

On the Estimation of Causality in a Bivariate Dynamic Probit Model on Panel Data with Stata Software: A Technical Review

Richard Moussa^{1,2*}, Eric Delattre¹

¹ThEMA, Université de Cergy-Pontoise, Cergy-Pontoise, France

²ENSEA, Abidjan, Côte D'Ivoire

Email: *moussa.kr@gmail.com, eric.delattre@u-cergy.fr

How to cite this paper: Moussa, R. and Delattre, E. (2018) On the Estimation of Causality in a Bivariate Dynamic Probit Model on Panel Data with Stata Software: A Technical Review. *Theoretical Economics Letters*, 8, 1257-1278.

<https://doi.org/10.4236/tel.2018.86083>

Received: January 25, 2018

Accepted: April 20, 2018

Published: April 24, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In order to assess causality between binary economic outcomes, we consider the estimation of a bivariate dynamic probit model on panel data that has the particularity to account the initial conditions of the dynamic process. Due to the intractable form of the likelihood function that is a two dimensions integral, we use an approximation method: The adaptative Gauss-Hermite quadrature method. For the accuracy of the method and to reduce computing time, we derive the gradient of the log-likelihood and the Hessian of the integrand. The estimation method has been implemented using the d1 method of Stata software. We made an empirical validation of our estimation method by applying on simulated data set. We also analyze the impact of the number of quadrature points on the estimations and on the estimation process duration. We then conclude that when exceeding 16 quadrature points on our simulated data set, the relative differences in the estimated coefficients are around 0.01% but the computing time grows up exponentially.

Keywords

Causality, Bivariate Dynamic Probit, Gauss-Hermite Quadrature, Simulated Likelihood, Gradient, Hessian

1. Introduction

Testing Granger causality has generated a large set of paper in the literature. The larger part of this literature concerns the case where we have continuous dependent variables. For binary outcomes, there is also a way to consider the causality problem. As described by [1] for a vector of dependant variables, the one order Granger causality can be analyse as a probability conditional indepen-

dence given a set of exogenous variables and the first order lagged dependent variables. And for a binary outcome in the dependent vector, one can use a probit probability that implies the use of latent variable.

For panel data case, as the one way fix effects model estimated on a finite sample has necessarily inconsistent estimators [2], the random effect model is used. Due to the fact that we aim to test for one order Granger causality, lagged dependent variables are included as explanatory variables. For the first wave of the panel, we do not have previous values for the dependent variables, and treating them casually or as exogenous leads to inconsistent estimators [2]. So we specify an other equation for initial conditions as described by [3]. The equation is allowed to have different explanatory variables and different idiosyncratic error terms from the dynamic equation.

This specification leads to a likelihood function with an intractable form that is a two dimensions integral with a large set of parameters to be estimated. The estimation of this likelihood function requires the use of numerical approximation of integral function such as maximum simulated likelihood (see [4] for more details) or Gauss-Hermite quadrature (for more details see [5] [6] [7]).

The main goal of this paper is to propose and to test a method for estimating a two equations system where the explanatory variables are binary in a panel data framework. To the extent of our knowledge, there is no program to do so, especially as we propose the calculation of the Hessian matrix and the gradient vector of our maximisation program.

In this paper, we discuss on the problem of testing Granger causality with a bivariate dynamic probit model taking into account the initial conditions. The organization of this paper is the following one. In Section 2 we explain the causality test method for bivariate probit model with panel data. In Section 3, we describe the estimation method available when the likelihood function has an intractable form (two dimensions integral in our case). Section 4 presents the calculation of the gradient with respect to the model parameters and the calculation of the Hessian matrix with respect to the random effects vector. In Section 5, we present a robustness analysis of our selected estimation method by doing some simulations¹.

2. Testing Causality with a Bivariate Dynamic Probit Model

This section aims to describe causality test method in the case of binary variables. We start by presenting the general approach in time series before introducing panel data case. We end this section by a discussion on the initial conditions problem.

2.1. Testing Causality: General Approach

Causality concept was introduced by [8] as a better predictability of a variable Y

¹For each section, specifics notations are down at the beginning of the section. Otherwise, in general $f(x)|_{x=a}$ denote the value of the function or the matrix f at the point a . When not specify, $|a|$ denote the integer part of the scalar a .

by the use of its lag values, the lag value of another variable Z and some controls X . In his paper, [8] distinguishes instantaneous causality that means Z_t is causing Y_t (if Z_t is included in the model it improves the predictability of Y_t than if not) from lag causality that means lag values of Z improve the predictability of Y_t . In this section, we rule out the instantaneous causality and deal with lag causality of one period.

The one period Granger causality can be rephrased in terms of conditional independence. Without loss of generality, we present the univariate case for time series. Let's Y_t and Z_t denote some dependent variables and X_t denote a set of controls variables. One period Granger non-causality from Z to Y is the conditional independence of Y_t from Z_{t-1} conditionally to X_t and Y_{t-1} . More clearly, Granger non-causality from Z to Y is:

$$f(Y_t | Y_{t-1}, X_t, Z_{t-1}) = f(Y_t | Y_{t-1}, X_t) \quad (1)$$

Note that the same kind of relationship can be written for Granger non-causality from Y to Z . As Y_t and Z_t are binary outcome variables, we can use latent variables (Y^* and Z^* respectively) and make the assumption that Y and Z have positive outcomes (equals to 1) if their latent variables are positive. The latent variables are defined as follows:

For the left side of the Equation (1) ($f(Y_t | Y_{t-1}, X_t, Z_{t-1})$):

$$Y_t^* = X_t \beta_1 + \delta_{11} Y_{t-1} + \delta_{12} Z_{t-1} + \epsilon_t^1 \quad (2)$$

$$Z_t^* = X_t \beta_2 + \delta_{21} Y_{t-1} + \delta_{22} Z_{t-1} + \epsilon_t^2 \quad (3)$$

For the right side of the Equation (1) ($f(Y_t | Y_{t-1}, X_t)$):

$$Y_t^* = X_t \beta_1 + \delta_{11} Y_{t-1} + \epsilon_t^1 \quad (4)$$

$$Z_t^* = X_t \beta_2 + \delta_{21} Z_{t-1} + \epsilon_t^2 \quad (5)$$

where

$$\begin{pmatrix} \epsilon_t^1 \\ \epsilon_t^2 \end{pmatrix} \rightsquigarrow N(0, \Sigma_\epsilon) \text{ with } \Sigma_\epsilon = \begin{pmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{pmatrix}$$

To fit the joint distribution of Y and Z conditionally to X (meaning that we estimate a bivariate model), we need to analyze four available situations that are $(Y = Z = 1)$, $(Y = Z = 0)$, $(Y = 1; Z = 0)$ and $(Y = 0; Z = 1)$. For each of these situations, we have:

$$\begin{aligned} & P(Y_t = 1, Z_t = 1 | X_t) \\ &= P(\epsilon_t^1 > -X_t \beta_1 - \delta_{11} Y_{t-1} - \delta_{12} Z_{t-1}, \epsilon_t^2 > -X_t \beta_2 - \delta_{21} Y_{t-1} - \delta_{22} Z_{t-1}) \\ & P(Y_t = 0, Z_t = 0 | X_t) \\ &= P(\epsilon_t^1 < -X_t \beta_1 - \delta_{11} Y_{t-1} - \delta_{12} Z_{t-1}, \epsilon_t^2 < -X_t \beta_2 - \delta_{21} Y_{t-1} - \delta_{22} Z_{t-1}) \\ & P(Y_t = 1, Z_t = 0 | X_t) \\ &= P(\epsilon_t^1 > -X_t \beta_1 - \delta_{11} Y_{t-1} - \delta_{12} Z_{t-1}, \epsilon_t^2 < -X_t \beta_2 - \delta_{21} Y_{t-1} - \delta_{22} Z_{t-1}) \\ & P(Y_t = 0, Z_t = 1 | X_t) \\ &= P(\epsilon_t^1 < -X_t \beta_1 - \delta_{11} Y_{t-1} - \delta_{12} Z_{t-1}, \epsilon_t^2 > -X_t \beta_2 - \delta_{21} Y_{t-1} - \delta_{22} Z_{t-1}) \end{aligned}$$

As we can see, by assuming $q_t^1 = 2Y_t - 1$ and $q_t^2 = 2Z_t - 1$, we can rewrite the probabilities above as:

$$P(Y_t, Z_t | X_t) = \Phi_2(q_t^1(X_t\beta_1 + \delta_{11}Y_{t-1} + \delta_{12}Z_{t-1}), q_t^2(X_t\beta_2 + \delta_{21}Y_{t-1} + \delta_{22}Z_{t-1}), q_t^1 q_t^2 \rho_\epsilon) \tag{6}$$

where $\Phi_2(\cdot)$ stands for the bivariate normal c.d.f.

Then testing Granger non-causality in this specification is testing $H0: \delta_{12} = 0$ for Z is not causing Y and testing $H0: \delta_{21} = 0$ for Y is not causing Z .

2.2. Testing Causality: Panel Data Case

For panel data case, two major approaches can be used. The first one is to consider that causal effect is not the same for all individuals in the panel ([9]). This approach is useful when individuals are heterogeneous or when the causal effect is not homogeneous. The specification for latent variables are:

$$Y_{it}^* = X_t^1 \beta_1 + \delta_{11,i} Y_{i,t-1} + \delta_{12,i} Z_{i,t-1} + \eta_i^1 + \zeta_{it}^1 \tag{7}$$

$$Z_{it}^* = X_t^2 \beta_2 + \delta_{21,i} Y_{i,t-1} + \delta_{22,i} Z_{i,t-1} + \eta_i^2 + \zeta_{it}^2 \tag{8}$$

where $(\eta_i^1, \eta_i^2)'$ denotes the individual random effects which are zero mean and covariance matrix Σ_η and $(\zeta_{it}^1, \zeta_{it}^2)'$ denote the idiosyncratic shocks which are zero mean and covariance matrix Σ_ζ with

$$\Sigma_\eta = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_\eta \\ \sigma_1 \sigma_2 \rho_\eta & \sigma_2^2 \end{pmatrix} \text{ and } \Sigma_\zeta = \begin{pmatrix} 1 & \rho_\zeta \\ \rho_\zeta & 1 \end{pmatrix}$$

In this approach, testing Granger non-causality is equivalent to test $\delta_{12,i} = 0, i = 1, \dots, N$ for Z is not causing Y and to test $\delta_{21,i} = 0, i = 1, \dots, N$ for Y is not causing Z .

The second approach (that is used in this paper) is to assume the causal effects, if they exist, are the same for all individuals in the panel. With the same notation that the previous case, the latent variables are:

$$Y_{it}^* = X_t^1 \beta_1 + \delta_{11} Y_{i,t-1} + \delta_{12} Z_{i,t-1} + \eta_i^1 + \zeta_{it}^1 \tag{9}$$

$$Z_{it}^* = X_t^2 \beta_2 + \delta_{21} Y_{i,t-1} + \delta_{22} Z_{i,t-1} + \eta_i^2 + \zeta_{it}^2 \tag{10}$$

Then testing Granger non-causality is equivalent to test $H0: \delta_{12} = 0$ for Z is not causing Y and to test $H0: \delta_{21} = 0$ for Y is not causing Z .

Finally, Equations (9) and (10) are the core of our problem. Since Y and Z are binary panel outcomes and each equation includes lag dependent variables, estimating jointly these two equations can be viewed as the estimation of a bivariate dynamic probit model.

2.3. Dealing with Initial Conditions

For the first wave of the panel (initial conditions), due to the fact that we do not have data for the previous state on Y and Z (no values for $Y_{i,0}$ and $Z_{i,0}$) we are not able to evaluate $P(Y_{i1}, Z_{i1} | Y_{i,0}, Z_{i,0}, X_i)$. By ignoring it in the individual likelihood, researchers also ignore the data generation process for the first wage

of the panel. This means that they assume the data generating process of the first wave of the panel to be exogenous or to be in equilibrium. These assumptions hold only if the individual random effects are degenerated. If this assumption is not fulfilled, the initial conditions (the first wave of the panel) are explained by the individual random effects and ignoring them leads to inconsistent parameter estimates [2].

The solution proposed by [2] for the univariate case and generalized by [3] is to estimate a static equation for the first wave of the panel (meaning that we do not introduce lagged dependent variables). In this static equation, the random effects are a linear combination of the random effects in the next wave of the panel and idiosyncratic error terms may have different structure from the idiosyncratic error terms in the dynamic equation. Formally, the latent variables for the first wave of the panel are defined as follows:

$$Y_{i,1}^* = X_i^1 \gamma_1 + \lambda_{11} \eta_i^1 + \lambda_{12} \eta_i^2 + \epsilon_i^1 \quad (11)$$

$$Z_{i,1}^* = X_i^2 \gamma_2 + \lambda_{21} \eta_i^1 + \lambda_{22} \eta_i^2 + \epsilon_i^2 \quad (12)$$

where $(\epsilon_i^1, \epsilon_i^2)'$ denotes the vector of idiosyncratic shocks which are zero mean and covariance matrix Σ_ϵ with $\Sigma_\epsilon = \begin{pmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{pmatrix}$.

As η^1 and η^2 are individual random effects respectively on Y and Z , λ_{12} and λ_{21} can be interpreted as the influence of the Y random individual effects (respectively Z random individual effects) on Z (respectively on Y) at the first wave of the panel.

3. Estimation Methods

Due to the fact that the likelihood function has an intractable form (an integral function), it is impossible to estimate this likelihood by usual methods. We then deal with numerical integration methods that are numerical approximation method for an integral. In this section we describe two major methods and argue for one of them to estimate our likelihood function.

3.1. Gauss-Hermite Quadrature Method

The Gauss-Hermite quadrature is a numerical approximation method use to close the value of an integral function. The default approach is related to an univariate integral of the form:

$$\int_{\mathbb{R}} f(x) \exp(-x^2) dx \quad (13)$$

where $\exp(-x^2)$ denotes the Gaussian factor². Then the integral above can be approximated using:

$$\int_{\mathbb{R}} f(x) \exp(-x^2) dx = \sum_{q=1}^Q w_q * f(x_q) \quad (14)$$

²Notice that even without this factor, one can use the Gauss-Hermite quadrature by using a straightforward transformation that is to multiply and divide the integrand $f(x)$ by a Gaussian factor $\exp(-x^2)$.

where $x_q, q=1, \dots, Q$ are nodes from the Hermite polynomial and $w_q, q=1, \dots, Q$ are corresponding weights.

This approximation supposes that the integrand can be well approximated by an $2Q+1$ order polynomial and that the integrand is sampled on a symmetric range centered on zero. So, for suitable results, these two assumptions must be taken into account.

We first assume that finding the optimal number of quadrature points can be achieved numerically. For the accuracy of the approximation, it is required to choose the optimal number of quadrature points. To do this, one can start with a number \bar{q} of quadrature points and increase it to assess if it significantly changes the result, and repeat this process until convergence in terms of overall likelihood value variation and estimated coefficients variation. But, it is also important to take into account the fact that increasing number of quadrature points also increases the computing time. An example of the impact of number of quadrature points on estimated results is given in Section 5.

For the problem of suitable sampling range, the solution of using the adaptative Gauss-Hermite quadrature was proposed by [5] and [6]. In this approach, instead of using $\exp(-x^2)$ as a Gaussian factor, they use a Gaussian density $\phi(\mu, \sigma^2)$ of mean μ and variance σ^2 . That means (see [5]):

$$\int_{\mathbb{R}} f(x) dx = \sum_{q=1}^Q w_q^* f(x_q^*) \tag{15}$$

Then the sampling range is transformed and the new nodes are $x_q^* = \mu + \sqrt{2}\sigma x_q$ and weights are $w_q^* = \sqrt{2}\sigma w_q \exp(x_q^2)$. For [5], one can choose the normal density with posterior mean and variance equal respectively to μ and σ^2 . For the implementation, we can start with $\mu=0$ and $\sigma=1$ and at each iteration of the likelihood maximization process, calculate the posterior weighted mean and variance of the quadrature points and use them to calculate the nodes and weights for the next iteration. For [6], one can choose μ to be the mode of the integrand $f(x)$ and σ to be the square of the Hessian of the log of integrand taken in the mode.

$$\sigma = \left(-\frac{\partial^2}{\partial x^2} \log(f(x)) \Big|_{x=\hat{x}} \right)^{-1/2} \tag{16}$$

For the multivariate integral case, the same approach is used. Without loss of generality, we discuss the bivariate case that can be apply to others multivariate cases. The function to approximate is written as follows:

$$\int_{\mathbb{R}^2} f(x, y) dx dy \tag{17}$$

With the assumption of independence between x and y (that can be overcome by using a Cholesky decomposition $x' = x$ and $y' = \rho x' + y$, see [5] or [7] for more precision on these Cholesky transformation or other transformations that can lead to similar results) the integral above can be approximated by:

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \sum_{q_1=1, q_2=1}^Q w_{q_1}^* w_{q_2}^* f(x_{q_1}^*, y_{q_1}^*) \tag{18}$$

And in this case, the nodes and weights are derived as follows:

$$\begin{pmatrix} x_{q_1}^* \\ y_{q_1}^* \end{pmatrix} = \hat{x} + \sqrt{2} * \left(-\frac{\partial^2}{\partial x^2} \log(f(x, y)) \Big|_{(x,y)=\hat{x}} \right)^{-1/2} * \begin{pmatrix} x_{q_1} \\ y_{q_1} \end{pmatrix} \quad (19)$$

and

$$\begin{pmatrix} w_{q_1}^* \\ w_{q_2}^* \end{pmatrix} = 2 * \left| -\frac{\partial^2}{\partial x^2} \log(f(x, y)) \Big|_{(x,y)=\hat{x}} \right|^{-1/2} * \begin{pmatrix} w_{q_1} \exp(x_{q_1}^2) \\ w_{q_2} \exp(x_{q_2}^2) \end{pmatrix} \quad (20)$$

where $|A|$ denotes the determinant of matrix A , and \hat{x} denotes the mode of the integrand $f(x, y)$.

Jackel (2005) also suggests that for the nodes with low weights (when contributions to the integral value are not significant) we can prune the range from those nodes in order to save calculation time. That means to set a scalar

$$\tau = \frac{w_1 w_{(Q+1)/2}}{Q} \text{ and drop all nodes with weights lower than this scalar.}$$

3.2. Maximum Simulated Likelihood Method

Maximum Simulated Likelihood method was introduced by [4] as a solution to maximization problems that have an integral as objective function. In this approach, the likelihood function is supposed to be defined as:

$$f(x, y) = \int_{\mathbb{R}^2} f^*(x, y, u_1, u_2) g(u_1, u_2) du_1 du_2 \quad (21)$$

where $g(u_1, u_2)$ is a probability distribution function, $f^*(x, y, u_1, u_2)$ is called simulator and denotes the function from which the mean value at some draws u_1 and u_2 gives an approximation of the overall likelihood. Without loss of generality, we only define the two dimensions case that can be generalized to fewer or larger dimensions integral. For this kind of likelihood function, [4] proposed as simulator the function $f^*(x, y, u_1, u_2)$ with u_1 and u_2 drawn from the same probability distribution function g (the probability distribution function of the individual random effects). Then the overall likelihood function can be approximated by (u_{1d} denotes the d^{th} draw from u_1 ; the same definition holds for u_{2d}):

$$f(x, y) = \frac{1}{D} \sum_{d=1}^D f^*(x, y, u_{1d}, u_{2d}) \quad (22)$$

where D denotes the number of draws.

To implement this method, we start by simulating a bivariate normal draw $N(0, I_2)$ and we give them the (u_1, u_2) covariance matrix structure. Then we calculate the value of the simulator at these transformed draws and we repeat D times. The overall likelihood is the mean of the simulator value at each transformed draw. At each iteration, once the random effects covariance matrix is calculated, we apply it to the simulated first normal draws to transform them in draws of the random effects and use them to calculate the likelihood. This process is repeated until convergence.

The simulated likelihood estimator is consistent and asymptotically equivalent to the likelihood estimator ([4]) if the number of draws tend to infinity faster than \sqrt{N} .

3.3. GHQ or MSL: What Method to Choose?

As described above, they are two main methods to estimate our likelihood function. To choose which method to implement, we deal with the accuracy and the computing time requirement.

For our estimations, we choose the adaptative Gauss-Hermite quadrature proposed by [7] for three main reasons.

- Our dataset is an unbalanced panel data with 10,569 individuals observed in mean over 26 years, that leads 255,206 observations. Due to the fact that the simulated likelihood method requires that the number of draw D be larger than the square of the number of observations, we do not use it to avoid waste of time in computing process.
- The Gauss-Hermite quadrature requires that we find the best number of quadrature Q that is the one for whom the integrand can be well approximated by an $2Q+1$ order polynomial. If Q is small, that reduces computing time. Our estimations are achieved in general for Q between 8 and 14. It means that at each iteration, for the likelihood value calculation, we do a weighted sum of between $8^2 = 64$ and $14^2 = 196$ terms.
- Using the Gauss-Hermite quadrature method reduces computing time but this computing time remains very long if the integrand is not sampled at the suitable range (meaning that the adaptative method has not been used). And in this case, the maximization process spends between two and three weeks before achieving convergence on an Intel Core i7 computer at 3.4 GHz with 8 GB of RAM memory. By applying the adaptative Gauss-Hermite quadrature, the computing time is significantly reduced and then, we spend between two and three days for achieving convergence on the same computer.

Note that the reduced convergence time mentioned above is in part due to the implementation of the first order derivatives of the likelihood function. Using the overall log-likelihood approximated by the Liu and Pierce adaptative Gauss-Hermite quadrature method, we can get derivatives with respect of all model parameters. The implementation of these derivative in the maximization process allows us to used the Stata's d1 method. The convergence time saved by this method is clearly huge. On our overall data set, with 8 quadratures points, when we use a non adaptative quadrature method, the convergence is not achieved: after 3 weeks of computation, the model underflows. When we use the [6] adaptative Gauss-Hermite quadrature, but without implementing the first order derivatives, the estimation process takes 11 days and 10 hours to achieve convergence. When we use the adaptative Gauss-Hermite quadrature in [6] with implemented the first order derivatives, the estimation process achieve convergence only after 1 day and 17 hours.

4. Chosen Method Requirements

In this section we describe some requirements of the selected method that is the adaptative Gauss-Hermite Quadrature. The first one is the fact that the adaptative Gauss-Hermite quadrature requires to derive the Hessian of the log of the integrand ([6]). The second one is that we derive the gradient of the overall likelihood function in order to use Stata’s d1 method (see [10]) for more accuracy and more speed in the calculations.

4.1. Gradient Vector Calculation

The gradient of the overall log-likelihood function has been calculated to speed up the maximization process. This will allow us to use the Stata’s d1 method that requires the implementation of the gradient vector in addition to the log-likelihood. The likelihood function for an individual i is:

$$L_i = \int_{\mathbb{R}^2} \Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon) \prod_{t=2}^{T_i} \Phi_2(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta) \phi(\eta_i, \Sigma_\eta) d\eta_i^1 d\eta_i^2 \quad (23)$$

where

$$\begin{aligned} q_{it}^1 &= 2y_{it}^1 - 1 \quad \forall i, t \\ q_{it}^2 &= 2y_{it}^2 - 1 \quad \forall i, t \\ h_i^0 &= Z_i^1 \gamma_1 + \lambda_{11} \eta_i^1 + \lambda_{12} \eta_i^2 \\ w_i^0 &= Z_i^2 \gamma_2 + \lambda_{21} \eta_i^1 + \lambda_{22} \eta_i^2 \\ \bar{h}_{it} &= X_{it}^1 \beta_1 + \delta_{11} h_{i,t-1} + \delta_{12} w_{i,t-1} + \eta_i^1 \\ \bar{w}_{it} &= X_{it}^2 \beta_2 + \delta_{21} h_{i,t-1} + \delta_{22} w_{i,t-1} + \eta_i^2 \end{aligned}$$

Using the adaptative Gauss-Hermite quadrature method by [6], the overall likelihood function is given by (we use the same notation that those used in Section 3):

$$\begin{aligned} L_i &= \sum_{k=1, j=1}^Q w_k^* w_j^* \Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon) \\ &\times \prod_{t=2}^{T_i} \Phi_2(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta) \phi(\eta_i, \Sigma_\eta) \Big|_{\eta_i^1 = x_k^*, \eta_i^2 = x_j^*} \end{aligned} \quad (24)$$

To get the gradient vector, the log-likelihood above must be derived with respect to 13 parameters that are: $\bar{\beta}_1 = (\beta_1, \delta_{11}, \delta_{12})'$, $\bar{\beta}_2 = (\beta_2, \delta_{21}, \delta_{22})'$, γ_1 , γ_2 , λ_{11} , λ_{12} , λ_{21} , λ_{22} , σ_1 , σ_2 , ρ_η , ρ_ζ , and ρ_ϵ .

Let’s l_{kj} denotes:

$$l_{kj} = \Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon) \prod_{t=2}^{T_i} \Phi_2(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta) \phi(\eta_i, \Sigma_\eta) \Big|_{\eta_i^1 = x_k^*, \eta_i^2 = x_j^*}$$

Then, the first order derivatives with respect to each α of the 13 parameters is given by:

$$\frac{\partial \log(L_i)}{\partial \alpha} = \sum_{k=1, j=1}^Q \frac{\partial l_{kj} / \partial \alpha}{L_i}$$

With respect to $\bar{\beta}_1$ the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \bar{\beta}_1} = l_{kj} \sum_{t=2}^{T_i} \frac{q_{it}^1 \phi(q_{it}^1 \bar{h}_{it}) \Phi_1 \left(\frac{q_{it}^2 \bar{w}_{it} - q_{it}^2 \rho_\zeta \bar{h}_{it}}{\sqrt{1 - \rho_\zeta^2}} \right)}{\Phi_2(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta)}$$

With respect to $\bar{\beta}_2$ the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \bar{\beta}_2} = l_{kj} \sum_{t=2}^{T_i} \frac{q_{it}^2 \phi(q_{it}^2 \bar{w}_{it}) \Phi_1 \left(\frac{q_{it}^1 \bar{h}_{it} - q_{it}^1 \rho_\zeta \bar{w}_{it}}{\sqrt{1 - \rho_\zeta^2}} \right)}{\Phi_2(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta)}$$

With respect to γ_1 the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \gamma_1} = l_{kj} \frac{q_{i0}^1 \phi(q_{i0}^1 h_i^0) \Phi_1 \left(\frac{q_{i0}^2 w_i^0 - q_{i0}^2 \rho_\epsilon h_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}$$

With respect to γ_2 the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \gamma_2} = l_{kj} \frac{q_{i0}^2 \phi(q_{i0}^2 w_i^0) \Phi_1 \left(\frac{q_{i0}^1 h_i^0 - q_{i0}^1 \rho_\epsilon w_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}$$

With respect to λ_{11} the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \lambda_{11}} = l_{kj} \frac{q_{i0}^1 x_k^* \phi(q_{i0}^1 h_i^0) \Phi_1 \left(\frac{q_{i0}^2 w_i^0 - q_{i0}^2 \rho_\epsilon h_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}$$

With respect to λ_{12} the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \lambda_{12}} = l_{kj} \frac{q_{i0}^1 x_j^* \phi(q_{i0}^1 h_i^0) \Phi_1 \left(\frac{q_{i0}^2 w_i^0 - q_{i0}^2 \rho_\epsilon h_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}$$

With respect to λ_{21} the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \lambda_{21}} = l_{kj} \frac{q_{i0}^2 x_k^* \phi(q_{i0}^2 w_i^0) \Phi_1 \left(\frac{q_{i0}^1 h_i^0 - q_{i0}^1 \rho_\epsilon w_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}$$

With respect to λ_{22} the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \lambda_{22}} = l_{kj} \frac{q_{i0}^2 x_j^* \phi(q_{i0}^2 w_i^0) \Phi_1 \left(\frac{q_{i0}^1 h_i^0 - q_{i0}^1 \rho_\epsilon w_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}$$

With respect to σ_1 the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \log(\sigma_1)} = l_{kj} * \left(-1 + \frac{(x_k^*/\sigma_1)^2 - \rho_\eta x_k^* x_j^*/(\sigma_1 \sigma_2)}{1 - \rho_\eta^2} \right)$$

With respect to σ_2 the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \log(\sigma_2)} = l_{kj} * \left(-1 + \frac{(x_j^*/\sigma_2)^2 - \rho_\eta x_k^* x_j^*/(\sigma_1 \sigma_2)}{1 - \rho_\eta^2} \right)$$

With respect to ρ_η the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \log\left(\frac{1 + \rho_\eta}{1 - \rho_\eta}\right)^{1/2}} = l_{kj} * \left(\rho_\eta - \frac{\rho_\eta \left((x_k^*/\sigma_1)^2 + (x_j^*/\sigma_2)^2 \right) - (1 + \rho_\eta^2) x_k^* x_j^*/(\sigma_1 \sigma_2)}{1 - \rho_\eta^2} \right)$$

With respect to ρ_ζ the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \log\left(\frac{1 + \rho_\zeta}{1 - \rho_\zeta}\right)^{1/2}} = l_{kj} \sum_{i=2}^{T_i} \frac{q_{ii}^1 q_{ii}^2 \phi(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta)}{\Phi_2(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta)}$$

With respect to ρ_ϵ the first order derivative is:

$$\frac{\partial l_{kj}}{\partial \log\left(\frac{1 + \rho_\epsilon}{1 - \rho_\epsilon}\right)^{1/2}} = l_{kj} \frac{q_{i0}^1 q_{i0}^2 \phi(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}$$

Remarks:

- For σ_1 , σ_2 , ρ_η , ρ_ζ , and ρ_ϵ , we used some transformations on parameters to insure that in the maximization process, each σ remains positive and each ρ remains between -1 and 1 at all iteration. For σ we use exponential transformation then in the derivation, we derive with respect to $\log(\sigma)$. For ρ we use arc-tangency transformation (*i.e.* $\frac{\exp(2\rho)-1}{\exp(2\rho)+1}$) then

in the derivation, we derive with respect to $\log\left(\frac{1 + \rho}{1 - \rho}\right)^{1/2}$.

- To easily derive a bivariate normal probability with zero mean, variance one and correlation ρ , we can transform it into an integral where the integrand is a product of an univariate normal density and an univariate normal probability as follows:

$$\Phi_2(x, y, \rho) = \int_{-\infty}^y \phi(v) \Phi\left(\frac{x - \rho v}{\sqrt{1 - \rho^2}}\right) dv = \int_{-\infty}^x \phi(u) \Phi\left(\frac{y - \rho u}{\sqrt{1 - \rho^2}}\right) du.$$

- Given the transformation above, the first order derivatives of $\Phi_2(x, y, \rho)$ with respect to x and y are respectively given by:

$$\frac{\partial \Phi_2(x, y, \rho)}{\partial x} = \phi(x) \Phi\left(\frac{y - \rho x}{\sqrt{1 - \rho^2}}\right)$$

$$\frac{\partial \Phi_2(x, y, \rho)}{\partial y} = \phi(y) \Phi\left(\frac{x - \rho y}{\sqrt{1 - \rho^2}}\right)$$

4.2. Hessian Matrix Calculation

For the requirement of the adaptative Gauss-Hermite quadrature method, we need to derive the Hessian matrix of the log of the integrand function with respect to the random effects vector³. From the individual likelihood function defined in Equation 23, the log of the integrand is:

$$\log(g(\eta_i^1, \eta_i^2)) = \log\left(\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon) \prod_{t=2}^{T_i} \Phi_2(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta) \phi(\eta_i, \Sigma_\eta)\right) \tag{25}$$

We derive from the log of the integrand in Equation (25) the Hessian matrix by calculating:

$$-\frac{\partial^2}{\partial (\eta_i^1)^2} \log(g(\eta_i^1, \eta_i^2))$$

$$-\frac{\partial^2}{\partial (\eta_i^2)^2} \log(g(\eta_i^1, \eta_i^2))$$

$$-\frac{\partial^2}{\partial \eta_i^1 \partial \eta_i^1} \log(g(\eta_i^1, \eta_i^2))$$

The first order derivatives are given by:

$$-\frac{\partial}{\partial \eta_i} \log(g) = -\frac{\Phi_2'(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)}{\Phi_2(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon)} - \sum_{t=2}^{T_i} \frac{\Phi_2'(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta)}{\Phi_2(q_{it}^1 \bar{h}_{it}, q_{it}^2 \bar{w}_{it}, q_{it}^1 q_{it}^2 \rho_\zeta)}$$

$$+ \frac{\eta_i^1 / \sigma_1^2 - \rho \eta_i^2 / (\sigma_1 \sigma_2)}{1 - \rho_\eta^2}$$

With respect to η_i^1 we have:

$$\Phi_{2\eta_i^1}'(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon) = q_{i0}^1 \lambda_{11} \phi(q_{i0}^1 h_i^0) \Phi_1\left(\frac{q_{i0}^2 w_i^0 - q_{i0}^2 \rho_\epsilon h_i^0}{\sqrt{1 - \rho_\epsilon^2}}\right)$$

$$+ q_{i0}^2 \lambda_{21} \phi(q_{i0}^2 w_i^0) \Phi_1\left(\frac{q_{i0}^1 h_i^0 - q_{i0}^1 \rho_\epsilon w_i^0}{\sqrt{1 - \rho_\epsilon^2}}\right)$$

³In this section, $\phi(x)$ denotes the univariate normal density function, $\phi(x, y, \rho)$ denote the bivariate normal density with correlation ρ , $\Phi_1(x)$ denote the univariate normal probability function, and $\Phi_2(x, y, \rho)$ denote the bivariate normal probability function with correlation ρ .

$$\Phi'_{2\eta_i^1} (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta) = q_i^1 \phi(q_i^1 \bar{h}_i) \Phi_1 \left(\frac{q_i^2 \bar{w}_i - q_i^2 \rho_\zeta \bar{h}_i}{\sqrt{1 - \rho_\zeta^2}} \right)$$

And with respect to η_i^2 we have:

$$\begin{aligned} \Phi'_{2\eta_i^2} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon) &= q_i^1 \lambda_{12} \phi(q_i^1 h_i^0) \Phi_1 \left(\frac{q_i^2 w_i^0 - q_i^2 \rho_\epsilon h_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \\ &\quad + q_i^2 \lambda_{22} \phi(q_i^2 w_i^0) \Phi_1 \left(\frac{q_i^1 h_i^0 - q_i^1 \rho_\epsilon w_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \\ \Phi'_{2\eta_i^2} (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta) &= \phi(q_i^2 \bar{w}_i) \Phi_1 \left(\frac{q_i^1 \bar{h}_i - q_i^1 \rho_\zeta \bar{w}_i}{\sqrt{1 - \rho_\zeta^2}} \right) \end{aligned}$$

The second order derivatives are given by:

$$\begin{aligned} -\frac{\partial^2}{\partial (\eta_i^1)^2} \log(g) &= -\frac{\Phi''_{2\eta_i^1} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon) \Phi_2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)}{\Phi_2^2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)} \\ &\quad + \frac{\Phi'^2_{2\eta_i^1} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)}{\Phi_2^2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)} \\ &\quad - \sum_{i=2}^{T_i} \left(\frac{\Phi''_{2\eta_i^1} (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta) \Phi_2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)}{\Phi_2^2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)} \right. \\ &\quad \left. - \frac{\Phi'^2_{2\eta_i^1} (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)}{\Phi_2^2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)} \right) + \frac{1}{\sigma_1^2 (1 - \rho_\eta^2)} \end{aligned} \tag{26}$$

$$\begin{aligned} -\frac{\partial^2}{\partial (\eta_i^2)^2} \log(g) &= -\frac{\Phi''_{2\eta_i^2} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon) \Phi_2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)}{\Phi_2^2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)} \\ &\quad + \frac{\Phi'^2_{2\eta_i^2} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)}{\Phi_2^2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)} \\ &\quad - \sum_{i=2}^{T_i} \left(\frac{\Phi''_{2\eta_i^2} (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta) \Phi_2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)}{\Phi_2^2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)} \right. \\ &\quad \left. - \frac{\Phi'^2_{2\eta_i^2} (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)}{\Phi_2^2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)} \right) + \frac{1}{\sigma_1^2 (1 - \rho_\eta^2)} \end{aligned} \tag{27}$$

$$\begin{aligned} -\frac{\partial^2}{\partial \eta_i^1 \partial \eta_i^2} \log(g) &= -\frac{\Phi''_{2\eta_i^1 \eta_i^2} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon) \Phi_2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)}{\Phi_2^2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)} \\ &\quad + \frac{\Phi'_{2\eta_i^1} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon) \Phi'_{2\eta_i^2} (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)}{\Phi_2^2 (q_i^1 h_i^0, q_i^2 w_i^0, q_i^1 q_i^2 \rho_\epsilon)} \\ &\quad - \sum_{i=2}^{T_i} \left(\frac{\Phi''_{2\eta_i^1 \eta_i^2} (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta) \Phi_2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)}{\Phi_2^2 (q_i^1 \bar{h}_i, q_i^2 \bar{w}_i, q_i^1 q_i^2 \rho_\zeta)} \right) \end{aligned}$$

$$\left. \frac{\Phi'_{2\eta^1} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) \Phi'_{2\eta^2} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right)}{\Phi_2^2 \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right)} \right) \quad (28)$$

$$-\frac{\rho_\eta}{\sigma_1 \sigma_2 (1 - \rho_\eta^2)}$$

where

$$\begin{aligned} & \Phi''_{2\eta^1} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) \\ &= -\bar{h}_i \Phi'_{2\eta^1} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) - \rho_\zeta \phi_{\eta^1} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) \\ & \Phi''_{2\eta^2} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) \\ &= -\bar{w}_i \Phi'_{2\eta^2} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) - \rho_\zeta \phi_{\eta^2} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) \\ & \Phi''_{2\eta^1 \eta^2} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) = q_{ii}^1 q_{ii}^2 \rho_\zeta \phi_{\eta^1} \left(q_{ii}^1 \bar{h}_i, q_{ii}^2 \bar{w}_i, q_{ii}^1 q_{ii}^2 \rho_\zeta \right) \\ & \Phi''_{2\eta^1} \left(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon \right) \\ &= \left(2\lambda_{11} \lambda_{21} - \rho_\epsilon \left(\lambda_{11}^2 + \lambda_{21}^2 \right) \right) \phi \left(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon \right) \\ & \quad - \lambda_{11}^2 h_i^0 \phi \left(q_{i0}^1 h_i^0 \right) \Phi_1 \left(\frac{q_{i0}^2 w_i^0 - \rho_\epsilon q_{i0}^2 h_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \\ & \quad - \lambda_{21}^2 w_i^0 \phi \left(q_{i0}^2 w_i^0 \right) \Phi_1 \left(\frac{q_{i0}^1 h_i^0 - \rho_\epsilon q_{i0}^1 w_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \\ & \Phi''_{2\eta^2} \left(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon \right) \\ &= \left(2\lambda_{12} \lambda_{22} - \rho_\epsilon \left(\lambda_{12}^2 + \lambda_{22}^2 \right) \right) \phi \left(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon \right) \\ & \quad - \lambda_{12}^2 h_i^0 \phi \left(q_{i0}^1 h_i^0 \right) \Phi_1 \left(\frac{q_{i0}^2 w_i^0 - \rho_\epsilon q_{i0}^2 h_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \\ & \quad - \lambda_{22}^2 w_i^0 \phi \left(q_{i0}^2 w_i^0 \right) \Phi_1 \left(\frac{q_{i0}^1 h_i^0 - \rho_\epsilon q_{i0}^1 w_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \\ & \Phi''_{2\eta^1 \eta^2} \left(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon \right) \\ &= q_{i0}^1 q_{i0}^2 \left(\lambda_{11} \lambda_{22} + \lambda_{12} \lambda_{21} - \rho_\epsilon \left(\lambda_{11} \lambda_{12} + \lambda_{21} \lambda_{22} \right) \right) \phi_2 \left(q_{i0}^1 h_i^0, q_{i0}^2 w_i^0, q_{i0}^1 q_{i0}^2 \rho_\epsilon \right) \\ & \quad - \lambda_{11} \lambda_{12} h_i^0 \phi \left(q_{i0}^1 h_i^0 \right) \Phi_1 \left(\frac{q_{i0}^2 w_i^0 - \rho_\epsilon q_{i0}^2 h_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \\ & \quad - \lambda_{21} \lambda_{22} w_i^0 \phi \left(q_{i0}^2 w_i^0 \right) \Phi_1 \left(\frac{q_{i0}^1 h_i^0 - \rho_\epsilon q_{i0}^1 w_i^0}{\sqrt{1 - \rho_\epsilon^2}} \right) \end{aligned}$$

Then, the Hessian matrix is given by:

$$H = \begin{pmatrix} -\frac{\partial^2}{\partial (\eta_i^1)^2} \log(g) & -\frac{\partial^2}{\partial \eta_i^1 \partial \eta_i^2} \log(g) \\ -\frac{\partial^2}{\partial \eta_i^1 \partial \eta_i^2} \log(g) & -\frac{\partial^2}{\partial (\eta_i^2)^2} \log(g) \end{pmatrix} \quad (29)$$

As described in Section 3.1, after having derived this Hessian matrix, we calculate its value at the mode of the integrand and use it to re-sample the integrand.

5. Robustness Analysis Based on Simulations

This section aims to insure that the implemented method gives suitable results. We consider that the implemented method give us suitable results if for a given relationship between variables, by applying the estimation method on these variables we find approximatively the same coefficients. To reach this goal, we perform a robustness analysis on the estimation method. This robustness analysis is an empirical one based on simulations. We use two different approaches for that.

The first approach is to simulate bivariate binary variables by specifying a relationship between some explanatory variables (it means that we set coefficients of explanatory variables) and estimate this relationship with the implemented method in order to compare the results with the relationship specified before. In the second approach, we introduce new variables (that were not used in the data generating process) when estimating the relationship with the implemented method and compare the new results with the first ones. The implemented method is robust when it is able to correctly estimate the relationship specified even if we introduce other variables and also to estimate non significant coefficients to those other variables. Finally, the method we make use of to check for the robustness is the same that in [11].

As the estimation method implemented is a numerical approximation method, the results will depend on the selected number of quadrature points. We deal with the incidence of number of quadrature points on results in the last part of this section. For a better analysis of the results we also add the standard errors of each estimated coefficients.

5.1. Simulated Relationship between Real Variables

In this section, we use variables from the French SIP (Santé et Itinéraire Professionnel) survey data set and we simulate error terms and a relationship between some selected variables. The subset of the database use for this section is an unbalanced panel of 1202 individuals with total waves per individual between 5 and 10 waves.

We set the error terms parameters as $\sigma_1 = 2.1$, $\sigma_2 = 3.1$, $\rho_\eta = 0.7$, $\rho_\zeta = 0.5$ and $\rho_\epsilon = 0.4$.

We simulate idiosyncratic errors vectors $\zeta = (\zeta_1, \zeta_2)'$ and $\epsilon = (\epsilon_1, \epsilon_2)'$ as bivariate normal variables with zero mean, variance equal to 1 and covariances respectively equal to ρ_ζ and ρ_ϵ . We also simulate individual random effects as bivariate normal variables with zero mean, covariance equals to ρ_η and variance equals to σ_1^2 for the first component of the random effects vector and equals to σ_2^2 for the second component of the random effects vector. It has been done as follows:

$$\begin{aligned} \epsilon_1 &= rnormal(0,1) \\ \epsilon_2 &= rnormal(0,1) * \sqrt{1-\rho_\epsilon^2} + \rho_\epsilon \epsilon_1 \\ \zeta_1 &= rnormal(0,1) \\ \zeta_2 &= rnormal(0,1) * \sqrt{1-\rho_\zeta^2} + \rho_\zeta \zeta_1 \end{aligned}$$

where $rnormal(\mu, \sigma)$ denote the random normal density with mean μ and standard deviation σ . As individuals effects are time invariant, we simulate η as follows:

$$\begin{aligned} \eta_1 &= rnormal(0, \sigma_1) \quad \text{if } t = 1 \\ \eta_2 &= rnormal(0, \sigma_2) * \sqrt{1-\rho_\eta^2} + \rho_\eta \frac{\sigma_2}{\sigma_1} \eta_1 \quad \text{if } t = 1 \\ \eta_1 &= \eta_1[t=1] \quad \text{if } t \neq 1 \\ \eta_2 &= \eta_2[t=1] \quad \text{if } t \neq 1 \end{aligned}$$

For the initial conditions ($t = 1$), the simulated relationship is the following:

$$\begin{aligned} y_1^* &= -0.2 + 0.3ill - 0.2unemp + 0.4\eta_1 - 0.5\eta_2 + \epsilon_1 \\ y_2^* &= 2 - 0.2ill - 0.08age + 0.3\eta_1 + 0.5\eta_2 + \epsilon_2 \\ y_1 &= \mathbb{I}(y_1^* > 0) \\ y_2 &= \mathbb{I}(y_2^* > 0) \end{aligned}$$

For $t > 1$, we specify the following relationship:

$$\begin{aligned} y_{1t}^* &= 1.9 + 0.3y_{1,t-1} + 0.1y_{2,t-1} - 0.05Male_t - 0.2unemp_t + \eta_1 + \zeta_{1t} \\ y_{2t}^* &= -0.4 - 0.1y_{1,t-1} + 0.4y_{2,t-1} + 0.05Male_t - 0.5dens_t + \eta_2 + \zeta_{2t} \\ y_{1t} &= \mathbb{I}(y_{1t}^* > 0) \\ y_{2t} &= \mathbb{I}(y_{2t}^* > 0) \end{aligned}$$

The variable *ill* denotes having an illness episode in the year, *unemp* denotes being out of labour marking during the year, *age* denotes the age of individual, and *Male* is 1 if individual is male and 0 otherwise. Estimation results for 16 quadrature points are displayed in **Table 1**. For all equations, we give the coefficients that are used in the DGP and those that are estimated by our program. As we can see, all the coefficients from the DGP are very closed from the estimates ones.

5.2. Simulated Relationship with Additional Variables

In this section, we keep the same DGP than in Section 5.1 and we add other variables in the model that we estimate in order to evaluate the robustness of the estimation method by the fact that all estimated coefficients for variables in the DGP should remain the same and the added variables coefficients should not

Table 1. Simulated data set estimation's results.

	Equation (1)		Equation (2)	
	<i>DGP</i>	<i>Estimated coef.</i>	<i>DGP</i>	<i>Estimated coef.</i>
	(1)	(2)	(1')	(2')
<i>Dynamic Equation</i>				
$y^1 - 1$	0.3	0.2195 ^{***} (0.05)	-0.4	-0.0051 (0.0567)
$y^2 - 1$	0.1	0.1267 ^{**} (0.0513)	0.4	0.4926 ^{***} (0.061)
Gender = Male	-0.05	-0.0554 (0.0521)	0.05	0.073 (0.0594)
Medical density	-	-	0.5	0.5687 (1.111)
Unemployment rate	-0.2	-0.1682 ^{***} (0.369)	-	-
Intercept	1.9	2.3113 ^{***} (0.2667)	-0.4	-0.4677 (2.122)
<i>Initial Conditions</i>				
Illness before prof: life	0.3	0.3032 ^{***} (0.0283)	-0.2	-0.1624 ^{***} (0.0221)
Age	-	-	-0.08	-0.093 ^{***} (0.0202)
Unemployment rate	-0.2	-0.144 ^{**} (0.057)	-	-
Intercept	-0.2	-0.7331 (0.6194)	2	2.6757 ^{***} (0.4591)
λ_1	0.4	0.2581 ^{***} (0.0651)	0.3	0.2660 ^{***} (0.0463)
λ_2	-0.5	-0.5168 ^{***} (0.0753)	0.5	0.7022 ^{***} (0.0598)
<i>Covariance matrix structure</i>				
	<i>DGP</i>		<i>Estimated coef.</i>	
	(4)	(5)	(4)	(5)
σ_1	2.1	2.4399 ^{***} (0.1054)		
σ_2	3.1	2.7649 ^{***} (0.1365)		
ρ_η	0.7	0.7188 ^{***} (0.0212)		
ρ_ζ	0.5	0.5290 ^{***} (0.0419)		
ρ_ε	0.4	0.6972 ^{***} (0.1378)		

Estimated standard deviations for estimated coefficients are given within parenthesis. ***: significant at the 1% level, **: significant at the 5% level.

significant. We introduce two variables *rural* and *nationality (not French)* in the dynamic equations of the regression.

Results are in **Table 2**. Columns 1 and 2 in **Table 2** are the same than corresponding columns in **Table 1**. We provide in **Table 2**, column 3, the new results with the additional variables in order to compare with previous estimates⁴. As we can see in the **Table 2**, the coefficients estimated (using again 16 quadrature points) for those variables are not significant and all initial coefficients in the model remain approximately the same.

⁴We do the same with columns 1', 2' of **Table 1** and **Table 2** (new results are in column 3') and with columns 4 and 5 of both tables (new results in column 6).

Table 2. Simulated data set with added variables estimation's results.

	Equation (1)			Equation (2)		
	DGP	coef:	coef:	DGP	coef:	coef:
	(1)	(2)	(3)	(1')	(2')	(3')
Dynamic Equation						
$y^1 - 1$	0.3	0.2195 ^{***} (0.05)	0.2184 ^{***} (0.05)	-0.1	-0.0051 (0.0567)	-0.0052 (0.0568)
$y^2 - 1$	0.1	0.1267 ^{**} (0.0513)	0.1283 ^{**} (0.0513)	0.4	0.4926 ^{***} (0.061)	0.4944 ^{***} (0.0612)
Gender = Male	-0.05	-0.0554 (0.0521)	-0.0571 (0.0521)	0.05	0.073 (0.0594)	0.0751 (0.0596)
Medical density	-	-	-	0.5	0.5687 (1.1111)	0.5567 (1.1112)
Unemployment rate	-0.2	-0.1682 ^{***} (0.0269)	-0.1698 ^{***} (0.0269)	-	-	-
Not French rural	-	-	0.1246 (0.0956)	-	-	0.0015 (0.1076)
	-	-	0.0743 (0.0628)	-	-	0.0283 (0.0719)
Intercept	1.9	2.3113 ^{***} (0.2667)	2.2994 ^{***} (0.2667)	-0.04	-0.4677 (2.122)	-0.4527 (2.1215)
Initial Conditions						
Illness before prof: life	0.3	0.3032 ^{***} (0.0283)	0.3032 ^{***} (0.0283)	-0.2	-0.1624 ^{***} (0.0221)	-0.1627 ^{***} (0.0221)
Age	-	-	-	-0.08	-0.093 ^{***} (0.0202)	-0.0932 ^{***} (0.0202)
Unemployment rate	-0.2	-0.144 ^{**} (0.057)	-0.144 ^{**} (0.057)	-	-	-
Intercept	-0.2	-0.7331 (0.6194)	-0.7335 (0.6195)	2	2.6757 ^{***} (0.4591)	2.6803 ^{***} (0.4595)
λ_1	0.4	0.2581 ^{***} (0.0651)	0.2582 ^{***} (0.0653)	0.3	0.266 ^{***} (0.0463)	0.267 ^{***} (0.0464)
λ_2	-0.5	-0.5168 ^{***} (0.0753)	-0.5171 ^{***} (0.0754)	0.5	0.7022 ^{***} (0.0598)	0.703 ^{***} (0.0599)
Covariance matrix structure						
	DGP	Estimated coef:	Estimated coef:			
	(4)	(5)	(6)			
σ_1	2.1	2.4399 ^{***} (0.1034)	2.4353 ^{***} (0.1032)			
σ_2	3.1	2.7649 ^{***} (0.1365)	2.763 ^{***} (0.1366)			
ρ_η	0.7	0.7188 ^{***} (0.0212)	0.7187 ^{***} (0.0212)			
ρ_ζ	0.5	0.529 ^{***} (0.0419)	0.5301 ^{***} (0.0419)			
ρ_ε	0.4	0.6972 ^{***} (0.1379)	0.697 ^{***} (0.1378)			

Estimated standard deviations for estimated coefficients are given within parenthesis. ***: significant at the 1% level. **: significant at the 5% level.

5.3. Impact of Number of Quadrature Points on Estimated Results

As the accuracy of the method depends on the number of quadrature points used for the likelihood calculation, we propose an assessment of how it affects the results when this number increases. For doing so, we fit the same model with different numbers of quadrature points and we calculate the relative difference in log-likelihood and in estimated parameters.

We fit some models by using the same simulated relationship between variables as in Section 5.1.

The results are displayed in the **Table 3** for dynamic equations and in the **Table 4** for initial conditions equations and errors terms covariance matrix structure.

As we can see from **Table 3** and **Table 4**, by increasing the number of quadrature points the changes in results decline and the relative differences are around 0.01% for significant coefficients and 0.1% or at most 1% for non significant coefficients. After 16 quadrature points, the relative differences in log-likelihood and in estimated coefficients become fewer as we increase the number of quadrature points. The estimations with 22 quadrature points are closer to those with 24 quadrature points than the others. So when we increase the number of quadrature points the changes in estimated coefficients are not significant but the computing time grows up exponentially. For these models, estimation time on an i5 core computer at 2.5 GHz with 6 GB of RAM memory for the different number of quadrature points are given in **Table 5**.

6. Conclusions

This paper describes the bivariate dynamic probit model with endogenous initial conditions starting by justifying the econometric specification of the model, giving the estimation method and its requirements and ending by presenting a robustness analysis. We calculate the derivatives of the log-likelihood function (the gradient) with respect to the 13 parameters in the model. This is the main contribution of our research as many programs use numerical computation of the gradient vector instead of encoding the mathematically derived expression of

Table 3. Impact of the number of quadrature points on estimation results. Part A.

	DGP	Q = 10	Q = 16	Q = 22	Q = 24
Log likelihood		-8212.05	-8211.26	-830.71	-8301.27
y^1		Dynamic equation			
$y^1 - 1$	0.3	0.2754*** (0.0489)	0.2195*** (0.05)	0.2206*** (0.052)	0.2131*** (0.0527)
$y^2 - 1$	0.1	0.1376*** (0.0483)	0.1267** (0.0513)	0.1196** (0.0554)	0.1010* (0.0568)
Gender = Male	-0.05	-0.0580 (0.0479)	-0.0554 (0.0521)	-0.0732 (0.058)	-0.0599 (0.0604)
Unemployment rate	-0.2	-0.1509*** (0.0262)	-0.1682*** (0.0269)	-0.1792*** (0.0273)	-0.1810*** (0.0275)
Intercept	1.9	2.3270*** (0.2598)	2.3113*** (0.2667)	2.3089*** (0.2726)	2.30*** (0.2753)
y^2		Dynamic equation			
y^1_{-1}	-0.1	0.0224 (0.0541)	-0.0051 (0.0567)	-0.0136 (0.0594)	-0.0191 (0.0605)
y^2_{-1}	0.4	0.5851*** (0.0596)	0.4926*** (0.0610)	0.4846*** (0.0642)	0.4752*** (0.0650)
Gender = Male	0.05	0.0570 (0.0542)	0.0730 (0.0594)	0.0817 (0.0650)	0.0725 (0.0673)
Medical density	0.5	1.3305 (1.0685)	0.5687 (1.1111)	0.4874 (1.1357)	0.3549 (1.1473)
Intercept	-0.4	-1.7595 (2.040)	-0.4677 (2.1220)	-0.4064 (2.1704)	-0.1492 (2.1936)

Estimated standard deviations for estimated coefficients are given within parenthesis. ***: significant at the 1% level. **: significant at the 5% level.

Table 4. Impact of the number of quadrature points on estimation results. Part B.

	DGP	Q = 10	Q = 16	Q = 22	Q = 24
y^1 Initial conditions					
Illness before prof: life	0.3	0.3005 ^{***} (0.0278)	0.3032 ^{***} (0.0283)	0.3022 ^{***} (0.0282)	0.3026 ^{***} (0.0284)
Unemployment rate	-0.2	-0.1592 ^{***} (0.0573)	-0.1440 ^{**} (0.0570)	-0.1437 ^{**} (0.0571)	-0.1431 ^{**} (0.0572)
Intercept	-0.2	-0.6120 (0.6197)	-0.7331 (0.6194)	-0.7065 (0.6187)	-0.7153 (0.6188)
λ_{11}	0.4	0.2608 ^{***} (0.0644)	0.2581 ^{***} (0.0651)	0.2584 ^{***} (0.0658)	0.2628 ^{***} (0.0664)
λ_{12}	-0.5	-0.5076 ^{***} (0.0723)	-0.5168 ^{***} (0.0753)	-0.5051 ^{***} (0.0744)	-0.5019 ^{***} (0.0741)
y^2 Initial conditions					
Age	-0.08	-0.0859 ^{***} (0.0196)	-0.0930 ^{***} (0.0202)	-0.0929 ^{***} (0.0205)	-0.0943 ^{***} (0.0207)
Illness before prof: life	-0.2	-0.1593 ^{***} (0.0221)	-0.1624 ^{***} (0.0221)	-0.1648 ^{***} (0.0225)	-0.1650 ^{***} (0.0226)
Intercept	2	2.7329 ^{***} (0.4483)	2.6757 ^{***} (0.4591)	2.5788 ^{***} (0.4644)	2.5904 ^{***} (0.4676)
λ_{21}	0.3	0.2689 ^{***} (0.0467)	0.2660 ^{***} (0.0463)	0.2691 ^{***} (0.0474)	0.2679 ^{***} (0.0475)
λ_{22}	0.5	0.7136 ^{***} (0.0607)	0.7022 ^{***} (0.0598)	0.7008 ^{***} (0.0625)	0.6932 ^{***} (0.0626)
Covariance matrix structure					
σ_1	2.1	2.5202 ^{***} (0.1053)	2.4399 ^{***} (0.1054)	2.3920 ^{***} (0.1047)	2.3898 ^{***} (0.1051)
σ_2	3.1	2.7012 ^{***} (0.1307)	2.7649 ^{***} (0.1365)	2.7928 ^{***} (0.1444)	2.8281 ^{***} (0.1468)
ρ_η	0.7	0.7380 ^{***} (0.0206)	0.7188 ^{***} (0.0212)	0.7143 ^{***} (0.0219)	0.7162 ^{***} (0.0219)
ρ_ζ	0.5	0.5451 ^{***} (0.0411)	0.5290 ^{***} (0.0419)	0.5225 ^{***} (0.0423)	0.5145 ^{***} (0.0424)
ρ_ε	0.4	0.6550 ^{***} (0.1394)	0.6972 ^{***} (0.1378)	0.6996 ^{***} (0.1381)	0.6944 ^{***} (0.1371)

Estimated standard deviations for estimated coefficients are given within parenthesis. ***: significant at the 1% level. **: significant at the 5% level. *: significant at the 10% level.

Table 5. Computing time for different number of quadrature points.

Quad. points	10	16	22	24
Comp. time (in min.)	83	190	450	480

the gradient. Furthermore, for the use of the adaptative Gauss-Hermite quadrature, we also calculate the Hessian matrix with respect to individual random effects vector.

The implementation has been done using Stata software. We wrote 2 ado-files for this purpose. We use Stata's d1 method for the maximization process. For the use of this method, we implement the gradient vector for the 13 parameters and we also implement the Hessian matrix with respect the random effects vector in order to use the adaptative Gauss-Hermite quadrature. We also wrote two others ado-files for the estimation of the bivariate probit for panel data and the bivariate dynamic probit without initial conditions for panel data. These ado-files are written using the same method (Stata's d1 method) with the adaptative Gauss-Hermite quadrature. These ado-files are available upon request.

Due to the fact that the integration is bi-dimensional, estimation time is very high and stills increasing when the number of quadrature points or the number of observation or the number of explanatory variable increase. For an estimated model, one should insure that when increasing the number of quadrature point, the computed results don't change significantly before using them. It means that the relative difference in the results must be around 0.1% or fewer, and if so, we can conclude that the results remain stable when increasing the number of quadrature points. And, it means that there is no need to increase the number of quadrature points that will increase computing time but will not improve significantly the results. However, increasing the number of quadrature points also increases the computation time. One way for major improvement of the program is the use of multi-core (parallel) computing scheme. This scheme allows to make the computation of the contributions to the likelihood (Equation (23)) at each quadrature point separately and simultaneously on several cores. It has the advantage to save time since the contributions are computed in the same time.

Finally, our method gives reasonable computing durations with real dataset. In [12], we make use of the full SIP data set with 10,569 individuals and 255,206 observations.

Acknowledgements

We acknowledge the Centre Maurice Halbwachs (Réseau Quetelet) for access to the SIP 2007 data set (Santé et itinéraire professionnel-2007. DARES producteur. Centre Maurice Halbwachs diffuseur).

References

- [1] Adams, P., Hurd, M.D., McFadden, D., Merrill, A. and Ribeiro, T. (2003) Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status. *Journal of Econometrics*, **112**, 3-56.
- [2] Heckman, J.J. (1979) The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating: A Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence. Graduate School of Business and Department of Economics, University of Chicago, Chicago.
- [3] Alessie, R., Hochguertel, S. and Van Soest, A. (2004) Ownership of Stocks and Mutual Funds: A Panel Data Analysis. *The Review of Economics and Statistics*, **86**, 783-796.
- [4] Gourieroux, C. and Monfort, A. (1996) Simulation-Based Econometric Methods. CORE Lectures, Oxford University Press, Oxford.
- [5] Naylor, J.C. and Smith, A.F.M. (1982) Applications of a Method for the Efficient Computation of Posterior Distributions. *Applied Statistics*, **31**, 214-225.
- [6] Liu, Q. and Pierce, D. (1994) A Note on Gauss-Hermite Quadrature. *Biometrika*, **83**, 624-629. <https://doi.org/10.2307/2337136>
- [7] Jäckel, P. (2005) A Note on Multivariate Gauss-Hermite Quadrature. ABN AMRO, London.

- [8] Granger, C.W.J. (1969) Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, **37**, 428-438.
<https://doi.org/10.2307/1912791>
- [9] Nair-Reichert, A. and Weinhold, D. (2000) Causality Tests for Cross-Country Panels: New Look at FDI and Economic Growth in Developing Countries. *Oxford Bulletin of Economics and Statistics*, **63**, 153-171.
- [10] Gould, W., Pitblado, J. and Poi, B. (2010) Maximum Likelihood Estimation with Stata. 4th Edition, Stata Press, College Station.
- [11] Miranda, A. (2007) Migrant Networks, Migrant Selection and High School Graduation in Mexico, IZA dp No. 3204, Institute for the Study of Labor, Bonn.
- [12] Delattre, E., Moussa, R. and Sabatier, M. (2015) Health Condition and Job Status Interactions: Econometric Evidence of Causality from a French Longitudinal Survey. THEMA Working Paper, No. 19, 1-28.