Scientific Research Publishing

Theoretical Economics Letters, 2020, 10, 69-89
https://www.scirp.org/journal/tel
ISSN Online: 2162-2086
ISSN Print: 2162-2078

Simultaneous Equations Model with Non-Linear and Linear Dependent Variables on Panel Data

Amélie Adeline[1], Richard K. Moussa[2]

[1]ThEMA, University of Cergy-Pontoise, Cergy-Pontoise, France
[2]Ecole Nationale Supérieure de Statistique et d'Economie Appliquée, Abidjan, Côte d'Ivoire
Email: amelie.adeline@yahoo.fr, moussakr@gmail.com

How to cite this paper: Adeline, A. and Moussa, R.K. (2020) Simultaneous Equations Model with Non-Linear and Linear Dependent Variables on Panel Data. *Theoretical Economics Letters*, **10**, 69-89.
https://doi.org/10.4236/tel.2020.101005

Received: December 6, 2019
Accepted: January 19, 2020
Published: January 22, 2020

Abstract

This paper provides an estimation approach for the multi-equations' systems in panel data. Multi-equations systems are at the heart of economic modeling. Researchers who want to establish causal links between two outcomes, often need to consider simultaneity between the latter, to overcome endogeneity issues (for instance when considering supply and demand equations). Difficulties arise when considering linear and non-linear outcomes at the same time and this is why Roodman [1] implemented the Stata module cmp for multidimensional models. In this paper, we further develop this technique to allow researchers to implement a simultaneous equations model in a panel dimension setting. Implemented under Stata, our method, xtcmp, is a Full Information Maximum Likelihood (FIML) estimator. This paper explains the associated theory (derivation of the log-likelihood function, the associated gradient and the Hessian matrices of the log-integrand function) and offers an application of t xtcmp, while making comparisons with cmp.

Keywords

Causality, Full Information Maximum Likelihood, Panel Data, Gauss-Hermite Quadrature, Gradient, Hessian

1. Introduction

In empirical economics, a common approach is to consider a linear data-generating process. However, non-linear outcomes are often present and important in research questions. This is due to the structure of the database where interviewers transcribe yes-no answers into binary outcomes. However, when researchers point out a project, they often have to take into account different variables, continuous and categorical, at the same time, while considering simul-

taneous equations framework in a dynamic setting, in which each dependent variable is endogenous in one (or more) equation(s) of the model. The advantage of simultaneous equations model is to consider the correlation between the error terms of each equation. More specifically, in a dynamic setup, such models allow researchers to consider different individual effects (which are part of error terms, the latter being decomposed into an individual effect fixed across time and an effect which depends on time) across equations. This is of importance since these terms are unobserved, specific to each outcome, and might imply endogeneity issues. For instance in health economics, when investigating causal relationship between health and income, this can run from income to health and from health to income such that both are endogenous to each other [2]. In this way, considering a dynamic simultaneous equation model allows to consider unobserved individual effects such as physical maturity (thanks to genetics for health) or intellectual abilities (for income).

The framework of multi-equations models has been widely used in the literature to address several issues including the case of an endogenous binary outcome. Greene [3] has reformulated the estimation of the impact of an endogenous treatment on a continuous outcome using a multi-equations model. This reformulation has been extended to the analysis of endogenous binary outcome by [4]. Thus, several papers have analyzed the effects of an endogenous treatment on diverse types of outcomes including continuous and count outcomes [5] [6] [7]. Also, some generalizations for the case of noncompliance and nonresponse have been introduced by [8]. However, all these methods focus on cross sectional data and do not account for panel data.

There is almost no automated estimation method in Stata software to estimate parameters in this multi-equations model. An exception is the cmp command which is the first general Stata tool for this class of models, and is written as a Seemingly Unrelated Regressions (SUR) estimator [1]. However, this command does not explicitly consider panel dimension of the data which might be an issue due to the effervescence of databases with a temporal dimension. Moreover, simple relationships among variables at a point in time do not capture adequately the dynamic interaction of changing humans in changing environments. Thus, there is a need to develop a command for simultaneous equations model for panel database.

As a result, we offer an extension of the cmp framework, in a case where there are either two equations (one linear and one binary outcomes) or three equations (either two linear dependent variables with one binary, or one linear and two binary dependent variables), while explicitly considering the panel dimension of the data. In this way, our command xtcmp is a Full Information Maximum Likelihood (FIML) estimator, taking into account time dimension of the data, as well as, linear and non-linear outcomes (which is not feasible with three-stage least squares, because the latter only takes into account linear dependent variables).

As a result, the likelihood function is a multidimensional integral, such that we use the adaptive Gauss-Hermite quadrature method as an approximation (as

proposed by Liu and Pierce [9]). For the accuracy of the method and to reduce computing time, we derive the gradient of the log-likelihood, and the Hessian of the respective integrand. The estimation method of xtcmp has been implemented using the *d*1 method of Stata software (see [10] for further details).

Section 2 derives the likelihood for a FIML estimator in a general setting, as well as, for three specific cases. Section 3 discusses the estimation requirements needed, and details the Hessian matrices, as well as the gradient vector according to the parameters. In Section 4, we give some examples on the use of xtcmp, while making comparisons with Roodman's [1] cmp results. Section 5 concludes the paper.

2. Likelihood for a FIML Estimator

Let Y denotes a d-dimensional vector of endogenous variables in a simultaneous equations model. Roodman [1] specifically discusses conditions for consistency and identification of such model. Let Y^k denotes the k^{th} component of Y such that the value of Y^k for individual i at period t is given by y_{it}^k. We assume that the first d_1 components of Y, where $d_1 < d$, are binary outcomes, and the others, $d_2 = d - d_1$, are continuous. Let ϵ denotes the vector of associated error terms. By assuming a panel random effects model, ϵ can be decomposed into two terms such that $\epsilon = \mu + \nu$, where μ is time-invariant. In this way, the error term for individual i at period t in the k^{th} equation is given by $\epsilon_{it}^k = \mu_i^k + \nu_{it}^k$. The full model can be written as follows:

$$\tilde{Y} = X\beta + \epsilon \tag{1}$$

where $\tilde{Y}$ contains the related latent variables for the first d_1 equation and the original continuous variables for the others. The explanatory matrix X is given by $X = diag\left(Z^k\right)$ where Z^k, with $k = 1, \cdots, d$, corresponds to the explanatory variables for the k^{th} equation. Similarly, the parameter vector is $\beta = \left(\beta_1, \cdots, \beta_d\right)'$ where β^k is the parameter vector of the k^{th} equation. We suppose that the classical hypotheses on independence between 1) error components, and 2) the error components and explanatory variables, are satisfied. Furthermore, let us assume that the error components are independent and identically distributed with zero means and covariance matrices Σ_μ and Σ_ν, the latter being defined as follows:

$$\Sigma_\nu = \begin{pmatrix} \Sigma_1 & \Sigma_3' \\ \Sigma_3 & \Sigma_2 \end{pmatrix} \tag{2}$$

in which Σ_1 is a d_1 dimension matrix with 1 over the diagonal (which corresponds to the covariance matrix structure for simultaneous equations with only binary outcomes), Σ_2 is a d_2 dimension matrix, and Σ_3 is a $d_2 * d_1$ dimension matrix. Thus, the overall individual likelihood is given by:

$$L_i = \int_{\mathbb{R}^d} \left\{ \prod_{t=1}^{T_i} f_\nu \left(\nu_{it}^1, \cdots, \nu_{it}^d \mid \mu_i^1, \cdots, \mu_i^d \right) \right\} f_\mu \left(\mu_i^1, \cdots, \mu_i^d \right) \mathrm{d}\mu_i^1 \cdots \mathrm{d}\mu_i^d \tag{3}$$

where $f_\mu \left(\mu_i^1, \cdots, \mu_i^d \right) = \dfrac{1}{\left(2\pi\right)^{d/2} \sqrt{\left(\det\left(\Sigma_\mu\right)\right)}} \exp\left(\dfrac{-1}{2} \mu' \Sigma_\mu^{-1} \mu \right)$. The d_1 first equ-

ations being related to binary outcomes, if we define $q_{it}^k = 2 * y_{it}^k - 1$, then, the density function $f_\nu\left(\nu_{it}^1, \cdots, \nu_{it}^d \mid \mu_i^1, \cdots, \mu_i^d\right)$ is given by The density function is denoted by ℓ_{it} for simplification:

$$\ell_{it} = f_\nu\left(\nu_{it}^1, \cdots, \nu_{it}^d \mid \mu_i^1, \cdots, \mu_i^d\right) = \int_{-\infty}^{q_{it}^1 z_{it}^1} \cdots \int_{-\infty}^{q_{it}^{d_1} z_{it}^{d_1}} \phi_d\left(\nu_{it}^1, \cdots, \nu_{it}^d\right) \mathrm{d}\nu_i^1 \cdots \mathrm{d}\nu_i^{d_1} \tag{4}$$

where $z_{it}^k = Z_{it}^k \beta^k + \mu_i^k$, with $k = 1, \cdots, d$ and $\nu_{it}^k = \tilde{y}_{it}^k - z_{it}^k$, with $k = d_1 + 1, \cdots, d$. The idiosyncratic error is $\nu = \left(\nu_1, \cdots, \nu_{d_1}, \nu_{d_1+1}, \cdots, \nu_d\right) \sim \mathcal{N}\left(0, \Sigma_\nu\right)$, such that $\left(\nu_{d_1+1}, \cdots, \nu_d\right) \sim \mathcal{N}\left(0, \Sigma_2\right)$ and $\left(\nu_1, \cdots, \nu_{d_1}\right) \mid \left(\nu_{d_1+1}, \cdots, \nu_d\right) \sim \mathcal{N}\left(m_{(\nu_1, \cdots, \nu_{d_1}) \mid (\nu_{d_1+1}, \cdots, \nu_d)}, \Sigma_{(\nu_1, \cdots, \nu_{d_1}) \mid (\nu_{d_1+1}, \cdots, \nu_d)}\right)$ with:

$$m_{d_1 \mid d_2} = m_{(\nu_1, \cdots, \nu_{d_1}) \mid (\nu_{d_1+1}, \cdots, \nu_d)} = \Sigma_3' \left(\Sigma_2\right)^{-1} \left(\nu_{d_1+1}, \cdots, \nu_d\right)' \tag{5}$$

and

$$\Sigma_{d_1 \mid d_2} = \Sigma_{(\nu_1, \cdots, \nu_{d_1}) \mid (\nu_{d_1+1}, \cdots, \nu_d)} = \Sigma_1 - \Sigma_3' \left(\Sigma_2\right)^{-1} \Sigma_3 \tag{6}$$

Thus, we have:

$$\ell_{it} = \phi_{d_2}\left(\nu_{it}^{d_1+1}, \cdots, \nu_{it}^d\right) \Phi_{d_1}\left(\left(q_{it}^1 z_{it}^1, \cdots, q_{it}^{d_1} z_{it}^{d_1}\right), m_{d_1 \mid d_2}, \Sigma_{d_1 \mid d_2}\right) \tag{7}$$

in which $\Phi_{d_1}\left(\left(q_{it}^1 z_{it}^1, \cdots, q_{it}^{d_1} z_{it}^{d_1}\right), m_{d_1 \mid d_2}, \Sigma_{d_1 \mid d_2}\right)$ denotes the cumulative distribution function of a multivariate normal function with mean $m_{d_1 \mid d_2}$ and a covariance matrix $\Sigma_{d_1 \mid d_2}$.

We now focus on two cases. The first one, related to a simultaneous equations model with two outcomes, one binary and the other one continuous, is treated in Section 2.1. Then, we focus on a case with three outcomes, composed of either one binary and two continuous variables (that is developed in Section 2.2), or two binary and one continuous variables (analyzed in Section 2.3).

2.1. Case with Two Outcomes: One Binary and One Continuous

Let us consider the two following equations:

$$\tilde{y}_{it}^1 = Z_{it}^1 \beta^1 + \epsilon_{it}^1 \tag{8}$$

$$\tilde{y}_{it}^2 = Z_{it}^2 \beta^2 + \epsilon_{it}^2 \tag{9}$$

in which y_{it}^1 is a binary variable equal to 1 if $\tilde{y}_{it}^1 > 0$, and $\tilde{y}_{it}^2 = y_{it}^2$ is a linear outcome. The associated variance/covariance matrices of error components are:

$$\Sigma_\nu = \begin{pmatrix} 1 & \rho_1 \sigma \\ \rho_1 \sigma & \sigma^2 \end{pmatrix}$$

$$\Sigma_\mu = \begin{pmatrix} \sigma_1^2 & \rho_{1,2} \sigma_1 \sigma_2 \\ \rho_{1,2} \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

By identification, we have $\Sigma_1 = 1$, $\Sigma_2 = \sigma^2$, and $\Sigma_3 = \rho_1 \sigma$. Thus, $m_{\nu_1 \mid \nu_2} = \frac{\rho_1}{\sigma} \nu_2$ and $\Sigma_{\nu_1 \mid \nu_2} = 1 - \rho_1^2$. In this way, the likelihood has the following form:

$$L_i = \int_{\mathbb{R}^2} \left\{ \prod_{t=1}^{T_i} f_\nu \left(\nu_{it}^1, \nu_{it}^2 \mid \mu \right) \right\} f_\mu \left(\mu_i^1, \mu_i^2 \right) \mathrm{d}\mu_i^1 \mathrm{d}\mu_i^2$$

$$= \int_{\mathbb{R}^2} \left\{ \prod_t \ell_{it} \right\} f_\mu \left(\mu_i^1, \mu_i^2 \right) \mathrm{d}\mu_i^1 \mathrm{d}\mu_i^2$$

in which ℓ_{it} is the individual likelihood:

$$\ell_{it} = \phi_1 \left(\nu_{it}^2, 0, \sigma^2 \right) \Phi_1 \left(\frac{q_{it}^1 \left(z_{it}^1 + \frac{\rho_1}{\sigma} \nu_{it}^2 \right)}{\sqrt{1-\rho_1^2}} \right) \tag{10}$$

2.2. Case with Three Outcomes: One Binary and Two Continuous

Let us consider the two previous equations (Equation (8) and Equation (9)), and a new one:

$$\tilde{y}_{it}^3 = Z_{it}^3 \beta^3 + \epsilon_{it}^3 \tag{11}$$

also corresponding to a linear equation with $\tilde{y}_{it}^3 = y_{it}^3$. The associated variance/covariance matrices of error components are:

$$\Sigma_\nu = \begin{pmatrix} 1 & \rho_1 \sigma_a & \rho_2 \sigma_b \\ \rho_1 \sigma_a & \sigma_a^2 & \rho_3 \sigma_a \sigma_b \\ \rho_2 \sigma_b & \rho_3 \sigma_a \sigma_b & \sigma_b^2 \end{pmatrix}$$

$$\Sigma_\mu = \begin{pmatrix} \sigma_1^2 & \rho_{1,2} \sigma_1 \sigma_2 & \rho_{1,3} \sigma_1 \sigma_3 \\ \rho_{1,2} \sigma_1 \sigma_2 & \sigma_2^2 & \rho_{2,3} \sigma_2 \sigma_3 \\ \rho_{1,3} \sigma_1 \sigma_3 & \rho_{2,3} \sigma_2 \sigma_3 & \sigma_3^2 \end{pmatrix}$$

By identification, we have $\Sigma_1 = 1$, $\Sigma_2 = \begin{pmatrix} \sigma_a^2 & \rho_3 \sigma_a \sigma_b \\ \rho_3 \sigma_a \sigma_b & \sigma_b^2 \end{pmatrix}$, and $\Sigma_3 = \left(\rho_1 \sigma_a, \rho_2 \sigma_b \right)'$. Thus, $m_{\nu_1 \mid (\nu_2, \nu_3)} = \dfrac{\left(\rho_1 - \rho_2 \rho_3 \right) \dfrac{\nu_2}{\sigma_a} + \left(\rho_2 - \rho_1 \rho_3 \right) \dfrac{\nu_3}{\sigma_b}}{1 - \rho_3^2}$ and $\Sigma_{\nu_1 \mid (\nu_2, \nu_3)} = \dfrac{\rho_1^2 + \rho_2^2 - 2\rho_1 \rho_2 \rho_3}{1 - \rho_3^2}$. As a result, we can write the following likelihood:

$$L_i = \int_{\mathbb{R}^3} \left\{ \prod_{t=1}^{T_i} f_\nu \left(\nu_{it}^1, \nu_{it}^2, \nu_{it}^3 \mid \mu \right) \right\} f_\mu \left(\mu_i^1, \mu_i^2, , \mu_i^3 \right) \mathrm{d}\mu_i^1 \mathrm{d}\mu_i^2 \mathrm{d}\mu_i^3$$

$$= \int_{\mathbb{R}^3} \left\{ \prod_t \ell_{it} \right\} f_\mu \left(\mu_i^1, \mu_i^2, , \mu_i^3 \right) \mathrm{d}\mu_i^1 \mathrm{d}\mu_i^2 \mathrm{d}\mu_i^3$$

in which ℓ_{it} is the individual likelihood, defined as:

$$\ell_{it} = \phi_2 \left(\left(\nu_{it}^2, \nu_{it}^3 \right), 0, \Sigma_2 \right) \Phi_1 \left(\frac{q_{it}^1 \left(z_{it}^1 + m_{\nu_1 \mid (\nu_2, \nu_3)} \right)}{\sqrt{\Sigma_{\nu_1 \mid (\nu_2, \nu_3)}}} \right) \tag{12}$$

2.3. Case with Three Outcomes: Two Binary and One Continuous

In order to derive the likelihood function for a case composed of two binary and

one continuous outcomes, let us consider Equation (8) for the first binary outcome and Equation (11) for the linear variable. To consider another binary outcome, let us redefine the following:

$$\tilde{y}_{it}^2 = Z_{it}^2\beta^2 + \epsilon_{it}^2 \tag{13}$$

where y_{it}^2 is a binary variable, equal to 1 if $\tilde{y}_{it}^2 > 0$. The associated variance/covariance matrices of error components are:

$$\Sigma_\nu = \begin{pmatrix} 1 & \rho_1 & \rho_2\sigma \\ \rho_1 & 1 & \rho_3\sigma \\ \rho_2\sigma & \rho_3\sigma & \sigma^2 \end{pmatrix}$$

$$\Sigma_\mu = \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \rho_{1,3}\sigma_1\sigma_3 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{2,3}\sigma_2\sigma_3 \\ \rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$$

By identification, we have $\Sigma_1 = \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}$, $\Sigma_2 = \sigma^2$, and $\Sigma_3 = \left(\rho_2\sigma, \rho_3\sigma\right)'$. Thus, $m_{(\nu_1,\nu_2)|\nu_3} = \left(\frac{\rho_2}{\sigma}\nu_3, \frac{\rho_3}{\sigma}\nu_3\right)'$ and $\Sigma_{(\nu_1,\nu_2)|\nu_3} = \begin{pmatrix} 1-\rho_2^2 & \rho_1-\rho_2\rho_3 \\ \rho_1-\rho_2\rho_3 & 1-\rho_3^2 \end{pmatrix}$. The likelihood has the following form:

$$\begin{aligned} L_i &= \int_{\mathbb{R}^3} \left\{ \prod_{t=1}^{T_i} f_\nu\left(\nu_{it}^1, \nu_{it}^2, \nu_{it}^3 \mid \mu\right) \right\} f_\mu\left(\mu_i^1, \mu_i^2, , \mu_i^3\right) \mathrm{d}\mu_i^1 \mathrm{d}\mu_i^2 \mathrm{d}\mu_i^3 \\ &= \int_{\mathbb{R}^3} \left\{ \prod_t \ell_{it} \right\} f_\mu\left(\mu_i^1, \mu_i^2, , \mu_i^3\right) \mathrm{d}\mu_i^1 \mathrm{d}\mu_i^2 \mathrm{d}\mu_i^3 \end{aligned}$$

in which ℓ_{it} is the individual likelihood:

$$\ell_{it} = \phi_1\left(\nu_{it}^3, 0, \sigma^2\right) \Phi_2\left(\frac{q_{it}^1\left(z_{it}^1 + \frac{\rho_2}{\sigma}\nu_{it}^3\right)}{\sqrt{1-\rho_2^2}}, \frac{q_{it}^2\left(z_{it}^2 - \frac{\rho_3}{\sigma}\nu_{it}^3\right)}{\sqrt{1-\rho_3^2}}; q_{it}^1 q_{it}^2 \rho \right) \tag{14}$$

with $\rho = \dfrac{\rho_1 - \rho_2\rho_3}{\sqrt{\left(1-\rho_2^2\right)\left(1-\rho_3^2\right)}}$.

3. Estimation Requirement

The likelihood function being a *d*-dimensional integral function, we use the Gauss-Hermite quadrature method (see Moussa and Delattre [11]). Implementing this method requires to 1) compute the mode $\hat{\mu}$ of the log-integrand $\log(f) = \log\left(\left\{\prod_{t=1}^{T_i} \ell_{it}\right\} f_\mu\left(\mu_i^1, \cdots, \mu_i^d\right)\right)$ in $\mu = \left(\mu_i^1, \cdots, \mu_i^d\right)$ and to derive the Hessian matrix H at $\hat{\mu}$ with respect to μ; and, 2) to derive the gradient of the overall likelihood function with respect to the parameters.

Let Q denotes the selected number of quadrature points, x denotes the Q dimension vector of quadrature nodes, and w denotes the Q dimension vector of quadrature weight. By applying the adaptive Gauss-Hermite quadrature [9], the

likelihood function in Equation (3) can be rewritten as:

$$L_i = \sum_{k_1=1}^{Q} \cdots \sum_{k_d=1}^{Q} w_{k_1}^* \cdots w_{k_d}^* \left\{ \prod_{t=1}^{T_i} \ell_{it} \right\} f_\mu \left(\mu_i^1, \cdots, \mu_i^d \right) \Big|_{\mu = x^*} \tag{15}$$

in which $x^* = \hat{\mu} + \sqrt{2} H^{-1/2} x$ and
$w^* = \left(w_1^* \cdots w_d^* \right)' = 2^{d/2} \det\left(H^{-1/2} \right) \cdot \operatorname{diag}\left(w' \cdot \exp\left(x^2 \right) \right)$.

The derivation of the Hessian matrix is explained in Section 3.1 while the gradient of the overall likelihood function is derived in Section 3.2.

3.1. Hessian Matrix at $\hat{\mu}$

Based on the expressions of ℓ_{it} for each case described in Section 2 (Equation (10), Equation (12), and Equation (14)), we first need to write the associated log-integrand $\log(f)$ corresponding to each three cases. Then, we can focus on the calculation of the Hessian matrices, where we need to derive $\dfrac{\partial^2}{\partial \eta_i^k \partial \eta_i^j} \log(f)$, in which $k, j = 1, \cdots, d$.

Focusing on the first case with two equations, we have the following log-integrand:

$$\log(f) = \log\left(f_\mu \left(\mu_i^1, \mu_i^2 \right) \right) + \sum_{t=1}^{T_i} \log\left(\phi_1 \left(v_{it}^2, 0, \sigma^2 \right) \right) + \sum_{t=1}^{T_i} \log\left(\Phi_1 \left(\frac{q_{it}^1 \left(z_{it}^1 + \dfrac{\rho_1}{\sigma} v_{it}^2 \right)}{\sqrt{1-\rho_1^2}} \right) \right)$$

with the notation $b_{it} = \dfrac{q_{it}^1 \left(z_{it}^1 + \dfrac{\rho_1}{\sigma} v_{it}^2 \right)}{\sqrt{1-\rho_1^2}}$, we find:

$$\frac{\partial^2}{\partial \left(\eta_i^1 \right)^2} \log(f) = -\frac{1}{\sigma_1^2 \left(1-\rho_{1,2}^2 \right)} + \frac{1}{1-\rho_1^2} \sum_{t=1}^{T_i} \frac{b_{it} \phi_1 \left(b_{it} \right) \Phi_1 \left(b_{it} \right) - \left(\phi_1 \left(b_{it} \right) \right)^2}{\left(\Phi_1 \left(b_{it} \right) \right)^2}$$

$$\frac{\partial^2}{\partial \left(\eta_i^2 \right)^2} \log(f) = -\frac{1}{\sigma_2^2 \left(1-\rho_{1,2}^2 \right)} - \frac{T_i}{\sigma^2} + \frac{\rho_1^2}{\sigma^2 \left(1-\rho_1^2 \right)} \sum_{t=1}^{T_i} \frac{b_{it} \phi_1 \left(b_{it} \right) \Phi_1 \left(b_{it} \right) - \left(\phi_1 \left(b_{it} \right) \right)^2}{\left(\Phi_1 \left(b_{it} \right) \right)^2}$$

$$\frac{\partial^2}{\partial \eta_i^1 \partial \eta_i^2} \log(f) = \frac{\rho_{1,2}}{\sigma_1 \sigma_2 \left(1-\rho_{1,2}^2 \right)} + \frac{\rho_1}{\sigma \left(1-\rho_1^2 \right)} \sum_{t=1}^{T_i} q_{it}^1 \frac{-b_{it} \phi_1 \left(b_{it} \right) \Phi_1 \left(b_{it} \right) + \left(\phi_1 \left(b_{it} \right) \right)^2}{\left(\Phi_1 \left(b_{it} \right) \right)^2}$$

Thus, the Hessian matrix is given by:

$$H = \begin{pmatrix} -\dfrac{\partial^2}{\partial \left(\eta_i^1 \right)^2} \log(f) & -\dfrac{\partial^2}{\partial \eta_i^1 \partial \eta_i^2} \log(f) \\ -\dfrac{\partial^2}{\partial \eta_i^1 \delta \eta_i^2} \log(f) & -\dfrac{\partial^2}{\partial \left(\eta_i^2 \right)^2} \log(f) \end{pmatrix} \tag{16}$$

Then, focusing on the case of three equations with one binary outcome, the log-integrand is given by:

$$\log(f)=\log\left(f_{\mu}\left(\mu_i^1,\mu_i^2,\mu_i^3\right)\right)+\sum_{t=1}^{T_i}\log\left(\phi_2\left(\left(\nu_{it}^2,\nu_{it}^3\right),0,\Sigma_2\right)\right)$$
$$+\sum_{t=1}^{T_i}\log\left(\Phi_1\left(\frac{q_{it}^1\left(z_{it}^1+m_{\nu_1|(\nu_2,\nu_3)}\right)}{\sqrt{\Sigma_{\nu_1|(\nu_2,\nu_3)}}}\right)\right)$$

The associated derivatives, assuming $b_{it}=\frac{q_{it}^1\left(z_{it}^1+m_{\nu_1|(\nu_2,\nu_3)}\right)}{\sqrt{\Sigma_{\nu_1|(\nu_2,\nu_3)}}}$, $a_1=\sqrt{1-\rho_{1,3}^2}$, $a_2=\sqrt{1-\rho_{2,3}^2}$, and $ra=\frac{\rho_{1,2}-\rho_{1,3}\rho_{2,3}}{a_1a_2}$, are given by:

$$\frac{\partial^2}{\partial\left(\eta_i^1\right)^2}\log(f)=-\frac{1}{\left(\sigma_1a_1\sqrt{1-ra^2}\right)^2}$$
$$-\frac{1}{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\sum_{t=1}^{T_i}\frac{b_{it}\phi_1\left(b_{it}\right)\Phi_1\left(b_{it}\right)+\left(\phi_1\left(b_{it}\right)\right)^2}{\left(\Phi_1\left(b_{it}\right)\right)^2}$$

$$\frac{\partial^2}{\partial\left(\eta_i^2\right)^2}\log(f)=-\frac{1}{\left(\sigma_2a_2\sqrt{1-ra^2}\right)^2}-\frac{T_i}{\left(1-\rho_3^2\right)\sigma_a^2}$$
$$-\frac{\left(\rho_1-\rho_2\rho_3\right)^2}{\left(1-\rho^2\right)\left(1-\rho_2^2\right)\left(\sigma_a\left(1-\rho_3^2\right)\right)^2}\sum_{t=1}^{T_i}\frac{b_{it}\phi_1\left(b_{it}\right)\Phi_1\left(b_{it}\right)+\left(\phi_1\left(b_{it}\right)\right)^2}{\left(\Phi_1\left(b_{it}\right)\right)^2}$$

$$\frac{\partial^2}{\partial\left(\eta_i^3\right)^2}\log(f)=-\frac{1-\rho_{1,2}^2}{\left(\sigma_3a_1a_2\sqrt{1-ra^2}\right)^2}-\frac{T_i}{\left(1-\rho_3^2\right)\sigma_b^2}$$
$$-\frac{\left(\rho_2-\rho_1\rho_3\right)^2}{\left(1-\rho^2\right)\left(1-\rho_2^2\right)\left(\sigma_b\left(1-\rho_3^2\right)\right)^2}\sum_{t=1}^{T_i}\frac{b_{it}\phi_1\left(b_{it}\right)\Phi_1\left(b_{it}\right)+\left(\phi_1\left(b_{it}\right)\right)^2}{\left(\Phi_1\left(b_{it}\right)\right)^2}$$

$$\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^2}\log(f)=\frac{\rho_{1,2}-\rho_{1,3}\rho_{2,3}}{\sigma_1\sigma_2\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$
$$-\frac{\rho_1-\rho_2\rho_3}{\sigma_a\left(1-\rho^2\right)\left(1-\rho_2^2\right)\left(1-\rho_3^2\right)}\sum_{t=1}^{T_i}q_{it}^1\frac{b_{it}\phi_1\left(b_{it}\right)\Phi_1\left(b_{it}\right)+\left(\phi_1\left(b_{it}\right)\right)^2}{\left(\Phi_1\left(b_{it}\right)\right)^2}$$

$$\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^3}\log(f)=\frac{\rho_{1,3}-\rho_{1,2}\rho_{2,3}}{\sigma_1\sigma_3\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$
$$-\frac{\rho_2-\rho_1\rho_3}{\sigma_b\left(1-\rho^2\right)\left(1-\rho_2^2\right)\left(1-\rho_3^2\right)}\sum_{t=1}^{T_i}q_{it}^1\frac{b_{it}\phi_1\left(b_{it}\right)\Phi_1\left(b_{it}\right)+\left(\phi_1\left(b_{it}\right)\right)^2}{\left(\Phi_1\left(b_{it}\right)\right)^2}$$

$$\frac{\partial^2}{\partial\eta_i^2\partial\eta_i^3}\log(f)=\frac{\rho_{2,3}-\rho_{1,2}\rho_{1,3}}{\sigma_2\sigma_3\left(a_1a_2\sqrt{1-ra^2}\right)^2}+\frac{\rho_3T_i}{\left(1-\rho_3^2\right)\sigma_a\sigma_b}$$
$$-\frac{\left(\rho_2-\rho_3\rho_3\right)\left(\rho_1-\rho_2\rho_3\right)}{\sigma_a\sigma_b\left(1-\rho^2\right)\left(1-\rho_2^2\right)\left(1-\rho_3^2\right)^2}\sum_{t=1}^{T_i}\frac{b_{it}\phi_1\left(b_{it}\right)\Phi_1\left(b_{it}\right)+\left(\phi_1\left(b_{it}\right)\right)^2}{\left(\Phi_1\left(b_{it}\right)\right)^2}$$

Thus, the Hessian matrix is given by:

$$H=\begin{pmatrix} -\frac{\partial^2}{\partial\left(\eta_i^1\right)^2}\log(f) & -\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^2}\log(f) & -\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^3}\log(f) \\ -\frac{\partial^2}{\partial\eta_i^1\delta\eta_i^2}\log(f) & -\frac{\partial^2}{\partial\left(\eta_i^2\right)^2}\log(f) & -\frac{\partial^2}{\partial\eta_i^2\partial\eta_i^3}\log(f) \\ -\frac{\partial^2}{\partial\eta_i^1\delta\eta_i^3}\log(f) & -\frac{\partial^2}{\partial\eta_i^2\partial\eta_i^3}\log(f) & -\frac{\partial^2}{\partial\left(\eta_i^3\right)^2}\log(f) \end{pmatrix} \tag{17}$$

Finally, for the three equations with two binary outcomes case, the log-integrand can be written as:

$$\log(f)=\log\left(f_\mu\left(\mu_i^1,\mu_i^2,\mu_i^3\right)\right)+\sum_{t=1}^{T_i}\log\left(\phi_1\left(v_{it}^3,0,\sigma^2\right)\right)$$
$$+\sum_{t=1}^{T_i}\log\left(\Phi_2\left(\frac{q_{it}^1\left(z_{it}^1+\frac{\rho_2}{\sigma}v_{it}^3\right)}{\sqrt{1-\rho_2^2}},\frac{q_{it}^2\left(z_{it}^2+\frac{\rho_3}{\sigma}v_{it}^3\right)}{\sqrt{1-\rho_3^2}};q_{it}^1q_{it}^2\rho\right)\right)$$

Then, considering a_1, a_2 and ra previously defined and the following notations:

$$b_{it}^1=\frac{q_{it}^1\left(z_{it}^1+\frac{\rho_2}{\sigma}v_{it}^3\right)}{\sqrt{1-\rho_2^2}}$$

$$b_{it}^2=\frac{q_{it}^2\left(z_{it}^2+\frac{\rho_3}{\sigma}v_{it}^3\right)}{\sqrt{1-\rho_3^2}}$$

$$r_n=\sqrt{1-\rho_2^2}$$

$$r_m=\sqrt{1-\rho_3^2}$$

$$\rho=\frac{\rho_1-\rho_2\rho_3}{r_nr_m}$$

$$p_{it}^1=q_{it}^1\phi\left(b_{it}^1\right)\Phi\left(\frac{b_{it}^2-q_{it}^1q_{it}^2\rho b_{it}^1}{\sqrt{1-\rho^2}}\right)$$

$$p_{it}^2=q_{it}^2\phi\left(b_{it}^2\right)\Phi\left(\frac{b_{it}^1-q_{it}^1q_{it}^2\rho b_{it}^2}{\sqrt{1-\rho^2}}\right)$$

$$p_{it}^3=\frac{\rho_3}{\sigma\sqrt{1-\rho^2}}\phi\left(b_{it}^2\right)\Phi\left(\frac{b_{it}^1-q_{it}^1q_{it}^2\rho b_{it}^2}{\sqrt{1-\rho^2}}\right)+\frac{\rho_2}{\sigma}\phi\left(b_{it}^1\right)\Phi\left(\frac{b_{it}^2-q_{it}^1q_{it}^2\rho b_{it}^1}{\sqrt{1-\rho^2}}\right)$$

We find the following:

$$\frac{\partial^2}{\partial\left(\eta_i^1\right)^2}\log(f)=-\frac{a_2^2}{\left(\sigma_1a_1a_2\sqrt{1-r_a^2}\right)^2}+\sum_{t=1}^{T_i}\frac{\frac{q_{it}^1}{r_n}\left(-b_{it}^1p_{it}^1-q_{it}^2\rho\phi\left(b_{it}^1\right)\phi\left(\frac{b_{it}^2-q_{it}^1q_{it}^2\rho b_{it}^1}{\sqrt{1-\rho^2}}\right)\right)\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)-\left(p_{it}^1\right)^2}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

$$\frac{\partial^2}{\partial\left(\eta_i^2\right)^2}\log(f)=-\frac{a_1^2}{\left(\sigma_2 a_1 a_2\sqrt{1-r_a^2}\right)^2}+\sum_{t=1}^{T_i}\frac{\frac{q_{it}^2}{r_m}\left(-b_{it}^2 p_{it}^2-q_{it}^1\rho\phi\left(b_{it}^2\right)\phi\left(\frac{b_{it}^1-q_{it}^1q_{it}^2\rho b_{it}^2}{\sqrt{1-\rho^2}}\right)\right)\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)-\left(p_{it}^2\right)^2}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

$$\frac{\partial^2}{\partial\left(\eta_i^3\right)^2}\log(f)=-\frac{1-\rho_1^2}{\left(\sigma_3 a_1 a_2\sqrt{1-r_a^2}\right)^2}-\frac{T_i}{\sigma^2}-\sum_{t=1}^{T_i}\frac{\left(p_{it}^3\right)^2}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

$$+\sum_{t=1}^{T_i}\frac{\frac{\rho_3}{\sigma a_2\sqrt{1-\rho^2}}\left(\left(\frac{\rho_2}{\sigma a_1}-q_{it}^1q_{it}^2\rho\frac{\rho_3}{\sigma a_2}\right)\phi\left(b_{it}^2\right)\phi\left(\frac{b_{it}^2-q_{it}^1q_{it}^2\rho b_{it}^1}{\sqrt{1-\rho^2}}\right)-\frac{\rho_3}{\sigma}b_{it}^2p_{it}^2\right)\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

$$+\sum_{t=1}^{T_i}\frac{\frac{\rho_2}{\sigma a_1\sqrt{1-\rho^2}}\left(\left(\frac{\rho_3}{\sigma a_2}-q_{it}^1q_{it}^2\rho\frac{\rho_2}{\sigma a_1}\right)\phi\left(b_{it}^1\right)\phi\left(\frac{b_{it}^1-q_{it}^1q_{it}^2\rho b_{it}^2}{\sqrt{1-\rho^2}}\right)-\frac{\rho_2}{\sigma}b_{it}^1p_{it}^1\right)\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

$$\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^2}\log(f)=\frac{\rho_1-\rho_2\rho_3}{\sigma_1\sigma_2\left(a_1a_2\sqrt{1-r_a^2}\right)^2}$$

$$+\sum_{t=1}^{T_i}\frac{\frac{q_{it}^1q_{it}^2}{r_nr_m\sqrt{1-\rho^2}}\left(\phi\left(b_{it}^1\right)\phi\left(\frac{b_{it}^2-q_{it}^1q_{it}^2\rho b_{it}^1}{\sqrt{1-\rho^2}}\right)\right)\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)-p_{it}^1p_{it}^2}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

$$\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^3}\log(f)=\frac{\rho_2-\rho_1\rho_3}{\sigma_1\sigma_3\left(a_1a_2\sqrt{1-r_a^2}\right)^2}$$

$$+\sum_{t=1}^{T_i}\frac{\frac{q_{it}^1}{r_n}\left(\frac{\frac{\rho_3}{\sigma r_m}-q_{it}^1q_{it}^2\rho\frac{\rho_2}{\sigma r_n}}{\sqrt{1-\rho^2}}\phi\left(b_{it}^1\right)\phi\left(\frac{b_{it}^2-q_{it}^1q_{it}^2\rho b_{it}^1}{\sqrt{1-\rho^2}}\right)-\frac{\rho_2}{\sigma}b_{it}^1p_{it}^1\right)\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)-p_{it}^1p_{it}^3}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

$$\frac{\partial^2}{\partial\eta_i^2\partial\eta_i^3}\log(f)=\frac{\rho_3-\rho_1\rho_2}{\sigma_2\sigma_3\left(a_1a_2\sqrt{1-r_a^2}\right)^2}$$

$$+\sum_{t=1}^{T_i}\frac{\frac{q_{it}^2}{r_m}\left(\frac{\frac{\rho_2}{\sigma r_n}-q_{it}^1q_{it}^2\rho\frac{\rho_3}{\sigma r_m}}{\sqrt{1-\rho^2}}\phi\left(b_{it}^2\right)\phi\left(\frac{b_{it}^1-q_{it}^1q_{it}^2\rho b_{it}^2}{\sqrt{1-\rho^2}}\right)-\frac{\rho_3}{\sigma}b_{it}^2p_{it}^2\right)\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)-p_{it}^2p_{it}^3}{\left(\Phi_2\left(b_{it}^1,b_{it}^2;q_{it}^1q_{it}^2\rho\right)\right)^2}$$

Thus, the Hessian matrix has the same form as the previous one (matrix 17). In other words, it is given by:

$$H = \begin{pmatrix} -\frac{\partial^2}{\partial\left(\eta_i^1\right)^2}\log(f) & -\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^2}\log(f) & -\frac{\partial^2}{\partial\eta_i^1\partial\eta_i^3}\log(f) \\ -\frac{\partial^2}{\partial\eta_i^1\delta\eta_i^2}\log(f) & -\frac{\partial^2}{\partial\left(\eta_i^2\right)^2}\log(f) & -\frac{\partial^2}{\partial\eta_i^2\partial\eta_i^3}\log(f) \\ -\frac{\partial^2}{\partial\eta_i^1\delta\eta_i^3}\log(f) & -\frac{\partial^2}{\partial\eta_i^2\partial\eta_i^3}\log(f) & -\frac{\partial^2}{\partial\left(\eta_i^3\right)^2}\log(f) \end{pmatrix} \quad (18)$$

3.2. Gradient Vector with Respect to the Parameters

Based on the likelihood function given by Equation (15), parameters to estimate are β^k, with $k = 1,\cdots,d$, and the associated covariance matrices Σ_μ and Σ_ν. Thus, the gradient has to be calculated with respect to these parameters. The first order derivative of the log-likelihood function with respect to a parameter α, in the set of parameters, is given by:

$$\frac{\partial\log(L_i)}{\partial\alpha} = \sum_{k_1=1}^{Q}\cdots\sum_{k_d=1}^{Q}\frac{\partial f/\partial\alpha}{L_i} \quad (19)$$

Focusing on the three cases, we apply this formula to compute derivatives with respect to each parameter.

First, considering the two outcomes case, we need to consider the six following parameters: β^1, β^2, σ, ρ_1, σ_1, σ_2, ρ_{12}. As in subsection 3.1, we consider the previously defined b_{it}, which is specific to the case with two outcomes, such that we have:

$$\frac{\partial f}{\partial\beta^1} = f*\sum_{t=1}^{T_i}\frac{q_{it}^1\phi_1(b_{it})}{\sqrt{1-\rho_1^2}\Phi_1(b_{it})}$$

$$\frac{\partial f}{\partial\beta^2} = f*\sum_{t=1}^{T_i}\left(\frac{v_{it}^2}{\sigma^2}-\frac{\rho_1 q_{it}^1\phi_1(b_{it})}{\sigma\sqrt{1-\rho_1^2}\Phi_1(b_{it})}\right)$$

$$\frac{\partial f}{\partial\log(\sigma)} = f*\sum_{t=1}^{T_i}\left(-1+\left(\frac{v_{it}^2}{\sigma}\right)^2-\frac{\rho_1 q_{it}v_{it}^2\phi_1(b_{it})}{\sigma\sqrt{1-\rho_1^2}\Phi_1(b_{it})}\right)$$

$$\frac{\partial f}{\partial\log\left(\frac{1+\rho_1}{1-\rho_1}\right)^{1/2}} = f*\sum_{t=1}^{T_i}\left(\rho_1 q_{it}z_{it}^1+\frac{q_{it}v_{it}^2}{\sigma}\right)$$

$$\frac{\partial f}{\partial\log(\sigma_1)} = f*\left(-1+\frac{\left(\frac{\mu_i^1}{\sigma_1}\right)^2-\rho_{1,2}\frac{\mu_i^1\mu_i^2}{\sigma_1\sigma_2}}{1-\rho_{1,2}^2}\right)$$

$$\frac{\partial f}{\partial\log(\sigma_2)} = f*\left(-1+\frac{\left(\frac{\mu_i^2}{\sigma_2}\right)^2-\rho_{1,2}\frac{\mu_i^1\mu_i^2}{\sigma_1\sigma_2}}{1-\rho_{1,2}^2}\right)$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_{12}}{1-\rho_{12}}\right)^{1/2}} = f\sum_{t=1}^{T_i}\frac{\left(q_{it}^1\rho_1 z_{it}^1 - \frac{v_{it}^2}{\sigma}\right)\phi_1\left(b_{it}\right)}{\sqrt{1-\rho_1^2}\Phi_1\left(b_{it}\right)}$$

$$= f*\left(\rho_{1,2} - \frac{\rho_{1,2}\left(\frac{\mu_i^1}{\sigma_1}\right)^2 + \rho_{1,2}\left(\frac{\mu_i^2}{\sigma_2}\right)^2 - \left(1+\rho_{1,2}^2\right)\frac{\mu_i^1\mu_i^2}{\sigma_1\sigma_2}}{1-\rho_{1,2}^2}\right)$$

Now, focusing on the case of three equations with one binary outcome, we consider notations associated to this case in subsection 3.1 (b_{it}, a_2, and ra). We compute derivatives with respect to β^1, β^2, β^3, σ_a, σ_b, ρ_1, ρ_2, ρ_3, σ_1, σ_2, σ_3, $\rho_{1,2}$, $\rho_{1,3}$ and $\rho_{2,3}$, such that:

$$\frac{\partial f}{\partial \beta^1} = f*\sum_{t=1}^{T_i}\frac{q_{it}^1\phi_1\left(b_{it}\right)}{\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}$$

$$\frac{\partial f}{\partial \beta^2} = f*\sum_{t=1}^{T_i}\left(\frac{\frac{v_{it}^2}{\sigma_a^2} - \rho_3\frac{v_{it}^3}{\sigma_a\sigma_b}}{1-\rho_3^2} - \frac{q_{it}^1\left(\rho_1-\rho_2\rho_3\right)\phi_1\left(b_{it}\right)}{\sigma_a\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}\right)$$

$$\frac{\partial f}{\partial \beta^3} = f*\sum_{t=1}^{T_i}\left(\frac{\frac{v_{it}^3}{\sigma_b^2} - \rho_3\frac{v_{it}^2}{\sigma_a\sigma_b}}{1-\rho_3^2} - \frac{q_{it}^1\left(\rho_2-\rho_1\rho_3\right)\phi_1\left(b_{it}\right)}{\sigma_b\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}\right)$$

$$\frac{\partial f}{\partial \log\left(\sigma_a\right)} = f*\sum_{t=1}^{T_i}\left(-1+\frac{\left(\frac{v_{it}^2}{\sigma_a}\right)^2 - \rho_3\frac{v_{it}^2v_{it}^3}{\sigma_a\sigma_b}}{1-\rho_3^2} - \frac{q_{it}^1\left(\rho_1-\rho_2\rho_3\right)v_{it}^2\phi_1\left(b_{it}\right)}{\sigma_a\left(1-\rho_3^2\right)\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}\right)$$

$$\frac{\partial f}{\partial \log\left(\sigma_b\right)} = f*\sum_{t=1}^{T_i}\left(-1+\frac{\left(\frac{v_{it}^3}{\sigma_b}\right)^2 - \rho_3\frac{v_{it}^2v_{it}^3}{\sigma_a\sigma_b}}{1-\rho_3^2} - \frac{q_{it}^1\left(\rho_2-\rho_1\rho_3\right)v_{it}^3\phi_1\left(b_{it}\right)}{\sigma_b\left(1-\rho_3^2\right)\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}\right)$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_1}{1-\rho_1}\right)^{1/2}} = f*\sum_{t=1}^{T_i}\left(1-\rho_1^2\right)\frac{\left(-\frac{q_{it}^1\left(\rho_3\frac{v_{it}^3}{\sigma_b} - \frac{v_{it}^2}{\sigma_a}\right)}{1-\rho_3^2} + b_{it}\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\frac{\rho^2}{\left(1-\rho^2\right)\left(\rho_1-\rho_2\rho_3\right)}\right)\phi_1\left(b_{it}\right)}{\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_2}{1-\rho_2}\right)^{1/2}} = f*\sum_{t=1}^{T_i}\left(\frac{\left(\frac{1-\rho_2^2}{1-\rho_3^2}q_{it}^1\left(-\rho_3\frac{v_{it}^2}{\sigma_a}+\frac{v_{it}^3}{\sigma_b}\right)+b_{it}\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\left(\rho_2+\frac{\left(\rho_1\rho_2-\rho_3\right)\rho^2}{\left(1-\rho^2\right)\left(\rho_1-\rho_2\rho_3\right)}\right)\right)\phi_1\left(b_{it}\right)}{\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}\right)$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_3}{1-\rho_3}\right)^{1/2}} = f*\sum_{t=1}^{T_i}\left(\rho_3\left(1-\frac{\left(\frac{v_{it}^2}{\sigma_a}\right)^2+\left(\frac{v_{it}^3}{\sigma_b}\right)^2}{1-\rho_3^2}\right)+\left(1+\rho_3^2\right)\frac{\frac{v_{it}^2v_{it}^3}{\sigma_a\sigma_b}}{1-\rho_3^2}\right)$$

$$+f*\sum_{t=1}^{T_i}\frac{q_{it}^1\left(-\frac{\rho_2v_{it}^2}{\sigma_a}-\frac{\rho_1v_{it}^3}{\sigma_b}+2\rho_3\frac{\left(\rho_1-\rho_2\rho_3\right)\frac{v_{it}^2}{\sigma_a}+\left(\rho_2-\rho_1\rho_3\right)\frac{v_{it}^3}{\sigma_b}}{1-\rho_3^2}+\frac{b_{it}\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\left(\rho_1\rho_3-\rho_2\right)\rho^2}{\left(1-\rho^2\right)\left(\rho_1-\rho_2\rho_3\right)}\right)\phi_1\left(b_{it}\right)}{\sqrt{\left(1-\rho^2\right)\left(1-\rho_2^2\right)}\Phi_1\left(b_{it}\right)}$$

$$\frac{\partial f}{\partial \log\left(\sigma_1\right)} = f*\left(-1+\frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_1^i}{\sigma_1}\right)^2-\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}-\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$\frac{\partial f}{\partial \log\left(\sigma_2\right)} = f*\left(-1+\frac{\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_2^i}{\sigma_2}\right)^2-\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}-\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$\frac{\partial f}{\partial \log\left(\sigma_3\right)} = f*\left(-1+\frac{\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_3^i}{\sigma_3}\right)^2-\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}-\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_{1,2}}{1-\rho_{1,2}}\right)^{1/2}} = f*\left(\frac{1-\rho_{1,2}^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\left(\rho_{1,2}\left(\frac{\mu_3^i}{\sigma_3}\right)^2+\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}-\rho_{2,3}\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}-\rho_{1,3}\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}\right)\right)$$

$$+f\frac{\left(1-\rho_{1,2}^2\right)\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\left(1-\frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_1^i}{\sigma_1}\right)^2+\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_2^i}{\sigma_2}\right)^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$+f\frac{\left(1-\rho_{1,2}^2\right)\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\left(\frac{-\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_3^i}{\sigma_3}\right)^2+2\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$+f\frac{\left(1-\rho_{1,2}^2\right)\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\left(2\frac{\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}+\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$\frac{\partial f}{\partial\log\left(\frac{1+\rho_{1,3}}{1-\rho_{1,3}}\right)^{1/2}}=f*\left(\frac{\left(1-\rho_{1,3}^2\right)\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)+\left(\rho_{1,3}\left(\frac{\mu_2^i}{\sigma_2}\right)^2-\rho_{2,3}\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}-\rho_{1,2}\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}\right)a_1^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$-f\left(\rho_{1,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{1,3}-\rho_{2,3}\right)\right)$$

$$*\frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_1^i}{\sigma_1}\right)^2+\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_2^i}{\sigma_2}\right)^2+\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_3^i}{\sigma_3}\right)^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

$$+2f\left(\rho_{1,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{1,3}-\rho_{2,3}\right)\right)$$

$$*\frac{\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}+\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

$$\frac{\partial f}{\partial\log\left(\frac{1+\rho_{2,3}}{1-\rho_{2,3}}\right)^{1/2}}=f*\left(\frac{\left(1-\rho_{2,3}^2\right)\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)+\left(\rho_{2,3}\left(\frac{\mu_1^i}{\sigma_1}\right)^2-\rho_{1,3}\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}-\rho_{1,2}\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}\right)a_2^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$-f\left(\rho_{2,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{2,3}-\rho_{1,3}\right)\right)$$

$$*\frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_1^i}{\sigma_1}\right)^2+\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_2^i}{\sigma_2}\right)^2+\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_3^i}{\sigma_3}\right)^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

$$+2f\left(\rho_{2,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{2,3}-\rho_{1,3}\right)\right)$$

$$*\frac{\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}+\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

Finally, we compute derivatives with respect to β^1, β^2, β^3, σ, ρ_1, ρ_2, ρ_3, σ_1, σ_2, σ_3, $\rho_{1,2}$, $\rho_{1,3}$ and $\rho_{2,3}$ for the three equations with two binary outcomes case. To do so, we consider notations defined for this case in subsection 3.1 concerning b_{it}^1, b_{it}^2, a_1, a_2, ra, r_n, r_m, ρ, p_{it}^1, and p_{it}^2, such that:

$$\frac{\partial f}{\partial \beta^1} = f * \sum_{t=1}^{T_i} \frac{p_{it}^1}{\Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)}$$

$$\frac{\partial f}{\partial \beta^2} = f * \sum_{t=1}^{T_i} \frac{p_{it}^2}{\Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)}$$

$$\frac{\partial f}{\partial \beta^3} = f * \sum_{t=1}^{T_i} \left(\frac{v_{it}^3}{\sigma^2} - \frac{\rho_3 p_{it}^2}{\sigma q_{it}^2 \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} - \frac{\rho_2 p_{it}^1}{\sigma q_{it}^1 \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} \right)$$

$$\frac{\partial f}{\partial \log(\sigma)} = f * \sum_{t=1}^{T_i} \left(-1 + \left(\frac{v_{it}^3}{\sigma}\right)^2 - \frac{\rho_3 v_{it}^3 p_{it}^2}{\sigma q_{it}^2 \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} - \frac{\rho_2 v_{it}^3 p_{it}^1}{\sigma q_{it}^1 \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} \right)$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_1}{1-\rho_1}\right)^{1/2}} = f * \sum_{t=1}^{T_i} \frac{q_{it}^1 q_{it}^2 \left(1-\rho_1^2\right) \phi\left(b_{it}^2\right) \phi\left(\frac{b_{it}^1 - q_{it}^1 q_{it}^2 \rho b_{it}^2}{\sqrt{1-\rho^2}} \right)}{r_n r_m \sqrt{1-\rho^2} \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)}$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_2}{1-\rho_2}\right)^{1/2}} = f * \sum_{t=1}^{T_i} \left(\frac{q_{it}^1 q_{it}^2 \left(\rho_1 \rho_2 - \rho_3\right) \phi\left(b_{it}^2\right) \phi\left(\frac{b_{it}^1 - q_{it}^1 q_{it}^2 \rho b_{it}^2}{\sqrt{1-\rho^2}} \right)}{r_n r_m \sqrt{1-\rho^2} \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} + \frac{\left(\rho_2 b_{it}^1 - \frac{r_n v_{it}^3}{\sigma} \right) p_{it}^1}{q_{it}^1 \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} \right)$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_3}{1-\rho_3}\right)^{1/2}} = f * \sum_{t=1}^{T_i} \left(\frac{q_{it}^1 q_{it}^2 \left(\rho_1 \rho_3 - \rho_2\right) \phi\left(b_{it}^2\right) \phi\left(\frac{b_{it}^1 - q_{it}^1 q_{it}^2 \rho b_{it}^2}{\sqrt{1-\rho^2}} \right)}{r_n r_m \sqrt{1-\rho^2} \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} + \frac{\left(\rho_3 b_{it}^2 - \frac{r_m v_{it}^3}{\sigma} \right) p_{it}^2}{q_{it}^2 \Phi_2\left(b_{it}^1, b_{it}^2; q_{it}^1 q_{it}^2 \rho\right)} \right)$$

$$\frac{\partial f}{\partial \log(\sigma_1)} = f * \left(-1 + \frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_i^1}{\sigma_1}\right)^2 - \left(\rho_{1,2} - \rho_{1,3}\rho_{2,3}\right)\frac{\mu_i^1 \mu_i^2}{\sigma_1 \sigma_2} - \left(\rho_{1,3} - \rho_{1,2}\rho_{2,3}\right)\frac{\mu_i^1 \mu_i^3}{\sigma_1 \sigma_3}}{\left(a_1 a_2 \sqrt{1-ra^2}\right)^2} \right)$$

$$\frac{\partial f}{\partial \log(\sigma_2)} = f * \left(-1 + \frac{\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_i^2}{\sigma_2}\right)^2 - \left(\rho_{1,2} - \rho_{1,3}\rho_{2,3}\right)\frac{\mu_i^1 \mu_i^2}{\sigma_1 \sigma_2} - \left(\rho_{2,3} - \rho_{1,2}\rho_{1,3}\right)\frac{\mu_i^2 \mu_i^3}{\sigma_2 \sigma_3}}{\left(a_1 a_2 \sqrt{1-ra^2}\right)^2} \right)$$

$$\frac{\partial f}{\partial \log(\sigma_3)} = f * \left(-1 + \frac{\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_i^3}{\sigma_3}\right)^2 - \left(\rho_{1,3} - \rho_{1,2}\rho_{2,3}\right)\frac{\mu_i^1 \mu_i^3}{\sigma_1 \sigma_3} - \left(\rho_{2,3} - \rho_{1,2}\rho_{1,3}\right)\frac{\mu_i^2 \mu_i^3}{\sigma_2 \sigma_3}}{\left(a_1 a_2 \sqrt{1-ra^2}\right)^2} \right)$$

$$\frac{\partial f}{\partial \log\left(\frac{1+\rho_{1,2}}{1-\rho_{1,2}}\right)^{1/2}} = f * \left(\frac{1-\rho_{1,2}^2}{\left(a_1 a_2 \sqrt{1-ra^2}\right)^2} \left(\rho_{1,2} \left(\frac{\mu_3^i}{\sigma_3}\right)^2 + \frac{\mu_1^i \mu_2^i}{\sigma_1 \sigma_2} - \rho_{2,3} \frac{\mu_1^i \mu_3^i}{\sigma_1 \sigma_3} - \rho_{1,3} \frac{\mu_2^i \mu_3^i}{\sigma_2 \sigma_3} \right) \right)$$

$$+f\frac{\left(1-\rho_{1,2}^2\right)\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\left(1-\frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_1^i}{\sigma_1}\right)^2+\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_2^i}{\sigma_2}\right)^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$+f\frac{\left(1-\rho_{1,2}^2\right)\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\left(\frac{-\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_3^i}{\sigma_3}\right)^2+2\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$+f\frac{\left(1-\rho_{1,2}^2\right)\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\left(2\frac{\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}+\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$\frac{\partial f}{\partial\log\left(\frac{1+\rho_{1,3}}{1-\rho_{1,3}}\right)^{1/2}}=f*\left(\frac{\left(1-\rho_{1,3}^2\right)\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)+\left(\rho_{1,3}\left(\frac{\mu_2^i}{\sigma_2}\right)^2-\rho_{2,3}\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}-\rho_{1,2}\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}\right)a_1^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$-f\left(\rho_{1,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{1,3}-\rho_{2,3}\right)\right)$$

$$*\frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_1^i}{\sigma_1}\right)^2+\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_2^i}{\sigma_2}\right)^2+\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_3^i}{\sigma_3}\right)^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

$$+2f\left(\rho_{1,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{1,3}-\rho_{2,3}\right)\right)$$

$$*\frac{\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}+\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

$$\frac{\partial f}{\partial\log\left(\frac{1+\rho_{2,3}}{1-\rho_{2,3}}\right)^{1/2}}=f*\left(\frac{\left(1-\rho_{2,3}^2\right)\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)+\left(\rho_{2,3}\left(\frac{\mu_1^i}{\sigma_1}\right)^2-\rho_{1,3}\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}-\rho_{1,2}\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}\right)a_2^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}\right)$$

$$-f\left(\rho_{2,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{2,3}-\rho_{1,3}\right)\right)$$

$$*\frac{\left(1-\rho_{2,3}^2\right)\left(\frac{\mu_1^i}{\sigma_1}\right)^2+\left(1-\rho_{1,3}^2\right)\left(\frac{\mu_2^i}{\sigma_2}\right)^2+\left(1-\rho_{1,2}^2\right)\left(\frac{\mu_3^i}{\sigma_3}\right)^2}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

$$+2f\left(\rho_{2,3}\left(a_1a_2\sqrt{1-ra^2}\right)^2+\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\left(\rho_{1,2}\rho_{2,3}-\rho_{1,3}\right)\right)$$

$$* \frac{\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)\frac{\mu_1^i\mu_2^i}{\sigma_1\sigma_2}+\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)\frac{\mu_1^i\mu_3^i}{\sigma_1\sigma_3}+\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right)\frac{\mu_2^i\mu_3^i}{\sigma_2\sigma_3}}{\left(a_1a_2\sqrt{1-ra^2}\right)^2}$$

4. Examples and Comparisons with Roodman's Command

In order to shed light on the advantages and the consistency of our method (xtcmp), we decide to implement examples using a dataset, previously used for xtsur[1], in Stata software. This database is an unbalanced panel database of 1672 observations, corresponding to 142 individuals followed between 1990 and 2003. All explanatory variables used are quantitative and do not contain any missing values.

We implement two cases: 1) a system of two equations with one linear and one binary dependent variables; 2) a system of three equations with two binary and one continuous outcomes. Indeed, let us consider the three following equations:

$$\tilde{y}_{it}^1 = \beta_0 + x1_{it}\beta_1 + x2_{it}\beta_2 + x3_{it}\beta_3 + x4_{it}\beta_4 + \mu_i^1 + \nu_{it}^1 \tag{20}$$

$$\tilde{y}_{it}^2 = \gamma_0 + x4_{it}\gamma_1 + x6_{it}\gamma_2 + x7_{it}\gamma_3 + \mu_i^2 + \nu_{it}^2 \tag{21}$$

$$\tilde{y}_{it}^3 = \alpha_0 + x7_{it}\alpha_1 + x9_{it}\alpha_2 + \mu_i^3 + \nu_{it}^3 \tag{22}$$

where y_{it}^1 and y_{it}^2 are binary variables equal to 1 if $\tilde{y}_{it}^1 > 0$ and $\tilde{y}_{it}^2 > 0$; and where $\tilde{y}_{it}^3 = y_{it}^3$ is a linear outcome.

When considering the first case (a system of two equations) with one continuous and one binary outcomes, we focus on the simultaneous estimation of Equation (20) and Equation (22). In this case, the associated variance/covariance matrices of the error components are given by:

$$\Sigma_\nu = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \text{ and } \Sigma_\mu = \begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

We consider four estimation techniques: 1) each equation is estimated separately as a single panel equation; 2) the two equations are estimated with cmp, while considering a pooled equation; 3) the two equations are estimated with cmp, with a posterior estimate of random effects; and, 4) the two equations are estimated with our method (xtcmp) presented in Section 2. Results are presented in Table 1.

Results suggest that first, estimating equations separately is misleading since the covariance between idiosyncratic errors and individual random effects is not considered; and, the significance of the coefficients appears to be false. On the other hand, considering equations simultaneously aims at obtaining consistent estimates among the three tests performed (coefficients are closer to each others in the last three columns). However, one should notice that, when using cmp with random effects, the estimation of the covariance matrices for both individual random effects and idiosyncratic errors is done after the estimation of the

[1]See Nguyen [12] for more detailed information.

Table 1. Two equations system with both a linear and a non-linear outcome.

Variables	Separate equations	Cmp on pooled data	Cmp with random effects	Our method xtcmp
		Binary outcome: $\tilde{y}_{it}^1$		
*x*1	0.0238 (0.0166)	−0.0056 (0.0062)	0.0021 (0.0224)	0.0119 (0.0082)
*x*2	0.2569*** (0.0238)	0.0297*** (0.0034)	0.0484*** (0.0045)	0.0204*** (0.0037)
*x*3	0.3171** (0.1228)	0.0053 (0.006)	0.0351 (0.0224)	0.0254 (0.0265)
*x*4	−0.5748*** (0.2182)	−0.0084 (0.0112)	−0.058 (0.0399)	−0.0463 (0.0471)
Intercept	−6.4382*** (0.375)	−0.7083*** (0.0379)	1.1074*** (0.1339)	−2.7151*** (0.0535)
		Continuous outcome: $\tilde{y}_{it}^3$		
*x*7	−0.1528*** (0.0207)	−0.0321* (0.0168)	−0.1044*** (0.015)	−0.0422*** (0.0158)
*x*9	0.0384*** (0.004)	0.1035*** (0.0047)	0.0361*** (0.004)	0.096*** (0.0043)
Intercept	15.9964*** (0.7339)	10.0568*** (0.5086)	17.0243*** (0.4477)	7.4308*** (0.4779)
		Covariance matrix: individual effects		
σ_1	3.8635*** (0.2135)	-	2.8431*** (0.2144)	2.0702*** (0.1263)
σ_2	6.024	-	5.706*** (0.1789)	2.2481*** (0.1421)
ρ_1	-	-	0.5135*** (0.0247)	0.8196*** (0.0417)
		Covariance matrix: idiosyncratic errors		
σ	2.6596	6.7719*** (0.1172)	2.6725*** (0.0481)	6.472*** (0.1182)
ρ	-	0.1525*** (0.0349)	0.0342 (0.0528)	0.5087*** (0.0418)

***: significant at the 1% level; **: significant at the 5% level; *: significant at the 10% level.

coefficients (post-estimation). Comparing with our results (xtcmp, last column), we can see that the variance of the individual effects seems to be overestimated in cmp's case.

Then, we offer an example for the second case, a three equations system with two binary outcomes and one continuous dependent variable, such that we consider Equations (20)-(22). In this case, the associated variance/covariance matrices of the error components are given by:

$$\Sigma_\nu = \begin{pmatrix} 1 & \rho_1 & \rho_2\sigma \\ \rho_1 & 1 & \rho_3\sigma \\ \rho_2\sigma & \rho_3\sigma & \sigma^2 \end{pmatrix} \text{ and } \Sigma_\mu = \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \rho_{1,3}\sigma_1\sigma_3 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{2,3}\sigma_2\sigma_3 \\ \rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$$

For this example, we cannot provide the third estimation technique where we used cmp command with a posterior estimate of the random effects, because this procedure does not converge. However, we provide the other estimation techniques: 1) each equation is estimated separately as a single panel equation; 2) the

three equations are estimated with cmp, considering a pooled database; and, 3) the three equations are estimated with our method (xtcmp). Results are provided in Table 2.

Table 2. Three equations system with two binary outcomes.

Variables	Separate equations	Cmp on pooled data	Our method xtcmp
	Binary outcome 1: $\tilde{y}_{it}^{1}$		
*x*1	0.0238 (0.0166)	−0.0067 (0.0063)	0.0399*** (0.0146)
*x*2	0.2569*** (0.0238)	0.0326*** (0.0034)	0.2313*** (0.0042)
*x*3	0.3171** (0.1228)	0.0065 (0.0072)	0.0445 (0.0379)
*x*4	−0.5748*** (0.2182)	−0.0107 (0.0132)	−0.0826 (0.0669)
Intercept	−6.4382*** (0.375)	−0.7176*** (0.0379)	−2.7205*** (0.1247)
	Binary outcome 2: $\tilde{y}_{it}^{2}$		
*x*4	0.1678*** (0.0480)	0.0076 (0.0078)	0.1345** (0.0584)
*x*6	−0.1068*** (0.0136)	−0.0499*** (0.0031)	−0.0201*** (0.0041)
*x*7	−0.0812*** (0.0193)	−0.0418*** (0.0040)	−0.0230*** (0.0071)
Intercept	8.8666*** (1.0978)	4.2916*** (0.2625)	1.8381*** (0.3885)
	Continuous outcome: $\tilde{y}_{it}^{3}$		
*x*7	−0.1528*** (0.0207)	−0.0343* (0.0168)	−0.0643*** (0.0123)
*x*9	0.0384*** (0.004)	0.1005*** (0.0049)	0.0608*** (0.0034)
Intercept	15.9964*** (0.7339)	10.2621*** (0.5172)	12.4296*** (0.3771)
	Covariance matrix: individual effects		
σ_1	3.8635*** (0.2135)	-	24.123*** (1.518)
σ_2	3.0296*** (0.3417)	-	6.5402*** (0.485)
σ_3	6.024	-	3.7135*** (0.235)
$\rho_{1,2}$	-	-	0.2091** (0.0853)
$\rho_{1,3}$	-	-	0.2379*** (0.0854)
$\rho_{2,3}$	-	-	0.6346*** (0.0565)
	Covariance matrix: idiosyncratic errors		
σ	2.6596	6.7756*** (0.1175)	4.6060*** (0.0870)
ρ_1	-	0.2468*** (0.0464)	0.8719*** (0.0681)
ρ_2	-	0.1132*** (0.0404)	0.8631*** (0.0274)
ρ_3	-	0.1501*** (0.0354)	0.5906*** (0.0565)

***: significant at the 1% level; **: significant at the 5% level; *: significant at the 10% level.

Results suggest that, as before, estimating equations separately leads to errors in the significance of coefficients, especially for the first outcome. Moreover, such method does not consider the covariance between idiosyncratic errors and individual random effects. On the other hand, considering equations simultaneously allows to obtain more consistent estimates (coefficient estimates seem closer in the last two columns, and significance of the latter is persistent along the two last columns).

5. Conclusions

xtcmp is a command implemented under Stata software. We focus on three main cases: 1) a simultaneous equations model with two equations (including one linear and one binary outcomes); 2) a case with three equations composed of two linear and one binary outcomes; and, 3) a three equations case with one linear and two binary dependent variables. This command further develops Roodman's [1] command cmp which does not explicitly consider the panel dimension of the data, nor simultaneous equations model since it is written as a SUR estimator. This technical note gives detailed description of the computations, namely likelihood functions, log-integrand associated, Hessian matrices and gradient vectors with respect to each parameter of interest, specific to the three cases described above.

xtcmp's estimation framework could be further developed in order to consider a broader range of non-linear outcomes (such as *ordered probit*, *multinomial probit* or *truncated framework*, for instance) or to consider much more equations simultaneously, in a dynamic setup.

Still, as it stands, xtcmp represents a significant development in Stata's commands. Indeed, it allows researchers to resolve endogeneity issues in a panel dimension context by analyzing correlation in the error terms of the equations and thus specific individual effects depending on the outcomes.

Acknowledgements

We thank E. Delattre for his helpful comments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Roodman, D. (2011) Fitting Fully Observed Recursive Mixed-Process Models with cmp. *The Stata Journal*, **11**, 159-206. https://doi.org/10.1177/1536867X1101100202

[2] Adeline, A. and Delattre, E. (2018) Health and Income: Testing for Causality on European Elderly People. THEMA Working Paper No. 2018-07.

[3] Williams Greene, H. (2012) Econometric Analysis. Seventh Edition, Prentice Hall, Upper Saddle River, NJ.

[4] Jeffrey Wooldridge, M. (2010) Econometric Analysis of Cross Section and Panel

Data. Second Edition, MIT Press, Cambridge, MA.

[5] Deb, P. and Trivedi, P.K. (2006) Maximum Simulated Likelihood Estimation of a Negative Binomial Regression Model with Multinomial Endogenous Treatment. *The Stata Journal*, **6**, 246-255. https://doi.org/10.1177/1536867X0600600206

[6] Kenkel, D.S. and Terza, J.V. (2001) The Effect of Physician Advice on Alcohol Consumption: Count Regression with an Endogenous Treatment Effect. *Journal of Applied Econometrics*, **16**, 165-184. https://doi.org/10.1002/jae.596

[7] Bratti, M. and Miranda, A. (2011) Endogenous Treatment Effects for Count Data Models with Endogenous Participation or Sample Selection. *Health Economics*, **20**, 1090-1109. https://doi.org/10.1002/hec.1764

[8] Esterling, K.M., Neblo, M.A. and Lazer, D.M.J. (2011) Estimating Treatment Effects in the Presence of Noncompliance and Nonresponse: The Generalized Endogenous Treatment Model. *Political Analysis*, **19**, 205-226. https://doi.org/10.1093/pan/mpr005

[9] Liu, Q. and Pierce, D. (1994) A Note on Gauss-Hermite Quadrature. *Biometrika*, **83**, 624-629. https://doi.org/10.2307/2337136

[10] Gould, W., Pitblado, J. and Poi, B. (2010) Maximum Likelihood Estimation with Stata. Fourth Edition, Stata Press, College Station, Texas.

[11] 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 Letter*, **8**, 1257-1278. https://doi.org/10.4236/tel.2018.86083

[12] Nguyen, M.C. (2008) XTSUR: Stata Module to Estimate Seemingly Unrelated Regression Model on Unbalanced Panel Data. Statistical Software Components S456953, Boston College Department of Economics.