N° 9414

MULTIVARIATE DISTRIBUTIONS FOR LIMITED DEPENDENT VARIABLE MODELS

C. GOURIEROUX*
F. JOUNEAU**

May 18, 1994

* CREST and CEPREMAP

** CREST and LILLE III

Abstract

Multivariate Distributions for Limited Dependent Variable Models

In limited dependent variable modelling the multivariate normal distribution is generally used for the latent models but in practice may lead to multiple integrals which have to be approximated by simulation techniques. In this paper we discuss multivariate distributions with tractable forms, given marginal distributions, and possible correlations. In particular we study some multivariate extensions of the logistic distribution and explain how it may be used to approximate gaussian limited dependent variable models.

Résumé

Lois multivariées pour les modèles à variables dépendantes limitées

Les modèles à variables dépendantes limitées sont généralement construits en supposant les erreurs de loi normale, mais, en pratique, ceci peut conduire à des intégrales multiples qui doivent être approchées par des techniques de simulation. Dans ce papier, nous introduisons des distributions multivariées, de formes maniables, de lois marginales données et compatibles avec la présence de corrélations. En particulier nous étudions des extensions multivariées de la loi logistique et expliquons comment elles peuvent être utilisées pour approcher la loi normale dans les modèles à variables dépendantes limitées.

keywords: Multivariate logistic distribution, limited dependent variable models, selectivity bias.

mots clés : Distribution logistique multivariée, modèles à variables dépendantes limitées, biais de sélection.

J.E.L: 211

1 Introduction

Limited dependent variable models are commonly introduced for analysing questions as different as multiple choices, intertemporal qualitative decisions, competing risks, disaggregated markets with rationing [see e.g. Mc Fadden (1973), (1976), Maddala (1983), Kiefer (1988), Laroque-Salanié (1989), Gouriéroux (1989)]. In such models the distribution of the observed variables $\tilde{Y}_t$, $t = 1, \ldots, T$, are defined in two steps. We first exhibit a relationship between the observed variables $\tilde{Y}_t$ and some latent variables Y_t : $\tilde{Y}_t = h(Y_t)$. In practice, the size of the latent vector is the size of the observable vector multiplied by a factor between two and more than fifty, and the transformation $h(.)$ often describes some threshold conditions (censoring, min conditions...). Secondly, we specify a parametric model for the latent variables. The observable model is derived from the latent one by applying the transformation h.

However this natural approach is not so easy to implement as soon as a large number of latent variables are correlated. It is necessary to introduce a latent distribution allowing for these correlations, generally a multivariate normal distribution, but even with this choice, the log-likelihood contains multiple integrals with large dimensions and is often numerically untractable.

This difficulty is solved in the literature by using estimation methods different from the Full Information Maximum Likelihood approach. These methods may be partial M.L. approaches, for instance in qualitative choice models, when the number of alternatives is artificially reduced for estimation purpose (this idea is the basis of the papers by Ashford-Sowden (1970) and Hausman-Mc Fadden (1984)). They may be Pseudo-Maximum Likelihood methods in which in a first step the correlations are constrained to be zero [Robinson (1982), Gouriéroux, Monfort and Trognon (1984), Dagenais (1982, 1986)]. Alternative strategies are provided by Prentice-Zhao [1990] or Liang-Zeger [1986], who specify tractable functions for the marginal moments depending on parameters which are in one to one relationship with the true parameters. Finally, a lot of the literature is currently devoted to simulated estimation methods [Mc Fadden (1989), Pakes-Pollard (1989), Gouriéroux-Monfort(1993)].

The approach of this paper is a completely different one. We directly introduce multivariate latent models leading to tractable log-likelihood functions.

In Section 2, we present the multivariate family with given marginal distributions and cross correlations initially introduced by Gumbel (1961), we discuss its properties concerning marginalisation and conditioning, and we exhibit the cross moments. In section 3 we consider different limited dependent variable models based on the previous family for analysing choices among non mutually exclusive alternatives, double hurdle type problems, competing risks, qualitative time series. In section 4 we consider the behavior of these models in a neighbourhood of the no correlation hypothesis, we exhibit the analogue of the Durbin Watson statistic and we propose a local estimator for the correlation parameter. Another class of multivariate family with given marginal distribution and cross correlation, more suitable for describing sequential processes, is introduced in section 5. It gives a simple way for correcting for the successive selectivity biases. Finally in section 6 we study the properties of these multivariate extensions considered as proxies for the multivariate normal distribution. Some proofs are gathered in the appendices.

2 Analysis of the multivariate family

2.1 Definition

Let us consider p univariate continuous distributions with p.d.f f_i, c.d.f F_i and survival functions S_i respectively.

Property 2.1

$$G(y_1,\ldots,y_p) = \prod_{j=1}^{p} F_j(y_j)\left(1+\sum_{i<j} a_{i,j} S_i(y_i) S_j(y_j)\right),$$

where the real numbers $a_{i,j}$, $i<j$ are such that $\sum_{i<j} \mid a_{i,j} \mid \leq 1$, is the c.d.f of a p-variate continuous distribution. The marginal distributions associated with G are the initial distributions F_i, $i=1,\ldots,p$, and the p.d.f corresponding to G is given by:

$$\begin{aligned}\frac{\partial^p G}{\partial y_1 \ldots \partial y_p}(y_1 \ldots y_p) &= g(y_1, \ldots, y_p) \\ &= \prod_{j=1}^{p} f_j(y_j) \left(1 + \sum_{i<j} a_{i,j}(2F_i(y_i) - 1)(2F_j(y_j) - 1)\right).\end{aligned}$$

Proof:
i) We have:

$$G(y_1, \ldots, y_p) = \prod_{j=1}^{p} F_j(y_j) + \sum_{i<j} a_{i,j} \left(F_i(y_i)S_i(y_i)\right)\left(F_j(y_j)S_j(y_j)\right) \prod_{k \neq i,j} F_k(y_k).$$

Therefore we deduce directly the expression of the derivative $\frac{\partial^p G}{\partial y_1 \ldots \partial y_p}(y_1, \ldots, y_p)$, since $\frac{\partial}{\partial y}\left[F(y)S(y)\right] = f(y)\left(2F(y) - 1\right)$.

ii) Since $F_i(y)$ lies between 0 and 1, $2F_i(y) - 1$ lies between -1 and 1. We conclude that the function g is non negative as soon as $\sum_{i<j} \mid a_{i,j} \mid \leq 1$.

iii) Finally we have to check some limit conditions. We get:

$$\lim_{y_i \to -\infty} G(y_1, \ldots, y_p) = 0,$$

$$\lim_{y_i \to +\infty} G(y_1, \ldots, y_p) = \prod_{j \neq i} F_j(y_j).$$

Therefore we deduce that G defines a p-variate distribution with marginal distributions F_i, $i = 1, \ldots, p$. *Q.E.D.*

This multivariate extension of univariate distributions has already been introduced in the literature, especially for defining multivariate logistic distributions [Gumbel (1961), Sawa (1984), Thelot (1985)]. The main advantages of this extension are the tractable form of the cumulative distribution function, the stability of the family by marginalizing, the symmetry conditions ... Its main drawback concerns some limitations of the admissible values of the correlations. We now discuss these different aspects.

Property 2.2 *If the marginal distributions F_i, $i = 1, \ldots, p$ are symmetric, the multivariate extension is also a symmetric distribution.*

Proof:
Indeed we have:

$$g(-y_1,\dots,-y_p)=\prod_{j=1}^{p} f_j(-y_j)\left(1+\sum_{i<j} a_{i,j}(2F_i(-y_i)-1)(2F_j(-y_j)-1)\right),$$

and the property follows from: $f_i(-y)=f_i(y)$, $2F_i(-y)-1=1-2F_i(y)$.
Q.E.D.

Property 2.3 *When $[Y_1,\dots,Y_p]$ follows the multivariate distribution extension G given in property 2.1, the subvector $[Y_1,\dots,Y_q]$, $q\le p$ follows the multivariate extension:*

$$G(y_1,\dots,y_q)=\prod_{j=1}^{q} F_j(y_j)\left(1+\sum_{i<j\le q} a_{i,j}S_i(y_i)S_j(y_j)\right).$$

Proof:
The proof is obvious by putting $y_{q+1}=\dots=y_p=+\infty$ in the expression of $G(y_1,\dots,y_p)$. *Q.E.D.*

This invariance by marginalizing implies that the conditional distributions have also a tractable form.

For instance, the p.d.f of $[Y_1,\dots,Y_q]$ given $[Y_{q+1},\dots,Y_p]$ is:

$$g(y_1,\dots,y_q\mid y_{q+1},\dots,y_p)=\frac{g(y_1,\dots,y_p)}{g(y_{q+1},\dots,y_p)}$$

$$=\prod_{j=1}^{q} f_j(y_j)\frac{1+\sum_{i<j} a_{i,j}(2F_i(y_i)-1)(2F_j(y_j)-1)}{1+\sum_{q+1\le i<j} a_{i,j}(2F_i(y_i)-1)(2F_j(y_j)-1)},$$

and the associated c.d.f is:

$$G(y_1,\dots,y_q\mid y_{q+1},\dots,y_p)=P[Y_1<y_1,\dots,Y_q<y_q\mid Y_{q+1}=y_{q+1},\dots,Y_p=y_p]$$

$$=\prod_{j=1}^{q} F_j(y_j)\frac{1+SS_j+SF_j+FF_j}{1+\sum_{q+1\le i<j} a_{i,j}(2F_i(y_i)-1)(2F_j(y_j)-1)},$$

where:

$$\begin{aligned}
SS_j &= \sum_{i<j\leq q} a_{i,j} S(y_i) S(y_j), \\
SF_j &= \sum_{i\leq q<j} a_{i,j} S(y_i)(2F_j(y_j)-1), \\
FF_j &= \sum_{q<i<j} a_{i,j}(2F_i(y_i)-1)(2F_j(y_j)-1).
\end{aligned}$$

The form of the extension is also suitable for the determination of cross moments. Indeed we have:

$$E\left[\prod_{j=1}^{p} Y_j^{c_j}\right] =$$
$$\prod_{j=1}^{p} E\left[Y_j^{c_j}\right] + \sum_{i<j} a_{i,j} E\left[Y_i^{c_i}(2F_i(Y_i)-1)\right] E\left[Y_j^{c_j}(2F_j(Y_j)-1)\right] \prod_{k\neq i,j} E\left[Y_k^{c_k}\right],$$

and these moments are known as soon as we have the expressions of the marginal moments $E[Y_j^{c_j}]$, $E\left[Y_j^{c_j}(2F_j(Y_j)-1)\right]$. From the previous formula, we directly deduce the expression of the correlations between the different variables. Note that the marginal moments of this distribution are known for all orders.

The last two results (*i.e* explicit expression of the moments and invariance by marginalization) emphasize the differences of our approach with that presented in papers by Prentice-Zhao [1990] or Liang-Zeger [1986]. They generalize the logisitc distribution to multivariate models by imposing special forms for the conditional distributions. However, marginal caracteristics of the distributions they propose are difficult to compute. For example they do not provide explicit expressions for the correlation between the latent variables of the model.

On the contrary, we directly deduce from the previous formula the expression of the correlation between Y_i and Y_j.

Property 2.4 *The correlations are given by:*

$$\begin{aligned}
\rho_{i,j} &= Corr(Y_i, Y_j) \\
&= a_{i,j} \frac{E[Y_i(2F_i(Y_i)-1)]E[Y_j(2F_j(Y_j)-1)]}{\sqrt{VY_i}\sqrt{VY_j}}.
\end{aligned}$$

If the reduced centered variables $[Y_i - EY_i][VY_i]^{-\frac{1}{2}}$ follow a same distribution F, we deduce that the correlations are given by:

$$\rho_{i,j} = a_{i,j} E_F \left(Y[2F(Y) - 1]\right)^2 = 4a_{i,j} E_F \left(YF(Y)\right)^2 . \tag{1}$$

They are proportional to the basic coefficient $a_{i,j}$ with a multiplicative factor depending on the form of the marginal distribution F. Since the coefficients $a_{i,j}$ verify the constraint $\sum_{i<j} \mid a_{i,j} \mid \leq 1$, the correlations satisfy:

$$\sum_{i<j} \mid \rho_{i,j} \mid < 4E_F \left(YF(Y)\right)^2 . \tag{2}$$

This constraint may be a strict constraint on the correlation imposing that the different correlations are smaller than a value stricly smaller than one. To have an idea of this effect we give below the values of the multiplicative factor $4E_F \left(YF(Y)\right)^2$ for some useful distributions (see appendix 1):

Table 1: Multiplicative factor

Marginal Distribution	Normal	Logistic	Exponential	Uniform
Maximal correlation	$\frac{1}{\pi}$	$\frac{3}{\pi^2}$	$\frac{1}{4}$	$\frac{1}{3}$

Some other interpretations of the coefficients $a_{i,j}$ in terms of correlations may also be found. For instance let us introduce the qualitative variables $Z_1(y_1) = \mathbf{1}_{Y_1<y_1}$, $Z_2(y_2) = \mathbf{1}_{Y_2<y_2}$. We have:

$$\begin{aligned} Cov\left(Z_1(y_1), Z_2(y_2)\right) &= G_2(y_1, y_2) - F_1(y_1)F_2(y_2) \\ &= a_{12}F_1(y_1)F_2(y_2)S_1(y_1)S_2(y_2). \end{aligned}$$

Therefore we directly get:

$$Corr\left(Z_1(y_1), Z_2(y_2)\right) = a_{1,2}. \tag{3}$$

The previous model is characterized by the constancy in y_1, y_2, of the correlations between the qualitative variables $Z_1(y_1), Z_2(y_2)$.

Another interpretation comes from the tail properties of the distribution g. Indeed if $\mid y_1 \mid \to +\infty, \ldots, \mid y_p \mid \to +\infty$, we get:

$$lim\ g(y_1, \ldots, y_p) = \prod_{j=1}^{p} f(y_j) \left(1 + \sum_{i<j} a_{i,j}(-1)^{\epsilon_i+\epsilon_j}\right),$$

where $\epsilon_j = 0$ if $y_j \to +\infty$ and $\epsilon_j = 1$ if $y_j \to -\infty$. Therefore the $a_{i,j}$ coefficients give the modifications of the tail of the distribution due to the correlation when the arguments become large.

The previous extension is not the only one which has been proposed in the literature to get multivariate distributions with given marginal distributions and possible dependence. For instance Genest-Rivets[1993] introduce Archimedean Copula defined by:

$$G(y_1, y_2) = \psi^{-1}\left[\psi(y_1) + \psi(y_2)\right],$$

where $\psi(.)$ is some convex decreasing function satisfying $\psi(1) = 0$ to extend marginal uniform distributions on $[0,1]$. This extension is such that : $U = \frac{\psi(Y_1)}{\psi(Y_1)+\psi(Y_2)}$ and $V = G(Y_1, Y_2)$ are independent variables.

Despite the limitations on the correlations, the extension which we have considered seems preferable because of its tractability but also for some interpretation in terms of correction for selectivity biases, a property which will be seen later on.

3 Application to limited dependent variable models

3.1 Choices among non mutually exclusive alternatives

When describing discrete choice models it is usual to assume mutually exclusive alternatives and to explain the decision through some comparison of underlying utility levels [see e.g. McFadden(1973)(1976), Maddala(1983), Gouriéroux (1989), for general presentations]. However this approach does not apply to the case of non mutually exclusive alternatives. Let us consider for instance the analysis of asset mix by the agents. From individual data it is easily seen that a given agent does not generally introduce in his portfolio every existing assets [see *e.g* Lollivier-Verger (1987) for a study on French data]. Even if this observation is not compatible with the usual portfolio choice theory, such a behavior is usually explained by the level of information of the agent, the transaction costs, the liquidity constraints...

Let us assume that if p different assets are available, the individual portfolios may or not contain a positive quantity of each asset j ($j = 1, \dots, p$). We denote by Z_j the indicator for the presence of asset j in the portfolio. To define the distribution of the Z variables, we first introduce some latent variables Y_j, $j = 1, \dots, p$ with an extended multivariate distribution G and such that:

$$Z_j = 1, \ if \ Y_j < y_j^0, \tag{4}$$

where y_j^0 is some threshold.

The following property is proved in appendix 2:

Property 3.1 *We have:*

$$P[Z_j = z_j, \ j = 1, \dots, p]$$

$$= \prod_{j=1}^{p} \left[F_j(y_j^0)^{z_j} S_j(y_j^0)^{1-z_j}\right] \left(1 + \textstyle\sum_{i<j\leq p} (-1)^{z_i+z_j} a_{i,j} S_i(y_i^0)^{z_i} F_i(y_i^0)^{1-z_i} S_j(y_j^0)^{z_j} F_j(y_j^0)^{1-z_j}\right).$$

Therefore these probabilities only depend on univariate integrals.

When explanatory variables and error terms are introduced in the latent model:

$$Y_j = X_j b_j + u_j,$$

where $(u_1, \dots, u_p)$ has an extended multivariate distribution based on the same centered distribution F, the elementary probabilities are:

$$P[Z_j = z_j, \ j = 1, \dots, p]$$

$$= \prod_{j=1}^{p} \left[F(-x_j b_j)^{z_j} S(-x_j b_j)^{1-z_j}\right] \left(1 + \textstyle\sum_{i<j\leq p} (-1)^{z_i+z_j} a_{i,j} S(-x_i b_i)^{z_i} F(-x_i b_i)^{1-z_i} S(-x_j b_j)^{z_j} F(-x_j b_j)^{1-z_j}\right). \tag{5}$$

3.2 Mixed qualitative-quantitative observations

Let us continue with the example of the portfolio choice. In the previous subsection we study the model associated with the kind of assets introduced in the portfolio. With more precise data we may know the quantities of these assets. Since some of these quantities are equal to zero, we have to consider models with a point mass at the origin. For instance, we may define the observations by censoring:

$$\tilde{Y}_j = \begin{cases} Y_j & if\ Y_j > y_j^0, \\ 0 & otherwise. \end{cases} \tag{6}$$

The form of the distribution of these observed variables is given in the following property, which is proved in appendix 2.

Property 3.2 *We have:*

$$\begin{aligned} &\Psi(y_1, \ldots, y_q, 0, \ldots, 0) \\ = &\lim_{\substack{h_1 \to 0 \\ \vdots \\ h_q \to 0}} \frac{P\left[y_1 < Y_1 < y_1 + h_1, \ldots, y_q < Y_q < y_q + h_q, Y_{q+1} < y_{q+1}^0, \ldots, Y_p < y_p^0\right]}{h_1 \ldots h_q} \\ = &\textstyle\prod_{j=1}^q f_j(y_j) \prod_{j=q+1}^p F_j(y_j^0) \quad \Big[1 + \sum_{i<j} a_{i,j} \left((2F_i(y_i) - 1)\mathbf{1}_{i \leq q} + S_i(y_i^0)\mathbf{1}_{q+1 \leq i}\right) \\ &\qquad\qquad\qquad\qquad\qquad \left((2F_j(y_j) - 1)\mathbf{1}_{j \leq q} + S_j(y_j^0)\mathbf{1}_{q+1 \leq j}\right)\Big]. \end{aligned}$$

3.3 Competing risks

The extended multivariate family may also be used in duration models with competing risks. Let us for instance consider the bivariate case where Y_1, Y_2 denote the two duration variables. $Y = \min(Y_1, Y_2)$ is the duration before the first occurrence, and $Z = \mathbf{1}_{Y_1 < Y_2}$ the indicator of the first risk occurred.

If the c.d.f of (Y_1, Y_2) is:

$$G(y_1, y_2) = F_1(y_1)F_2(y_2)\left(1 + aS_1(y_1)S_2(y_2)\right),$$

the hazard function associated with the first variable Y_1 is given by:

$$
\begin{aligned}
\tilde{\lambda}_1(y) &= \lim_{h\to 0}\frac{1}{h}P\left[y < Y_1 < y+h \mid Y_1 > y, Y_2 > y\right] \\
&= \lim_{h\to 0}\frac{1}{h}\frac{P\left[y < Y_1 < y+h, Y_2 > y\right]}{P\left[Y_1 > y, Y_2 > y\right]} \\
&= \frac{\frac{\partial}{\partial y_1}\left(P\left[Y_1 < y_1, Y_2 > y\right]\right)_{y_1=y}}{P\left[Y_1 > y, Y_2 > y\right]} \\
&= \frac{\frac{\partial}{\partial y_1}\left(F_1(y_1)S_2(y)\left(1 - aS_1(y_1)F_2(y)\right)\right)_{y_1=y}}{S_1(y)S_2(y)\left(1 + aF_1(y)F_2(y)\right)} \\
&= \frac{f_1(y)S_2(y)\left(1 + a(2F_1(y)-1)F_2(y)\right)}{S_1(y)S_2(y)\left(1 + aF_1(y)F_2(y)\right)}.
\end{aligned}
$$

Thus,

$$
\tilde{\lambda}_1(y) = \lambda_1(y)\frac{1 + a(2F_1(y)-1)F_2(y)}{1 + aF_1(y)F_2(y)}, \tag{7}
$$

where λ_1 is the marginal hazard function associated with Y_1. The correcting factor takes into account the censoring effect due to Y_2. This term equals:

$$
1 + a\frac{(F_1(y)-1)F_2(y)}{1 + aF_1(y)F_2(y)},
$$

tends to 1 asymptotically ($\mid y \mid \to +\infty$), and has always the same position with respect to one: larger than one if a is negative, smaller otherwise.

Let us now assume that we observe the data Y with censoring d and the risk characteristics Z of the first occurrence. To compute the log-likelihood function, we have to distinguish three regimes:

i) the observation is not censored and corresponds to risk number 1. If we denote:

$$\Psi(y,1) = \lim_{h\to 0} \frac{1}{h} P\left[y < Y_1 < y+h \ , \ Z=1\right], \ y < d,$$

we get:

$$\begin{aligned}\Psi(y,1) &= \lim_{h\to 0} \frac{1}{h} P\left[y < Y_1 < y+h, y < Y_2\right] \\ &= \tfrac{\partial}{\partial y_1} \left(P\left[Y_1 < y_1, y < Y_2\right]\right)_{y_1=y} \\ &= f_1(y)S_2(y)\left(1 + a(2F_1(y)-1)F_2(y)\right);\end{aligned}$$

ii) the observation is not censored and corresponds to risk number 2. If we denote:

$$\Psi(y,0) = \lim_{h\to 0} \frac{1}{h} P\left[y < Y_1 < y+h \ , \ Z=0\right], \ y < d,$$

we get:

$$\Psi(y,0) = S_1(y)f_2(y)\left(1 + aF_1(y)(2F_2(y)-1)\right);$$

iii) finally, the observation may be censored and the contribution to the log-likelihood corresponds to:

$$\begin{aligned}P[Y > d] &= P\left[Y_1 > d, Y_2 > d\right] \\ &= S_1(d)S_2(d)\left(1 + aF_1(d)F_2(d)\right).\end{aligned}$$

Then the log-likelihood function corresponding to one observation has the tractable form:

$$\begin{aligned}L &= \mathbf{1}_{y<d}\ z \log\left(f_1(y)S_2(y)\left(1 + a(2F_1(y)-1)F_2(y)\right)\right) \\ &+\mathbf{1}_{y<d}\ (1-z)\log\left(S_1(y)f_2(y)\left(1 + a(2F_2(y)-1)F_1(y)\right)\right) \\ &+\mathbf{1}_{y>d}\ \log\left(S_1(d)S_2(d)\left(1 + aF_1(d)F_2(d)\right)\right).\end{aligned}$$

$$
\begin{aligned}
L \; &= \mathbf{1}_{y_1<\min(y_2,d)} \log f_1(y) + \left(1 - \mathbf{1}_{y_1<\min(y_2,d)}\right) \log S_1(\min(y_2,d)) \\
&+\mathbf{1}_{y_2<\min(y_1,d)} \log f_2(y) + \left(1 - \mathbf{1}_{y_2<\min(y_1,d)}\right) \log S_2(\min(y_1,d)) \\
&+\mathbf{1}_{y_1<\min(y_2,d)} \log\left(1 + aF_2(y_1)(2F_1(y_1)-1)\right) \qquad (8) \\
&+\mathbf{1}_{y_2<\min(y_1,d)} \log\left(1 + aF_1(y_2)(2F_2(y_2)-1)\right) \\
&+\mathbf{1}_{d<\min(y_1,y_2)} \log\left(1 + aF_1(d)F_2(d)\right).
\end{aligned}
$$

4 The models in a neighbourhood of the no correlation hypothesis

4.1 Testing for no correlation

It is interesting to study if the previous models have nice properties for testing the no correlation hypothesis $H_0 = \{a_{i,j} = 0, \ \forall\, i < j\}$. This question is of course strongly related to the problem of selectivity bias [Heckman (1974),(1979), Melino (1982), Olsen (1980), Vella(1992),(1993)].

Let us consider the Lagrange multiplier approach applied to a qualitative model of the type introduced in 3.1 . If we have T independent observations $(z_{1,t},\dots,z_{p,t})$, $t = 1,\dots,T$, the log-likelihood function is given by:

$$
\begin{aligned}
L(a,b) \; &= \sum_t \log P\left[Z_{j,t} = z_{j,t}, \ j = 1,\dots,p\right] \\
&= \sum_t \log\left[F(-x_{j,t}b_j)^{z_{j,t}} S(-x_{j,t}b_j)^{1-z_{j,t}}\right] \\
&+ \log[1 + \sum_{1<j\leq p} (-1)^{z_{i,t}+z_{j,t}} a_{i,j} \\
&F(-x_{i,t}b_i)^{1-z_{i,t}} S(-x_{i,t}b_i)^{z_{i,t}} F(-x_{j,t}b_j)^{1-z_{j,t}} S(-x_{j,t}b_j)^{z_{j,t}}].
\end{aligned}
$$

The score associated with the correlation parameter $a_{i,j}$ and evaluated at $a = 0$ is given by:

$\frac{\partial L}{\partial a_{i,j}}(0,b) =$

$$
\begin{aligned}
&\sum_t (-1)^{z_{i,t}+z_{j,t}} a_{i,j} F(-x_{i,t}b_i)^{1-z_{i,t}} S(-x_{i,t}b_i)^{z_{i,t}} F(-x_{j,t}b_j)^{1-z_{j,t}} S(-x_{j,t}b_j)^{z_{j,t}} \\
&= \sum_t z_{i,t} z_{j,t} S(-x_{i,t}b_i) S(-x_{j,t}b_j) + \sum_t (1-z_{i,t})(1-z_{j,t}) F(-x_{i,t}b_i) F(-x_{j,t}b_j) \\
&+ \sum_t -z_{i,t}(1-z_{j,t}) S(-x_{i,t}b_i) F(-x_{j,t}b_j) + \sum_t -z_{j,t}(1-z_{i,t}) S(-x_{j,t}b_j) F(-x_{i,t}b_i) \\
&= \sum_t \left(z_{i,t} S(-x_{i,t}b_i) + (1-z_{i,t}) F(-x_{i,t}b_i)\right) \left(z_{j,t} S(-x_{j,t}b_j) + (1-z_{j,t}) F(-x_{j,t}b_j)\right) \\
&= \sum_t \left(z_{i,t} - F(-x_{i,t}b_i)\right) \left(z_{j,t} - F(-x_{j,t}b_j)\right).
\end{aligned}
$$

Therefore the score evaluated at $a = 0$ and $b = \hat{b}_0$, where $\hat{b}_0$ is the M.L. estimator under the null hypothesis is the second order empirical cross moment of the qualitative residuals:

$$
\frac{\partial L}{\partial a_{i,j}}(0, \hat{b}_0) = \sum_t \hat{v}_{i,t}^0 \hat{v}_{j,t}^0, \tag{9}
$$

where $\hat{v}_{i,t}^0 = z_{i,t} - F(-x_{i,t}\hat{b}_{i,0})$.

The same kind of results applies to the other kinds of censored models. For instance in the case of duration models for competing risks, the individual score evaluated under the null hypothesis $\{a = 0\}$ is (see 8):

$$
\begin{aligned}
\frac{\partial L}{\partial a}(0) \quad &= \mathbf{1}_{Y_1 < \min(d, Y_2)} \left(2F_1(Y_1) - 1\right) F_2(Y_1) \\
&+ \mathbf{1}_{Y_2 < \min(Y_1, d)} \left(2F_2(Y_2) - 1\right) F_1(Y_2) \\
&+ \mathbf{1}_{d < \min(Y_1, Y_2)} F_1(d) F_1(d).
\end{aligned}
$$

However, as usual [see Gouriéroux, Monfort and Trognon (1984)], the interpretation in terms of residual is now lost, because of the constraints defining the regimes. Indeed these constraints do not have a cross form $(Y_1 \in A_1) \times (Y_2 \in A_2)$ in the two variables, but depend on some transformed functions such as $Y_1 - Y_2$.

4.2 Local estimators of the correlation coefficients

Before determining the maximum likelihood estimator $\hat{a}$ of the dependence coefficient a, it may be interesting to introduce an approximated estimator $\tilde{a}$, which is asymptotically equivalent to the unconstrained one $\hat{a}$ in the neighbourhood of the null hypothesis.

If $\hat{b}_0$ denotes the constrained estimator of the other parameters, such an approximation is given by:

$$\tilde{a} = -\left(\frac{\partial^2 L}{\partial a \partial a'}(0, \hat{b}_0)\right)^{-1} \frac{\partial L}{\partial a}(0, \hat{b}_0). \tag{10}$$

For instance for the qualitative model, we get:

$$\frac{\partial L}{\partial a_{i,j}}(a, b) =$$

$$\sum_t (-1)^{z_{i,t}+z_{j,t}} F(-x_{i,t}b_i)^{1-z_{i,t}} S(-x_{i,t}b_i)^{z_{i,t}} F(-x_{j,t}b_j)^{1-z_{j,t}} S(-x_{j,t}b_j)^{z_{j,t}}$$

$$+\left[1 + \sum_{1<j\leq p} (-1)^{z_{i,t}+z_{j,t}} a_{i,j} F(-x_{i,t}b_i)^{1-z_{i,t}} S(-x_{i,t}b_i)^{z_{i,t}} F(-x_{j,t}b_j)^{1-z_{j,t}} S(-x_{j,t}b_j)^{z_{j,t}}\right]^{-1}.$$

Therefore:

$$\frac{\partial^2 L}{\partial a_{i,j} \partial a_{k,l}}(0, b) =$$

$$\sum_t (-1)^{z_{i,t}+z_{j,t}+z_{k,t}+z_{l,t}+1} F(-x_{i,t}b_i)^{1-z_{i,t}} S(-x_{i,t}b_i)^{z_{i,t}} F(-x_{j,t}b_j)^{1-z_{j,t}} S(-x_{j,t}b_j)^{z_{j,t}}$$
$$F(-x_{k,t}b_k)^{1-z_{k,t}} S(-x_{k,t}b_i)^{z_{k,t}} F(-x_{l,t}b_l)^{1-z_{l,t}} S(-x_{l,t}b_j)^{z_{l,t}}.$$

In the particular case of two alternatives 1,2, this approximation $\tilde{a}$ is given by:

$$\tilde{a} = \frac{\sum_{\mathcal{I}_{11}} \hat{S}_{1,t}\hat{S}_{2,t} + \sum_{\mathcal{I}_{00}} \hat{F}_{1,t}\hat{F}_{2,t} - \sum_{\mathcal{I}_{10}} \hat{S}_{1,t}\hat{F}_{2,t} - \sum_{\mathcal{I}_{01}} \hat{F}_{1,t}\hat{S}_{2,t}}{\sum_{\mathcal{I}_{11}} \hat{S}_{1,t}^2\hat{S}_{2,t}^2 + \sum_{\mathcal{I}_{00}} \hat{F}_{1,t}^2\hat{F}_{2,t}^2 - \sum_{\mathcal{I}_{10}} \hat{S}_{1,t}^2\hat{F}_{2,t}^2 - \sum_{\mathcal{I}_{01}} \hat{F}_{1,t}^2\hat{S}_{2,t}^2}, \tag{11}$$

where $\mathcal{I}_{11}, \mathcal{I}_{00}, \mathcal{I}_{10}, \mathcal{I}_{01}$ is the partition of the individuals along their choices, *i.e* $\mathcal{I}_{11} = \{t : \; z_{1,t} = z_{2,t} = 1\}$, and the arguments such as $\hat{S}_{1,t}$ are $S(-x_{1,t}\hat{b}_{10})$.

It is also equal to :

$$\tilde{a} = \frac{\sum_t \left[\hat{S}_{1,t} z_{1,t} - \hat{F}_{1,t}(1 - z_{1,t})\right] \left[\hat{S}_{2,t} z_{2,t} - \hat{F}_{2,t}(1 - z_{2,t})\right]}{\sum_t \left[\hat{S}^2_{1,t} z_{1,t} - \hat{F}^2_{1,t}(1 - z_{1,t})\right] \left[\hat{S}^2_{2,t} z_{2,t} - \hat{F}^2_{2,t}(1 - z_{2,t})\right]}.$$

5 Sequential Generalized Logistic Distribution

Some other extended multivariate distributions may also be defined. In this section we introduce a generalized logistic distribution leading to simple interpretations for sequential decision processes.

5.1 The bivariate case

Let us consider a bidimensional vector (Y_1, Y_2) with marginal logistic distributions: $F_1(y) = F_2(y) = F(y) = (1 - exp(-y))^{-1}$, and the multivariate distribution family discussed in the previous subsections:

$$\begin{aligned} G(y_1, y_2) &= F(y_1)F(y_2)\left(1 + aS(y_1)S(y_2)\right) \\ &= F(y_1)\left(F(y_2) + af(y_2)S(y_1)\right), \end{aligned}$$

where f stands for the logistic p.d.f. If a is close to zero, we have:

$$G(y_1, y_2) \simeq F(y_1)F\left(y_2 + aS(y_1)\right).$$

Definition 5.1 *In the bivariate case, a Sequential Generalized Logistic distribution is associated with the c.d.f:*

$$G^\star(y_1, y_2) = F(y_1)F\left(y_2 + aS(y_1)\right),$$

where a is the dependence coefficient, $\mid a \mid < 1$.

It is easily seen that:

$$\lim_{y_1 \to +\infty} G^\star(y_1, y_2) = F(y_2),$$

$$\lim_{y_2 \to +\infty} G^\star(y_1, y_2) = F(y_1),$$

$$\lim_{y_1, y_2 \to -\infty} G^\star(y_1, y_2) = 0.$$

Moreover, we have:

$$\begin{aligned} \frac{\partial^2 G^\star}{\partial y_1 \partial y_2}(y_1, y_2) &= \frac{\partial}{\partial y_1}\left\{F(y_1) f(\tilde{y}_2)\right\} \\ &= f(y_1) f(\tilde{y}_2)\left[1 + aF(y_1)\left(2F(\tilde{y}_2) - 1\right)\right], \end{aligned}$$

with $\tilde{y}_2 = y_2 + aS(y_1)$.

Therefore this cross derivative is non negative as soon as the dependence coefficient has a modulus smaller than one.

Definition 1 has a clear recursive aspect. In particular a symmetrical formulation might also have been introduced, i.e:

$$G^{\star\star}(y_1, y_2) = F(y_2) F(y_1 + aS(y_2)),$$

and would correspond to another bivariate logistic distribution with the same expansion as $G^\star$ or G in a neighbourhood of $a = 0$. These sequential definitions appear suitable for modelling sequential selectivity sampling. Indeed if $G^\star(y_1, y_2) = F(y_1) F(y_2 + aS(y_1))$, we directly deduce that the distribution of Y_2 conditional to the endogenous sampling $Y_1 < y_1$ is the logistic distribution $F(y_2 + aS(y_1))$. It is transformed from the logistic distribution F by introducing a correction term: $aS(y_1) = a\frac{f(y_1)}{F(y_1)}$, which may be interpreted as a "Mill's ratio" applied to distribution F [see Heckman (1974),(1979)].

The dependence parameter a is directly linked with the correlation coefficient between the two variables Y_1, Y_2. More precisely it is proved in appendix 3 that:

$$Corr(Y_1, Y_2) = a\frac{E\left[YF(Y)\right]\left(2E\left[YF(Y)\right] + 1\right)}{VY}. \tag{12}$$

As previously the correlation is constrained. Since $\mid a \mid < 1$, we deduce from the calculus performed in appendix 3 for the logistic distribution that the sequential model may only be compatible with correlations of modulus smaller than $\frac{3}{\pi^2}$.

5.2 The multivariate case

The extension to any dimension is based on the same principle, i.e the correction for the successive selectivity biases by means of factors measuring the magnitude of selectivity.

Definition 5.2 *A Sequential Generalized Logistic distribution is associated with the c.d.f:*

$$G^{\star}(y_1,\ldots,y_p)=\prod_{j=1}^{p}F\left[y_j+\sum_{i<j}a_{j,i}S(\tilde{y}_i)\right],$$

where $\tilde{y}_j=y_j+\sum_{i<j}a_{j,i}S(\tilde{y}_i)$ *and* $\tilde{y}_1=y_1$.

For instance, for $p=3$ we get:

$$G^{\star}(y_1,y_2,y_3)=F(y_1)F(y_2+a_{21}S(y_1))F(y_3+a_{31}S(y_1)+a_{32}S(y_2+a_{21}S(y_1))).$$

It is easily checked that in a neighbourhood of $(a_{j,i}=0,\ \forall\, i<j)$, $G^{\star}$ and G are first order equivalent.

5.3 Discrete duration models

The previous family may be useful for studying discrete duration models allowing for duration-type dependence effects. Let us for instance consider job search models. The duration of unemployment D results from a sequence of qualitative decisions: accept or reject the job at date j. Let us denote $Z_j=1$ if the job is rejected at j, $Z_j=0$ otherwise, and let us define Z_j as a function of some latent variable:

$$\begin{aligned} Z_j &= 1 \quad if\ Y_j<0, \\ &= 0 \quad otherwise, \end{aligned}$$

where $Y_j=x_jb_j+u_j$.

If $(u_1,\ldots,u_p)'$ has a sequential generalized distribution, the distribution of the duration is:

$$P[D=j] = \prod_{i=1}^{j-1} F\left(-x_i b_i + \sum_{k<i} a_{i,k} S(\tilde{\alpha}_k)\right) S\left(-x_j b_j + \sum_{k<j} a_{j,k} S(\tilde{\alpha}_k)\right)$$

where:

$$\begin{aligned} \tilde{\alpha}_k &= -x_k b_k + \sum_{l<k} a_{k,l} S(\tilde{\alpha}_l) \\ \tilde{\alpha}_1 &= -x_1 b_1. \end{aligned}$$

The rate of exit is given by:

$$\lambda_j = S\left(-x_j b_j + \sum_{k<j} a_{j,k} S(\tilde{\alpha}_k)\right).$$

It is not constant in j as soon as the dependence coefficients are non zero and it takes into account the conditions $\tilde{\alpha}_k,\ k<j$ of the previous choices.

Moreover this rate has a tractable form which has to be compared with the expressions which would have been derived under a normality assumption of the error term u_j and would involve $j-1$ multidimensional integrals [see e.g. Kiefer (1988), Narendranathan-Elias (1990), Devine-Kiefer (1991), Muhleisen (1993)]. Therefore such formulations will avoid the use of numerical algorithms [Butler-Moffit (1982)] or of modified estimation methods based on simulations.

6 Extended Logistic Distributions as Approximations of the Multivariate Normal Distribution

In the previous sections we propose to use directly the multivariate extensions of the marginal distributions in the modelling procedure. In this section we are interested in examples based on the multivariate normal distribution and in the use of the previous multivariate extensions of the logistic distribution to derive some approximated formulations.

In the univariate case it is well known [see e.g. Morimune (1976)] that the one dimensional standard normal distribution is well approximated by a logistic distribution with zero mean and unit variance , whose c.d.f is given by $F\left(\frac{\pi}{\sqrt{3}}\right)$, where $F(y) = [1 + exp - y]^{-1}$.

6.1 An approximation of the c.d.f of the multivariate normal distribution

Les us consider a multivariate normal distribution with mean m and a variance-covariance matrix Σ. We decompose this matrix to see the variance terms σ_i^2 and the correlation terms $\rho_{i,j}$. R denotes the correlation matrix. The c.d.f of this distribution is:

$$\begin{aligned} G(y_1,\ldots,y_p) &= P(Y_1 < y_1,\ldots,Y_p < y_p) \\ &= P(\tfrac{Y_1-m_1}{\sigma_1} < \tfrac{y_1-m_1}{\sigma_1},\ldots,\tfrac{Y_p-m_p}{\sigma_p} < \tfrac{y_p-m_p}{\sigma_p}) \\ &= \int_{-\infty}^{\frac{y_1-m_1}{\sigma_1}} \cdots \int_{-\infty}^{\frac{y_p-m_p}{\sigma_p}} \frac{1}{(2\pi)^{p/2}\sqrt{detR}} exp - \frac{1}{2}u'R^{-1}u \; du. \end{aligned}$$

We are going to examine this distribution in a neighbourhood of the no correlation hypothesis, *i.e.* when the correlations $\rho_{i,j}$, $i \neq j$ are uniformly small. We only consider the first order expansion. Since the first non zero term in the expansion of $detR - 1$ is of order 2 in the components of R, we get:

$$\begin{aligned} \tfrac{1}{(2\pi)^{p/2}\sqrt{detR}} exp - \tfrac{1}{2}u'R^{-1}u &\simeq \tfrac{1}{(2\pi)^{p/2}} exp - \tfrac{1}{2}u'(Id + R - Id)^{-1}u \\ &\simeq \tfrac{1}{(2\pi)^{p/2}} exp - \tfrac{1}{2}\left[u'u - u'(R-Id)u\right] \\ &\simeq \tfrac{1}{(2\pi)^{p/2}} \left[1 + \tfrac{1}{2}u'(R-Id)u\right] exp - \tfrac{1}{2}u'u \\ &= \prod_{i=1}^{p} \phi(u_i) \left[1 + \sum_{i<j} \rho_{ij}u_iu_j\right], \end{aligned}$$

where ϕ stands for the p.d.f of the standard normal.

Therefore the first order expansion of the c.d.f of the multivariate normal distribution is:

$$
\begin{aligned}
G(y_1,\ldots,y_p) &\simeq \int_{-\infty}^{\frac{y_1-m_1}{\sigma_1}} \ldots \int_{-\infty}^{\frac{y_p-m_p}{\sigma_p}} \prod_{i=1}^{p} \phi(u_i) \left[1+\sum_{i<j} \rho_{ij} u_i u_j\right] du \\
&= \prod_{i=1}^{p} \Phi(\frac{y_i-m_i}{\sigma_i}) \left[1+\sum_{i<j} \rho_{i,j} \frac{\phi(\frac{y_j-m_j}{\sigma_j})\phi(\frac{y_i-m_i}{\sigma_i})}{\Phi(\frac{y_j-m_j}{\sigma_j})\Phi(\frac{y_i-m_i}{\sigma_i})}\right],
\end{aligned}
$$

where Φ stands for the c.d.f of the standard normal.

In this expression the factor of the correlation $\rho_{i,j}$ is equal to a product of Mill's ratios.

Let us now approximate in this expression the standard normal distribution by a logistic one:

$$
\Phi(y) \simeq F(\frac{\pi y}{\sqrt{3}}),\ \phi(y) \simeq \frac{\pi}{\sqrt{3}} F(\frac{\pi y}{\sqrt{3}}) \left[1-F(\frac{\pi y}{\sqrt{3}})\right].
$$

We get:

$$
\begin{aligned}
&G(y_1,\ldots,y_p) \\
&\simeq \prod_{i=1}^{p} F(\frac{\pi}{\sqrt{3}}\frac{y_i-m_i}{\sigma_i}) \left[1+\sum_{i<j} \rho_{i,j} \frac{\pi^2}{3} \left(1-F(\frac{\pi}{\sqrt{3}}\frac{y_j-m_j}{\sigma_j})\right) \left(1-F(\frac{\pi}{\sqrt{3}}\frac{y_i-m_i}{\sigma_i})\right)\right] \\
&= \prod_{i=1}^{p} F(\frac{\pi}{\sqrt{3}}\frac{y_i-m_i}{\sigma_i}) \left[1+\sum_{i<j} \rho_{i,j} \frac{\pi^2}{3} S(\frac{\pi}{\sqrt{3}}\frac{y_j-m_j}{\sigma_j}) S(\frac{\pi}{\sqrt{3}}\frac{y_i-m_i}{\sigma_i})\right].
\end{aligned}
$$

This is nothing else than one of the multivariate extensions of the logistic distribution applied to the transformed variables $\frac{\pi}{\sqrt{3}}\frac{y_i-m_i}{\sigma_i}$, $i=1,\ldots,p$ and with the dependence coefficients : $a_{i,j} = \frac{\pi^2}{3}\rho_{i,j}$.

Remark 6.1 :

When $p=2$ we have a unique dependence coefficient $a = \frac{\pi^2}{3}\rho$ which has to be constrained by one to give a meaning to the bivariate extension of the logistic distribution. We can see that the relationship $a = \frac{\pi^2}{3}\rho$ valid in a neighbourhood of $\rho = 0$ seems to be also satisfied approximately for large values of the correlation. Indeed $\mid a \mid < 1$ implies with this relationship $\mid \rho \mid < \frac{3}{\pi^2}$, which is nothing else than the maximal correlation derived in section 2 for the bivariate logistic distribution.

Remark 6.2 :

In multinomial probit models we have generally to evaluate the c.d.f of the multivariate normal distribution at point $y_i = 0,\ i = 1, \ldots, p$. Moreover the mean has a linear form: $m_i = x_i b_i$, and the variance may be constrained to be one for identification purposes. Then the components of b_i are interpretable up to a multiplicative factor. For such a problem the transformed variables are :

$$\frac{\pi}{\sqrt{3}} \left(\frac{y_i - m_i}{\sigma_i} \right) = \frac{\pi}{\sqrt{3}} x_i b_i = x_i \beta_i,$$

where $\beta_i = \frac{\pi}{\sqrt{3}} b_i$ is also interpretable up to a multiplicative factor. Therefore in practice and before the estimation step, the only correction resulting from the replacement of the normal distribution by the extended logistic one concerns the dependence parameter.

6.2 The quality of the approximation for large correlations

We have seen in the previous subsection that the multivariate logistic distribution is a good approximation of the multivariate normal distribution when $\rho_{i,j} \simeq 0,\ \forall\ i < j$. What about this fit when the $\rho_{i,j}$ are no longer close to zero ?

To discuss this point we consider the example of a probit model with endogenous selectivity as it is used for credit scoring [Boyes-Hoffman-Low(1989)]. The model is based on two latent variables satisfying a linear specification:

$$\left\{ \begin{array}{ll} Y_{i,1}^{\star} & = X_i' b_1 + u_{i,1}, \\ Y_{i,2}^{\star} & = X_i' b_2 + u_{2,1}, \end{array} \right. \tag{13}$$

where $\begin{pmatrix} u_{i,1} \\ u_{i,2} \end{pmatrix}$ are i.i.d variables, with distribution $\mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$.
To these latent variables correspond qualitative variables:

$$
\begin{cases}
Y_{i,1} = \mathbf{1}_{Y_{i,1}^\star > 0} & = \begin{cases} 1 & \text{, if the credit demander is not risky} \\ & \text{(no potential default),} \\ 0 & \text{, if the credit demander is risky} \\ & \text{(potential default),} \end{cases} \\
\\
Y_{i,2} = \mathbf{1}_{Y_{i,2}^\star > 0} & = \begin{cases} 1 & \text{, the demand is granted,} \\ 0 & \text{, otherwise.} \end{cases}
\end{cases}
\tag{14}
$$

The observable variables are the decisions $Y_{2,i}$ of the creditmen and the qualities of the credits $Y_{1,i}$, if $Y_{2,i} = 1$. Therefore the model for the observable variables is a bivariate probit model with partial observability [Poirier (1980)]. In such a problem we are mainly interested in the potential risk summarized through the scoring function $X'b_1$.

Several estimation methods may be considered.

- Method 1: We only retain granted credits and apply a univariate probit model to the endogenous variable $Y_{i,1}$. As soon as there is endogenous selectivity (*i.e.* $\rho \neq 0$), this approach leads to an inconsistent estimator of the scoring function.

- Method 2: We apply M.L approach based on the bivariate probit model with partial observability. The estimator is consistent and asymptotically efficient.

- Method 3: We apply a two steps approach after replacement of the bivariate probit model by the associated logit model. This model corresponds to probabilities such that:

$$
\begin{aligned}
&P[Y_2 = 1] = F[X\gamma_2], \\
&P[Y_1 = 1 \mid Y_2 = 1] = F\left[X\gamma_1 + a(1 - F[X\gamma_2])\right].
\end{aligned}
$$

The correction for selectivity bias simply consists in adding the selection rate as an additional explanatory variable. We first estimate γ_2 by a dichotomous logit model. Then we apply to Y_1 and for granted credits a dichotomous logit model with Y_1 and $1 - F[X\hat{\gamma}_2]$ as explanatory variables, where $\hat{\gamma}_2$ stands for the first step estimator. This method is not consistent under the normality assumption, but the error may be small.

To compare these different approaches, we perform some Monte Carlo studies. In the credit scoring the explanatory variables are generally qualitative ones to be able to define classes of risk. We consider 15 dichotomous explanatory variables plus the constant. These variables are assumed to be independent with Bernoulli distributions $\mathcal{B}(1, p_j)$, $j = 1, \ldots, 15$.

Table 2: Values of p_j

.4	.7	.3	.5	.2	.4	.3	.8	.7	.6	.4	.3	.6	.4	.5	.8

The values of the coefficients retained for the data generating process are given in table 3.

Table 3: Values of the coefficients b

b_1	.5	.3	.5	-.3	-.4	.3	-.6	.6	-.5	-.2	.4	.2	-.4	.5	.3	-.2
b_2	.5	-.2	.2	.3	.4	.2	-.6	.3	-.4	.3	-.3	.5	.3	.3	-.6	-.3

We give below the results of two rather less favorable simulations since we introduce large values ($\rho = 0.9$, $\rho = 0.7$) for the correlation coefficient.

In the tables we successively provide the estimates and the standard errors evaluated by the three methods for the different parameters b_2, and for the correlation coefficient on the last line. For the last method the estimates are corrected by the suitable factors to allow a direct comparison. It is immediately seen that the logistic approximation provides good estimates of the scoring function and a rather good idea of the correlation.

In practice, the quality of the scoring function essentially depends on the quality of the classification associated with the scoring function $S_i = X_{1,i}b_1$. This quality may be summarized by the performance curve (see Gouriéroux[1992]). This curve is obtained in the following way: we first classify the individuals by increasing estimated score $\hat{S}_i$. Then for each possible

Table 4: Estimations with $\rho = 0.7$

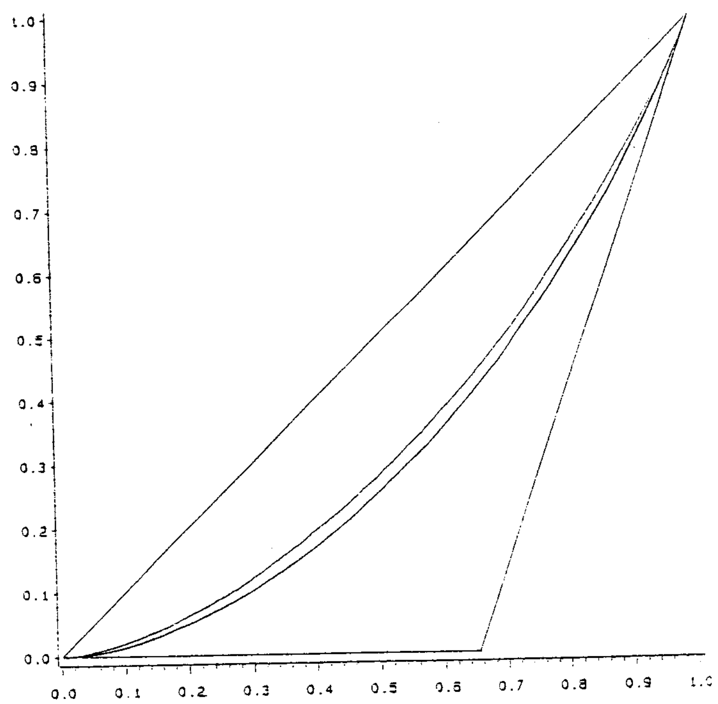
True values	M1 est.	M1 std.	M2 est.	M2 std.	M3 est.	M3 std.
0.5	1.0468	0.0351	0.598	0.0641	0.636	0.0981
-0.2	-0.321	0.0168	-0.223	0.0206	-0.222	0.0343
0.2	0.0745	0.0154	0.220	0.0233	0.1913	0.0372
0.3	0.436	0.0147	0.346	0.0182	0.339	0.0301
0.4	0.624	0.0218	0.502	0.0263	0.500	0.045
0.2	0.121	0.0147	0.206	0.0179	0.186	0.0297
-0.6	-0.464	0.0160	-0.660	0.0285	-0.603	0.0447
0.3	0.0946	0.0193	0.285	0.0299	0.251	0.0475
-0.4	-0.266	0.0157	-0.411	0.0235	-0.374	0.0376
0.3	0.4013	0.0145	0.345	0.0160	0.332	0.0270
-0.3	0.471	0.0146	-0.352	0.0203	-0.349	0.0328
0.5	0.491	0.0165	0.551	0.0181	0.518	0.0312
0.3	0.482	0.0146	0.364	0.0204	0.357	0.0329
0.3	0.159	0.0146	0.3212	0.0243	0.283	0.0384
-0.6	-0.787	0.0151	-0.696	0.0186	-0.677	0.0313
-0.3	-0.313	0.0186	-0.378	0.02015	-0.354	0.0347
0.7	.	.	0.820	0.0984	0.596	0.0236

Table 5: Estimations with $\rho = 0.9$

True values	M1 est.	M1 std.	M2 est.	M2 std.	M3 est.	M3 std.
0.5	1.225	0.124	0.866	0.219	0.918	0.340
-0.2	-0.388	0.0591	-0.305	0.073	-0.314	0.123
0.2	0.0158	0.0536	0.122	0.076	0.094	0.122
0.3	0.469	0.052	0.398	0.063	0.039	0.106
0.4	0.632	0.078	0.502	0.102	0.526	0.173
0.2	0.124	0.0518	0.196	0.063	0.171	0.105
-0.6	-0.552	0.057	-0.724	0.104	-0.662	0.165
0.3	0.008	0.068	0.173	0.108	0.142	0.173
-0.4	-0.224	0.055	-0.347	0.083	-0.317	0.133
0.3	0.397	0.051	0.357	0.055	0.353	0.094
-0.3	-0.504	0.0512	-0.409	0.07	-0.416	0.115
0.5	0.475	0.058	0.51	0.061	0.489	0.107
0.3	0.541	0.051	0.465	0.065	0.458	0.107
0.3	0.130	0.052	0.2490	0.076	0.212	0.124
-0.6	-0.843	0.053	-0.775	0.063	-0.766	0.108
-0.3	-0.326	0.065	-0.375	0.070	-0.356	0.122
0.9	.	.	0.661	0.334	0.450	0.818

value of this score s we compute the proportion of individuals with $\hat{S}_1 < s$ ($x(s)$ say), and the ratio of the proportion of risky among credits individuals with $\hat{S}_1 < s$ to the proportion of the risky credits in the whole population ($y(s)$ say). The performance curve is the set of points $(x(s), y(s))$, s varying. A scoring function is well chosen if such a curve is increasing, convex and small (notice that it is constrained to pass by the points $(0,0)$ and $(1,1)$). We give below the performance curves corresponding to the three previous approachs. It is seen that the bivariate logistic based scoring and the ML probit approach lead to identical performance curves, and to better discrimination than the crude scoring on granted credits.

Figure 1: Performance curves



6.3 Pseudo true values for some approximations of the bivariate normal distribution

In this subsection we are interested in pseudo true values of the parameters for several Extended Logistic Distributions as approximations to the centered normalized bivariate distribution. We consider the normal bivariate distribution $\mathcal{N}_\rho$ whose c.d.f is:

$$\phi(\rho, x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{x} \int_{-\infty}^{y} exp\left(-\frac{x^2 + y^2 - 2\rho xy}{2} \right) dxdy.$$

The associated c.d.f will be denoted $\Phi(x, y, \rho)$.

We will determine the "best" approximations of this distribution by a distribution $F(m_x, m_y, \sigma_x, \sigma_y, a)$ which belongs to one of the classes defined above. We consider three kinds of distributions:

$$\begin{aligned}
&(\mathrm{I}) \quad F(x,y,m_x,m_y,\sigma_x,\sigma_y,a) &&= \frac{\Phi(\frac{x-m_x}{\sigma_x})\Phi(\frac{y-m_y}{\sigma_y})}{\sigma_x\sigma_y} \left(1 + a(1-\Phi(\tfrac{x-m_x}{\sigma_x}))(1-\Phi(\tfrac{y-m_y}{\sigma_y}))\right), \\
&(\mathrm{II}) \quad F(x,y,m_x,m_y,\sigma_x,\sigma_y,a) &&= \frac{L(\frac{x-m_x}{\sigma_x})L(\frac{y-m_y}{\sigma_y})}{\sigma_x\sigma_y} \left(1 + a(1-L(\tfrac{x-m_x}{\sigma_x}))(1-L(\tfrac{y-m_y}{\sigma_y}))\right), \\
&(\mathrm{III}) \quad F(x,y,m_x,m_y,\sigma_x,\sigma_y,a) &&= \frac{\Phi(\frac{x-m_x}{\sigma_x})\Phi(\frac{y-m_y+a(1-\Phi(\frac{x-m_x}{\sigma_x})}{\sigma_y})}{\sigma_x\sigma_y},
\end{aligned}$$

where $\Phi(.)$ and $L(.)$ stand for the (centered and reduced) normal and logistic distributions respectively.

The first and second distributions belong to the family described in section 2 (for two different choices of the underlying univariate distributions) while the third one belongs to the family described in section 5.

The usual criterion to compare two distributions is the Kullback Leibler Information Criterion (KLIC). Given ρ the parameters $m_x, m_y, \sigma_x, \sigma_y, a$ minimizing the KLIC:

$$K(m_x, m_y, \sigma_x, \sigma_y, a, \rho) = \int\int_{\mathbb{R}^2} \phi(x, y, \rho) Log(\frac{\varphi(x, y, \rho)}{f(x, y, m_x, m_y, \sigma_x, \sigma_y, a)})dxdy,$$

where φ and f are the associated pdf, are known as the pseudo true values of the parameters [see for instance Gouriéroux-Monfort (1989)]. The minimization of the KLIC has been performed under the constraint $\mid a \mid< 1$ to provide admissible values of the p.d.f.

Since we know that the pseudo trues value are $(0, 0, 1, 1, 0)$ when $\rho = 0$ (for all three families of distributions) we are more interested in the pseudo true values when ρ is far from 0. Except when $\rho = 0$ the minimization can not be performed explicitely, and we give numerical approximations of these quantities as ρ varies between 0 and 1 (computations have been performed on a 486-66MHz, they only take a few minutes).

Figure 2: Approximation (I) of the bivariate normal distribution. Pseudo true values of a and $\sigma_x = \sigma_y = \sigma$, and values of the KLIC K.

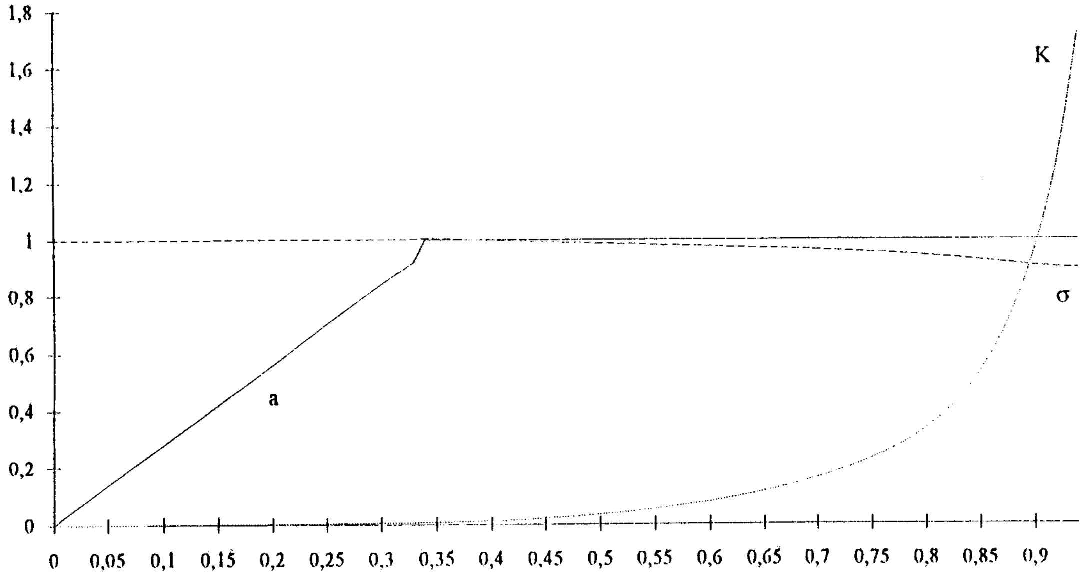


Figure 3: Approximation (II) of the bivariate normal distribution. Pseudo true values of a and $\sigma_x = \sigma_y = \sigma$, and values of the KLIC K.

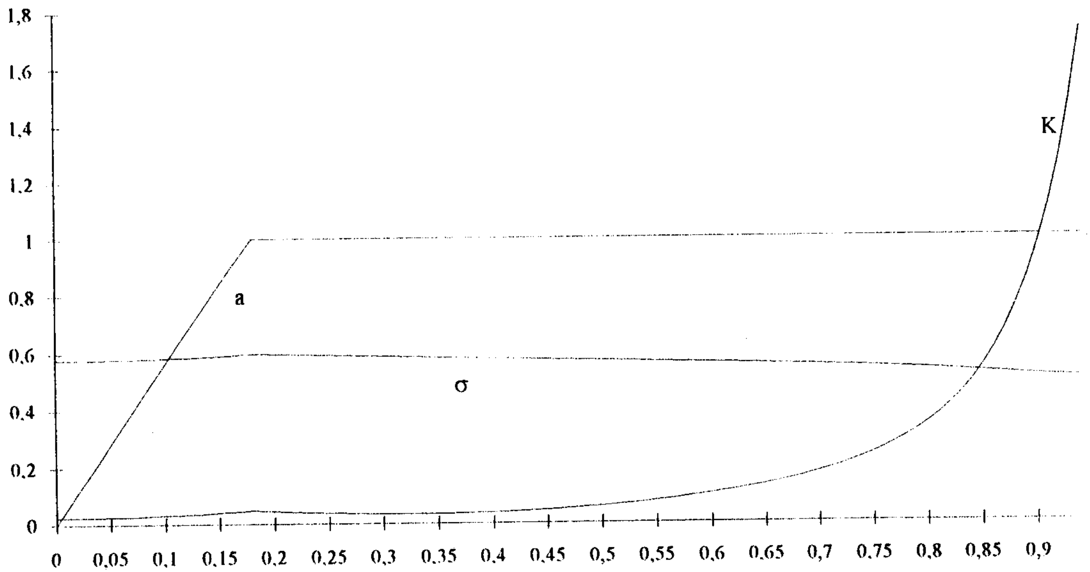


Figure 4: Approximation (III) of the bivariate normal distribution. Pseudo true values of $a, m_x, \sigma_x, m_y, \sigma_y$, and values of the KLIC K.

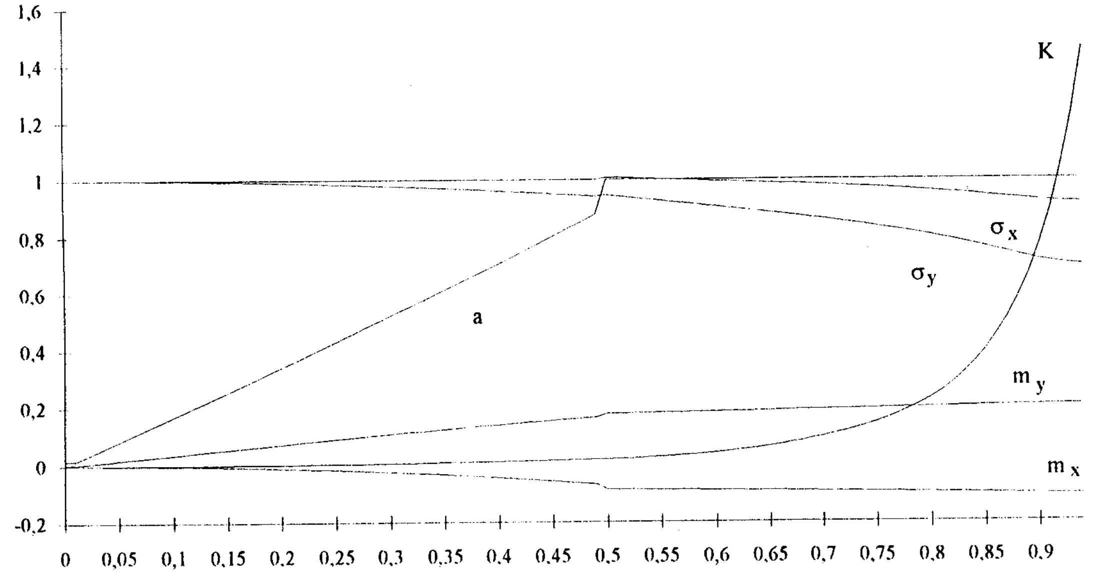
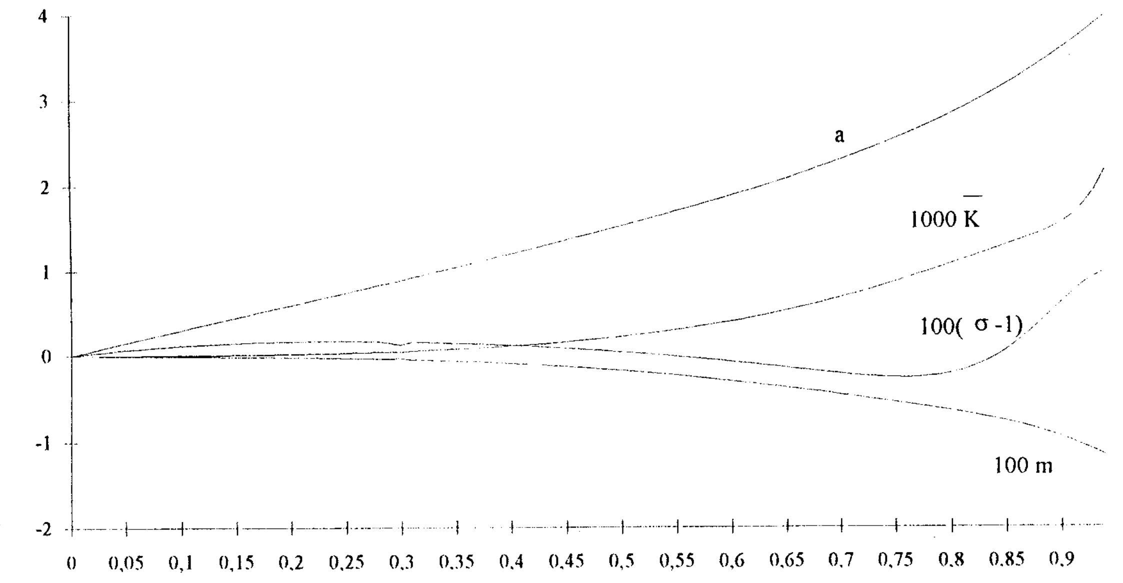


Figure 5: Approximation (I bis) of the bivariate normal distribution. Pseudo true values of $a, m_x = my = m, \sigma_x = \sigma_y = \sigma$, and values of the average KLIC $\overline{K}$.



First, we note that for the approximation (II), the KLIC is not equal to zero when ρ is null, since the approximation of the normal distribution by a logistic distribution is not exact.

For approximations (I) and (II) the pseudo true values of the location parameters are equal to zero, and the pseudo true values for σ_x and σ_y are equal. We note that in these two cases the pseudo true values of σ are close to the true value 1. Moreover the pseudo true value of a is an increasing function of the (true) correlation coefficient ρ.

The behavior of the pseudo true value for parameter a as a function of ρ is interesting, since the function is almost linear even for large values of ρ. We know from subsection 6.1 that it is true in the neighbourhood of $\rho = 0$ and that the slope of this linear function is equal to the inverse of the maximum admissible correlation ρ_m between Y_1 and Y_2, when these (centered and reduced) random variables follow an Extended Logistic Distribution.

In fact, the behavior of the pseudo true values as functions of ρ depends on the position of ρ with respect to ρ_m.

When $\rho < \rho_m$, the pseudo true value of the parameter σ is very close to one (perhaps exactly, but it seems really hard to prove), and a is (almost) a linear function of ρ, whose slope equals the inverse of the maximum available correlation. When $\rho > \rho_m$, the pseudo true value of a equals 1 (recall a cannot be greater than 1) and the pseudo true value σ is no longer equal to 1.

Unfortunately, in this last case, it is not clear if the pseudo true value σ is or not a monotonic function of ρ from $\rho = \rho_m$ to $\rho = 0.9$. The pseudo true values of σ first decrease but they seem to increase when ρ is close to one. It may be due to a computation artefact, because as ρ approach 1 the minimization requires very precise computations). Thus, if $a = 1$, the pseudo true value σ may not be an invertible function of ρ for large values of this parameter. If we want to use an approximation of a true c.d.f as a proxy model (for indirect inference procedures for example, see Gouriéroux, Monfort, Renault [1993]), we have to be very careful if the estimate of a is close to 1.

Finally, with family (III), which is the simplest one to use, we note that the pseudo true values of the location parameters are no longer equal to zero, and the pseudo true values of σ_x and σ_y no longer coincide. The pseudo true values of the parameters related to y (the variable used for the sequential aspect of the distribution), varie more rapidly as ρ increases than that related to x.

These pseudo true values are useful if we believe that the sample is drawn from an underlying normal bivariate model. The previous functions give the corrections to be used when a proxy model of the form (I), (II) or (III) is used to calibrate the exact model. For family (I) and (II) the correction depends on whether we find a close to 1 or not.

If the previous distributions families are used as proxy models in the case of qualitative data, we are more likely interested in the best proxy for the observed (qualitative) variable model than in a good proxy for the latent variable model. A relevant objective function for such an approximation may be defined as an "average qualitative KLIC":

$$\overline{K}(m_x, m_y, \sigma_x, \sigma_y, a, \rho) = \int\int_{\mathbb{R}^2} \phi(x, y, \rho) K(x, y, m_x, m_y, \sigma_x, \sigma_y, a, \rho) dx dy,$$

where:

$$K(x, y, m_x, m_y, \sigma_x, \sigma_y, a, \rho) = \Phi(x, y, \rho) Log \frac{\Phi(x, y, \rho)}{F(x, y, m_x, m_y, \sigma_x, \sigma_y, a)}$$

$$+ (1 - \Phi(x, y, \rho)) \, Log \frac{(1 - \Phi(x, y, \rho))}{(1 - F(x, y, m_x, m_y, \sigma_x, \sigma_y, a))}.$$

We compute the pseudo true values of $a, m_x, m_y, \sigma_x, \sigma_y$ for this criterion using the qualitative model associated with family (I) as proxy model. Again, m_x and $m_y(\text{resp}\sigma_x, \sigma_y)$ coincide and In the previous formulation, the value of a does not need to be constrained for computational purposes, thus the following minimization gives the unconstrained pseudo true value of a.

We note that the pseudo true value of a is an increasing (but no longer linear) function of ρ and it is not always smaller than 1. Thus, the estimation of the parameter a in the (qualitative) proxy model is sufficient to infer indirectly the value of ρ. We also note substantial changes in the behaviours of the pseudo true values of σ and m as functions of ρ if ρ exceed ρ_m. It indicates that the constraint imposed on a for the computation of the KLIC do have an effect on the pseudo true values of the location and scale parameters.

7 Conclusion

In this paper we propose two families of distributions suitable for limited dependent variable models. We study the properties of these distributions from both empirical and theoretical points of view. We show that these distributions only involve computations of univariate integrals either by conditioning or marginalizing. We study the simplifications induced when using these distributions instead of the multivariate normal distributions for some usual qualitative problems (non mutually exclusive alternatives, double hurdle type problems, competing risks etc.).

For these distributions, the correlations among the variables are constrained. So they cannot be used for problems in which the correlation is high. On the contrary they are shown to be suitable for testing the no correlation hypothesis. Finally, we discuss the approximation of the bivariate normal distribution provided by these distributions. The empirical behavior of the estimates after asymptotic correction, turn to be quite good when compared to exact techniques. We also provide numerical approximations of some pseudo true values for both qualitative and latent models.

Appendix 1 : Computation of

$\beta = E_F\left[\frac{Y-E[Y]}{VY^{\frac{1}{2}}}F\left(\frac{Y-E[Y]}{VY^{\frac{1}{2}}}\right)\right]$

1 Normal Distribution

We get : $\beta = \int_{-\infty}^{+\infty} y\phi(y)\Phi(y)dy$, where ϕ and Φ are respectively the p.d.f. and the c.d.f. of the standard normal distribution.

$$\begin{aligned}
\beta &= -\int_{-\infty}^{+\infty}\Phi(y)d\phi(y) \\
&= [-\phi(y)\Phi(y)]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty}\phi^2(y)dy \\
&= \tfrac{1}{2\pi}\int_{-\infty}^{+\infty} exp(-y^2)dy \\
&= \tfrac{1}{2\sqrt{\pi}}.
\end{aligned}$$

Therefore $4\beta^2 = \frac{1}{\pi}$.

2 : Logistic distribution

Let:

$$I(M) = \int_{-M}^{M}\frac{xexp(-x)dx}{(1+exp(-x))^3}.$$

Let $u = exp(-x)$, we get :

$$\begin{aligned}
I(M) &= -\int_{exp(-M)}^{expM}\frac{ln(u)du}{(1+u)^3} \\
&= \left[\tfrac{ln(u)}{2(1+u)^2}\right]_{exp(-M)}^{expM} - \tfrac{1}{2}\int_{exp(-M)}^{expM}\frac{du}{u(1+u)^2},
\end{aligned}$$

and thus:

$$
\begin{aligned}
I(M) &= \tfrac{M}{2(1+exp(M))^2} + \tfrac{M}{2(1+exp(-M))^2} - \tfrac{1}{2}\int_{exp(-M)}^{expM}\left(\frac{du}{u} - \frac{du}{1+u} - \frac{du}{(1+u)^2}\right) \\
&= \tfrac{M}{2(1+exp(M))^2} + \tfrac{M}{2(1+exp(-M))^2} - M + \tfrac{ln(exp(M)+1)}{2} - \tfrac{ln(exp(-M)+1)}{2} \\
&- \tfrac{1}{2(1+exp(M))} + \tfrac{1}{2(1+exp(-M))},
\end{aligned}
$$

and, as M goes to $+\infty$, we easily obtain:

$$\lim_{M\to+\infty} I(M) = \frac{1}{2}.$$

Now, since $V(Y) = \frac{\pi^2}{3}$ for the logistic distribution, we deduce that:

$$4\beta^2 = 4(\frac{1}{2})^2\frac{3}{\pi^2} = \frac{3}{\pi^2}.$$

3 : Exponential distribution

$F(y) = 1 - exp - y$

$$
\begin{aligned}
E\left(Y[2F(Y)-1]\right) &= E\left(Y[1-2exp-Y]\right) \\
&= EY - 2E\left(Yexp-Y\right) = 1 - 2\int_0^{+\infty} yexp(-2y)dy = \frac{1}{2}.
\end{aligned}
$$

Therefore $4\beta^2 = \frac{1}{4}$.

4 Uniform distribution

$$F(y) = y,\ 0 \leq y \leq 1,$$

$$E\left(Y[2F(Y)-1]\right) = 2EY^2 - EY = \tfrac{1}{6}.$$

Therefore $4\beta^2 = \frac{(1/6)^2}{1/12} = \frac{1}{3}$.

Appendix 2 Determination of the distributions of the observed variables

1 Qualitative case

We have to compute the different probabilities $P[Y_1 < y_1^0, \ldots, Y_q < y_q^0, Y_{q+1} > y_{q+1}^0, \ldots, Y_p > y_p^0]$, for any p and q, since the other ones are deduced by permutation of variables.

The result is obtained by induction. We assume that the formula of property 3.1 is valid for any $\overline{p} \leq \overline{q}$, such that $\overline{p} < p$ or $\overline{q} \leq q$ and $\overline{p} = p$. Then we will prove it for $\overline{p} = p, \overline{q} = q + 1$. We have:

$$P[Y_1 < y_1^0, \ldots, Y_q < y_q^0, Y_{q+1} < y_{q+1}^0, Y_{q+2} > y_{q+2}^0, \ldots, Y_p > y_p^0]$$

$$= P[Y_1 < y_1^0, \ldots, Y_q < y_q^0, Y_{q+2} > y_{q+2}^0, \ldots, Y_p > y_p^0]$$

$$-P[Y_1 < y_1^0, \ldots, Y_q < y_q^0, Y_{q+1} > y_{q+2}^0, \ldots, Y_p > y_p^0]$$

$$= \prod_{j=1}^{q} F_j(y_j^0) \prod_{j=q+2}^{p} S_j(y_j^0)$$
$$\left\{1 + \sum_{i<j\leq q} a_{i,j} S_i(y_i^0) S_j(y_j^0) + \sum_{q+2\leq i<j\leq p} a_{i,j} F_i(y_i^0) F_j(y_j^0) - \sum_{i\leq q<q+2\leq j} a_{i,j} S_i(y_i^0) F_j(y_j^0)\right\}$$
$$- \prod_{j=1}^{q} F_j(y_j^0) S_{q+1}(y_{q+1}^0) \prod_{j=q+2}^{p} S_j(y_j^0)$$
$$\left\{1 + \sum_{i<j\leq q} a_{i,j} S_i(y_i^0) S_j(y_j^0) + \sum_{q+2\leq i<j\leq p} a_{i,j} F_i(y_i^0) F_j(y_j^0) - \sum_{i\leq q<q+2\leq j} a_{i,j} S_i(y_i^0) F_j(y_j^0)\right.$$
$$\left. - \sum_{i\leq q} a_{i,q+1} S_i(y_i^0) F_{q+1}(y_{q+1}^0) + \sum_{q+2\leq j} a_{q+1,j} F_j(y_j^0) F_{q+1}(y_{q+1}^0)\right\}$$

$$
\begin{aligned}
&= \prod_{j=1}^{q+1} F_j(y_j^0) \prod_{j=q+2}^{p} S_j(y_j^0) \\
&\left\{ 1 + \sum_{i<j\leq q} a_{i,j} S_i(y_i^0) S_j(y_j^0) + \sum_{q+2\leq i<j\leq p} a_{i,j} F_i(y_i^0) F_j(y_j^0) - \sum_{i\leq q<q+2\leq j} a_{i,j} S_i(y_i^0) F_j(y_j^0) \right\} \\
&+ \prod_{j=1}^{q} F_j(y_j^0) \prod_{j=q+2}^{p} S_j(y_j^0) F_{q+1}(y_{q+1}^0) S_{q+1}(y_{q+1}^0) \left\{ \sum_{i\leq q} a_{i,q+1} S_i(y_i^0) - \sum_{q+2\leq j} a_{q+1,j} F_j(y_j^0) \right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \prod_{j=1}^{q+1} F_j(y_j^0) \prod_{j=q+2}^{p} S_j(y_j^0) \\
&\left\{ 1 + \sum_{i<j\leq q+1} a_{i,j} S_i(y_i^0) S_j(y_j^0) + \sum_{q+2\leq i<j\leq p} a_{i,j} F_i(y_i^0) F_j(y_j^0) - \sum_{i\leq q+1<q+2\leq j} a_{i,j} S_i(y_i^0) F_j(y_j^0) \right\}
\end{aligned}
$$

2 The mixed quantitative qualitative case

We get:

$$
\Psi(y_1, \ldots, y_q, 0, \ldots, 0) = \frac{\partial^q}{\partial y_1 \ldots \partial y_q} G(y_1, \ldots, y_q, y_{q+1}^0, \ldots, y_p^0)
$$

$$
\begin{aligned}
= \frac{\partial^q}{\partial y_1 \ldots \partial y_q} \quad & \left\{ \prod_{j=1}^{q} F_j(y_j) \prod_{j=q+1}^{p} F_j(y_j^0) + \sum_{i<j\leq q} a_{i,j} F_i(y_i) S_i(y_i) F_j(y_j) S_j(y_j) \prod_{k\neq i,j} F_k(y_k) \right. \\
& + \sum_{q+1\leq i<j} a_{i,j} F_i(y_i^0) S_i(y_i^0) F_j(y_j^0) S_j(y_j^0) \prod_{k\neq i,j} F_k(y_k) \\
& \left. + \sum_{i\leq q<q+1\leq j} a_{i,j} F_i(y_i) S_i(y_i) F_j(y_j^0) S_j(y_j^0) \prod_{k\neq i,j} F_k(y_k) \right\}
\end{aligned}
$$

$$= \prod_{j=1}^{q} f_j(y_j) \prod_{j=q+1}^{p} F_j(y_j^0) + \sum_{i<j\leq q} a_{i,j}[2F_i(y_i)-1][2F_j(y_j)-1] \prod_{j=1}^{q} f_j(y_j) \prod_{j=q+1}^{p} F_j(y_j^0)$$
$$+ \sum_{q+1\leq i<j} a_{i,j} S_i(y_i^0) S_j(y_j^0) \prod_{j=1}^{q} f_j(y_j) \prod_{j=q+1}^{p} F_j(y_j^0)$$
$$+ \sum_{i\leq q<q+1\leq j} a_{i,j}[2F_i(y_i)-1] S_j(y_j^0) \prod_{j=1}^{q} f_j(y_j) \prod_{j=q+1}^{p} F_j(y_j^0)$$

$$= \prod_{j=1}^{q} f_j(y_j) \prod_{j=q+1}^{p} F_j(y_j^0)$$
$$\left\{1 + \sum_{i<j} a_{i,j} \left\{[2F_i(y_i)-1]\mathbf{1}_{i\leq q} + S_i(y_i^0)\mathbf{1}_{q+1\leq i}\right\} \left\{[2F_j(y_j)-1]\mathbf{1}_{j\leq q} + S_j(y_j^0)\mathbf{1}_{q+1\leq j}\right\}\right\}$$

Appendix 3 The correlation for the bivariate sequential generalized logistic distribution

We know that the marginal distributions are logistic ones. Therefore:

$$EY_1 = EY_2 = 0,\ VY_1 = VY_2 = \frac{\pi^2}{3}.$$

Moreover, we have:

$$\begin{aligned} E\left[Y_1(Y_2 + aS(Y_1)\right] &= E\left[Y_1 Y_2\right] + aE\left[Y_1\left(1 - F(Y_1)\right)\right] \\ &= E\left[Y_1 Y_2\right] + aE\left[Y_1\right] + aE\left[Y_1 F(Y_1)\right]. \end{aligned}$$

Now, by taking into account the particular form of the p.d.f., we have:

$$EY_1Y_2 - aE\left[Y_1 F(Y_1)\right] = 2aE\left[Y_1 F(Y_1)\right] E\left[Y_2 F(Y_2)\right].$$

Thus,

$$E\left[Y_1(Y_2 + aS(Y_1)\right] = E[Y_1] + E[Y_2] + aE\left[Y_1 F(Y_1)\right] E\left[Y_2\left(2F(Y_2) - 1\right)\right],$$

and we get:

$$EY_1Y_2 = aE\left[YF(Y)\right]\left(2E\left[YF(Y)\right] + 1\right).$$

References

[1] Ali M., Mikhail N., and M. Haq (1978): *A Class of Bivariate Distributions Including the Bivariate Logistic*, Journal of Multivariate Analysis, **8**, 405-412.

[2] Ashford J. and R. Sowden (1970): *Multivariate Probit Analysis*, Biometrika, **26**, 535-546.

[3] Boyes W., Hoffman D. and S. Low (1989): *An Econometric Analysis of the Bank Credit Scoring Problem*, Journal of Econometrics, **40**, 3-14.

[4] Butler J.S. and R. Moffit (1982): *A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model*, Econometrica, **50**, 761-764.

[5] Dagenais M. (1982): *The Tobit Model with Serial Correlation*, Economic Letters,

[6] Dagenais M. (1986): *Estimating the Tobit Model with Serial Correlation: an Operational Approach*, Univ. of Montreal, DP 8615.

[7] Devine T.J. and N.M. Kiefer (1991): *Empirical Labor Economics: The Search Approach*, New York, Oxford, O.U.P.

[8] Genest C. and L.P. Rivest (1993): *Statistical Inference Procedures for the Bivariate Archimedean Copula*, Journal of the American Statistical Association **88**, 1034-1043.

[9] Gouriéroux C. (1989): *Econométrie des variables qualitatives*, Economica, 2nd edition.

[10] Gouriéroux C. (1992): *Courbes de Performance, de Selection et de Discrimination*, Annales d'Economie et de Statistiques, **28**, 107-123.

[11] Gouriéroux C. and A. Monfort (1993): *Simulation Based Inference: A Survey with Special Reference to Panel Data Models*, Journal of Econometrics, **59**, 5-33.

[12] Gouriéroux C., Monfort A., and A., Trognon (1984): *Estimation and Test in Probit Models with Serial Correlation*, in Florens, Mouchart, Raoult, Simar, Alternative Approaches to Time Series Analysis, Univ. St Louis, Brussels.

[13] Gumbel E.J (1961): *Bivariate Logistic Distributions*, Journal of the American Statistical Association, **56**.

[14] Hausman J. and D. Mc Fadden (1984): *A Specification Test for the Multinomial Logit Model*,Econometrica, **52**, 1219-1240.

[15] Heckman J. (1974): *Shadows Prices, Market Wages , and Labor Supply*, Econometrica, **46**, 679-694.

[16] Heckman J. (1979): *Sample Selection Bias as a Specification Error*, Econometrica, **47**, 153-161.

[17] Kiefer N.M. (1988): *Economic Duration Data and Hazard Functions*, Journal of Economic Literature, **26**, 646-679.

[18] Laroque G. and B. Salanié (1989): *Estimation of Multimarket Fix-price Models: An Application of Pseudo Maximum Likelihood Methods*, Econometrica, **57**, 831-860.

[19] Lee L.F. (1982): *Some Approaches to the Correction of Selectivity Bias*, Review of Economic Studies, **49**, 355-372.

[20] Liang K-Y. and S.L. Zeger (1986): *Longitudinal Data Analysis Using Generalized Linear Models*, Biometrika, **73**, 13-22.

[21] Liang K-Y. and S.L. Zeger (1989): *A Class of Regression Models for Multivariate Binary Time Series*, Journal of the American Statistical Association **84**, 447-451.

[22] Lollivier S. and D. Verger (1987): *Les Comportements en Matière d'Epargne et de Patrimoine*, Economie et Statistiques, **202**, 65-78.

[23] Mc.Fadden D. (1973): *Conditional Logit Analysis of Qualitative Choice Behavior*, in Frontiers in Econometrics, Ed. Zarembka, Academic Press, 105-142.

[24] Mc.Fadden D. (1976): *Quantal Choice Analysis: A Survey*, Annals of Economic and Social Measurement, **5** , 363-390.

[25] Mc.Fadden D. (1989): *A Method for Simulated Moments for Estimation of Discrete Response Models without Numerical Integration*, Econometrica, **57**, 995-1026.

[26] Maddala G.S. (1983): *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.

[27] Melino A. (1982): *Testing for Sample Selection Bias*, Review of Economic Studies, **49**, 151-153.

[28] Morimune K. (1976): *Some Minimum Chi-Square Estimators and Comparison of Normal and Logistic Models in Qualitative Response Analysis*, Stanford DP208.

[29] Muhleisen M. (1993): *Simulation Estimation of State Dependence Effects in Unemployment* , University of Munich.

[30] Narendranathan W. and P. Elias (1990): *Influences of Past History of the Incidence of Youth Unemployment: Empirical Findings for the U.K*, Warwick Economic Research Papers, 369.

[31] Olsen T. (1980): *A Least Squares Correction for Selectivity Bias*, Econometrica, **48**, 1815-1820.

[32] Pakes A. and D. Pollard (1989): *Simulation and the Asymptotics of the Optimization Estimators* , Econometrica, **57**, 1027-1058.

[33] Poirier D. (1980): *Partial Observability in the Bivariate Models*, Journal of Econometrics, **12**, 209-217.

[34] Prentice R.L. and L.P. Zhao (1990): *Correlated Binary Regression Using a Quadratic Exponential Model*, Biometrika, **77**, 642-648.

[35] Robinson P. (1982): *On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables*, Econometrica, **50**, 27-41.

[36] Sawa T. (1984): *A Generalization of the Univariate Logit Model and its Bivariate Extension*, London School of Economics.

[37] Thelot C. (1985): *Lois Logistiques à Deux Dimensions*, Annales de l'Insee, **58** , 123-150.

[38] Vella F. (1992): *Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models,* Journal of Applied Econometrics, **7** , 413-421.

[39] Vella F. (1993): *A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors*, International Economic Review (forthcoming).