phi_{M,Z}(z, \Sigma(s))}{\partial \sigma_{ij}(s)}$$

where the derivative of the cdf with respect to any element of Σ is given by Lemma 1 and 2, such that:

$$\frac{\partial \Phi_{M,Z}(z, \Sigma)}{\partial \sigma_{ij}} = \varphi_{2,Z_{ij}}(z_{ij}, \Sigma_{ij}) \Phi_{M-2,Z_{-ij}}\left(\widetilde{z}_{-ij}(s), \widetilde{\Sigma}_{-ij}(s)\right)$$

with $\widetilde{z}_{-ij} = \Delta_{-ij}^{-1}\left(z_{-ij} - \Gamma_{ij}\Sigma_{ij}^{-1} z_{ij}\right)$ and $\widetilde{\Sigma}_{-ij} = \Delta_{-ij}^{-1}\left(\Sigma_{-ij} - \Gamma_{ij}\Sigma_{ij}^{-1}\Gamma'_{ij}\right)\Delta_{-ij}^{-1}$.
Moreover, it may easily be shown further that the elements of $\widetilde{z}_{-ij}$ and $\widetilde{\Sigma}_{-ij}$ are explicitly given by:

$$\begin{aligned}
[\widetilde{z}_{-ij}]_k &= \frac{\left(1-\sigma_{ij}^2\right) z_k - (\sigma_{ik} - \sigma_{ij}\sigma_{jk}) z_i - (\sigma_{jk} - \sigma_{ij}\sigma_{ik}) z_j}{\sqrt{\left(1 - \sigma_{ij}^2 - \sigma_{ik}^2 - \sigma_{jk}^2 + 2\sigma_{ij}\sigma_{ik}\sigma_{jk}\right)\left(1-\sigma_{ij}^2\right)}} \\
\left[\widetilde{\Sigma}_{-ij}\right]_{kl} &= \frac{\left(1-\sigma_{ij}^2\right)\sigma_{kl} - (\sigma_{ik}\sigma_{il} + \sigma_{jk}\sigma_{jl}) + (\sigma_{ij}\sigma_{ik}\sigma_{jl} + \sigma_{ij}\sigma_{il}\sigma_{jk})}{\sqrt{\left(1 - \sigma_{ij}^2 - \sigma_{ik}^2 - \sigma_{jk}^2 + 2\sigma_{ij}\sigma_{ik}\sigma_{jk}\right)\left(1 - \sigma_{ij}^2 - \sigma_{il}^2 - \sigma_{jl}^2 + 2\sigma_{ij}\sigma_{il}\sigma_{jl}\right)}} \\
\left[\widetilde{\Sigma}_{-ij}\right]_{kk} &= 1.
\end{aligned}$$

And thus we have:

$$\Phi_{M,Z}(z, \Omega) = \prod_{k=1}^{M} \Phi(z_k) + \int_0^1 \sum_{i<j} \omega_{ij} \varphi_{2,Z_{ij}}(z_{ij}, \Sigma_{ij}(s)) \Phi_{M-2,Z_{-ij}}\left(\widetilde{z}_{-ij}(s), \widetilde{\Sigma}_{-ij}(s)\right) ds.$$

$\square$

Appendix 5: Proof of Lemma 4

By definition, the likelihood of observation i is given by:

$$\begin{aligned}
L_i\left(y_i \mid x_i; \beta, \Omega\right) &= \Pr\left(-w_{i1} y_{i1}^* \leq 0, \ldots, -w_{iM} y_{iM}^* \leq 0\right) \\
&= \Pr\left(-w_{i1}\epsilon_{i1} \leq w_{i1} c_{i1}, \ldots, -w_{iM}\epsilon_{iM} \leq w_{iM} c_{iM}\right) \\
&= \Phi_{M,-W_i\mathcal{E}_i}\left(W_i c_i \mid 0_M, \Omega\right) \\
&= \int_{-\infty}^{w_{iM}c_{iM}} \cdots \int_{-\infty}^{w_{i1}c_{i1}} \varphi_{M,-W_i\mathcal{E}_i}\left(W_i\epsilon_i, \Omega\right) \prod_{m=1}^{M} d\epsilon_{im}.
\end{aligned}$$

Since each w_{im} takes only the values $\{-1, 1\}$, it is straightforward to show that $W_i = W_i^{-1}$ and $|W_i \Omega W_i| = |\Omega|$. Moreover, the density of an M-variate standardized normal vector $-W_i\mathcal{E}_i$ with correlation matrix Ω may be re-written as the density of an M-variate standardized normal vector $\mathcal{E}_i$ with correlation matrix $W_i\Omega W_i$:

$$\begin{aligned}
\varphi_{M,-W_i\mathcal{E}_i}\left(W_i\epsilon_i, \Omega\right) &= \left|2\pi\Omega\right|^{\frac{-1}{2}} \exp\left\{\frac{-1}{2}\left(-W_i\epsilon_i\right)' \Omega^{-1}\left(-W_i\epsilon_{it}\right)\right\} \\
&= \left|2\pi\left(W_i\Omega W_i\right)\right|^{\frac{-1}{2}} \exp\left\{\frac{-1}{2}\epsilon_i'\left(W_i\Omega W_i\right)^{-1}\epsilon_i\right\} \\
&= \varphi_{M,\mathcal{E}_i}\left(\epsilon_i, W_i\Omega W_i\right).
\end{aligned}$$

Therefore, the likelihood of observation i is given by:

$$\begin{aligned}
L_i\left(y_i \mid x_i; \beta, \Omega\right) &= \int_{-\infty}^{w_{iM}c_{iM}} \cdots \int_{-\infty}^{w_{i1}c_{i1}} \varphi_{M,\mathcal{E}_i}\left(\epsilon_i, W_i\Omega W_i\right) \prod_{m=1}^{M} d\epsilon_{im} \\
&= \Phi_{M,\mathcal{E}_i}\left(W_i c_i, W_i\Omega W_i\right).
\end{aligned}$$

$\square$

Technical report

The technical report provides additional results. More specifically,

1. The decomposition of a bivariate, trivariate, and quadrivariate normal integral;
2. The decomposition of a singular multivariate probit model;
3. The derivation of the score vector using the ML and the EML estimators in the case of a trivariate probit models;
4. Some Monte-Carlo simulations in a simultaneous bivariate probit equations model.

Applications of Proposition 2

In this section, we report the decomposition of the bivariate, trivariate, and quadrivariate normal cumulative distribution function. In particular, we discuss the positiveness conditions of the correlation matrix in the trivariate case.[39]

- Bivariate Normal cumulative distribution function

In the bivariate case, the cumulative distribution function $\Phi_2(z,\Omega) = \Phi_2(z,\omega_{12})$ may be decomposed, according to Proposition 2, in the following way:

$$\begin{aligned}\Phi_2(z,\Omega) &= \Phi(z_1)\Phi(z_2) + \int_0^{\omega_{12}} \varphi_2(z,\lambda_{12})\,d\lambda_{12}\\ &= \Phi(z_1)\Phi(z_2) + \Psi_2(z_1,z_2,\omega_{12}).\end{aligned}$$

Interestingly, this decomposition allows to derive very easily the particular case of z being a zero vector, as $\Phi(0) = \frac{1}{2}$:

$$\begin{aligned}\Phi_2(\mathbf{0},\Omega) &= \Phi^2(0) + \Psi_2(0,0,\omega_{12})\\ &= \frac{1}{4} + \frac{\arcsin\omega_{12}}{2\pi}\\ &= \frac{1}{2} - \frac{\arccos\omega_{12}}{2\pi}\end{aligned}$$

- Trivariate Normal cumulative distribution function

In the trivariate case, Proposition 2 yields:

$$\begin{aligned}\Phi_3(z,\Omega) = {} & \Phi(z_1)\Phi(z_2)\Phi(z_3)\\ &+\Phi(z_3)\int_0^{\omega_{12}} \varphi_2(z_1,z_2,\lambda_{12})\,d\lambda_{12}\\ &+\Phi(z_2)\int_0^{\omega_{13}} \varphi_2(z_1,z_3,\lambda_{13})\,d\lambda_{13}\\ &+\Phi(z_1)\int_0^{\omega_{23}} \varphi_2(z_2,z_3,\lambda_{23})\,d\lambda_{23}\\ &+\int_0^{\omega_{12}}\int_0^{\omega_{13}} \frac{\partial\varphi_3(z,\lambda_{12},\lambda_{13},0)}{\partial z_1}d\lambda_{12}d\lambda_{13}\\ &+\int_0^{\omega_{12}}\int_0^{\omega_{23}} \frac{\partial\varphi_3(z,\lambda_{12},0,\lambda_{23})}{\partial z_2}d\lambda_{12}d\lambda_{23}\\ &+\int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial\varphi_3(z,0,\lambda_{13},\lambda_{23})}{\partial z_3}d\lambda_{13}d\lambda_{23}\\ &+\int_0^{\omega_{12}}\int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial^3\varphi_3(z,\Lambda)}{\partial z_1\partial z_2\partial z_3}d\lambda_{12}d\lambda_{13}d\lambda_{23}.\end{aligned}$$

[39]Higher decompositions are available upon request.

The last integral can be decomposed in a non-unique way as follows:

$$\begin{aligned}
&\int_0^{\omega_{12}}\int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial^3 \varphi_3(z,\Lambda)}{\partial z_1 \partial z_2 \partial z_3} d\lambda_{12} d\lambda_{13} d\lambda_{23} \\
&\qquad = \int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial \varphi_3(z,\omega_{12},\lambda_{13},\lambda_{23})}{\partial z_3} d\lambda_{13} d\lambda_{23} \\
&\qquad - \int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial \varphi_3(z,0,\lambda_{13},\lambda_{23})}{\partial z_3} d\lambda_{13} d\lambda_{23} \\
&\qquad = \int_0^{\omega_{12}}\int_0^{\omega_{23}} \frac{\partial \varphi_3(z,\lambda_{12},\omega_{13},\lambda_{23})}{\partial z_2} d\lambda_{12} d\lambda_{23} \\
&\qquad - \int_0^{\omega_{12}}\int_0^{\omega_{23}} \frac{\partial \varphi_3(z,\lambda_{12},0,\lambda_{23})}{\partial z_2} d\lambda_{12} d\lambda_{23} \\
&\qquad = \int_0^{\omega_{12}}\int_0^{\omega_{13}} \frac{\partial \varphi_3(z,\lambda_{12},\lambda_{13},\omega_{23})}{\partial z_1} d\lambda_{12} d\lambda_{13} \\
&\qquad\qquad - \int_0^{\omega_{12}}\int_0^{\omega_{13}} \frac{\partial \varphi_3(z,\lambda_{12},\lambda_{13},0)}{\partial z_1} d\lambda_{12} d\lambda_{13}.
\end{aligned}$$

Therefore, the cumulative distribution function writes:

$$\begin{aligned}
\Phi_3(z,\Omega) \quad = \quad & \Phi(z_1)\Phi(z_2)\Phi(z_3) \\
& +\Phi(z_3)\Psi_2(z_1,z_2,\omega_{12}) \\
& +\Phi(z_2)\Psi_2(z_1,z_3,\omega_{13}) \\
& +\Phi(z_1)\Psi_2(z_2,z_3,\omega_{23}) \\
& +\Psi_3(z_1,z_2,z_3,\omega_{12},\omega_{13},0) \\
& +\Psi_3(z_2,z_3,z_1,\omega_{23},\omega_{12},0) \\
& +\Psi_3(z_3,z_1,z_2,\omega_{13},\omega_{23},\omega_{12}).
\end{aligned}$$

Note however that it is only one of the three possible specifications due to the non-unique decomposition of the three-dimension integral exposed above. In this example, the provision concerning the positive definiteness of each matrix in the sum is of particular relevance. Assume that $\omega_{12} = 0.2$, $\omega_{13} = 0.7$ and $\omega_{23} = 0.8$. With the exception of Ω_{11}, all matrices Ω_{pq} defined above are positive definite. This means that both $\int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial \varphi_3(z,0,\lambda_{13},\lambda_{23})}{\partial z_3} d\lambda_{13} d\lambda_{23}$ and $\int_0^{\omega_{12}}\int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial^3 \varphi_3(z,\Lambda)}{\partial z_1 \partial z_2 \partial z_3} d\lambda_{12} d\lambda_{13} d\lambda_{23}$ are not well defined at some of the boundary values of the integration domain. Consequently, Proposition 1 is no more applicable, but Proposition 2 still is, at least if expressed in the right way. Indeed, the element $\int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{\partial \varphi_3(z,\omega_{12},\lambda_{13},\lambda_{23})}{\partial z_3} d\lambda_{13} d\lambda_{23}$ in the above application of Proposition 2 is well defined over the whole range of integration, as are all other elements in that sum. In other words, while the application of Proposition 1 requires to impose the three following constraints, $\omega_{12}^2 + \omega_{13}^2 < 1$, $\omega_{12}^2 + \omega_{23}^2 < 1$, and $\omega_{13}^2 + \omega_{23}^2 < 1$, Proposition 2 only requires to impose two of those. It turns out that one may choose the expression of the cumulative distribution function by using the smallest (in absolute value) of the elements ω_{12}, ω_{13}, and ω_{23} as the one entering as non-zero sixth argument in Ψ_3.

As a final remark, if z is a zero vector, the last three terms of the previous expression equal zero and we obtain:

$$\begin{aligned}
\Phi_3(\mathbf{0},\Omega) &= \Phi^3(0) + \Phi(0)\Psi_2(0,0,\omega_{12}) + \Phi(0)\Psi_2(0,0,\omega_{13}) + \Phi(0)\Psi_2(0,0,\omega_{23}) \\
&= \frac{1}{8} + \frac{\arcsin\omega_{12} + \arcsin\omega_{13} + \arcsin\omega_{23}}{4\pi} \\
&= \frac{1}{2} - \frac{\arccos\omega_{12} + \arccos\omega_{13} + \arccos\omega_{23}}{4\pi}.
\end{aligned}$$

The computation of the trivariate standardized normal probabilities is then straightforward.

- Quadrivariate Normal cumulative distribution function

In the quadrivariate case ($n = 4$), Proposition 1 yields a sum of 64 elements, which will not be listed here. Instead, according to proposition 2, we can reduce this number to 26 elements. In that respect, one of these possible sums is given by:

$$\begin{aligned}
& \Phi_4\left(z,\omega_{12},\omega_{13},\omega_{14},\omega_{23},\omega_{24},\omega_{34}\right) \\
= & \Phi\left(z_1\right)\Phi\left(z_2\right)\Phi\left(z_3\right)\Phi\left(z_4\right) \\
& +\Phi\left(z_1\right)\Phi\left(z_2\right)\Psi_2\left(z_3,z_4,\omega_{34}\right)+\Phi\left(z_1\right)\Phi\left(z_3\right)\Psi_2\left(z_2,z_4,\omega_{24}\right)+\Phi\left(z_1\right)\Phi\left(z_4\right)\Psi_2\left(z_2,z_3,\omega_{23}\right) \\
& +\Phi\left(z_2\right)\Phi\left(z_3\right)\Psi_2\left(z_1,z_4,\omega_{14}\right)+\Phi\left(z_2\right)\Phi\left(z_4\right)\Psi_2\left(z_1,z_3,\omega_{13}\right)+\Phi\left(z_3\right)\Phi\left(z_4\right)\Psi_2\left(z_1,z_2,\omega_{12}\right) \\
& +\Psi_2\left(z_1,z_4,\omega_{14}\right)\Psi_2\left(z_2,z_3,\omega_{23}\right) \\
& +\Phi\left(z_4\right)\left[\Psi_3\left(z_2,z_3,z_1,\omega_{23},\omega_{12},0\right)+\Psi_3\left(z_3,z_1,z_2,\omega_{13},\omega_{23},0\right)+\Psi_3\left(z_1,z_2,z_3,\omega_{12},\omega_{13},\omega_{23}\right)\right] \\
& +\Phi\left(z_3\right)\left[\Psi_3\left(z_1,z_2,z_4,\omega_{12},\omega_{14},0\right)+\Psi_3\left(z_4,z_1,z_2,\omega_{14},\omega_{24},0\right)+\Psi_3\left(z_2,z_4,z_1,\omega_{24},\omega_{12},\omega_{14}\right)\right] \\
& +\Phi\left(z_2\right)\left[\Psi_3\left(z_4,z_1,z_3,\omega_{14},\omega_{34},0\right)+\Psi_3\left(z_1,z_3,z_4,\omega_{13},\omega_{14},0\right)+\Psi_3\left(z_3,z_4,z_1,\omega_{34},\omega_{13},\omega_{14}\right)\right] \\
& +\Phi\left(z_1\right)\left[\Psi_3\left(z_3,z_4,z_2,\omega_{34},\omega_{23},0\right)+\Psi_3\left(z_2,z_3,z_4,\omega_{23},\omega_{24},0\right)+\Psi_3\left(z_4,z_2,z_3,\omega_{24},\omega_{34},\omega_{23}\right)\right] \\
& +\Psi_{41}\left(z_1,z_2,z_3,z_4,\omega_{12},\omega_{13},\omega_{14},\omega_{23},\omega_{24},\omega_{34}\right) \\
& +\Xi_{41}\left(z_1,z_3,z_2,z_4,\omega_{13},\omega_{12},\omega_{14},\omega_{23},\omega_{34},\omega_{24}\right) \\
& +\Xi_{42}\left(z_1,z_4,z_2,z_3,\omega_{14},\omega_{12},\omega_{13},\omega_{24},\omega_{34},\omega_{23}\right) \\
& +\Psi_{42}\left(z_1,z_2,z_3,z_4,\omega_{12},\omega_{13},\omega_{14}\right)+\Psi_{42}\left(z_2,z_3,z_4,z_1,\omega_{23},\omega_{24},\omega_{12}\right) \\
& +\Psi_{42}\left(z_3,z_4,z_1,z_2,\omega_{34},\omega_{13},\omega_{23}\right)+\Psi_{42}\left(z_4,z_1,z_2,z_3,\omega_{14},\omega_{24},\omega_{34}\right)
\end{aligned}$$

where the functions with the same integration variables are grouped, with $z'=\{z_1,z_3,z_2,z_4\}$:

$$\begin{aligned}
& \Xi_{41}\left(z',\omega_{13},\omega_{12},\omega_{14},\omega_{23},\omega_{34},\omega_{24}\right) \\
= & \Psi_{41}\left(z',\omega_{13},\omega_{12},\omega_{14},\omega_{23},0,\omega_{24}\right) \\
& +\Psi_{41}\left(z',\omega_{13},0,\omega_{14},\omega_{23},\omega_{34},\omega_{24}\right) \\
& -\Psi_{41}\left(z',\omega_{13},0,\omega_{14},\omega_{23},0,\omega_{24}\right) \\
= & \int_0^{\omega_{13}}\int_0^{\omega_{24}}\left[\varphi_4\left(z',\lambda_{13},\omega_{12},\omega_{14},\omega_{23},0,\lambda_{24}\right)\right. \\
& \qquad +\varphi_4\left(z',\lambda_{13},0,\omega_{14},\omega_{23},\omega_{34},\lambda_{24}\right) \\
& \qquad \left.-\varphi_4\left(z',\lambda_{13},0,\omega_{14},\omega_{23},0,\lambda_{24}\right)\right]d\lambda_{13}d\lambda_{24}
\end{aligned}$$

and, with $z''=\{z_1,z_4,z_2,z_3\}$:

$$\begin{aligned}
& \Xi_{42}\left(z'',\omega_{14},\omega_{12},\omega_{13},\omega_{24},\omega_{34},\omega_{23}\right) \\
= & \Psi_{41}\left(z'',\omega_{14},\omega_{12},\omega_{13},0,0,\omega_{23}\right)+\Psi_{41}\left(z'',\omega_{14},\omega_{12},0,\omega_{24},0,\omega_{23}\right) \\
& +\Psi_{41}\left(z'',\omega_{14},0,\omega_{13},0,\omega_{34},\omega_{23}\right)+\Psi_{41}\left(z'',\omega_{14},0,0,\omega_{24},\omega_{34},\omega_{23}\right) \\
& -\Psi_{41}\left(z'',\omega_{14},\omega_{12},0,0,0,\omega_{23}\right)-\Psi_{41}\left(z'',\omega_{14},0,\omega_{13},0,0,\omega_{23}\right) \\
& -\Psi_{41}\left(z'',\omega_{14},0,0,\omega_{24},0,\omega_{23}\right)-\Psi_{41}\left(z'',\omega_{14},0,0,0,\omega_{34},\omega_{23}\right) \\
= & \int_0^{\omega_{14}}\int_0^{\omega_{23}}\left[\varphi_4\left(z'',\lambda_{14},\omega_{12},\omega_{13},0,0,\lambda_{23}\right)+\varphi_4\left(z'',\lambda_{14},\omega_{12},0,\omega_{24},0,\lambda_{23}\right)\right. \\
& \quad +\varphi_4\left(z'',\lambda_{14},0,\omega_{13},0,\omega_{34},\lambda_{23}\right)+\varphi_4\left(z'',\lambda_{14},0,0,\omega_{24},\omega_{34},\lambda_{23}\right) \\
& \quad -\varphi_4\left(z'',\lambda_{14},\omega_{12},0,0,0,\lambda_{23}\right)-\varphi_4\left(z'',\lambda_{14},0,\omega_{13},0,0,\lambda_{23}\right) \\
& \quad \left.-\varphi_4\left(z'',\lambda_{14},0,0,\omega_{24},0,\lambda_{23}\right)-\varphi_4\left(z'',\lambda_{14},0,0,0,\omega_{34},\lambda_{23}\right)\right]d\lambda_{14}d\lambda_{23}.
\end{aligned}$$

Here, setting the particular case of z being a zero vector is not as simple as for the dimensions 2 and 3, as only the odd-order derivatives are then zero, and all other double and triple integrals have still to be computed.

Singular multivariate probit model

In this section, we analyze the singular case. By singularity, we mean here some equicorrelation across the error terms. For sake of simplicity, we consider a bivariate probit model with a positive unit correlation between the error terms. The model is given by:

$$\begin{cases} y_{i1}^* = c_{i1}+\epsilon_{i1} \\ y_{i2}^* = c_{i2}+\epsilon_{i2} \end{cases}$$

$$\begin{cases} y_{i1} = \mathbb{1}_{\left(y_{i1}^*>0\right)} \\ y_{i2} = \mathbb{1}_{\left(y_{i2}^*>0\right)} \end{cases}$$

$$\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \omega_{12} \\ \omega_{12} & 1 \end{pmatrix}\right]$$

with $c_{i1} = x_{i1}\beta_1$, $c_{i2} = x_{i2}\beta_2$ and $\omega_{12} = 1$. It is worth noticing that $\omega_{12} = 1$ implies that $\epsilon_{i1} = \epsilon_{i2} = \epsilon_i \sim N(0,1)$ and thus:

$$\begin{cases} y_{i1}^* = c_{i1} + \epsilon_i \\ y_{i2}^* = c_{i2} + \epsilon_i. \end{cases}$$

It is then straightforward to obtain the likelihood of the outcomes ($L_i^{kl} = \Pr(y_{i1} = k, y_{i2} = l | x_{i1}, x_{i2}; \beta_1, \beta_2)$; $k, l = 0, 1$):

$$\begin{aligned} L_i^{00} &= \Pr(\epsilon_i \leq -c_{i1}, \epsilon_i \leq -c_{i2} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \Pr(\epsilon_i \leq \min\{-c_{i1}, -c_{i2}\} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \int_{-\infty}^{\min\{-c_{i1}, -c_{i2}\}} \varphi(t)\, dt \\ &= \Phi(\min\{-c_{i1}, -c_{i2}\}) \end{aligned}$$

$$\begin{aligned} L_i^{11} &= \Pr(\epsilon_i > -c_{i1}, \epsilon_i > -c_{i2} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \Pr(\epsilon_i > \max\{-c_{i1}, -c_{i2}\} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \Pr(\epsilon_i > -\min\{c_{i1}, c_{i2}\} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \int_{-\infty}^{\min\{c_{i1}, c_{i2}\}} \varphi(t)\, dt \\ &= \Phi(\min\{c_{i1}, c_{i2}\}) \end{aligned}$$

$$\begin{aligned} L_i^{01} &= \Pr(\epsilon_i \leq -c_{i1}, \epsilon_i > -c_{i2} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \mathbb{1}_{(-c_{i1} \geq -c_{i2})} \Pr(-c_{i2} < \epsilon_i \leq -c_{i1} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \mathbb{1}_{(-\min\{-c_{i1}, c_{i2}\} \leq \max\{-c_{i1}, c_{i2}\})} \Pr(-\min\{-c_{i1}, c_{i2}\} < \epsilon_i \leq \max\{-c_{i1}, c_{i2}\} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \mathbb{1}_{(-\min\{-c_{i1}, c_{i2}\} \leq \max\{-c_{i1}, c_{i2}\})} \int_{-\min\{-c_{i1}, c_{i2}\}}^{\max\{-c_{i1}, c_{i2}\}} \varphi(t)\, dt \\ &= \mathbb{1}_{(-\min\{-c_{i1}, c_{i2}\} \leq \max\{-c_{i1}, c_{i2}\})} \left[\Phi(\max\{-c_{i1}, c_{i2}\}) - \Phi(-\min\{-c_{i1}, c_{i2}\})\right] \end{aligned}$$

$$\begin{aligned} L_i^{10} &= \Pr(\epsilon_i > -c_{i1}, \epsilon_i \leq -c_{i2} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \mathbb{1}_{(-c_{i1} \leq -c_{i2})} \Pr(-c_{i1} < \epsilon_i \leq -c_{i2} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \mathbb{1}_{(-\min\{c_{i1}, -c_{i2}\} \leq \max\{c_{i1}, -c_{i2}\})} \Pr(-\min\{c_{i1}, -c_{i2}\} < \epsilon_i \leq \max\{c_{i1}, -c_{i2}\} | x_{i1}, x_{i2}; \beta_1, \beta_2) \\ &= \mathbb{1}_{(-\min\{c_{i1}, -c_{i2}\} \leq \max\{c_{i1}, -c_{i2}\})} \int_{-\min\{c_{i1}, -c_{i2}\}}^{\max\{c_{i1}, -c_{i2}\}} \varphi(t)\, dt \\ &= \mathbb{1}_{(-\min\{c_{i1}, -c_{i2}\} \leq \max\{c_{i1}, -c_{i2}\})} \left[\Phi(\max\{c_{i1}, -c_{i2}\}) - \Phi(-\min\{c_{i1}, -c_{i2}\})\right]. \end{aligned}$$

Therefore,

$$L_i = \begin{cases} \Phi(\min\{w_{i1}c_{i1}, w_{i2}c_{i2}\}) & \text{if} \quad y_{i1} = y_{i2} \\ J_i\left[\Phi(\max\{w_{i1}c_{i1}, w_{i2}c_{i2}\}) - \Phi(-\min\{w_{i1}c_{i1}, w_{i2}c_{i2}\})\right] & \text{if} \quad y_{i1} \neq y_{i2} \end{cases}$$

with $w_{im} = 2y_{im} - 1$ and $J_i = \mathbb{1}(-\min\{w_{i1}c_{i1}, w_{i2}c_{i2}\} \leq \max\{w_{i1}c_{i1}, w_{i2}c_{i2}\})$. Consequently, L_i is not continuously differentiable with respect to the arguments c_{i1} and c_{i2}. In contrast, if one uses the decomposition in Proposition 2, the likelihood of observation i is given by:

$$L_i = \Pr(-w_{i1}\epsilon_i \leq w_{i1}c_{i1}, -w_{i2}\epsilon_i \leq w_{i2}c_{i2} | x_{i1}, x_{i2}; \beta_1, \beta_2).$$

Noting that $\begin{pmatrix} -w_{i1}\epsilon_i \\ -w_{i2}\epsilon_i \end{pmatrix} \rightsquigarrow N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & w_{i1}w_{i2} \\ w_{i1}w_{i2} & 1 \end{pmatrix}\right]$ yields:

$$\begin{aligned} L_i &= \Phi_{2,W_i c_i}(w_{i1}w_{i2}) \\ &= \Phi(w_{i1}c_{i1})\,\Phi(w_{i2}c_{i2}) + \int_0^{w_{i1}w_{i2}} \varphi_{2,W_i c_i}(\lambda)\, d\lambda \\ &= \Phi(w_{i1}c_{i1})\,\Phi(w_{i2}c_{i2}) + w_{i1}w_{i2} \int_0^1 \varphi_{2,c_i}(\lambda)\, d\lambda. \end{aligned}$$

We thus obtain the following definition.

Definition 2 (Definition of singular cdf) *The regular bivariate normal cumulative density function is given by:*

$$\Phi_{2,z}(\omega) = \int_{-\infty}^{z_1}\int_{-\infty}^{z_2} \varphi_{2,t}(\omega)\, dt_1 dt_2$$

for $\omega \in (-1,1)$, *which can be exactly decomposed as:*

$$\Phi_{2,z}(\omega) = \Phi(z_1)\Phi(z_2) + \int_0^{\omega} \varphi_{2,z}(\lambda)\, d\lambda$$

and the singular bivariate normal cumulative density function, for $\omega \in \{-1,1\}$, *is given by:*

$$\begin{aligned}
\Phi_{2,z}(+1) &= \Pr(t \leq z_1, t \leq z_2) \\
&= \int_{(-\infty,z_1]\cap(-\infty,z_2]} \varphi(t)\, dt = \int_{-\infty}^{\min\{z_1,z_2\}} \varphi(t)\, dt \\
&= \Phi(\min\{z_1,z_2\}) \\
\Phi_{2,z}(-1) &= \Pr(t \leq z_1, -t \leq z_2) = \Pr(t \leq z_1, t \geq -z_2) \\
&= \int_{(-\infty,z_1]\cap[-z_2,\infty)} \varphi(t)\, dt = \mathbb{1}_{-\min\{z_1,z_2\}\leq\max\{z_1,z_2\}} \int_{-\min\{z_1,z_2\}}^{\max\{z_1,z_2\}} \varphi(t)\, dt \\
&= \mathbb{1}_{-\min\{z_1,z_2\}\leq\max\{z_1,z_2\}} \left[\Phi(\max\{z_1,z_2\}) - \Phi(-\min\{z_1,z_2\})\right]
\end{aligned}$$

which can still be exactly decomposed as:

$$\begin{aligned}
\Phi_{2,z}(\pm 1) &= \lim_{\omega\to\pm 1}\left[\Phi(z_1)\Phi(z_2) + \int_0^{\omega} \varphi_{2,z}(\lambda)\, d\lambda\right] \\
&= \Phi(z_1)\Phi(z_2) + \int_0^{\pm 1} \varphi_{2,z}(\lambda)\, d\lambda.
\end{aligned}$$

□

For a given singular correlation matrix, this definition may be generalized to any multivariate probit model.

Derivations of the ML and EML score vector in a trivariate probit model

In this section, we show the derivations of the score vector in a trivariate probit model using both the usual maximum likelihood estimator and our Proposition 2. Results for higher multivariate probit models are available upon request.

To simplify notations, we omit the i index. In the usual case (without Proposition 2), the likelihood of observation i is given by:

$$\begin{aligned}
P &= \Phi_3(w_1c_1, w_2c_2, w_3c_3, w_1w_2\omega_{12}, w_1w_3\omega_{13}, w_2w_3\omega_{23}) \\
&= \int_{-\infty}^{w_1c_1}\int_{-\infty}^{w_2c_2}\int_{-\infty}^{w_3c_3} \varphi_3(z_1,z_2,z_3,w_1w_2\omega_{12},w_1w_3\omega_{13},w_2w_3\omega_{23})\, dz_1dz_2dz_3.
\end{aligned}$$

This may be written as:

$$\begin{aligned}
P &= \int_{-\infty}^{w_1c_1}\int_{-\infty}^{w_2c_2} \varphi_2(z_1,z_2,\omega_{12})\,\Phi\left(w_3\frac{(1-\omega_{12}^2)c_3 - w_1(\omega_{13}-\omega_{12}\omega_{23})z_1 - w_2(\omega_{23}-\omega_{12}\omega_{13})z_2}{\sqrt{|\Omega|(1-\omega_{12}^2)}}\right) dz_1dz_2dz_3 \\
&= \int_{-\infty}^{w_1c_1}\int_{-\infty}^{w_3c_3} \varphi_2(z_1,z_3,\omega_{13})\,\Phi\left(w_2\frac{(1-\omega_{13}^2)c_2 - w_1(\omega_{12}-\omega_{13}\omega_{23})z_1 - w_3(\omega_{23}-\omega_{12}\omega_{13})z_3}{\sqrt{|\Omega|(1-\omega_{13}^2)}}\right) dz_1dz_2dz_3 \\
&= \int_{-\infty}^{w_2c_2}\int_{-\infty}^{w_3c_3} \varphi_2(z_2,z_3,\omega_{23})\,\Phi\left(w_1\frac{(1-\omega_{23}^2)c_1 - w_2(\omega_{12}-\omega_{13}\omega_{23})z_2 - w_3(\omega_{13}-\omega_{12}\omega_{23})z_3}{\sqrt{|\Omega|(1-\omega_{23}^2)}}\right) dz_1dz_2dz_3
\end{aligned}$$

or equivalently as:

$$\begin{aligned}
P &= \int_{-\infty}^{w_1c_1} \varphi(z_1)\,\Phi_2\left(w_2\frac{c_2 - w_1\omega_{12}z_1}{\sqrt{1-\omega_{12}^2}}, w_3\frac{c_3 - w_1\omega_{13}z_1}{\sqrt{1-\omega_{13}^2}}, w_2w_3\frac{\omega_{23}-\omega_{12}\omega_{13}}{\sqrt{(1-\omega_{12}^2)(1-\omega_{13}^2)}}\right) dz_1 \\
&= \int_{-\infty}^{w_2c_2} \varphi(z_2)\,\Phi_2\left(w_3\frac{c_3 - w_2\omega_{23}z_2}{\sqrt{1-\omega_{23}^2}}, w_1\frac{c_1 - w_2\omega_{12}z_2}{\sqrt{1-\omega_{12}^2}}, w_3w_1\frac{\omega_{13}-\omega_{23}\omega_{12}}{\sqrt{(1-\omega_{23}^2)(1-\omega_{12}^2)}}\right) dz_2 \\
&= \int_{-\infty}^{w_3c_3} \varphi(z_3)\,\Phi_2\left(w_1\frac{c_1 - w_3\omega_{13}z_3}{\sqrt{1-\omega_{13}^2}}, w_2\frac{c_2 - w_3\omega_{23}z_3}{\sqrt{1-\omega_{23}^2}}, w_1w_2\frac{\omega_{12}-\omega_{13}\omega_{23}}{\sqrt{(1-\omega_{13}^2)(1-\omega_{23}^2)}}\right) dz_3.
\end{aligned}$$

It is straightforward to show that:

$$
\begin{aligned}
\frac{\partial}{\partial c_1}P &= w_1\varphi(c_1)\,\Phi_2\left(w_2\frac{c_2-\omega_{12}c_1}{\sqrt{1-\omega_{12}^2}}, w_3\frac{c_3-\omega_{13}c_1}{\sqrt{1-\omega_{13}^2}}, w_2w_3\frac{\omega_{23}-\omega_{12}\omega_{13}}{\sqrt{(1-\omega_{12}^2)(1-\omega_{13}^2)}}\right)\\
\frac{\partial}{\partial c_2}P &= w_2\varphi(c_2)\,\Phi_2\left(w_3\frac{c_3-\omega_{23}c_2}{\sqrt{1-\omega_{23}^2}}, w_1\frac{c_1-\omega_{12}c_2}{\sqrt{1-\omega_{12}^2}}, w_3w_1\frac{\omega_{13}-\omega_{23}\omega_{12}}{\sqrt{(1-\omega_{23}^2)(1-\omega_{12}^2)}}\right)\\
\frac{\partial}{\partial c_3}P &= w_3\varphi(c_3)\,\Phi_2\left(w_1\frac{c_1-\omega_{13}c_3}{\sqrt{1-\omega_{13}^2}}, w_2\frac{c_2-\omega_{23}c_3}{\sqrt{1-\omega_{23}^2}}, w_1w_2\frac{\omega_{12}-\omega_{13}\omega_{23}}{\sqrt{(1-\omega_{13}^2)(1-\omega_{23}^2)}}\right)\\
\frac{\partial}{\partial \omega_{12}}P &= w_1w_2\varphi_2(c_1,c_2,\omega_{12})\,\Phi\left(w_3\frac{(1-\omega_{12}^2)c_3-(\omega_{13}-\omega_{12}\omega_{23})c_1-(\omega_{23}-\omega_{12}\omega_{13})c_2}{\sqrt{|\Omega|(1-\omega_{12}^2)}}\right)\\
\frac{\partial}{\partial \omega_{13}}P &= w_1w_3\varphi_2(c_1,c_3,\omega_{13})\,\Phi\left(w_2\frac{(1-\omega_{13}^2)c_2-(\omega_{12}-\omega_{13}\omega_{23})c_1-(\omega_{23}-\omega_{12}\omega_{13})c_3}{\sqrt{|\Omega|(1-\omega_{13}^2)}}\right)\\
\frac{\partial}{\partial \omega_{23}}P &= w_2w_3\varphi_2(c_2,c_3,\omega_{23})\,\Phi\left(w_1\frac{(1-\omega_{23}^2)c_1-(\omega_{12}-\omega_{13}\omega_{23})c_2-(\omega_{13}-\omega_{12}\omega_{23})c_3}{\sqrt{|\Omega|(1-\omega_{23}^2)}}\right).
\end{aligned}
$$

Expressions of the second order partial derivatives follow.

We now turn to the decomposition in Proposition 2. As explained in the first section of the technical report, the likelihood of observation i may be written as:

$$
\begin{aligned}
P &= \Phi_3(w_1c_1, w_2c_2, w_3c_3, w_1w_2\omega_{12}, w_1w_3\omega_{13}, w_2w_3\omega_{23})\\
&= \Phi(w_1c_1)\,\Phi(w_2c_2)\,\Phi(w_3c_3)\\
&\quad + w_1w_2\Phi(w_3c_3)\,\Psi_2(c_1,c_2,\omega_{12})\\
&\quad + w_1w_3\Phi(w_2c_2)\,\Psi_2(c_1,c_3,\omega_{13})\\
&\quad + w_2w_3\Phi(w_1c_1)\,\Psi_2(c_2,c_3,\omega_{23})\\
&\quad + w_1w_2w_3\Psi_3(c_3,c_1,c_2,\omega_{13},\omega_{23},0)\\
&\quad + w_1w_2w_3\Psi_3(c_2,c_3,c_1,\omega_{23},\omega_{12},0)\\
&\quad + w_1w_2w_3\Psi_3(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23})
\end{aligned}
$$

where:

$$
\begin{aligned}
\Psi_2(c_1,c_2,\omega_{12}) &= \int_0^{\omega_{12}} \varphi_2(c_1,c_2,\lambda_{12})\,d\lambda_{12}\\
\Psi_2(c_1,c_3,\omega_{13}) &= \int_0^{\omega_{13}} \varphi_2(c_1,c_3,\lambda_{13})\,d\lambda_{13}\\
\Psi_2(c_2,c_3,\omega_{23}) &= \int_0^{\omega_{23}} \varphi_2(c_2,c_3,\lambda_{23})\,d\lambda_{23}
\end{aligned}
$$

with:

$$
\begin{aligned}
\Psi_3(c_3,c_1,c_2,\omega_{13},\omega_{23},0) &= \int_0^{\omega_{13}}\int_0^{\omega_{23}} \frac{-c_3+\lambda_{13}c_1+\lambda_{23}c_2}{1-\lambda_{13}^2-\lambda_{23}^2}\varphi_3(c_3,c_1,c_2,\lambda_{13},\lambda_{23},0)\,d\lambda_{13}d\lambda_{23}\\
\Psi_3(c_2,c_3,c_1,\omega_{23},\omega_{12},0) &= \int_0^{\omega_{23}}\int_0^{\omega_{12}} \frac{-c_2+\lambda_{23}c_3+\lambda_{12}c_1}{1-\lambda_{23}^2-\lambda_{12}^2}\varphi_3(c_2,c_3,c_1,\lambda_{23},\lambda_{12},0)\,d\lambda_{23}d\lambda_{12}\\
\Psi_3(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}) &= \int_0^{\omega_{12}}\int_0^{\omega_{13}} \frac{-(1-\omega_{23}^2)c_1+(\lambda_{12}-\lambda_{13}\omega_{23})c_2+(\lambda_{13}-\lambda_{12}\omega_{23})c_3}{1-\lambda_{12}^2-\lambda_{13}^2-\omega_{23}^2+2\lambda_{12}\lambda_{13}\omega_{23}}\\
&\qquad \cdot\varphi_3(c_1,c_2,c_3,\lambda_{12},\lambda_{13},\omega_{23})\,d\lambda_{12}d\lambda_{13}.
\end{aligned}
$$

It is then straightforward to show that the first-order partial derivatives are as follows:

$$
\begin{aligned}
\frac{\partial}{\partial c_1}P &= w_1\varphi\left(c_1\right)\Phi\left(w_2c_2\right)\Phi\left(w_3c_3\right)\\
&+w_1w_2\Phi\left(w_3c_3\right)\frac{\partial}{\partial c_1}\Psi_2\left(c_1,c_2,\omega_{12}\right)\\
&+w_1w_3\Phi\left(w_2c_2\right)\frac{\partial}{\partial c_1}\Psi_2\left(c_1,c_3,\omega_{13}\right)\\
&+w_1w_2w_3\varphi\left(c_1\right)\Psi_2\left(c_2,c_3,\omega_{23}\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_1}\Psi_3\left(c_3,c_1,c_2,\omega_{13},\omega_{23},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_1}\Psi_3\left(c_2,c_3,c_1,\omega_{23},\omega_{12},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_1}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial c_2}P &= w_2\varphi\left(c_2\right)\Phi\left(w_1c_1\right)\Phi\left(w_3c_3\right)\\
&+w_1w_2\Phi\left(w_3c_3\right)\frac{\partial}{\partial c_2}\Psi_2\left(c_1,c_2,\omega_{12}\right)\\
&+w_1w_2w_3\varphi\left(c_2\right)\Psi_2\left(c_1,c_3,\omega_{13}\right)\\
&+w_2w_3\Phi\left(w_1c_1\right)\frac{\partial}{\partial c_2}\Psi_2\left(c_2,c_3,\omega_{23}\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_2}\Psi_3\left(c_3,c_1,c_2,\omega_{13},\omega_{23},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_2}\Psi_3\left(c_2,c_3,c_1,\omega_{23},\omega_{12},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_2}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial c_3}P &= w_3\varphi\left(c_3\right)\Phi\left(w_1c_1\right)\Phi\left(w_2c_2\right)\\
&+w_1w_2w_3\varphi\left(c_3\right)\Psi_2\left(c_1,c_2,\omega_{12}\right)\\
&+w_1w_3\Phi\left(w_2c_2\right)\frac{\partial}{\partial c_3}\Psi_2\left(c_1,c_3,\omega_{13}\right)\\
&+w_2w_3\Phi\left(w_1c_1\right)\frac{\partial}{\partial c_3}\Psi_2\left(c_2,c_3,\omega_{23}\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_3}\Psi_3\left(c_3,c_1,c_2,\omega_{13},\omega_{23},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_3}\Psi_3\left(c_2,c_3,c_1,\omega_{23},\omega_{12},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial c_3}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \omega_{12}}P &= w_1w_2\Phi\left(w_3c_3\right)\frac{\partial}{\partial \omega_{12}}\Psi_2\left(c_1,c_2,\omega_{12}\right)\\
&+w_1w_2w_3\frac{\partial}{\partial \omega_{12}}\Psi_3\left(c_2,c_3,c_1,\omega_{23},\omega_{12},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial \omega_{12}}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \omega_{13}}P &= w_1w_3\Phi\left(w_2c_2\right)\frac{\partial}{\partial \omega_{13}}\Psi_2\left(c_1,c_3,\omega_{13}\right)\\
&+w_1w_2w_3\frac{\partial}{\partial \omega_{13}}\Psi_3\left(c_3,c_1,c_2,\omega_{13},\omega_{23},0\right)\\
&+w_1w_2w_3\frac{\partial}{\partial \omega_{13}}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right)
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial}{\partial \omega_{23}} P &= w_2 w_3 \Phi(w_1 c_1) \frac{\partial}{\partial \omega_{23}} \Psi_2(c_2, c_3, \omega_{23}) \\
&+ w_1 w_2 w_3 \frac{\partial}{\partial \omega_{23}} \Psi_3(c_3, c_1, c_2, \omega_{13}, \omega_{23}, 0) \\
&+ w_1 w_2 w_3 \frac{\partial}{\partial \omega_{23}} \Psi_3(c_2, c_3, c_1, \omega_{23}, \omega_{12}, 0) \\
&+ w_1 w_2 w_3 \frac{\partial}{\partial \omega_{23}} \Psi_3(c_1, c_2, c_3, \omega_{12}, \omega_{13}, \omega_{23})
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial}{\partial c_1} \Psi_3(c_3, c_1, c_2, \omega_{13}, \omega_{23}, 0) &= \int_0^{\omega_{23}} \int_0^{\omega_{13}} \frac{\partial}{\partial \lambda_{13}} \varphi_3(c_3, c_1, c_2, \lambda_{13}, \lambda_{23}, 0)\, d\lambda_{13} d\lambda_{23} \\
&= \int_0^{\omega_{23}} \varphi_3(c_3, c_1, c_2, \omega_{13}, \lambda_{23}, 0)\, d\lambda_{23} \\
\frac{\partial}{\partial c_2} \Psi_3(c_3, c_1, c_2, \omega_{13}, \omega_{23}, 0) &= \int_0^{\omega_{13}} \int_0^{\omega_{23}} \frac{\partial}{\partial \lambda_{23}} \varphi_3(c_3, c_1, c_2, \lambda_{13}, \lambda_{23}, 0)\, d\lambda_{23} d\lambda_{13} \\
&= \int_0^{\omega_{13}} \varphi_3(c_3, c_1, c_2, \lambda_{13}, \omega_{23}, 0)\, d\lambda_{13} \\
\frac{\partial}{\partial c_3} \Psi_3(c_3, c_1, c_2, \omega_{13}, \omega_{23}, 0) &= \int_0^{\omega_{13}} \int_0^{\omega_{23}} \left[(c_3 - \lambda_{13} c_1 - \lambda_{23} c_2)^2 - \left(1 - \lambda_{13}^2 - \lambda_{23}^2\right) \right] \\
&\quad \cdot \frac{1}{(1 - \lambda_{13}^2 - \lambda_{23}^2)^2} \varphi_3(c_3, c_1, c_2, \lambda_{13}, \lambda_{23}, 0)\, d\lambda_{13} d\lambda_{23} \\
\frac{\partial}{\partial \omega_{13}} \Psi_3(c_3, c_1, c_2, \omega_{13}, \omega_{23}, 0) &= \int_0^{\omega_{23}} \frac{-c_3 + \omega_{13} c_1 + \lambda_{23} c_2}{1 - \omega_{13}^2 - \lambda_{23}^2} \varphi_3(c_3, c_1, c_2, \omega_{13}, \lambda_{23}, 0)\, d\lambda_{23} \\
\frac{\partial}{\partial \omega_{23}} \Psi_3(c_3, c_1, c_2, \omega_{13}, \omega_{23}, 0) &= \int_0^{\omega_{13}} \frac{-c_3 + \lambda_{13} c_1 + \omega_{23} c_2}{1 - \lambda_{13}^2 - \omega_{23}^2} \varphi_3(c_3, c_1, c_2, \lambda_{13}, \omega_{23}, 0)\, d\lambda_{13}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial c_1} \Psi_3(c_2, c_3, c_1, \omega_{23}, \omega_{12}, 0) &= \int_0^{\omega_{23}} \int_0^{\omega_{12}} \frac{\partial}{\partial \lambda_{12}} \varphi_3(c_2, c_3, c_1, \lambda_{23}, \lambda_{12}, 0)\, d\lambda_{12} d\lambda_{23} \\
&= \int_0^{\omega_{23}} \varphi_3(c_2, c_3, c_1, \lambda_{23}, \omega_{12}, 0)\, d\lambda_{23} \\
\frac{\partial}{\partial c_2} \Psi_3(c_2, c_3, c_1, \omega_{23}, \omega_{12}, 0) &= \int_0^{\omega_{23}} \int_0^{\omega_{12}} \left[(c_2 - \lambda_{23} c_3 - \lambda_{12} c_1)^2 - \left(1 - \lambda_{23}^2 - \lambda_{12}^2\right) \right] \\
&\quad \cdot \frac{1}{(1 - \lambda_{23}^2 - \lambda_{12}^2)^2} \varphi_3(c_2, c_3, c_1, \lambda_{23}, \lambda_{12}, 0)\, d\lambda_{23} d\lambda_{12} \\
\frac{\partial}{\partial c_3} \Psi_3(c_2, c_3, c_1, \omega_{23}, \omega_{12}, 0) &= \int_0^{\omega_{12}} \int_0^{\omega_{23}} \frac{\partial}{\partial \lambda_{23}} \varphi_3(c_2, c_3, c_1, \lambda_{23}, \lambda_{12}, 0)\, d\lambda_{23} d\lambda_{12} \\
&= \int_0^{\omega_{12}} \varphi_3(c_2, c_3, c_1, \omega_{23}, \lambda_{12}, 0)\, d\lambda_{12} \\
\frac{\partial}{\partial \omega_{12}} \Psi_3(c_2, c_3, c_1, \omega_{23}, \omega_{12}, 0) &= \int_0^{\omega_{23}} \frac{-c_2 + \lambda_{23} c_3 + \omega_{12} c_1}{1 - \lambda_{23}^2 - \omega_{12}^2} \varphi_3(c_2, c_3, c_1, \lambda_{23}, \omega_{12}, 0)\, d\lambda_{23} \\
\frac{\partial}{\partial \omega_{23}} \Psi_3(c_2, c_3, c_1, \omega_{23}, \omega_{12}, 0) &= \int_0^{\omega_{12}} \frac{-c_2 + \omega_{23} c_3 + \lambda_{12} c_1}{1 - \omega_{23}^2 - \lambda_{12}^2} \varphi_3(c_2, c_3, c_1, \omega_{23}, \lambda_{12}, 0)\, d\lambda_{12}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial}{\partial c_1}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right) &= \int_0^{\omega_{12}}\int_0^{\omega_{13}}\Big\{\left[\left(1-\omega_{23}^2\right)c_1-\left(\lambda_{12}-\lambda_{13}\omega_{23}\right)c_2-\left(\lambda_{13}-\lambda_{12}\omega_{23}\right)c_3\right]^2 \\
&\qquad -\left(1-\omega_{23}^2\right)\left(1-\lambda_{12}^2-\lambda_{13}^2-\omega_{23}^2+2\lambda_{12}\lambda_{13}\omega_{23}\right)\Big\} \\
&\quad \cdot\frac{1}{\left(1-\lambda_{12}^2-\lambda_{13}^2-\omega_{23}^2+2\lambda_{12}\lambda_{13}\omega_{23}\right)^2}\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\lambda_{13},\omega_{23}\right)d\lambda_{12}d\lambda_{13} \\
\frac{\partial}{\partial c_2}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right) &= \int_0^{\omega_{13}}\int_0^{\omega_{12}}\frac{\partial}{\partial\lambda_{12}}\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\lambda_{13},\omega_{23}\right)d\lambda_{12}d\lambda_{13} \\
&= \int_0^{\omega_{13}}\varphi_3\left(c_1,c_2,c_3,\omega_{12},\lambda_{13},\omega_{23}\right)d\lambda_{13} \\
\frac{\partial}{\partial c_3}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right) &= \int_0^{\omega_{12}}\int_0^{\omega_{13}}\frac{\partial}{\partial\lambda_{13}}\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\lambda_{13},\omega_{23}\right)d\lambda_{13}d\lambda_{12} \\
&= \int_0^{\omega_{12}}\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\omega_{13},\omega_{23}\right)d\lambda_{12} \\
\frac{\partial}{\partial \omega_{12}}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right) &= \int_0^{\omega_{13}}\frac{-\left(1-\omega_{23}^2\right)c_1+\left(\omega_{12}-\lambda_{13}\omega_{23}\right)c_2+\left(\lambda_{13}-\omega_{12}\omega_{23}\right)c_3}{1-\omega_{12}^2-\lambda_{13}^2-\omega_{23}^2+2\omega_{12}\lambda_{13}\omega_{23}} \\
&\quad \cdot\varphi_3\left(c_1,c_2,c_3,\omega_{12},\lambda_{13},\omega_{23}\right)d\lambda_{13} \\
\frac{\partial}{\partial \omega_{13}}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right) &= \int_0^{\omega_{12}}\frac{-\left(1-\omega_{23}^2\right)c_1+\left(\lambda_{12}-\omega_{13}\omega_{23}\right)c_2+\left(\omega_{13}-\lambda_{12}\omega_{23}\right)c_3}{1-\lambda_{12}^2-\omega_{13}^2-\omega_{23}^2+2\lambda_{12}\omega_{13}\omega_{23}} \\
&\quad \cdot\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\omega_{13},\omega_{23}\right)d\lambda_{12} \\
\frac{\partial}{\partial \omega_{23}}\Psi_3\left(c_1,c_2,c_3,\omega_{12},\omega_{13},\omega_{23}\right) &= \int_0^{\omega_{12}}\int_0^{\omega_{13}}\frac{\partial^2}{\partial c_2\partial\lambda_{13}}\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\lambda_{13},\omega_{23}\right)d\lambda_{12}d\lambda_{13} \\
&= \int_0^{\omega_{12}}\frac{-\left(1-\omega_{13}^2\right)c_2+\left(\lambda_{12}-\omega_{13}\omega_{23}\right)c_1+\left(\omega_{23}-\lambda_{12}\omega_{13}\right)c_3}{1-\lambda_{12}^2-\omega_{13}^2-\omega_{23}^2+2\lambda_{12}\omega_{13}\omega_{23}} \\
&\quad \cdot\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\omega_{13},\omega_{23}\right)d\lambda_{12} \\
&= \int_0^{\omega_{12}}\int_0^{\omega_{13}}\frac{\partial^2}{\partial c_3\partial\lambda_{12}}\varphi_3\left(c_1,c_2,c_3,\lambda_{12},\lambda_{13},\omega_{23}\right)d\lambda_{12}d\lambda_{13} \\
&= \int_0^{\omega_{13}}\frac{-\left(1-\omega_{12}^2\right)c_3+\left(\lambda_{13}-\omega_{12}\omega_{23}\right)c_1+\left(\omega_{23}-\omega_{12}\lambda_{13}\right)c_2}{1-\omega_{12}^2-\lambda_{13}^2-\omega_{23}^2+2\omega_{12}\lambda_{13}\omega_{23}} \\
&\quad \cdot\varphi_3\left(c_1,c_2,c_3,\omega_{12},\lambda_{13},\omega_{23}\right)d\lambda_{13}.
\end{aligned}
$$

Additional Monte-Carlo evidence

We consider the following class of models

$$
\begin{aligned}
y_{i1} &= \mathbb{1}\left(y_{i1}^*>0\right) \\
y_{i2} &= \mathbb{1}\left(y_{i2}^*>0\right)
\end{aligned}
$$

where:

$$
\begin{aligned}
y_{i1}^* &= x_{i1}'\beta_1+u_{i1} && (11) \\
y_{i2}^* &= x_{i2}'\beta_2+y_{i1}\gamma_{21}+y_{i1}^*\alpha_{21}+u_{i2} && (12)
\end{aligned}
$$

$$
\begin{aligned}
x_{i1} &= \begin{pmatrix} 1 & x_{i1,0} & x_{i1,1}\end{pmatrix}' \\
x_{i2} &= \begin{pmatrix} 1 & x_{i2,0} & x_{i2,2}\end{pmatrix}'
\end{aligned}
$$

and:

$$
\begin{pmatrix} u_{i1} \\ u_{i2}\end{pmatrix}\sim\mathcal{N}\left(0,\Sigma\right),\ \Sigma=\begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1\end{pmatrix}
$$

Three models are studied by imposing some restrictions in the second equation: $M1)$ $\alpha_{21}=0$, $(M2)$ $\gamma_{21}=0$, and $(M3)$ $\gamma_{21}\neq 0$ and $\alpha_{21}\neq 0$.

Table 1: Bivariate probit Model ($M1$)

		EML			MSL (10 draws)			MSL (100 draws)		
True values		Estimates	Bias	RMSE	Estimates	Bias	RMSE	Estimates	Bias	RMSE
$N = 1,000$										
β_{1c}	1	1.0112	0.0112	0.0716	1.0117	0.0117	0.0721	1.0116	0.0116	0.0719
β_{11}	-0.5	-0.5046	-0.0046	0.0329	-0.5058	-0.0058	0.0331	-0.5052	-0.0052	0.0331
β_{2c}	-2	-2.0367	-0.0367	0.2132	-1.9291	0.0709	0.2179	-2.0109	-0.0109	0.2093
β_{22}	1	1.0093	0.0093	0.0648	1.0208	0.0208	0.0683	1.0120	0.0120	0.0654
γ_{21}	1	1.0403	0.0403	0.2363	0.8747	-0.1253	0.2486	1.0000	0.0000	0.2251
σ_{12}	-0.4	-0.4179	-0.0179	0.1757	-0.2294	0.1706	0.2252	-0.3735	0.0265	0.1709
$\sigma_{\beta_{1c}}$	0.0674	0.0677	0.0003	0.0045	0.0679	0.0005	0.0045	0.0678	0.0004	0.0045
$\sigma_{\beta_{11}}$	0.0312	0.0314	0.0002	0.0027	0.0314	0.0002	0.0027	0.0314	0.0002	0.0027
$\sigma_{\beta_{2c}}$	0.2004	0.1956	-0.0049	0.0223	0.1917	-0.0087	0.0193	0.1944	-0.0060	0.0222
$\sigma_{\beta_{22}}$	0.0664	0.0660	-0.0005	0.0076	0.0655	-0.0009	0.0077	0.0659	-0.0005	0.0077
$\sigma_{\gamma_{21}}$	0.2260	0.2179	-0.0082	0.0256	0.2011	-0.0249	0.0292	0.2137	-0.0123	0.0259
$\sigma_{\sigma_{12}}$	0.1740	0.1572	-0.0168	0.0340	0.1275	-0.0465	0.0494	0.1517	-0.0223	0.0349
Computing time			0.01			0.04			0.27	
$N = 10,000$										
β_{1c}	1	0.9978	-0.0022	0.0208	0.9985	-0.0015	0.0207	0.9979	-0.0021	0.0207
β_{11}	-0.5	-0.4997	0.0003	0.0096	-0.5010	-0.0010	0.0095	-0.5000	0.0000	0.0096
β_{2c}	-2	-2.0040	-0.0040	0.0696	-1.8859	0.1141	0.1315	-1.9926	0.0074	0.0696
β_{22}	1	1.0012	0.0012	0.0206	1.0105	0.0105	0.0228	1.0023	0.0023	0.0206
γ_{21}	1	1.0050	0.0050	0.0787	0.8308	-0.1692	0.1825	0.9876	-0.0124	0.0783
σ_{12}	-0.4	-0.4004	-0.0004	0.0565	-0.2128	0.1872	0.1916	-0.3822	0.0178	0.0575
$\sigma_{\beta_{1c}}$	0.0211	0.0211	0.0000	0.0004	0.0211	0.0000	0.0004	0.0211	0.0000	0.0004
$\sigma_{\beta_{11}}$	0.0100	0.0100	0.0000	0.0002	0.0100	0.0000	0.0002	0.0100	0.0000	0.0002
$\sigma_{\beta_{2c}}$	0.0639	0.0640	0.0000	0.0024	0.0624	-0.0016	0.0025	0.0638	-0.0001	0.0023
$\sigma_{\beta_{22}}$	0.0206	0.0206	0.0000	0.0008	0.0205	-0.0001	0.0008	0.0206	0.0000	0.0008
$\sigma_{\gamma_{21}}$	0.0714	0.0713	-0.0001	0.0027	0.0650	-0.0064	0.0066	0.0707	-0.0007	0.0027
$\sigma_{\sigma_{12}}$	0.0523	0.0518	-0.0005	0.0031	0.0406	-0.0117	0.0118	0.0510	-0.0013	0.0033
Computing time			0.10			0.47			9.36	

Note: The number of simulations is 1,000.

Table 2: Bivariate probit Model ($M2$)

		EML			MSL (10 draws)			MSL (100 draws)		
True values		Estimates	Bias	RMSE	Estimates	Bias	RMSE	Estimates	Bias	RMSE
$N = 1,000$										
β_{1c}	1	1.0042	0.0042	0.0629	1.0049	0.0049	0.0631	1.0044	0.0044	0.0629
β_{11}	-0.5	-0.5046	-0.0046	0.0336	-0.5048	-0.0048	0.0338	-0.5047	-0.0047	0.0337
β_{2c}	-2	-2.0227	-0.0227	0.1625	-1.8654	0.1346	0.1927	-1.8284	0.1716	0.2171
β_{22}	1	1.0105	0.0105	0.0646	0.9317	-0.0683	0.0910	0.9141	-0.0859	0.1043
α_{21}	0.2	0.2032	0.0032	0.0498	0.1863	-0.0137	0.0439	0.1825	-0.0175	0.0434
σ_{12}	-0.6	-0.6036	-0.0036	0.1007	-0.4819	0.1181	0.1471	-0.5795	0.0205	0.1011
$\sigma_{\beta_{1c}}$	0.0667	0.0665	-0.0002	0.0036	0.0667	0.0000	0.0037	0.0666	-0.0001	0.0036
$\sigma_{\beta_{11}}$	0.0319	0.0319	0.0000	0.0027	0.0320	0.0001	0.0027	0.0319	0.0000	0.0027
$\sigma_{\beta_{2c}}$	0.1594	0.1596	0.0002	0.0159	0.1480	-0.0114	0.0182	0.1445	-0.0149	0.0204
$\sigma_{\beta_{22}}$	0.0631	0.0633	0.0001	0.0070	0.0586	-0.0045	0.0079	0.0574	-0.0058	0.0086
$\sigma_{\alpha_{21}}$	0.0502	0.0498	-0.0004	0.0050	0.0461	-0.0041	0.0057	0.0451	-0.0051	0.0064
$\sigma_{\sigma_{12}}$	0.1130	0.1074	-0.0055	0.0134	0.1013	-0.0117	0.0156	0.1069	-0.0061	0.0142
Computing time			0.01			0.04			0.25	
$N = 10,000$										
β_{1c}	1	0.9996	-0.0004	0.0212	1.0002	0.0002	0.0212	0.9997	-0.0003	0.0212
β_{11}	-0.5	-0.5002	-0.0002	0.0103	-0.5004	-0.0004	0.0103	-0.5002	-0.0002	0.0103
β_{2c}	-2	-1.9964	0.0036	0.0508	-1.8459	0.1541	0.1600	-1.8060	0.1940	0.1983
β_{22}	1	1.0005	0.0005	0.0208	0.9251	-0.0749	0.0774	0.9052	-0.0948	0.0967
α_{21}	0.2	0.1990	-0.0010	0.0161	0.1836	-0.0164	0.0214	0.1799	-0.0201	0.0241
σ_{12}	-0.6	-0.6016	-0.0016	0.0305	-0.4830	0.1170	0.1206	-0.5930	0.0070	0.0308
$\sigma_{\beta_{1c}}$	0.0212	0.0212	0.0000	0.0004	0.0212	0.0000	0.0004	0.0212	0.0000	0.0004
$\sigma_{\beta_{11}}$	0.0098	0.0098	0.0000	0.0003	0.0099	0.0000	0.0003	0.0098	0.0000	0.0003
$\sigma_{\beta_{2c}}$	0.0495	0.0494	-0.0002	0.0017	0.0458	-0.0037	0.0040	0.0447	-0.0048	0.0051
$\sigma_{\beta_{22}}$	0.0202	0.0202	0.0000	0.0008	0.0187	-0.0014	0.0016	0.0183	-0.0019	0.0020
$\sigma_{\alpha_{21}}$	0.0161	0.0161	-0.0001	0.0005	0.0149	-0.0012	0.0013	0.0145	-0.0016	0.0016
$\sigma_{\sigma_{12}}$	0.0352	0.0349	-0.0003	0.0014	0.0325	-0.0027	0.0030	0.0347	-0.0004	0.0014
Computing time			0.10			0.48			10.00	

Note: The number of simulations is 1,000.

Table 3: Bivariate probit Model ($M3$)

True values		EML			MSL (10 draws)			MSL (100 draws)		
		Estimates	Bias	RMSE	Estimates	Bias	RMSE	Estimates	Bias	RMSE
$N = 10,000$										
β_{1c}	1	1.0015	0.0015	0.0249	1.0023	0.0023	0.0249	1.0019	0.0019	0.0249
β_{11}	-2	-2.0005	-0.0005	0.0431	-2.0023	-0.0023	0.0430	-2.0012	-0.0012	0.0431
β_{2c}	-1	-0.9897	0.0103	0.0944	-0.7165	0.2835	0.2932	-0.8719	0.1281	0.1539
β_{22}	2	2.0085	0.0085	0.0672	1.9094	-0.0906	0.1073	1.8269	-0.1731	0.1824
γ_{21}	0.5	0.4718	-0.0282	0.1774	0.0060	-0.4940	0.5107	0.3737	-0.1263	0.2076
α_{21}	0.2	0.2059	0.0059	0.0398	0.2666	0.0666	0.0738	0.1959	-0.0041	0.0355
σ_{12}	-0.6	-0.5901	0.0099	0.0672	-0.3845	0.2155	0.2208	-0.5639	0.0361	0.0769
$\sigma_{\beta_{1c}}$	0.0250	0.0251	0.0000	0.0007	0.0251	0.0001	0.0007	0.0251	0.0000	0.0007
$\sigma_{\beta_{11}}$	0.0412	0.0413	0.0000	0.0014	0.0413	0.0001	0.0014	0.0413	0.0000	0.0014
$\sigma_{\beta_{2c}}$	0.1006	0.0996	-0.0010	0.0121	0.0790	-0.0216	0.0220	0.0902	-0.0104	0.0150
$\sigma_{\beta_{22}}$	0.0635	0.0631	-0.0004	0.0041	0.0556	-0.0078	0.0085	0.0566	-0.0069	0.0077
$\sigma_{\gamma_{21}}$	0.1921	0.1888	-0.0033	0.0272	0.1317	-0.0604	0.0610	0.1694	-0.0227	0.0328
$\sigma_{\alpha_{21}}$	0.0400	0.0393	-0.0007	0.0045	0.0299	-0.0101	0.0103	0.0352	-0.0048	0.0061
$\sigma_{\sigma_{12}}$	0.0775	0.0765	-0.0010	0.0134	0.0567	-0.0208	0.0214	0.0765	-0.0010	0.0131
Computing time			0.11			0.50			9.88	

Note: The number of simulations is 1,000.

Figure 1: Density function of α_{21} in the bivariate probit model ($M4$)

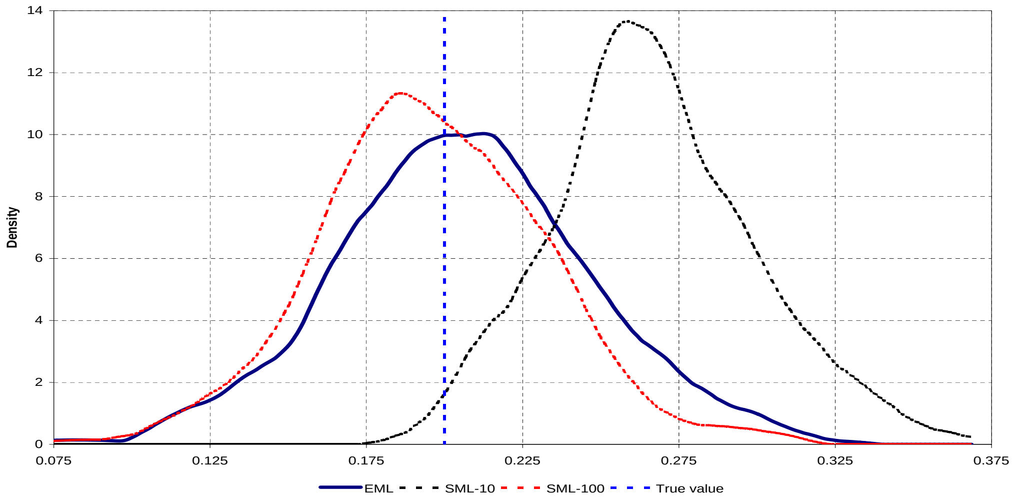


Figure 2: Density function of $\sigma_{\alpha_{21}}$ in the bivariate probit model ($M4$)

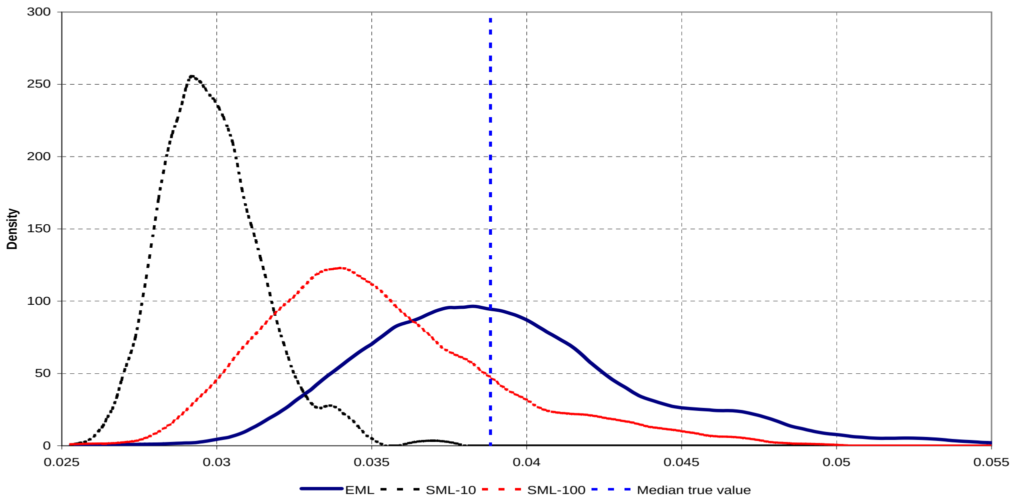


Figure 3: Density function of γ_{21} in the bivariate probit model ($M4$)

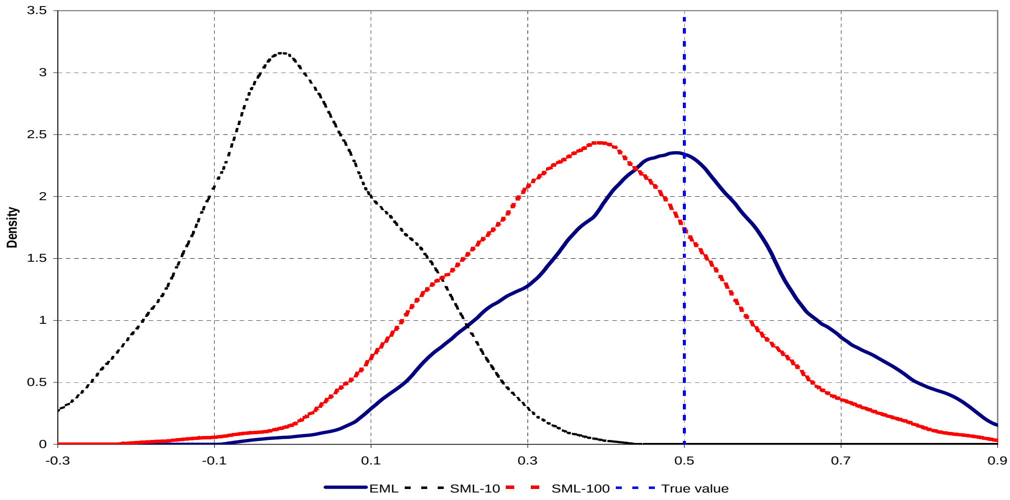
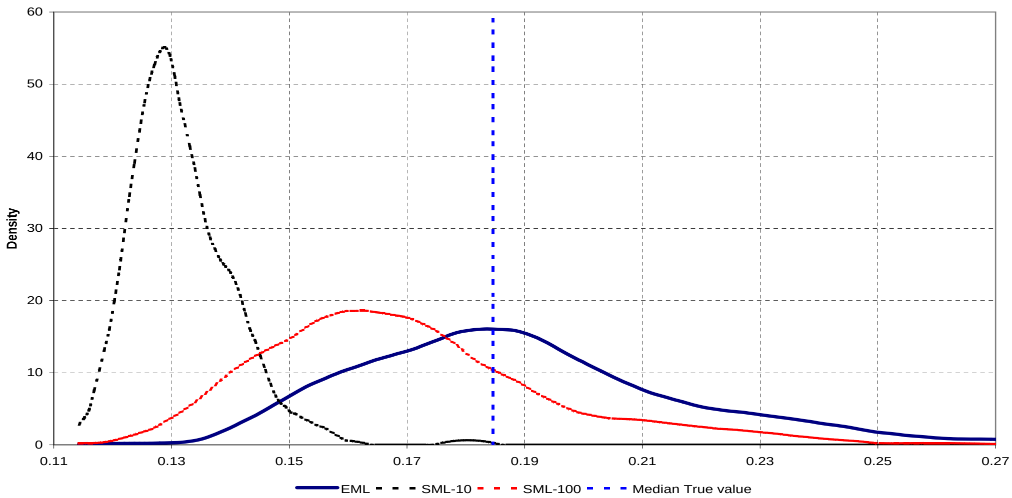


Figure 4: Density function of $\sigma_{\gamma_{21}}$ in the bivariate probit model ($M4$)



Institute of Health Economics and Management (IEMS)
UNIL Dorigny
Extranef
1015 Lausanne
Switzerland
Phone +41 (0)21 692 33 20
Fax +41 (0)21 692 36 55