MRE
MONTPELLIER
RECHERCHE
ECONOMIE

DOCUMENT DE RECHERCHE
WORKING PAPER

Benoît MULKAY

« Bivariate Probit Estimation for Panel Data:
a two-step Gauss-Hermite Quadrature Approach
with an application to product and process
innovations for France »

WP MRE 2018.1

Montpellier Recherche en Economie EA 7491 – Faculté d'Economie
Université de Montpellier - MUSE « Montpellier Université d'Excellence »
Contact : alain.marciano@umontpellier.fr

Bivariate Probit Estimation for Panel Data: a two-step Gauss-Hermite Quadrature Approach with an application to product and process innovations for France

Benoît MULKAY
Université de Montpellier - MRE*

October 2017

Abstract

This paper describes a method for computing a bivariate probit model on panel data with correlated random effects. Instead of an approach using simulated maximum likelihood, an alternative method based on a two-step Gauss-Hermite quadrature in order to evaluate the likelihood function is proposed in this article. A simulation shows the importance to estimate the correlation in random effects and the correlation between both equations. Finally an application is performed to estimate the determinants of product or process innovations on a large panel of French firms covering the period 2000-2013. It shows a positive and very large correlation between unobserved individual characteristics of firms, as well as a positive correlation between the idiosyncratic shocks. We show also small differences in the the determinants of product or process innovations.

JEL code : C33, C35, O31.

*Univ Montpellier, MRE. Address : Montpellier Recherche en Economie, Faculté d'Economie, Université de Montpellier, Avenue Raymond Dugrand, CS 79606, 34960 Montpellier Cedex 2. E-mail : benoit.mulkay@umontpellier.fr .The author thanks Jacques Mairesse and Wladimir Raymond for helpful comments on an earlier version of the paper.

1 Introduction

The estimation of a probit model on panel data is now usual. Many softwares propose such method of estimation which relies on individual random effects because the fixed effects approach is not valid due to the incidental parameters problem in the non-linear panel data model. In a seminal paper, Butler and Moffitt (1982) suggested to integrating the conditional density over the distribution of the individual effects in order to eliminate them by taking an average density. They proposed to use a Gauss-Hermite Quadrature to compute this integral for each individual in the panel.

On the other hand, many empirical problems imply two binary variables. The classic bivariate probit model is now common for cross-section data, but no usual procedure is available for panel data where there is individual random effects in each equation. Sometimes it is interesting to estimate the correlation between the individual effects of the two equations, because it shows how the unoberved heterogeneity between individuals is correlated accross equations, while there is still a correlation between the idiosyncratic error terms in the two equations.

Therefore a joint estimation of a bivariate probit model implies, as in the method of Butler and Moffitt, to integrate the conditional density over the bivariate distribution of individual random effects. Lee and Oguzoglu (2007) and Kano (2008) have proposed a simulated maximum likelihood approach where the individuals effects are integrated out by computing the double integral by simulation. But this procedure could be very time-consuming even with fast modern computer. In this article, an alternative approach based on a two-step Gauss-Hermite Quadrature is used in order to compute this double integral. Such a method has been already investigated in the context of a Heckman selection model on panel data by Raymond et al (2007, 2010). This paper adapts the method in the case of a bivariate panel data model in the section 2. A similar method has been proposed by Moussa and Delattre (2015) in the context of a causality analysis in a bivariate dynamic probit model.

A simulation analysis is done in Section 3 in order to show the importance of taking account individual effects in estimation of a probit model on panel data. The separated estimation of the two probit models shows clearly that they are consistent due to the fact that the model is correctly specified and that the correlations between the individual effects or between the error terms are only of second order. In fact like in a seemingly unrelated regression equations model, there is only a gain in efficiency of taking account of the covariance structure of the error terms composed of an individual effect and a idiosyncratic error. However the estimation of the correlations is of interest in order to assess the effects of unobserved heterogeneity on each equation.

Finally in Section 4, we present an application of this procedure in the case of the estimation of the determinants of product and process innovations on a large panel of 7 651 French firms during the period 2000 - 2013. The annual French data indicate only whether a firm, with positive R&D expenditures, introduces a product or a process innovation during the year. The model explaining the product or process innovations is simple because it depends only on the variables coming from the French R&D surveys : size of the firm and on the R&D intensity, characteristics of the R&D. Some conclusions are drawn for the estimation of the bivariate model for product and process innovation. We found some weak differences in the determinants of product or process innovation behavior. It seems that the differences betwwen them are diffuclt to assess within a firm or that a product innovation is always linked. The introduction of a new product leads to the introduction of a new process of production. The lack of information on the market on which the firms operates, or the level of competiton on this market could explained the difficulty to assess a difference between the two innovations.

2 The random effect bivariate probit

2.1 The bivariate probit model

Here we present briefly the bivariate probit model[1]. Thismodel is composed by 2 latent variables y_1^* and y_2^* which are explained by exogenous variables x_1 and x_2 and by possibly correlated error terms ε_1 and ε_2, normally distributed with unit variances[2]. and correlation coefficient τ :

$$\begin{cases} y_1^* = x_1'\beta_1 + \varepsilon_1 \\ y_2^* = x_2'\beta_2 + \varepsilon_2 \end{cases} \qquad \text{where } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \approx i.i.d.N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix} \right]$$

If the data are observed on several individuals only, we obtain the classical bivariate probit model when the observed variables y_1 and y_2 are defined as :

$$\begin{cases} y_1 = 1\left(y_1^* > 0\right) \\ y_2 = 1\left(y_2^* > 0\right) \end{cases}$$

where $1(...)$ is the indicator function with value one if the expression in parenthesis is true, and zero otherwise. The maximum likelihood estimator is then simply obtained with the classical transformation :

$$q_j = 2y_j - 1, \qquad j = 1 \text{ or } 2$$

such that the probability of a given choice between the 4 possible configurations of choice is :

$$\Pr\left(Y_1 = y_1, Y_2 = y_2 | x_1, x_2; \beta_1, \beta_2, \tau\right) = \Phi_2\left[q_1\left(x_1'\beta_1\right), q_2\left(x_2'\beta_2\right), q_1 q_2 \tau\right]$$

[1] See for example : Greene (2008, Section XXI.6).

[2] The classical normalization of variances to unity is done here because only the signs of the latent variables are observed. Therefore the scale does not matter.

with $\Phi_2(\bullet)$ is the cumulative density fonction of the bivariate standard normal distribution :

$$\begin{aligned}\Phi_2[u_1, u_2; \tau] &= \int_{-\infty}^{u_1}\int_{-\infty}^{u_2} \phi_2(z_1, z_2; \tau)\, dz_1 dz_2 \\ &= \int_{-\infty}^{u_1}\int_{-\infty}^{u_2} \frac{1}{2\pi}\frac{1}{\sqrt{1-\tau^2}} \exp\left\{-\frac{z_1^2 + z_2^2 - 2\tau z_1 z_2}{2(1-\tau^2)}\right\} dz_1 dz_2\end{aligned}$$

where $\phi_2(\bullet)$ is its probability density function of a bivariate standard normal variable with correlation τ. As the N observations of the sample are independent, the log-likelihood function is given by :

$$\ln \pounds = \sum_{i=1}^{n} \ln \Phi_2\left[q_{1,i}x'_{1,i}\beta_1, q_{2,i}x'_{2,i}\beta_2; q_{1,i}q_{2,i}\tau\right]$$

which should be maximized to obtain the maximum likelihood estimator of the bivariate probit model[3]. Greene (2008) gives the analytic first and second order conditions of the estimation problem.

When the observations come from a panel of individuals observed during a given time period (supposed here for simplicity to be the same for all individuals such that the panel is balanced), there is often an individual effect to take account of the unobserved heterogenity of the individuals. However with a non-linear model, like the probit model, if the individual effects are treated as fixed or correlated with the explanatory variables, there is an incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000; or Cameron and Trivedi; 2006). Thus we need to assume that individual effects are not correlated with the explanatory variables, and we use a random effects model with a specified distribution. These random effects are then eliminated by integrating over the distribution.

The univariate probit case has been first studied by Butler and Moffitt (1982) and Skrondal and Rabe-Hasketh (2004). We generalize this univariate random effect probit model to the case of two latent variables for $i = 1, ..., N$ individuals and $t = 1, ..., T$ time periods :

$$\begin{cases} y^*_{1,it} = x'_{1,it}\beta_1 + \alpha_{1,i} + \varepsilon_{1,it} \\ y^*_{2,it} = x'_{2,it}\beta_2 + \alpha_{2,i} + \varepsilon_{2,it} \end{cases} \quad \text{for } i = 1, ..., N \text{ and } t = 1, ..., T.$$

$$\text{where : } \begin{cases} \varepsilon_{it} = \begin{pmatrix} \varepsilon_{1,it} \\ \varepsilon_{2,it} \end{pmatrix} \approx i.i.d.N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix}\right] \\ \alpha_i = \begin{pmatrix} \alpha_{1,i} \\ \alpha_{2,i} \end{pmatrix} \approx i.i.d.N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right] \end{cases}$$

Here we assume implicitly that the observations are independant over time and accross indivisuals. The explanatory variables are exogenous with respect to

[3]See for example the estimation procedure `biprobit` in **Stata** (Hardin, 1996).

the error terms and with the individual random effects. This last hypothesis could be relaxed by introducing the average values of the regressors along the lines proposed by Mundlak (1978) if the individual effect can be decomposed on a linear combination of the averaged regressors plus an uncorrelated effects. The observed model is:

$$\begin{cases} y_{1,it} = 1\left(y_{1,it}^{*} > 0\right) \\ y_{2,it} = 1\left(y_{2,it}^{*} > 0\right) \end{cases}$$

Let us define the classical transformation of the observed variables :

$$\begin{cases} q_{1,it} = 2y_{1,it} - 1 \\ q_{2,it} = 2y_{2,it} - 1 \end{cases}$$

2.2 The individual joint density function

Because of the independance of observations over time, the conditional joint density for the T observations of the i^{th} individual is:

$$f_i\left(y_i | X_i, \alpha_i, \beta, \tau\right) = \prod_{t=1}^{T} f_{it}\left(y_{it} | X_{it}, \alpha_i, \beta, \tau\right)$$

As we have assumed a normal distribution for the error terms in the latent model, the density for an observation is given as in the bivariate probit model above:

$$\begin{aligned} & \Pr\left(Y_1 = y_1, Y_2 = y_2 | x_1, x_2; \alpha_i, \beta, \tau\right) \\ = \; & \Phi_2\left(q_1\left(x_1'\beta_1 + \alpha_1\right), q_2\left(x_2'\beta_2 + \alpha_2\right); q_1 q_2 \tau\right) \end{aligned}$$

where the individual random effects are added up to the conventional observable parts of the latent functions. The joint density for an individual, conditional to the vector of the individual random effects $\alpha_i = (\alpha_{1,i}, \alpha_{2,i})$, is then:

$$f_i\left(y_i | X_i, \alpha_i, \beta, \tau\right) = \prod_{t=1}^{T} \Phi_2\left(q_{1,it}\left(x_{1,it}'\beta_1 + \alpha_{1,i}\right), q_{2,it}\left(x_{2,it}'\beta_2 + \alpha_{2,i}\right); q_{1,it}q_{2,it}\tau\right)$$

Assuming a normal disribution for these individual random effects with variances σ_1^2 and σ_2^2 respectively and a correlation coefficient ρ, the density function for the individual effects is given by :

$$\begin{aligned} g_i\left(\alpha_i | \sigma_1^2, \sigma_2^2, \rho\right) = \; & \frac{1}{2\pi} \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2 \left(1 - \rho^2\right)}} \times \\ & \exp\left\{ \frac{-1}{2\left(1 - \rho^2\right)} \left[\left(\frac{\alpha_{1,i}}{\sigma_1}\right)^2 - 2\rho\left(\frac{\alpha_{1,i}}{\sigma_1}\right)\left(\frac{\alpha_{2,i}}{\sigma_2}\right) + \left(\frac{\alpha_{2,i}}{\sigma_2}\right)^2 \right] \right\} \end{aligned}$$

This density function does not depend on observables but on the three parameters which should be estimated. The unconditional (to the individual random effects) joint density for the i^{th} individual is obtained by averaging over the distribution of these individual effects :

$$\begin{aligned}
&\ell_i \left(y_i | \, X_i, \beta, \tau, \sigma_1^2, \sigma_2^2, \rho \right) \\
= &\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_i \left(y_i | \, X_i, \alpha_i, \beta, \tau \right) \times g_i \left(\alpha_i | \, \sigma_1^2, \sigma_2^2, \rho \right) d\alpha_{1,i} d\alpha_{2,i} \\
= &\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[\prod_{t=1}^{T} \Phi_2 \left(q_{1,it} \left(x'_{1,it} \beta_1 + \alpha_{1,i} \right), q_{2,it} \left(x'_{2,it} \beta_2 + \alpha_{2,i} \right); q_{1,it} q_{2,it} \tau \right) \right] \times \\
&\quad \frac{1}{2\pi} \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2 \left(1 - \rho^2 \right)}} \times \qquad (1) \\
&\quad \exp \left\{ \frac{-1}{2 \left(1 - \rho^2 \right)} \left[\left(\frac{\alpha_{1,i}}{\sigma_1} \right)^2 - 2\rho \left(\frac{\alpha_{1,i}}{\sigma_1} \right) \left(\frac{\alpha_{2,i}}{\sigma_2} \right) + \left(\frac{\alpha_{2,i}}{\sigma_2} \right)^2 \right] \right\} d\alpha_{1,i} d\alpha_{2,i}
\end{aligned}$$

2.3 Decomposition of the double integral

The evaluation of the individual likelihood function (1) requires the computation of a double integral. Lee and Oguzoglu (2007) and Kano (2008) have proposed a method of computation by simulation where $\alpha_{1,i}$ and $\alpha_{2,i}$ are randomly drawn in the bivariate normal distribution[4]. The individual joint density (unconditional to the individual random effects is approximated by :

$$\begin{aligned}
&\ell_i \left(y_i | \, X_i, \beta, \tau, \sigma_1^2, \sigma_2^2, \rho \right) \\
= &\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_i \left(y_i | \, X_i, \alpha_i, \beta, \tau \right) \times g \left(\alpha_i | \, \sigma_1^2, \sigma_2^2, \rho \right) d\alpha_{1,i} d\alpha_{2,i} \\
\simeq &\frac{1}{R} \sum_{r=1}^{R} \left[\prod_{t=1}^{T} \Phi_2 \left(q_{1,it} \left(x'_{1,it} \beta_1 + a_{1,i}^{(r)} \right), q_{2,it} \left(x'_{2,it} \beta_2 + a_{2,i}^{(r)} \right), q_{1,it} q_{2,it} \tau \right) \right] \qquad (2)
\end{aligned}$$

where $\left(a_{1,i}^{(r)} \right)$ and $\left(a_{2,i}^{(r)} \right)$ are R random draws in a bivariate normal distribution:

$$\left(\begin{array}{c} a_{1,i}^{(r)} \\ a_{1,i}^{(r)} \end{array} \right) \sim i.i.d.N \left[\left(\begin{array}{c} 0 \\ 0 \end{array} \right); \left(\begin{array}{cc} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_1^2 \end{array} \right) \right]$$

However the computation should be very time-consuming and imprecise even though we use modern simulator like GHK or Halton simulators, because we need to compute R cumulative density function with a large value of R in order to obtain sufficient precision in the log-likelihood function.

Instead we use the two-step Gauss-Hermite quadrature technique originally proposed in a couple of papers by Raymond *et al.* (2007, 2010) in the case of

[4] Miranda (2010) suggests the same procedure in an unpublished paper presented at the Mexican Stata Conference in 2010.

an Heckman sample selection model on panel data. There is also a paper by Moussa and Delattre (2015) which use an adaptative Gauss-Hermite procedure for a bivariate dynamic probit model estimated on panel data.This method relies on a decomposition of the two-dimensional normal distribution for the individual effects into a one-dimensional marginal distribution and a one-dimensional conditional distribution.

The unconditional joint density for the i^{th} individual is rewritten as:

$$\begin{aligned}
& \ell_i \left(y_i | X_i, \beta, \tau, \sigma_1^2, \sigma_2^2, \rho \right) \\
= & \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_i \left(y_i | X_i, \alpha_i, \beta, \tau \right) \times \frac{1}{2\pi} \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2 \left(1-\rho^2\right)}} \times \\
& \exp \left\{ \frac{-1}{2\left(1-\rho^2\right)} \left[\left(\frac{\alpha_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{\alpha_1}{\sigma_1} \right) \left(\frac{\alpha_2}{\sigma_2} \right) + \left(\frac{\alpha_2}{\sigma_2} \right)^2 \right] \right\} d\alpha_1 d\alpha_2 \\
= & \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_i \left(y_i | X_i, \alpha_i, \beta, \tau \right) \times \frac{1}{2\pi} \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2 \left(1-\rho^2\right)}} \times \\
& \exp \left\{ \frac{-1}{2\left(1-\rho^2\right)} \left[\left(\frac{\alpha_{1,i}}{\sigma_1} \right)^2 - 2\rho \left(\frac{\alpha_{1,i}}{\sigma_1} \right) \left(\frac{\alpha_{2,i}}{\sigma_2} \right) \right] \right\} \times \exp \left[-\frac{1}{2} \frac{\left(\alpha_{2,i} / \sigma_2 \right)^2}{1-\rho^2} \right] d\alpha_{1,i} d\alpha_{2,i}
\end{aligned}$$

which can be in turn rewritten as:

$$\ell_i \left(y_i | X_i, \beta, \sigma_1^2, \sigma_2^2, \rho \right) = \int_{-\infty}^{+\infty} H_i \left(\alpha_{2,i} \right) \times \exp \left[-\frac{1}{2} \frac{\left(\alpha_{2,i} / \sigma_2 \right)^2}{1-\rho^2} \right] d\alpha_{2,i} \qquad (3)$$

with

$$\begin{aligned}
H_i \left(\alpha_{2,i} \right) = & \frac{1}{2\pi} \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2 \left(1-\rho^2\right)}} \int_{-\infty}^{+\infty} f_i \left(y_i | X_i, \alpha_i, \beta, \tau \right) \times \\
& \exp \left\{ \frac{-1}{2\left(1-\rho^2\right)} \left[\left(\frac{\alpha_{1,i}}{\sigma_1} \right)^2 - 2\rho \left(\frac{\alpha_{1,i}}{\sigma_1} \right) \left(\frac{\alpha_{2,i}}{\sigma_2} \right) \right] \right\} d\alpha_{1,i}
\end{aligned}$$

Let us evaluate this last function by using a gauss-Hermite Quadrature by doing a change in variable such that $\left(\alpha_1 / \sigma_1 \right) = z_1 \sqrt{2\left(1-\rho^2\right)}$ with $d\alpha_1 =$

$\sigma_1\sqrt{2\left(1-\rho^2\right)}dz_1$ such that[5]:

$$\begin{aligned}
H\left(\alpha_2\right) &= \frac{1}{2\pi}\frac{\sigma_1\sqrt{2\left(1-\rho^2\right)}}{\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}}\int_{-\infty}^{+\infty}\ell_i\left(\left.y_i\right|X_i, z_1\sigma_1\sqrt{2\left(1-\rho^2\right)}, \alpha_2,\beta,\tau\right)\times \\
&\quad \exp\left\{-\frac{1}{2}\frac{z_1^2 2\left(1-\rho^2\right)}{1-\rho^2}\right\}\times\exp\left\{\frac{\rho}{\left(1-\rho^2\right)}z_1\sqrt{2\left(1-\rho^2\right)}\left(\frac{\alpha_2}{\sigma_2}\right)\right\}dz_1 \\
&= \frac{1}{\pi\sqrt{2\sigma_2^2}}\int_{-\infty}^{+\infty}f_i\left(\left.y_i\right|X_i, z_1\sigma_1\sqrt{2\left(1-\rho^2\right)}, \alpha_2,\beta,\tau\right)\times \\
&\qquad \exp\left\{\frac{\rho\sqrt{2}}{\sqrt{1-\rho^2}}\left(\frac{\alpha_2}{\sigma_2}\right)z_1\right\}\times\exp\left\{-z_1^2\right\}dz_1.
\end{aligned}$$

This is a Gaussian integral which can be approximated by a Gauss-Hermite quadrature with weights ω_m and abscissas a_m for M integration points ($m = 1,...,M$)[6] :

$$\int_{-\infty}^{+\infty}f\left(z\right)e^{-z^2}dz \simeq \sum_{m=1}^{M}\omega_m f\left(a_m\right)$$

Thus the $H\left(\alpha_2\right)$ function is approximated by :

$$H_i\left(\alpha_{2,i}\right)\simeq\frac{1}{\pi\sqrt{2\sigma_2^2}}\sum_{m=1}^{M}\omega_m f_i\left(\left.y_i\right|X_i, a_m\sigma_1\sqrt{2\left(1-\rho^2\right)}, \alpha_{2,i},\beta,\tau\right)\exp\left[\frac{\rho\sqrt{2}}{\sqrt{1-\rho^2}}\left(\frac{\alpha_{2,i}}{\sigma_2}\right)a_m\right].$$

Now the second step of the procedure is to introduce this function in the individual joint density $\ell_i\left(\left.y_i\right|X_i,\beta,\tau,\sigma_1^2,\sigma_2^2,\rho\right)$ above (3) with a second change in variables $\left(\alpha_2/\sigma_2\right) = z_2\sqrt{2\left(1-\rho^2\right)}$ with $d\alpha_2 = \sigma_2\sqrt{2\left(1-\rho^2\right)}dz_2$ to obtain:

[5] We drop the individual index i for the clarity of the exposition.

[6] The more the number of points, the more precise is the approximation. Genrally the number of points is set to $8, 12$ or 16 (see Cameron and Trivedi, 2005, Section XII.3.1). The values of weights ω_m and abscissas a_m can be found in mathematical textbooks.

$$
\begin{aligned}
& \ell_i\left(\left.y_i\right| X_i, \beta, \tau, \sigma_1^2, \sigma_2^2, \rho\right) \\
= & \int_{-\infty}^{+\infty} H\left(\alpha_2\right) \exp \left[-\frac{1}{2} \frac{\left(\alpha_2 / \sigma_2\right)^2}{1-\rho^2}\right] d \alpha_{2, i} \\
= & \frac{1}{\pi \sqrt{2 \sigma_2^2}} \int_{-\infty}^{+\infty} \sum_{m=1}^M \omega_m f_i\left(\left.y_i\right| X_i, a_m \sigma_1 \sqrt{2\left(1-\rho^2\right)}, \alpha_{2,} \beta, \tau\right) \\
& \qquad \times \exp \left[\frac{\rho \sqrt{2}}{\sqrt{1-\rho^2}}\left(\frac{\alpha_2}{\sigma_2}\right) a_m\right] \times \exp \left[-\frac{1}{2} \frac{\left(\alpha_2 / \sigma_2\right)^2}{1-\rho^2}\right] d \alpha_2 \\
= & \frac{\sigma_2 \sqrt{2\left(1-\rho^2\right)}}{\pi \sqrt{2 \sigma_2^2}} \int_{-\infty}^{+\infty} \sum_{m=1}^M \omega_m f_i\left(\left.y_i\right| X_i, a_m \sigma_1 \sqrt{2\left(1-\rho^2\right)}, z_2 \sigma_2 \sqrt{2\left(1-\rho^2\right)}, \beta, \tau\right) \\
& \qquad \times \exp \left[\frac{\rho \sqrt{2}}{\sqrt{1-\rho^2}} z_2 \sqrt{2\left(1-\rho^2\right)} a_m\right] \times \exp \left[-\frac{1}{2} \frac{z_2^2 2\left(1-\rho^2\right)}{1-\rho^2}\right] d z_2 \\
= & \frac{\sqrt{1-\rho^2}}{\pi} \int_{-\infty}^{+\infty} \sum_{m=1}^M \omega_m \ell_i\left(\left.y_i\right| X_i, a_m \sigma_1 \sqrt{2\left(1-\rho^2\right)}, z_2 \sigma_2 \sqrt{2\left(1-\rho^2\right)}, \beta, \tau\right) \\
& \qquad \times \exp \left[2 \rho a_m z_2\right] \times \exp \left[-z_2^2\right] d z_2
\end{aligned}
$$

A second Gauss-Hermite quadrature can be used to compute this Gaussian integral. For P integration points ($p=1, \ldots, P$), we have the weights ω_p and the abscissas a_p. Finally the individual joint density unconditional to the individual effects can be approximated by :

$$
\begin{aligned}
& \ell_i\left(\left.y_i\right| X_i, \beta, \tau, \sigma_1^2, \sigma_2^2, \rho\right) \qquad (4) \\
\simeq & \frac{\sqrt{\left(1-\rho^2\right)}}{\pi} \sum_{p=1}^P \sum_{m=1}^M \omega_p \omega_m \exp \left[2 \rho a_m a_p\right]\left(\prod_{t=1}^T \Phi_2\left(q_1 u_{1, m} ; q_2 u_{2, p} ; q_1 q_2 \tau\right)\right)
\end{aligned}
$$

where the arguments of the bivariate cumulative density function are :

$$
\begin{aligned}
u_{1, m} & =x_1^{\prime} \beta_1+a_m \sigma_1 \sqrt{2\left(1-\rho^2\right)} \\
u_{2, p} & =x_2^{\prime} \beta_2+a_p \sigma_2 \sqrt{2\left(1-\rho^2\right)}
\end{aligned}
$$

Finally as the individuals are independent, the log-likelihood function should be expressed as:

$$
\begin{aligned}
\ln \pounds &= \sum_{i=1}^{N} \ln \ell_i \left(y_i | \, X_i, \beta, \tau, \sigma_1^2, \sigma_2^2, \rho \right) \\
&= -N \ln (\pi) + \frac{N}{2} \ln \left(1 - \rho^2 \right) + \\
&\quad \sum_{i=1}^{N} \ln \left(\sum_{p=1}^{P} \sum_{m=1}^{M} \omega_p \omega_m \exp \left[2 \rho a_m a_p \right] \prod_{t=1}^{T} \Phi_2 \left(q_{1,i} u_{1,m,i} ; q_{2,i} u_{2,m,i} ; q_{1,i} q_{2,i} \tau \right) \right) \quad (5)
\end{aligned}
$$

In order to maximize this log-likelihood function, we can use the usual transformations for the correlation coefficients:

$$
\begin{cases} \rho^* = a \tanh \rho = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \\ \tau^* = a \tanh \tau = \frac{1}{2} \ln \left(\frac{1+\tau}{1-\tau} \right) \end{cases}
$$

or

$$
\begin{cases} \rho = \frac{\exp(2\rho^*)-1}{\exp(2\rho^*)+1} \\ \tau^* = \frac{\exp(2\tau^*)-1}{\exp(2\tau^*)+1} \end{cases}
$$

At each evaluation of the likelihood function, it is necessary to compute $N \times M \times P$ cumulative density functions of the bivariate normal variables Φ_2 with this two-step quadrature, which seems much more reasonable relative to the computation of $N \times R^2$ cumulative density functions for the simulated method. In fact we should have a sufficiently good approximation with $M = P = 12$ points in the Gauss-Hermite quadrature, even though we shoud take at least $R = 200$ points for the computation by simulation with less precision. The two procedures of estimation of the bivariate probit model by maximum likelihood have been written in a Stata program either with the simulated maximum likelihood or with the Gauss-Hermite quadrature[7].

3 A simulation

A simulation of the procedures for the estimation of the bivariate probit model has been performed in order to assess the effect of neglecting the correlation between the two equations, and between the unobserved heterogeneity in each equation. A set of observations for N individuals during T periods has been generated for a bivariate latent process:

$$
\begin{cases} y_1^* = \beta_{1,0} + \beta_{1,1} x_1 + \beta_{1,2} x_2 + \alpha_1 + \varepsilon_1 \\ y_2^* = \beta_{2,0} + \beta_{2,1} x_1 + \beta_{2,2} x_2 + \alpha_2 + \varepsilon_2 \end{cases}
$$

$$
\text{where } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \approx i.i.d.N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix} \right]
$$

[7]These programs uses the maximum likelihood procedures in Stata by Gould *et al.* (2010).

where the exogenous variables x_1 and x_2 have been drawn independently for each observations in a standard normal distribution. The individual effects α_1 and α_2 have been also drawn into a bivariate normal distribution with correlation ρ:

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \approx i.i.d.N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right].$$

Then the observable dependent variables are constructed on the basis of the sign of the corresponding latent variables:

$$\begin{cases} y_1 = 1\left(y_1^* > 0\right) \\ y_2 = 1\left(y_2^* > 0\right) \end{cases}$$

In the following simulations, the number of individuals has been set to 1 000 with 10 periods for each individuals, such that there are 10 000 observations in the panel data set which corresponds to the usual size of such data. The true structural parameters in the model are the following : $\beta_{1,0} = 0.50, \beta_{1,1} = 1.00, \beta_{1,2} = 0.00$ and $\beta_{2,0} = -0.50, \beta_{2,1} = -0.50, \beta_{2,2} = 1.00$. Therefore the second explanatory variable appears only in the second equation. The correlation coefficient of the error terms has been set to $\tau = 0.50$, the same value has the correlation coefficient between the individual random effects : $\rho = 0.50$, while the standard deviation of these individual effects are the same: $\sigma_1 = \sigma_2 = 2.00$.

The observed patterns of response in this simulated model is the shown in the Table 1.This simulated data set exhibits an association between both dependent variable with a Kendall's-t_b measure of association of 0.222 with a standard error of 0.010, as well a Pearson Chi-squared of 491.55 showing clearly a positive significant association between the two observed dependent variables. Moreover the tetrachoric correlation is 0.349 with a standard error 0.015 which is less than the assumed correlation between the error terms in the latent model.

		y_2		
		0	1	Total
y_1	0	29.7 %	12.4 %	42.0 %
	1	28.1 %	29.9 %	58.0%
	Total	57.7 %	42.3 %	100 %

Table 1 : Contingency table of the binary variables in the simualted model.

This simulated data set exhibits an association between both dependent variable with a Kendall's-t_b measure of association of 0.222 with a standard error of 0.010, as well a Pearson Chi-squared of 491.55 showing clearly a positive significant association between the two observed dependent variables. Moreover the tetrachoric correlation is 0.349 with a standard error 0.015 which is less than the assumed correlation between the error terms in the latent model.

The model is estimated by a pooled bivariate probit method where there are no individual effects as a benchmark for estimations. Then it is estimated using the Gauss-Hermite Quadrature (with 12 points) allowing individual effects. We proceed to four estimations : the first one (estimation 1) with the individual random effects but with a zero correlation between error terms ($\tau = 0$) and a zero correlation between the individual effects ($\rho = 0$), the second estimation (2) allows for an estimated correlation between the error terms (τ), wile the third estimation (3) allows only a correlation between the individual effects (ρ). Finally the last estimation (4) is the complete model where both correlations must be estimated. The standard likelihood ratio tests are performed in order to verify the assumption about the individual effects and the correlations in the model.

The Gauss-Hermite Quadrature procedure with 12 integration points is here faster by 40 % than the simulated maximum likelihhod procedures performed on the same dataset and on the same computer. Even though the convergence is quite fast in three or four iterations starting with the initial values from the two univariate panel probit estimations, it takes hovever between 7 minutes (for the first estimation) to 14 minutes (for the last estimation) to perform such a regression[8] on 10 000 observations for a model with only 3 parameters in each equation !

The benchmark estimation is clearly biased for the structural parameters of each equation because there is no individual effects. Only a correlation between the idiosyncratic error terms is estimated with an estimated value (0.511) close to the theoretical correlation (0.50). Let us remark that the parameter estimates are less than the half of their theoretical values. A likelihood ratio test rejects clearly this hypothesis of no individual random effects. Introducing individual random effects in the estimation but with no correlation is equivalent to two distinct estimation of a random effect probit model for each equation. The structural parameter estimates are now close to their theoretical value, taking account their standard errors. This is rather the case for the second equation, while the first one presents estimates a litle bit smaller than theitr theoretical values. However the estimated standard deviations of the individual effects are lower than expected for both equations. The likelihood ratio tests of the correlations between individual effects and/or between the error terms in the model clearly accept the presence of such correlations in the estimations. Moreover these estimated correlations have a very small estimated standard error, even though they are non-linear transformations of the estimated parameters in constructing interval confidence foor these correlations.

[8] The estimation are performed on a Dell OptiPlex 9010 with a i7 Intel processor running at 3.4 Ghz. The procedures are written in a standard code for maximum likelihhod estimation with Stata 12 software.

	Benchmark	(1)	(2)	(3)	(4)
	Equation 1				
$\beta_{1,0}$	0.212	0.429	0.440	0.436	0.446
[= 0.50]	*(0.013)*	*(0.047)*	*(0.047)*	*(0.052)*	*(0.051)*
$\beta_{1,1}$	0.408	0.928	0.927	0.942	0.939
[= 1.00]	*(0.014)*	*(0.027)*	*(0.026)*	*(0.027)*	*(0.027)*
$\beta_{2,1}$	−0.005	−0.005	−0.004	−0.004	−0.003
[= 0.00]	*(0.013)*	*(0.020)*	*(0.020)*	*(0.020)*	*(0.020)*
	Equation 2				
$\beta_{1,0}$	−0.218	−0.523	−0.507	−0.545	−0.511
[= −0.50]	*(0.013)*	*(0.067)*	*(0.061)*	*(0.067)*	*(0.065)*
$\beta_{1,1}$	−0.220	−0.489	−0.493	−0.495	−0.497
[= −0.50]	*(0.014)*	*(0.022)*	*(0.022)*	*(0.023)*	*(0.022)*
$\beta_{2,1}$	0.465	1.016	1.008	1.029	1.022
[= 1.00]	*(0.014)*	*(0.028)*	*(0.028)*	*(0.028)*	*(0.028)*
	Standard Error of Individual Effects				
σ_1	0	1.777	1.737	2.002	1.995
[= 2.00]	–	*(0.098)*	*(0.095)*	*(0.128)*	*(0.127)*
σ_2	0	1.668	1.637	1.902	1.895
[= 2.00]	–	*(0.092)*	*(0.090)*	*(0.124)*	*(0.122)*
	Correlations				
τ	0.511	0	0.534	0	0.476
[= 0.50]	*(0.014)*	–	*(0.046)*	–	*(0.036)*
ρ	0	0	0	0.550	0.536
[= 0.50]	–	–	–	*(0.027)*	*(0.027)*
Log Likelihood	−11972.8	−7896.7	−7816.4	−7762.9	−7688.3

Standard errors of estimates in parenthesis.
True value of parameters in squared brackets in first column.

Table 2 : Simulation Results

If a correlation between the error terms in both equations is allowed ($\tau \neq 0$), the estimated results are closer from the theoretical values, while the standard deviation of the individual effect are again under-estimated. In the opposite if only a correlation between individual random effects is allowed in the estimation ($\rho \neq 0$), there are small changes in the structural parameters estimates, even though the estimated standard deviations of the individual effects are now close from their theoretical values. The same conclusions are obtained in the full model where both correlations are estimated. All estimated parameters are now very close from their theoretical values, and the hypothesis of no correlations between individual effects and between error terms is clearly rejected by the likelihood ratio tests.

4 An application to product and process innovations

In this section, we investigate the behavior of product and process innovations on a panel of French firms on the period 1999 - 2007. The data comes from the annual R&D surveys conllected each year by the Ministry of Research. The 1999 reform of the R&D surveys in France introduced two new question s about the product or the process innovations. These question are stated as :

"During the year, did your enterprise or your group introduce new or significantly improved goods coming from the R&D activity of your firm?"
(Yes or No)
"During the year, did your enterprise or your group introduce new or significantly improved methods of manufacturing or producing goods or services coming from the R&D activity of your firm?"
(Yes or No)

These questions are slightly different from the usual Community Innovation Survey (CIS) questionnaire because in the latter the time period is prolonged over 3 years. For examples in the CIS 2004 questions, the first words are replaced by "During the three years 2002 to 2004,...". Moreover in the French R&D surveys, only innovations coming from the R&D done by the firm are considered. That excludes the innovations which were introduced without any R&D effort. On the other hand, the product or process innovations can be done by another firm in the group. This is why the answers to the CIS surveys and the R&D surveys are not directly comparable. But the most important difference is that in CIS surveys, the innovations are accounted for on the three years period.

A second problem arises from the fact that firms has many difficulties to disentangle product or process innovations, even though the definitions from the Oslo manual are quite precise (see the discussion in Mairesse and Mohnen, 2001). When a firm introduces a new product on the market, it changes and improves also the methods of production. Therefore, the product and process innovations is linked at the firm level. Even though this problem of measurement is a serious one, we will consider both types of innovations in the following. While there are some firms which innovates only in product or in process, the statistical difference between both types of innovations are thin. There are also cross relationships between product and process innovations.

4.1 Data and Descriptive Statistics

The sample of the French R&D surveys covers a 14 years period : from 2000 to 2013. We use only data coming from the annual R&D surveys. This avoids the losses of many observation due to the merge with other sources like the financial data of the firms. Only firms with at least 3 consecutive years of data are retained in the sample. There are 7 651 firms, corresponding to 37 847 observations in the unbalanced sample[9]. 69.4 % of firms report an innovation in a new product during the year, while there are 65.7 % of firms indicating a process innovation. But large firms are more innovative than smaller firms. When the share of innovators are weighted by employment, the rate of innovation rises to 81 % for both product and process innovations. In fact about 60 % of small and medium-sized firms report an innovation, either in product or in process, while 85 % of large firms (more than 2000 employees) introduce an innovation during a given year.

		Process Innovation		
		NO	YES	TOTAL
Product	NO	20.1 %	10.5 %	30.6 %
Innovation	YES	14.3 %	55.1 %	69.4 %
	TOTAL	34.4 %	65.6 %	100 %

37 847 observations, 7 651 firms, 2000 - 2013.

Table 3 : Share of Product and Process Innovators in France.

There is a positive and large association between product and process association. More than a half of the observations in the sample shows both types of innovation, while routhly one fifth of the sample reports no innovation at all, neither in product nor in process, even though the firms are doing R&D during the year. Finally 10.5 % of observations show only a product innovation, while 14.3 % only a process innovation. The Kendall's-τ_B measure of association is 0.437 with an asymptotic standard error of 0.005 showing a large and positive association between both types of innovations. Finally the tetrachoric correlation is 0.649 with a standard error 0.006. This clearly demonstrates the link between both type of innovations at the firm level. But this high correlation can be due to the unobserved characteristics of the firm, or rather to an idiosyncratic shock affecting both innovations at each period. We will estimate a simple bivariate probit model determining each type of innovations at the firm level to illustrate which correlations are the most important at the firm level.

In the estimation of the bivariate probit model for both types of innovation, only variables found in the annual R&D surveys are used. This includes the size of the firm, measured by the total employment (L) and its square $\left(L^2\right)$ to capture the positive effect of size on the declared innovations. The second main variable is the R&D intensity: R/Y, i.e. the total R&D expenditure divided by the total turnover of the firm. The squared value of the R&D intensity $(R/Y)^2$ is also introduced in the model in order to capture a non linear effect. The other explicative variables are : the logarithm of the productivity log(Prod), measured as the turnover per employee; the share of the external R&D (*Ext. R&D*); and the share of applied R&D (*Applied R&D*) or experimental R&D[10] (*Experim*

[9] The detailed composition of the sample is given in Appendix A.

[10] The share of basic R&D has been excluded to avoid multicolinearity, as these shares sum up to unity.

R&D) in the internal R&D. We also have in the surveys the total amount of R&D which is publicly financed by direct aid (grants, contract, subsidies,...). We use in the regression the share of R&D publicly financed (*Public R&D*) as well as a dummy if the firm has received a public aid to R&D in the year (*Subsidies*). A dummy variable (*R&D Lab*)for the existence of a formal research laboratory inside the firm is also available. Finally firms are asked to declare the share of their internal R&D which is devoted to new technologies (*Hitech*) which is decomposed in several categories : computer (*Computer*), environment (*Environment*), new materials (*New Mat.*) and biotechnologies (*Biotech*). The descriptive statistics on these variables are given in the Table 4 where t-tests for the equality of means and F-tests of equality of variances are performed between product / process innovators and non-innovators. The difference in variance is nearly always significant at 1% level. The equality of variance is only accepted in for R&D lab dummies (at 10% level), and for environment share of R&D for process innovations. Therefore the t-test is adapted to take account of a different variances between the groups.

The average size of firms is quite large with 841 employees. But the size distribution is skewed with a median of 120 employees. This is also a lot of small firms because the the first quartile of the distribution is only 35 employees. The larger the firm is, the more innovative in product or in process, because the avrage size of product innovator is 951 employees against 590 employees for non-innovators. The process innovators are larger again with an average size of 1002 employees. These differences in size between innovators and non-innovators are highly significant. The R&D intensity (Total R&D over the total turnover) is also skewed because the median R&D intensity is 5.2% while the average is 17.3% of total turnover. However, even though the differences between innovators and non-innovators for the R&D instensity are significant, theya re quite small. In fact the product innovators have a smaller R&D intensity (-0.8%) than the non-innovators, while for process innovators, they have an average R&D intensity slightly larger by 1.3%.

The characteristics of R&D asked in the French R&D surveys show that the share of external R&D which is done by a subsidiary, a firm within the group, a public or a private R&D laboratory is on avarge 9.1% of total firm's R&D. Once more there are only 53% of observations which have an external R&D, and 12.5% of them with a share of external R&D above 25%. The R&D is mainly devoted to applied R&D (44%) rather than for experimental R&D or development (44%). Only 4% of total R&D is devoted to basic or fundamental R&D. There are no significant differences between innovators and non-innovators with respect of the types of R&D : basic, applied or experimental.

31% of observations exhibit a public support to R&D by grants or subsidies[11]. But the average rate of subsidies is small with 5.3% of total R&D. In fact only one quarter of firms receives a subsidies larger than 1% of their total

[11] We don't take account of the R&D tax credit for Franch firms because they are general, from the fisrt euro of R&D since a first reform in 2004. A second reform in 2008 has increased the indirect support to &D by the r&D tax credit which is now one of the most generous in the world.

R&D. The direct public support for R&D is highly concentrated on a few large firms. There is a significant difference for product or process innovators which are more often aided by government for their R&D. However there is no significant difference in the rate of subsidies for the product innovators, while the difference is significant, but weak in favor of process innovators (+0.7%).

	All Sample		**Product Innovation**				**Process Innovation**			
Variable	*Mean*	*Std.Dev.*	*Dif. Mean*		*Var. Ratio*		*Dif. Mean*		*Var. Ratio*	
Employment	841	740	362	***	2.874	***	468	***	4.569	***
R/Y	0.173	0.291	−0.008	**	0.819	***	0.0129	***	0.968	**
log(*prod*)	5.190	0.774	0.0251	**	0.830	***	−0.0357	***	0.882	***
Ext. R&D	9.1%	16.4%	−1.16%	***	0.671	***	−1.18%	***	0.713	***
Basic R&D	4.0%	13.5%	−0.31%	*	0.735	***	0.13%		0.873	***
Applied R&D	51.8%	39.2%	0.47%		0.908	***	−0.21%		0.911	***
Experim. R&D	44.2%	39.4%	−0.17%		0.925	***	0.08%		0.925	***
Subsidies	30.8%	46.1%	6.39%	***	1.135	***	6.76%	***	1.141	***
Public R&D	5.3%	15.0%	0.25%		0.926	***	0.69%	***	1.042	***
R&D Lab	55.2%	49.7%	13.71%	***	0.972	*	9.94%	***	0.971	*
Hitech	48.9%	45.3%	7.81%	***	0.940	***	10.55%	***	0.971	*
Computer	22.7%	38.5%	6.43%	***	1.212	***	6.72%	***	1.244	***
Environment	5.4%	16.8%	0.46%	**	0.912	***	0.62%	***	0.993	
NewMat	12.9%	29.4%	3.24%	***	1.170	***	4.12%	***	1.256	***
Biotech	7.8%	25.3%	−2.32%	***	0.749	***	−0.90%	**	0.873	***

37 847 observations, 7 651 firms, 2000 – 2013.
Product innovators : 26 267 observations (69.4%), Process innovators : 24 869 observations (65.7%)
Dif. Mean : Difference in means between innovators and non-innovators, with associated t-test.
Var. Ratio : Variance Ratio between innovators and non-innovators, with associated F-test
*** : significant at 1% level, ** : significant at 5% level, * : significant at 10% level.

Table 4 : Descritive Statistics and tests on the variables.

A formal research laboratory inside the firm seems to have a large impact on the innovation process. Even though 55% of firms declare on average to have such a research laboratory, the share is 59% for product innovators and process innovators. The differences with the non-innovators are here large and very significant. Finally, the share of R&D in new techhnologies which is splited in four categories : computer science, environment, new materials and biotechnologies[12]. On average, about 23% of R&D is done in computer science (hardware,

[12] There are two other categories : the nanotechnologies category which appears only since 2007, with less than 1.4% of R&D on average. We have pooled this category with the new materials category. Finaly, we have excluded the category "social sciences and humanities" which represents 1.2% of total R&D, but which could not be considered as a hitech R&D.

software, networks, applications). while the second most important category is the research in new materials (including nanotechnologies) with 13%. The biotechnologies represent on average less than 8% of the total R&D which is quite small. But this is concentrated on a few industries (chemicals, pharmaceuticals). There is only 5% of R&D in environment (green product, new methods of production protecting the environment, reduction in pollution,...). For these categories of R&D, the differences between (product / process) innovators and non-innovators are significant in favour of innovators which devoted a larger share of their R&D to these new technologies. However there is an exception with the biotechnologies where the innovators do less R&D in this category than non-innovators. This can be explaines by the fact that biotechnologies are used only in a few industries which account for a small part of firms in the sample. The innovators in other industries do more R&D in other categories which leads to this effects.

4.2 A simple model

A first simple model with only the effects of the size of firms and the R&D instensity has been estimated with different estimators. First we consider a probit estimator without individual effects, separately for the two types of innovations: product or process. We performed also a bivariate probit estimation, still without any individual effects. This allows to estimate the correlation between between the idiosyncratic shocks in both types of innovations. Then we estimate the probit model with individual random effects : separately for each type of innovation, and finaly with a possible correlation between the idiosyncratic errors terms (like in the bivariate standard probit model) and also a possible correlation betwen the individual effects standing for the unobserved characteristcics of firms. This is the method developped in the section 2 of the paper. The Table 5 shows the results of these estimations, with a full set of time dummies which are are always jointly significant, but are not reported here.

All the parameters estimates are highly significant because the sample size is large with 37 847 observations, while there are up to 36 structural parameters to estimate in the full bivariate model plus the two standard errors of the individual effects and the two correlation coefficients. Moreover the likelihood ratio tests reject clearly the assumptions of the absence of individual effects, as well as the zero correlations between these individual effects or between the equations. The standard errors of the individual effects are quite the same for both types of innovations. They are only slightly larger in the case of correlated individual effects. This correlation between the individual effects is very large with an estimates of $\widehat{\rho} = 0.78$. Therefore the unobserved individual characteristics of the firm affect in the same way the probability to innovate in product or in process. A given set of firm characteristics leads to both types of innovations or none of these innovations. Finally the bivariate panel probit estimation allows to disentangle the correlation between the individual effects from the correlation in the idiosyncratic error terms between the two equations, which is precisely estimated with $\widehat{\tau} = 0.49$. This is smaller relative to the estimated correlation in a model without individual effects ($\widehat{\tau} = 0.66$). This suggests that a part of the correlation between innovations is due to the unobserved firm's characteristics.

	PROBIT		**PANEL PROBIT**	
	UNIVARIATE	BIVARIATE	UNIVARIATE	BIVARIATE
		PRODUCT INNOVATION		
L	0.020	0.022	0.026	0.020
	(0.006)	(0.006)	(0.007)	(0.005)
$L^2/1000$	−0.081	−0.089	−0.109	−0.084
	(0.0027)	(0.028)	(0.034)	(0.027)
(R/Y)	0.155	0.174	0.515	0.460
	(0.098)	(0.098)	(0.110)	(0.108)
$(R/Y)^2$	−0.238	−0.246	−0.373	−0.332
	(0.070)	(0.071)	(0.078)	(0.077)
σ_1			0.548	0.598
			(0.017)	(0.021)
		PROCESS INNOVATION		
L	0.031	0.033	0.026	0.028
	(0.008)	(0.008)	(0.009)	(0.006)
$L^2/1000$	−0.126	−0.098	−0.096	−0.103
	(0.035)	(0.029)	(0.036)	(0.027)
(R/Y)	0.480	0.487	0.741	0.784
	(0.096)	(0.096)	(0.106)	(0.106)
$(R/Y)^2$	−0.357	−0.357	−0.437	−0.435
	(0.068)	(0.069)	(0.076)	(0.075)
σ_2			0.526	0.559
			(0.017)	(0.019)
		CORRELATIONS		
τ (Errors)		0.656		0.488
		(0.010)		(0.010)
ρ (Ind. Eff.)				0.784
				(0.008)
Log.Lik.	−46270	−42817	−40422	−37962

37 847 observations, 7 651 firms, 2000 - 2013.
Asymptotic standard errors in parenthesis.

Table 5 : Simple Model - Parameter Estimates

The results exhibit a clear non-linear effect of the size, even though the non-linearity has a significant effect on large firms with more than 20 000 employees. The maximum effect of the size is reached for a total employment of about 120 000 employees for product innovation, and about 130 000 employees for process innovation. This threshold (120 000 employees) concerns only 59 observations (0.16%) on only 8 different firms ! The effect of the size on innovation is alwways smaller for the product innovation than for the process innovation, whatever the considered model. In fact a similar non-linear effect is found with or without individual effects in the estimation.

On the othe hand, the introduction of individual effects rises the non-linear effect of the R&D intensity on the innovations in product or in process. The estimates are larger in magnitude for the level and quadartic coefficients for both types of innovations. The effect of the R&D intensity is positive but decreasing In consequence the effect of R&D intensity on product innovation increases up to a maximum which is routhly 35% of the total firm's turnover without individual effects, while the maximum effect is found for a R&D instensity by 70% when individual effects are introduced in the model. The two thresholds for process innovations are higher with 68% without individual effects and 92% with the individual effects. Once again, the R&D intensity effects is larger for process innovations than for product innovations.

4.3 An extended model

The Table 6 presents the results for an extended model where the other explicative variables are introduced in the estimation. Only the estimates for the full bivariate probit model with correlated effects are presented, corrsponding to the fourth columns in Table5 for the simple model. We have also computed the average marginal effects for each variable in the model which measure the average impact of each explanatory variable on the probability to innovate in product or in process. The standard errors of the individual effects are roughly the same as in the simple model with only a very slight reduction to $\sigma_1 = 0.55$ and $\sigma_2 = 0.54$. The additional explanatory variables do not into account a large share of the individual firm's characteristics. It still has a large unobserved heterogeneity in firm's innovation behavior. The correlation between the idiosyncratic error terms ($\widehat{\tau} = 0.48$) and between the individual effects ($\widehat{\rho} = 0.78$) are quite unchanged with these additional variables. They do not bring specific elements in the product or process innovation behavior.

The non-linear effects of the firm's size or the R&D intensity is the same as above with a maximum effect at a large value : 120 000 employees for the product innovations and 130 000 employees for the process innovations. Relative to a firm with zero employment, a firm at the median employment (173 workers) has a 0.50% rise in the probability to innovate in product, and a 0.84% rise in the probability to innovate in process. For the R&D instensity, the same incrteasing and convex effects is found for both types of innovations, with a maximum effect at a R&D intensity rate of respectively 68% and 90% for product or process innovations. Relative to a firm with zero R&D, a firm at the median R&D instensity value (5.2%) shows an increase in the probability to innovate by 1.51% for product innovation, and 2.91% for process innovations. The effect of the size and the R&D intensity are quite small for a large proportion of firms in the sample.

	ESTIMATES		MARGINAL EFFECTS	
	PRODUCT	PROCESS	PRODUCT	PROCESS
L	0.022***	0.031***	0.0066***	0.0100***
	(0.006)	(0.006)	(0.0017)	(0.0019)
$L^2/1000$	−0.093***	−0.117***	−0.028***	−0.038***
	(0.028)	(0.027)	(0.008)	(0.009)
(R/Y)	0.337***	0.572***	0.1002***	0.1859***
	(0.119)	(0.117)	(0.0354)	(0.0380)
$(R/Y)^2$	−0.249***	−0.320***	−0.0740***	−0.1039***
	(0.080)	(0.078)	(0.0237)	(0.0255)
$\log(prod)$	0.056***	0.015	0.0165***	0.0049
	(0.020)	(0.019)	(0.0058)	(0.0062)
Ext. R&D	−0.125*	−0.193***	−0.0373*	−0.0626***
	(0.069)	(0.068)	(0.0207)	(0.0221)
Applied R&D	0.120	−0.086	0.0357	−0.0281
	(0.079)	(0.078)	(0.0297)	(0.0252)
Experim. R&D	0.100	−0.037	0.0297	−0.0121
	(0.079)	(0.078)	(0.0235)	(0.0253)
Subsidies Index	0.177***	0.131***	0.0526***	0.0426***
	(0.028)	(0.027)	(0.0082)	(0.0086)
Subsidies Rate	−0.131*	−0.091	−0.0390*	−0.0296
	(0.079)	(0.020)	(0.0236)	(0.0252)
R&D Lab	0.431***	0.337***	0.1283***	0.1094***
	(0.020)	(0.020)	(0.0060)	(0.0064)
Computer	0.289***	0.365***	0.0860***	0.1184***
	(0.038)	(0.038)	(0.0114)	(0.0122)
Environment	0.058	0.247***	0.0173	0.0801***
	(0.069)	(0.067)	(0.0204)	(0.0217)
New.Mat.	0.271***	0.404***	0.0805***	0.1313***
	(0.040)	(0.039)	(0.0120)	(0.0127)
Biotech	−0.123**	0.049	−0.0367**	0.0160
	(0.052)	(0.051)	(0.0153)	(0.0165)

37 847 observations, 7 651 firms, 2000 - 2013.
Asymptotic standard errors in parenthesis.
*** : significant at 1% level, ** : significant at 5% level, * : significant at 10% level.

Table 6 : Estimates of the Full model with Average marginal effects.

The effect of the firm's productivity is positive for both types of innovation, but only significant for product innovations with a marginal effect of 0.0165. This means that a firm which is twice more productive than another firm's has an average increase in the probability to innovate in product by 1.1%. This effect is again small in magnitude. There is a negative effect of the share of external R&D on the probability to innovate. The probbaility to innovate is smaller in a firm with a large share of external R&D. If a firm has only external R&D, this reduces the probability to innovate by 3.7% in product and by 6.3% in process. Once again here, the effect is larger for process than for product innovations. In the other hand the composition of internal R&D between basic, applied or experimental R&^{2}D has no significant effect on the probability to innovate for both types of innovations. The public support to the private R&D by direct aid (grants, subsidies,...) has a positive effect on the probability to innovate but this effect is only significant when a dummy variable for such aid is introduced in the model. The subsidy rate (the share of R&D publicly supported) has even a negative, but not significant, effect on both types of innovations. The fact to received a direct aid to R&D from the government can be considered as a signal that the R&D could lead more probably to an innovation in product (+5.3%) or in process (+4.3%). even though the marginal effects is larger for product innovations, the difference with the product innovation is not significant. The intensity of the aid has then a decreasing effect on the probability to innovate, even though it remains always positive.

A formal research laboratory inside the firm has a very important and significant effect on the probability to innovate, relative to a diffuse research made by the firm. A research laboratory increases the probability to innovate by 12.8% for product innovation and by 10.9% for process innovation. The difference between the types of innovation is again not significant. May be when the firm installs a formal research laboratory, it expects larger results in terms in innovation. Finally the categories of R&D in high technologies has jointly a positive effects on the probability to innovate. If R&D is devoted only to computer science, the probability to innovate is increased by 8.6% for product and 11.8% for process innovations. The effect is larger for process innovations, even though the difference with product innovation is not really significant. This means that R&D in computer science has a greater effect in firm's production than in the introduction to new products. The same effect are found for the R&D in new materials which has an larger impact on the probability to innovate in process (+13.1%) than on the probability to introduce a new product (+8.1%). Here the difference between product and process innovation is significant. The R&D in environment has only a positive effect on the probability to innovate in process. It seems that it is not the way for a product innovation, but it should implies more sustainable techniques of production by limiting pollution or saving energy. Finally the R&D in biotechnologies has a negative and signficant effect on product innovation, even though it has no effect on process innovation. The interpretation of this effect could be linked to the descriptive statistics above where we found that only a few firms has R&D in biotechnologies, and that they are concentrated on only a few industries.

This shows that the differences in the innovation behavior is not globally different betwwen the product or the process innovation. The individual characteristics of the firms are a main determinants, while the explanatory variables has the same effects on product or proceess innovation, except that they have a larger effect on the probability to innovate in process than in products. The lack of information of the firm's market, its competitive position, and the strength of competition in its industry could be explained this difficulty to assess the product innovation behavior. For the process innovation, ther is a lack of information on the quality of workers or the degrees or machine use in the production process.

5 Conclusions

In this article, an alternative metod of estimation of a bivariate probit model on panel data is presented. In such bivariate probit model on panel data, the likelihood function implies to integrate the density conditional over the distribution of the individual random effects in order to eliminate them by taking an average density. In the literature, some papers use simulations in order to compute the double integral. The alernative method relies on a double Gauss-Hermite quadrature procedure in order to evaluate the double integral. This paper develops the log-likelihood function in this case and a program is written in Stata to estimate such model. This program should be optimized in the future in order to reduce estimation time, may be by using an adaptative Gauss-Hermite procedure.

On panel data , it is important to introduce individial specific effects in order to avoid the omitted variable bias. This is shown in a simulation exercise where the pooled bivariate probit model is clearly rejected when there is no individual effects in estimation. The separated estimation of the two probit models is clearly consistent due to the fact that the model is correctly specified and that the correlations between the individual effects or between the error terms are only of second order. However a bivariate probit model allows also to estimate consistently the correlation between the individual random effect and between the idiosyncratic error terms in the 2 equations model. But the procedure should be long even thouh the number of iterations is reduced.

This procedure is applied in the case of the estimation of the determinants of product and process innovations on a panel of French firms during the period 2000 - 2013. There is a positive correlation between the idiosyncratic errors terms for both innovations that a shock affects in the same way both types of innovations. Here the model explaining the product or process innovations showd that the size of the firm and positively on the R&D intensity.

The unobserved heterogeneity also affects both product and process innovations with a very high positive correlation of 78 %, which can be due to the fact that our model is very simple. The firm's unobserved characteristics may lead to a firm's innovative behaviour for both innovations, but these characteristics should come from the internal organization of the firm or from the market on which it operates. A further investigation of these determinants should be on the next agenda of research. The large correlated effects could be also the sign of a high persistence of innovative behaviour at the firm's level. The firm's characteristics can also affet persistently the product or process innovations. We should investigate the persistence of this innovative behaviour in a following paper.

There is a positive non-linear convex effect of the size of the firm of the same magnitude for both types of innovations, while there is also a non-linear positive effect of the R&D intensity, but these effects is larger for process than for product innovation. We found some weak differences in the determinants of product or process innovation behavior. It seems that the differences betwwen them are diffuclt to assess within a firm or that a product innovation is always linked. The introduction of a new product leads to the introduction of a new process of production. The lack of information on the market on which the firms operates, or the level of competiton on this market could explained the difficulty to assess a difference between the two innovations.

References

BUTLER, J. S. and Robert MOFFITT (1982) : "A computationnaly Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model", *Econometrica*, 50(3), May 1982, 761-764.

CAMERON Colin A. and Pravin K. TRIVEDI (2005) : *Microeconometrics: Methods and Applications*, Cambridge University Press.

GOULD, William W., Jeffrey S. PITBLADO, and Brian P. POI (2010) : *Maximum Likelihood Estimation with Stata (4th edition)*, College

Station, TX: Stata Press.

GREENE, William (2003) : *Econometric Analysis* (5th edition), Prentice-Hall - Pearson Education.

HARDIN, James W. (1996) : "sg61: Bivariate probit models", *Stata Technical Bulletin Reprints*, vol. 6, pp. 152–158.

KANO Shigeki (2008) : "Like Husband, Like Wife: A Bivariate Dynamic Probit Analysis of Spousal Obesities", mimeo, Osaka Prefecture University.

LANCASTER, Tony (2000) : "The Incidental Parameter Problem since 1948", *Journal of Econometrics*, 95, 391-413.

LEE, Wang-Sheng and Umut OGUZOGLU (2007) : "Well-Being and Ill-Being: A Bivariate Panel Data Analysis", IZA Discussion Paper Series No. 3108, October 2007.

MAIRESSE, Jacques and Pierre MOHNEN (2001) : "Using innovation Surveys for Econometric Analysis", NBER Working paper No. 15857.

MIRANDA, Alfonso (2010) : "Dynamic Bivariate Probit Models for Panel Data", paper presented at the Mexican Users Group Meeting, available at http://fmwww.bc.edu/repec/msug2010/mex10sug_miranda.pdf

MOUSSA, Richard and Eric DELATTRE (2015) : "On the estimation of causality in a bivariate dynamic probit model on panel data with Stata software. A technical review", THEMA Working Paper n°2015-04, Université de Cergy-Pontoise;

MUNDLAK, Yairn (1978) : "On the Pooling of Time Series and Cross Section Data", *Econometrica*, 46, 69-85.

NEYMAN, J. and Elizabeth SCOTT (1948) : "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16(1), January 1948, 1-32.

RAYMOND Wladimir, Pierre MOHNEN, Franz PALM, and Sybrand SCHIM VAN DER LOEFF (2007) : "The Behavior of the Maximum Likelihood Estimator of Dynamic Panel Data Sample Selection Models", CESIfo Working paper No. 1992.

RAYMOND Wladimir, Pierre MOHNEN, Franz PALM, and Sybrand SCHIM VAN DER LOEFF (2010) : "Persistence of Innovation in Dutch Manufacturing: Is it Spurious ?", *Review of Economics and Statistics*, 92(3), August 2010, 495-504.

SKRONDAL, Anders and Sophia RABE-HASKETH (2004) : *Generalized Latent Variable Modelling: Multilevel, Longitudinal and Structural Equation Models*, Boca Raton, FL, Chapman and Hall.

Appendix A : Characteristics of the sample

Year	Number	Percent
2000	1 379	3.6 %
2001	1 552	4.1 %
2002	1 899	5.0 %
2003	2 588	6.8 %
2004	3 065	8.1 %
2005	2 988	7.9 %
2006	3 818	7.4 %
2007	2 438	6.4 %
2008	2 847	7.5 %
2009	3 086	8.2 %
2010	3 058	8.1 %
2011	3 161	8.4 %
2012	3 654	9.7 %
2013	3 304	8.7 %
TOTAL	**37 847**	**100 %**

Table A1 : Number of Observation by Year

Industry	**Number**	**Percent**
Agriculture	57	0.7 %
Energy	69	0.9 %
Food, Beverage	349	4.6 %
Textile, Cloths	171	2.2 %
Wood, Furnitures	105	1.4 %
Chemicals	441	5.8 %
Pharamceuticals	165	2.2 %
Rubber, Plastics	376	4.9 %
Fabricated Metals	411	5.4 %
Electrical equipments	601	7.9 %
Electronic Goods	262	3.4 %
Machinery	553	7.2 %
Transport Equipments	241	3.1 %
Other Manufacturing	338	4.4 %
Building	97	1.3 %
Trade	470	6.1 %
Edition, Television	471	6.2 %
Software, Computing	877	11.5 %
Finance, Insurance	110	1.4 %
Management	863	11.3 %
R&D	400	5.2%
Other Business Services	187	2.4 %
Personal Services	37	0.5 %
TOTAL	**7 651**	**100 %**

Table A2 : Industry composition of the sample