Estimation of simultaneous-equations models with panel data and censored endogenous variables

Cantu-Bazaldua, Fernando

This publication URL: https://archive-ouverte.unige.ch/unige:28386
Publication DOI: 10.13097/archive-ouverte/unige:28386

Estimation of simultaneous-equations models with panel data and censored endogenous variables

THÈSE

présentée à la Faculté des sciences économiques et sociales
de l'Université de Genève

par

Fernando Cantú-Bazaldúa

sous la direction de

Prof. Jaya Krishnakumar

pour l'obtention du grade de

Docteur ès sciences économiques et sociales
mention économétrie

Membres du jury de thèse

Prof. Jaya Krishnakumar, directrice de thèse, Université de Genève
Prof. Tobias Mueller, président du jury, Université de Genève
M. Jean-Paul Chaze, Université de Genève
Prof. Michael Lechner, Université de Saint-Gall

Thèse No. 797
Genève, 22 mars 2013

La Faculté des sciences économiques et sociales, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par là, émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 22 mars 2013

Le doyen

Prof. Bernard MORARD

Impression d'après le manuscrit de l'auteur

Abstract

In this research we develop an estimation methodology for a system of simultaneous equations where the endogenous variables are subject to censorship and where the data follows a panel structure. The likelihood function of such a model presents several complications, so that traditional optimization procedures cannot be employed. We propose the application of a simulation-based estimator that mimics the Expectation-Maximization (EM) algorithm and also inherits its likelihood-maximizing properties. Simulation exercises for both random-effects and fixed-effect models verify that this estimation methodology performs remarkably well in comparison to traditional methods, without a high cost in terms of loss of efficiency. The same idea is then extended to other types of data limitations and to a dynamic model.

Resumé

Dans cette recherche on développe une méthode d'estimation pour un système d'équations simultanées avec des variables endogènes censurées et des données en panel. La fonction de vraisemblance de ce modèle présente des difficultés et les méthodes traditionnelles d'optimisation ne peuvent être utilisées. On propose alors l'application d'un estimateur basé sur l'algorithme Espérance-Maximisation (EM) construit avec des simulations et qui hérite ses propriétés concernant la maximisation de la vraisemblance. Des exercices de simulation pour un modèle à effets aléatoires et un modèle à effets fixes vérifient que cette méthode d'estimation a une bonne performance par rapport aux méthodes traditionnelles avec un coût modéré du point de vue de la perte d'efficacité. La même idée peut aussi être utilisée pour des autres types de variables limitées et pour des modèles dynamiques.

Acknowledgments
Agradecimientos

I would like to thank everyone that helped me with this work through constant guidance, suggestions, support and motivation. I would also like to acknowledge my thesis director, Prof. Jaya Krishnakumar, and all the committee for their valuable advice, patience and support.

Esta trabajo va especialmente dedicado a mi padre, quien siempre tuvo el sueño de obtener un doctorado, y a mi madre por su incondicional apoyo.

I would finally like to dedicate this thesis to TheWeeknd, who accompanied me through the countless hours of work.

Contents

1 Introduction **1**

2 Literature review **5**

2.1 Simultaneous equation models and censored dependent variables . . . 5
2.2 Simultaneous equation models and panel data . . . 7
2.3 Censored dependent variables and panel data . . . 8
2.4 Intersection of the three areas . . . 11

3 Description of the model **13**

3.1 Panel with individual random effects . . . 16
3.1.1 Four possible cases . . . 18
3.1.2 Constructing the likelihood function . . . 23
3.1.3 Gradient vector . . . 29
3.1.4 Marginal effects of continuous exogenous variables . . . 30
3.1.5 Effects of dichotomous exogenous variables . . . 47
3.2 Panel with individual fixed effects . . . 53
3.2.1 Construction of the likelihood . . . 54
3.2.2 Gradient vector . . . 58
3.2.3 Marginal effects of exogenous continuous variables . . . 59
3.2.4 Effects of dichotomous exogenous variables . . . 62
3.3 Consistency and identification constraints . . . 65
3.3.1 Consistency . . . 65
3.3.2 Identification . . . 67

4 Optimization methods **69**

4.1 Standard gradient methods . . . 70
4.2 Simulated maximum likelihood and simulated score method . 72
4.3 Metaheuristic algorithms . . . 75
4.3.1 Genetic algorithm . . . 76
4.3.2 Simulated annealing . . . 78

4.4 EM algorithm 80
4.5 Gibbs sampler 82
4.6 MCECM algorithm 87

5 Simulation study 93
5.1 Simulation study for the random-effects model 93
5.1.1 Robustness to different distributions of the exogenous variables 109
5.1.2 Robustness to different distributions of the error term 115
5.1.3 Robustness to different functional forms 122
5.1.4 Results with different sample sizes 127
5.1.5 Other simulations 133
5.1.6 Summary of the simulation study for the random-effects model 140
5.2 Simulation study for the fixed-effects model 142

6 Extensions 155
6.1 Probit-probit model 155
6.1.1 Four possible cases 156
6.1.2 The likelihood function and its optimization 160
6.2 Tobit type I-II model 161
6.2.1 Four possible cases 162
6.2.2 The likelihood function and its optimization 170
6.3 Dynamic models 172

7 Conclusions 177

Appendix A Results from the truncated bivariate normal distribution 181

Appendix B Gradient of the log likelihood function 187
B.1 Random effects model 187
B.2 Fixed effects model 223

Appendix C Computational considerations 233
C.1 Ascent-based version of the MCECM algorithm 234
C.2 Simulating from a truncated normal distribution 238
C.3 Parallel computing 243

List of Tables

3.1 Summary of the reduced form of the model 65

5.1 Comparison of the estimation results for the simulation study of the random-effects model, base simulation 101

5.2 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, base simulation 107

5.3 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, exogenous variables generated with a normal distribution . . 110

5.4 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, exogenous variables generated with a log-normal distribution 112

5.5 Comparison of the average effects for the simulation study of the random-effects model, values multiplied by 10^2, one dummy exogenous variable 115

5.6 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, error terms drawn from a mixture of normal distributions . . 117

5.7 Comparison of the estimation results for the simulation study of the random-effects model, error terms drawn from a mixture of normal distributions 120

5.8 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, error terms drawn from a skew-normal distribution 121

5.9 Comparison of the estimation results for the simulation study of the random-effects model, error terms drawn from a skew-normal distribution 124

5.10 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, functional form quadratic in the exogenous variables 125

5.11 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, functional form logarithmic in the exogenous variables 128
5.12 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, sample size $N = 25$, $T = 5$ 129
5.13 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, sample size $N = 40$, $T = 10$ 131
5.14 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, one additional exogenous variable 134
5.15 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, weak instrument 137
5.16 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, less censorship 138
5.17 Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, more censorship 139
5.18 Comparison of the average marginal effects for the simulation study of the fixed-effects model, values multiplied by 10^2 . . . 150

C.1 Parameters of the ascent-based Monte Carlo Expectation Conditional Maximization (MCECM) algorithm 236

List of Figures

5.1 Comparison of the estimation results of β_{11}, β_{12}, β_{21} and β_{22} for the simulation study of the random-effects model, base simulation . . . 103
5.2 Comparison of the estimation results of γ_1 and γ_2 for the simulation study of the random-effects model, base simulation 103
5.3 Comparison of the estimation results of elements of the variance-covariance matrices of the error components for the simulation study of the random-effects model, base simulation . . . 105
5.4 Comparison of the average marginal effects for the simulation study of the random-effects model, base simulation 108
5.5 Comparison of the average marginal effects for the simulation study of the random-effects model, exogenous variables generated with a normal distribution . . . 111
5.6 Comparison of the average marginal effects for the simulation study of the random-effects model, exogenous variables generated with a log-normal distribution . . . 113
5.7 Comparison of the average effects for the simulation study of the random-effects model, one dummy exogenous variable . . 116
5.8 Comparison of the average marginal effects for the simulation study of the random-effects model, error terms drawn from a mixture of normal distributions . . . 118
5.9 Comparison of the average marginal effects for the simulation study of the random-effects model, error terms drawn from a skew-normal distribution . . . 123
5.10 Comparison of the average marginal effects for the simulation study of the random-effects model, functional form quadratic in the exogenous variables . . . 126
5.11 Comparison of the average marginal effects for the simulation study of the random-effects model, functional form logarithmic in the exogenous variables . . . 128

5.12 Comparison of the average marginal effects for the simulation study of the random-effects model, sample size $N = 25$, $T = 5$ 130
5.13 Comparison of the average marginal effects for the simulation study of the random-effects model, sample size $N = 40$, $T = 10$ 132
5.14 Comparison of the average marginal effects for the simulation study of the random-effects model, one additional exogenous variable . . . 135
5.15 Comparison of the average marginal effects for the simulation study of the random-effects model, weak instrument . . . 137
5.16 Comparison of the average marginal effects for the simulation study of the random-effects model, less censorship . . . 139
5.17 Comparison of the average marginal effects for the simulation study of the random-effects model, more censorship . . . 140
5.18 Comparison of the average marginal effects for the simulation study of the fixed-effects model . . . 152

List of Algorithms

1 Genetic algorithm 77
2 Simulated annealing 79
3 EM algorithm 81
4 Gibbs sampler 85
5 MCECM algorithm 88
6 Ascent-based MCECM algorithm 235

List of Acronyms

2SLS	Two-Stage Least Squares
3SLS	Three-Stage Least Squares
EM	Expectation-Maximization
FIML	Full Information Maximum Likelihood
DGP	Data Generating Process
GMM	Generalized Method of Moments
GLS	Generalized Least Squares
GLS-RE	Generalized Least Squares for a Random-Effects Model
LDV	Limited Dependent Variable
LSDV	Least Squares Dummy Variables
OLS	Ordinary Least Squares
MCECM	Monte Carlo Expectation Conditional Maximization
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimator
MSE	Mean Squared Error
SEM	Simultaneous Equation Model
SEM-RE	Simultaneous Equation Model with Random Effects
SEM-FE	Simultaneous Equation Model with Fixed Effects
SUR	Seemingly Unrelated Regressions
TOB	Tobit Model

TOB-RE Tobit Model with Random Effects

TOB-FE Tobit Model with Fixed Effects

Chapter 1

Introduction

Consider the following two situations:

1. An agricultural production function defines the supply of a particular output in terms of the inputs employed for its production (i.e., land, labor, capital and "technology"). In an empirical study of this relationship, the sample consists of M crops produced by N agricultural firms over T periods of time. However, not all firms exhibit a positive production for the M products at all T periods; it is sometimes observed that, even though some production factors have been engaged, harvest does not occur at every sampling period because of seasonal or other factors. In this case, the database consists of MNT observations for which production can be either zero or a positive quantity.

2. Health expenditures can be an important component of the household expenditure at a given period of time. In a developing country where important sectors of the population have informal jobs and do not have access to health insurance, this type of expenditure can represent a significant outlay for a household if one of its members falls seriously ill. An equation can attempt to model this type of expenditure based

on some explanatory variables, which may include the household's income. But expenses on healthcare are not observed continuously: only when some major medical treatment is needed would the household incur these kinds of expenditures. At the same time, the income of the household depends on many factors, including the health situation of its members, and some cases with zero income are expected to be observed for those households with no income during a certain period (because of the loss of employment due to the inadequate health status of the earning member, to mention one reason). In general, these simultaneous relationships can be estimated through a panel of N households observed over T periods.

The two cases cited above refer to a set of M related, possibly simultaneous relationships observed for N individuals and T periods of time when the data is not observed continuously. Explicitly, this configuration lies at the intersection of three important areas of study in econometrics: models for panel data, Limited Dependent Variable (LDV) models and Simultaneous Equation Models (SEMs). Until now, few studies have jointly considered the three areas. However, as illustrated by the examples described above, there are situations that can be modeled in such a setting and a specific estimation procedure would be advantageous.

In this thesis we develop an estimation procedure based on the likelihood principle for this type of model. In this setting, however, the likelihood function becomes intractable even for the simplest cases. Because of its complexity, traditional optimization solutions (either analytical or numerical) cannot be employed to find the Maximum Likelihood Estimator (MLE). We propose the use of an Expectation-Maximization (EM) algorithm based on simulation as an alternative and that is closely linked to the method of simulated scores, which showed favorable results in a simulation exercise. Even though we concentrate on a model with censored endogenous variables or type-I Tobit model, it will be shown in this research that the same methodology can be extended to other forms of limitation (binary variables, truncation or sample selection, for example).

The rest of the thesis is organized in the following way. First, Chapter 2 briefly reviews previous methodological and applied studies similar to the one studied here. Chapter 3 then presents the theoretical model when either random or fixed effects are considered. The different optimizations procedures that were attempted to estimate it are summarized in the following chapter. We describe in Chapter 5 the detailed results from two simulation studies undertaken to evaluate the performance of the selected optimization algorithm, one with random effects and a second for a fixed-effects model. Chapter 6 shows, through two fully developed examples, how the methodology applied here can be extended to other types of data limitations; it also includes a word about extending the model to a dynamic setting. Lastly, a final chapter summarizes the main results and presents a general conclusion.

Chapter 2

Literature review

Very few studies have considered a model that lies at the intersection of the three problems described in the introduction (simultaneous equations, censored dependent variables and panel data). However, pairwise combinations of the three areas have produced a rich body of literature. This chapter will briefly present some of the methodological results and empirical applications. The material listed here is far from exhaustive and only intends to provide an idea of the direction that research has taken in each area.

2.1 Simultaneous equation models and censored dependent variables

The interaction between SEMs and censored dependent variables was first explored by Amemiya (1974), who extended the classic Tobit model to a system of multivariate regressions or simultaneous equations and proposed a consistent, albeit inefficient, estimator obtained from the properties of the truncated multivariate normal distribution and an indirect-least-squares type estimator; he then employed this methodology to study the joint determination of husband and wife work hours. Nelson and Olson (1978) developed a two-stage least squares estimator for a simpler version of the SEM where the simultaneity depends on the unobserved, latent variables instead of the observed values of the limited variables; they applied their estimator to model decisions about post-secondary vocational school training. Remaining in this same setting, Amemiya (1979) proposed a generalized, more efficient version of the estimator. Another consistent estimator, this time based on conditional maximum likelihood of the reduced-form parameters, was developed by Smith and Blundell (1986). Newey (1987) studied the statistical properties of these estimators and used them to derive an asymptotically efficient estimator. Later, Blundell and Smith (1989, 1994) elaborated on the conditional MLE and provided an iterative method for computing an asymptotically efficient joint MLE; they applied their algorithm to a model where female labor supply and other household income are jointly determined.

Vella (1993) developed a general framework that encompassed many types of LDVs (censored variables, several types of sample selection and others); he then proposed a consistent two-step estimator based on the generalized residuals and applied it in modeling the trade-off between wages and fringe benefits. Huang et al. (1987) estimated a bivariate regression model[1] with censored dependent variables through the EM algorithm and then applied their method to a model of life-health insurance and pension

[1] Multivariate regression models are also called Seemingly Unrelated Regressions (SUR) models.

benefits. On the other hand, Cornick et al. (1994) obtained a MLE for a SUR model for censored data and used it to estimate the household consumption of different diary products; the authors remarked on the computational complexity of their estimator. This idea was subsequently extended by Chavas and Kim (2004) to equations that include lagged dependent variables. Huang (1999) followed an estimation method based on a simulated version of the EM algorithm for a bivariate SUR with an application to stock and cash dividend payments. Finally, a semi-parametric estimation approach for a SEM was developed by Lee (1995) (with an application to the labor supply of married women) and later by Chen and Zhou (2011).

2.2 Simultaneous equation models and panel data

The estimation of a SUR model with panel data was first exposed by Avery (1977), who extended the traditional GLS estimator of a single-equation random-effects model to a SUR setting; he also proposed a feasible version based on estimated values of the variance of the error components. However, Baltagi (1980) proved that this estimator was asymptotically inefficient and proposed an alternative, asymptotically efficient estimator. Baltagi (1981) expanded the error components literature to a SEM, developed equivalents of the Two-Stage Least Squares (2SLS) and Three-Stage Least Squares (3SLS) estimators in an error components context, and derived the asymptotic properties of such estimators. Prucha (1985) then developed a Full Information Maximum Likelihood (FIML) estimator assuming normality of the error components and showed that this estimator has an instrumental variable representation that can be used to generate a wide class of estimators that are computationally simpler. Balestra and Varadharajan-Krishnakumar (1987) proposed alternative, "generalized" specifications of the 2SLS and 3SLS estimators; they also developed the FIML estimator of structural parameters and derived the limiting distribution of both the coefficient estimators and the covariance estimators. An extensive survey of these methods is presented in Krishnakumar (1988).

Not only an error-component structure has been studied; for instance, Cornwell et al. (1992) studied a SEM with either fixed individual effects or random effects that are correlated with the exogenous variables. An empirical application of a panel SUR model with random effects was undertaken by Wan et al. (1992) for a set of agricultural production functions. Finally, Kinal and Lahiri (1993) followed a computational perspective to propose efficient algorithms to calculate some of the estimators cited above; they applied their procedures to a model of foreign trade in developing countries.

2.3 Censored dependent variables and panel data

Of the possible pairwise combinations of models, that which has been studied more extensively is the intersection of regressions with censored dependent variables and the use of data with a panel structure. The body of literature was triggered by two influential articles: the study by Hausman and Wise (1979) of attrition in social experimentation and the work by Heckman and Macurdy (1980) about female labor supply. The former modeled attrition in a panel data with random individual effects and obtained MLEs for the parameters of the model; they then used this method to estimate a model based on the Gary experiment of potential labor supply and earning responses to possible income maintenance plans. The latter extended the Tobit model to a panel with fixed individual effects and estimated the parameters through an iterative method; the estimation procedure was then applied to a life cycle model of female labor supply. These articles were included in a precursory review of panel data models with LDVs (Maddala, 1987) and a comparison between the random and fixed individual effects methodologies was carried out by Jakubson (1988) in an application to the life cycle model of labor supply by married women. An application of the random-effects specification was presented by Kim and Maddala (1992) when studying a model of dividend with both time and individual error components.

However, as noted by Honoré in his seminal article (1992), the estimator proposed by Heckman and Macurdy is inconsistent because of the incidental parameters problem. By exploiting the symmetry in the distribution of the latent data, he proposed the use of a trimmed sample for truncated and censored variable models with fixed effects. The resulting estimator, calculated through a Generalized Method of Moments (GMM) approach, is both consistent and asymptotically normal. To mention one empirical application of this model, Grootendorst (1997) employed it to study expenditure in prescription drugs from insurance claims data. Honoré's solution was later extended to equations with lagged dependent variables by Honoré (1993) and Hu (2002), while Honoré et al. (2000) used the same idea to estimate sample selection models and other types of LDVs. Honoré and Hu (2004) then modified the method of moments to allow for predetermined and endogenous regressors, while Alan et al. (2008) transformed the model to account for two-sided censoring. An alternative to Honoré's method was devised by Kalwij (2003), who devised a MLE based on the first difference of the latent dependent variable; an application to the determination of paid overtime hours was presented by Gregory and Kalwij (2000). Finally, Bover and Arellano (1997) developed a two-step within-groups estimator for a dynamic censored panel model with random effects.

Due to the significant estimation difficulties encountered when estimating these models, a number of articles has proposed alternative procedures. For example, Keane (1993, 1994) proposed a method of simulated moments for a panel data model with LDVs, while Hajivassiliou (1994) employed the method of simulated maximum likelihood and simulated scores to estimate a model of external debt crises in developing countries. These developments were included in a comprehensive survey by Arellano and Honoré (2001).

In a study slightly distinct from the rest of this group, Vella and Verbeek (1999) deal with the estimation of a panel data model with censored endogenous variables and sample selection[2]. They undertook an empirical application of their estimation method in a study of the impact of hours worked on the offered hourly wage rate, where the hours worked was modeled as a LDV that depends on a second equation of exogenous determinants. Finally, Greene (2004a,b) studied the extent of the incidental parameters problem in the MLE of a fixed effects panel data model with several types of LDVs and, in particular, censored dependent variables.

[2]This is a case of LDV where the censorship of a variable depends on the value taken by an auxiliary equation based on exogenous variables.

2.4 Intersection of the three areas

Of the studies included in this literature review, only the note by Ai and Chen (1992) considers the intersection of the three areas of study mentioned here. They extended Honoré's trimmed estimator for fixed-effects panel data with censored or truncated dependent variables to a bivariate (SUR) setting. However, their approach cannot be generalized to more than two equations or when such equations are defined simultaneously.

Chapter 3

Description of the model

This chapter will outline the model under study in this thesis: a system of simultaneous equations with censored dependent variables and data with a panel structure[3]. This model can be written as the following system of M structural equations

$$y^*_{mit} = \sum_{\substack{j=1 \\ j \neq m}}^{M} \gamma_{jm} y_{jit} + \beta'_m X_{it} + u_{mit} \tag{3.1}$$

$$u_{mit} = \alpha_{mi} + \lambda_{mt} + \epsilon_{mit} \tag{3.2}$$

for $m = 1, \ldots, M$ equations, $i = 1, \ldots, N$ individuals and $t = 1, \ldots, T$ periods. In these equations, the y_{jit} represent the endogenously determined dependent variables, X_{it} is a K-dimensional vector of exogenous regressors, while γ_{jm} and β_m are unknown coefficients. The error term u_{mit} has the structure specified in (3.2): an individual-specific component α_{mi}, a time-specific component λ_{mt} and a separate error component ϵ_{mit}. Depending on the properties taken by the individual and temporal components, we will be working with either an error-components model or a fixed-effects panel

[3]This chapter took as its starting point the master's thesis by Criton (2007).

data model.

y^*_{mit} is not an observed quantity but rather a latent construction. The observed variable is y_{mit} and it is related to its latent counterpart by the following function

$$y_{mit} = \psi(y^*_{mit}) \tag{3.3}$$

$\psi(\cdot)$ is the function that defines the limited character of the dependent variables.

In this research we will will concentrate on a particular version of the model described above. Specifically, we will consider a model with the following characteristics.

1. $\psi(\cdot)$ defines a one-sided censorship for the endogenous variables; that is,

$$y_{mit} = \psi(y^*_{mit}) = \begin{cases} y^*_{mit} & \text{if } y^*_{mit} > 0 \\ 0 & \text{otherwise} \end{cases}$$

2. there are only two equations in the system (i.e., $M = 2$).

3. only individual effects are present (i.e., $\lambda_{mt} = 0$, $\forall m, t$) so that equation (3.2) becomes

$$u_{mit} = \alpha_{mi} + \epsilon_{mit}$$

Putting all the pieces together, the model becomes a one-sided censored bivariate simultaneous system with individual effects. That is,

$$y^*_{1it} = \gamma_1 y_{2it} + \beta_1' X_{it} + \alpha_{1i} + \epsilon_{1it} \tag{3.4}$$

$$y^*_{2it} = \gamma_2 y_{1it} + \beta_2' X_{it} + \alpha_{2i} + \epsilon_{2it} \tag{3.5}$$

$$y_{1it} = \begin{cases} y^*_{1it} & \text{if } y^*_{1it} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$y_{2it} = \begin{cases} y^*_{2it} & \text{if } y^*_{2it} > 0 \\ 0 & \text{otherwise} \end{cases}$$

We call this configuration a Tobit-Tobit model[4]. We chose this particular setting because censored regressions are suitable for empirical econometric applications in many areas of economics and the Tobit type-I model is the basic setting in this context so that, once this is properly analysed and understood, it becomes easier to go on to generalisations such as Tobit type-II or sample selection models. Some of these extensions will be presented and developed in detail in Chapter 6.

The following sections will develop the likelihood function for this model, first when the individual effects are random quantities uncorrelated to the exogenous variables and then when they are taken as parameters that may be correlated with the exogenous variables. For both cases we will develop the model extensively and calculate the full marginal effects to changes in the exogenous variables; then we will obtain the likelihood function and will present the derivation of its gradient vector.

[4]In equations (3.4) and (3.5), the endogenous variables appear in their *observed* form in the right-hand side. This is the standard specification proposed by Amemiya (1974) and that has been used more frequently in the literature. The justification to do so is that the variables are censored for economic or logical reasons (for example, production cannot be smaller than zero or minimum wages are enforced by a country's regulation). Since the latent variables have no sense over the censored interval, the observed variables must be used. However, following Nelson and Olson (1978), it is also possible to include the endogenous variables in their *latent* form in the right-hand side:

$$y_{1it}^* = \gamma_1 y_{2it}^* + \beta_1' X_{it} + \alpha_{1i} + \epsilon_{1it}$$
$$y_{2it}^* = \gamma_2 y_{1it}^* + \beta_2' X_{it} + \alpha_{2i} + \epsilon_{2it}$$

Nelson and Olson argue that this specification is appropriate when the censorship is simply an artifact of data measurement or reporting. Even though this second configuration simplifies considerably the calculation of the reduced form, we sustain than Amemiya's specification is more suitable for most applied work in economics, so we will keep it in this research.

3.1 Panel with individual random effects

The vector X_{it} contains all the exogenous regressors of the model. It can be partitioned as follows:

$$X_{it} = \begin{bmatrix} x_{1it} \\ x_{2it} \\ x_{it} \end{bmatrix}$$

where x_{mit} is the subvector of exogenous variables included only in equation m and x_{it} represents the subvector of exogenous variables common to both equations. The coefficient vectors can be partitioned accordingly.

$$\beta_1 = \begin{bmatrix} \beta_{11} \\ 0 \\ \beta_{12} \end{bmatrix} \qquad \beta_2 = \begin{bmatrix} 0 \\ \beta_{21} \\ \beta_{22} \end{bmatrix}$$

Under this change of variables, equations (3.4) and (3.5) become

$$y^*_{1it} = \gamma_1 y_{2it} + \beta'_{11} x_{1it} + \beta'_{12} x_{it} + \alpha_{1i} + \epsilon_{1it} \tag{3.4b}$$

$$y^*_{2it} = \gamma_2 y_{1it} + \beta'_{21} x_{2it} + \beta'_{22} x_{it} + \alpha_{2i} + \epsilon_{2it} \tag{3.5b}$$

We make the following standard assumptions about the error components.

i) The error components have the following joint normal distribution $\forall i, t$

$$\begin{pmatrix} \alpha_i \\ \epsilon_{it} \end{pmatrix} = \begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^\alpha_{11} & \sigma^\alpha_{12} & 0 & 0 \\ \sigma^\alpha_{12} & \sigma^\alpha_{22} & 0 & 0 \\ 0 & 0 & \sigma^\epsilon_{11} & \sigma^\epsilon_{12} \\ 0 & 0 & \sigma^\epsilon_{12} & \sigma^\epsilon_{22} \end{pmatrix} \right]$$

The iid assumption implies that $\mathrm{E}(\alpha_i \alpha_j) = 0 \; \forall i \neq j$ and that $\mathrm{E}(\epsilon_{it} \epsilon_{js}) = 0 \; \forall i \neq j$ or $\forall t \neq s$. Note also that the individual random effects and the error terms are uncorrelated; i.e.

$$\mathrm{E}(\alpha_{mi}\epsilon_{kjt}) = 0, \ \forall m, k, i, j, t$$

ii) The explanatory variables and the error components are independent.

From assumption (i), it follows that two observations at different periods for the same individual are not independent since they share the common error component α_i.

According to the limited character of the endogenous variables each individual i and period t can belong to one of four cases[5].

I. Both variables show a positive value:

$$\begin{cases} y_{1it} > 0 \Rightarrow y^*_{1it} = y_{1it} \\ y_{2it} > 0 \Rightarrow y^*_{2it} = y_{2it} \end{cases}$$

II. Only the second variable is censored:

$$\begin{cases} y_{1it} > 0 \Rightarrow y^*_{1it} = y_{1it} \\ y_{2it} = 0 \Rightarrow y^*_{2it} \leq 0 \end{cases}$$

III. Only the first variable is censored:

$$\begin{cases} y_{1it} = 0 \Rightarrow y^*_{1it} \leq 0 \\ y_{2it} > 0 \Rightarrow y^*_{2it} = y_{2it} \end{cases}$$

IV. Both variables are censored:

$$\begin{cases} y_{1it} = 0 \Rightarrow y^*_{1it} \leq 0 \\ y_{2it} = 0 \Rightarrow y^*_{2it} \leq 0 \end{cases}$$

[5]This is true only in this bivariate setting. In general, there are 2^M possible cases.

In order to build the complete likelihood function, we need to consider the reduced form for each of these cases separately. We will study each one of them in the following subsection.

3.1.1 Four possible cases

Case I: $y_{1it} > 0$ **and** $y_{2it} > 0$

Since both variables are observed, it is possible to compute the reduced form by inserting (3.5b) into (3.4b) and viceversa, so that we obtain

$$\begin{aligned} y^*_{1it} &= \gamma_1\left(\gamma_2 y_{1it} + \beta'_{21}x_{2it} + \beta'_{22}x_{it} + \alpha_{2i} + \epsilon_{2it}\right) + \beta'_{11}x_{1it} + \beta'_{12}x_{it} + \alpha_{1i} + \epsilon_{1it} \\ &= \frac{1}{1-\gamma_1\gamma_2}\left[\begin{pmatrix} \beta_{11} \\ \gamma_1\beta_{21} \\ \beta_{12}+\gamma_1\beta_{22} \end{pmatrix}' \begin{pmatrix} x_{1it} \\ x_{2it} \\ x_{it} \end{pmatrix} + (\alpha_{1i}+\gamma_1\alpha_{2i}) + (\epsilon_{1it}+\gamma_1\epsilon_{2it})\right] \end{aligned}$$

and

$$y^*_{2it} = \frac{1}{1-\gamma_1\gamma_2}\left[\begin{pmatrix} \gamma_2\beta_{11} \\ \beta_{21} \\ \gamma_2\beta_{12}+\beta_{22} \end{pmatrix}' \begin{pmatrix} x_{1it} \\ x_{2it} \\ x_{it} \end{pmatrix} + (\alpha_{2i}+\gamma_2\alpha_{1i}) + (\epsilon_{2it}+\gamma_2\epsilon_{1it})\right]$$

Define the following quantities

$$\begin{gathered} \Gamma = 1-\gamma_1\gamma_2 \\ \delta_{1i} = \alpha_{1i}+\gamma_1\alpha_{2i} \qquad \delta_{2i} = \alpha_{2i}+\gamma_2\alpha_{1i} \\ \nu_{1it} = \epsilon_{1it}+\gamma_1\epsilon_{2it} \qquad \nu_{2it} = \epsilon_{2it}+\gamma_2\epsilon_{1it} \\ \Pi_1 = \begin{pmatrix} \beta_{11} \\ \gamma_1\beta_{21} \\ \beta_{12}+\gamma_1\beta_{22} \end{pmatrix} \quad \Pi_2 = \begin{pmatrix} \gamma_2\beta_{11} \\ \beta_{21} \\ \gamma_2\beta_{12}+\beta_{22} \end{pmatrix} \end{gathered}$$

so that the system can be written in the following reduced form

$$y^*_{1it} = \frac{1}{\Gamma}\left(\Pi'_1 X_{it} + \delta_{1i} + \nu_{1it}\right) \tag{3.6}$$

$$y^*_{2it} = \frac{1}{\Gamma}\left(\Pi'_2 X_{it} + \delta_{2i} + \nu_{2it}\right) \tag{3.7}$$

Now we derive the statistical properties of the new variables, a straightforward task since they are linear combinations of normal variables.

$$
\begin{aligned}
\mathrm{E}(\delta_{1i}) &= \mathrm{E}(\alpha_{1i} + \gamma_1\alpha_{2i}) = 0\\
\mathrm{E}(\delta_{2i}) &= \mathrm{E}(\alpha_{2i} + \gamma_2\alpha_{1i}) = 0\\
\mathrm{Var}(\delta_{1i}) &= \mathrm{Var}(\alpha_{1i} + \gamma_1\alpha_{2i}) = \sigma_{11}^{\alpha} + \gamma_1^2\sigma_{22}^{\alpha} + 2\gamma_1\sigma_{12}^{\alpha} = \sigma_{11}^{\delta}\\
\mathrm{Var}(\delta_{2i}) &= \mathrm{Var}(\alpha_{2i} + \gamma_2\alpha_{1i}) = \sigma_{22}^{\alpha} + \gamma_2^2\sigma_{11}^{\alpha} + 2\gamma_2\sigma_{12}^{\alpha} = \sigma_{22}^{\delta}\\
\mathrm{E}(\delta_{1i}\delta_{2j}) &= \sigma_{12}^{\delta} = \begin{cases} (1+\gamma_1\gamma_2)\sigma_{12}^{\alpha} + \gamma_2\sigma_{11}^{\alpha} + \gamma_1\sigma_{22}^{\alpha} & \text{if } i = j\\ 0 & \text{otherwise}\end{cases}\\
\mathrm{E}(\nu_{1it}) &= \mathrm{E}(\epsilon_{1it} + \gamma_1\epsilon_{2it}) = 0\\
\mathrm{E}(\nu_{2it}) &= \mathrm{E}(\epsilon_{2it} + \gamma_2\epsilon_{1it}) = 0\\
\mathrm{Var}(\nu_{1it}) &= \mathrm{Var}(\epsilon_{1it} + \gamma_1\epsilon_{2it}) = \sigma_{11}^{\epsilon} + \gamma_1^2\sigma_{22}^{\epsilon} + 2\gamma_1\sigma_{12}^{\epsilon} = \sigma_{11}^{\nu}\\
\mathrm{Var}(\nu_{2it}) &= \mathrm{Var}(\epsilon_{2it} + \gamma_2\epsilon_{1it}) = \sigma_{22}^{\epsilon} + \gamma_2^2\sigma_{11}^{\epsilon} + 2\gamma_2\sigma_{12}^{\epsilon} = \sigma_{22}^{\nu}\\
\mathrm{E}(\nu_{1it}\nu_{2js}) &= \sigma_{12}^{\nu} = \begin{cases} (1+\gamma_1\gamma_2)\sigma_{12}^{\epsilon} + \gamma_2\sigma_{11}^{\epsilon} + \gamma_1\sigma_{22}^{\epsilon} & \text{if } i = j,\, t = s\\ 0 & \text{otherwise}\end{cases}\\
\mathrm{E}(\delta_{mit}\nu_{kjs}) &= 0, \forall m, k, i, j, t
\end{aligned}
$$

Therefore,

$$
\begin{pmatrix} \delta_{1i}\\ \delta_{2i}\\ \nu_{1it}\\ \nu_{2it}\end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\delta} & \sigma_{12}^{\delta} & 0 & 0\\ \sigma_{12}^{\delta} & \sigma_{22}^{\delta} & 0 & 0\\ 0 & 0 & \sigma_{11}^{\nu} & \sigma_{12}^{\nu}\\ 0 & 0 & \sigma_{12}^{\nu} & \sigma_{22}^{\nu}\end{pmatrix}\right]
$$

As a consequence of these developments,

$$
\begin{pmatrix} y_{1it}^{*}\\ y_{2it}^{*}\end{pmatrix} \sim \mathbb{N}\left[\frac{1}{\Gamma}\begin{pmatrix}\Pi_1' X_{it}\\ \Pi_2' X_{it}\end{pmatrix}, \frac{1}{\Gamma^2}\begin{pmatrix} \sigma_{11}^{\delta} + \sigma_{11}^{\nu} & \sigma_{12}^{\delta} + \sigma_{12}^{\nu}\\ \sigma_{12}^{\delta} + \sigma_{12}^{\nu} & \sigma_{22}^{\delta} + \sigma_{22}^{\nu}\end{pmatrix}\right]
$$

It can be noted that the correlation coefficient between the two variables is given by

$$
\rho_I = \frac{\sigma_{12}^{\delta} + \sigma_{12}^{\nu}}{\sqrt{(\sigma_{11}^{\delta} + \sigma_{11}^{\nu})(\sigma_{22}^{\delta} + \sigma_{22}^{\nu})}}
$$

Case II: $y_{1it} > 0$ **and** $y_{2it} = 0$

Since the second endogenous variable is censored, the simultaneous term drops from (3.4) and this equation becomes its own reduced form. For the second equation, we have that

$$\begin{aligned} y^*_{2it} &= \gamma_2 \left(\beta'_{11}x_{1it} + \beta'_{12}x_{it} + \alpha_{1i} + \epsilon_{1it}\right) + \beta'_{21}x_{2it} + \beta'_{22}x_{it} + \alpha_{2i} + \epsilon_{2it} \\ &= \begin{pmatrix} \gamma_2\beta_{11} \\ \beta_{21} \\ \gamma_2\beta_{12} + \beta_{22} \end{pmatrix}' \begin{pmatrix} x_{1it} \\ x_{2it} \\ x_{it} \end{pmatrix} + (\alpha_{2i} + \gamma_2\alpha_{1i}) + (\epsilon_{2it} + \gamma_2\epsilon_{1it}) \leq 0 \end{aligned}$$

Consequently, the reduced form of the system becomes

$$y^*_{1it} = \beta'_{11}x_{1it} + \beta'_{12}x_{it} + \alpha_{1i} + \epsilon_{1it} \tag{3.8}$$

$$y^*_{2it} = \Pi'_2 X_{it} + \delta_{2i} + \nu_{2it} \leq 0 \tag{3.9}$$

The distribution of these variables is the same as before, so now we only need to derive the covariance between the error components.

$$\mathrm{E}(\alpha_{1i}\delta_{2j}) = \sigma^{\alpha,\delta}_{12} = \begin{cases} \sigma^{\alpha}_{12} + \gamma_2\sigma^{\alpha}_{11} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{E}(\epsilon_{1it}\nu_{2js}) = \sigma^{\epsilon,\nu}_{12} = \begin{cases} \sigma^{\epsilon}_{12} + \gamma_2\sigma^{\epsilon}_{11} & \text{if } i = j,\ t = s \\ 0 & \text{otherwise} \end{cases}$$

Summarizing,

$$\begin{pmatrix} \alpha_{1i} \\ \delta_{2i} \\ \epsilon_{1it} \\ \nu_{2it} \end{pmatrix} \sim \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{\alpha}_{11} & \sigma^{\alpha,\delta}_{12} & 0 & 0 \\ \sigma^{\alpha,\delta}_{12} & \sigma^{\delta}_{22} & 0 & 0 \\ 0 & 0 & \sigma^{\epsilon}_{11} & \sigma^{\epsilon,\nu}_{12} \\ 0 & 0 & \sigma^{\epsilon,\nu}_{12} & \sigma^{\nu}_{22} \end{pmatrix} \right]$$

In this case, the joint distribution of the endogenous variables is

$$\begin{pmatrix} y_{1it}^* \\ y_{2it}^* \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} \beta_{11}' x_{1it} + \beta_{12}' x_{it} \\ \Pi_2' X_{it} \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon} & \sigma_{12}^{\alpha,\delta} + \sigma_{12}^{\epsilon,\nu} \\ \sigma_{12}^{\alpha,\delta} + \sigma_{12}^{\epsilon,\nu} & \sigma_{22}^{\delta} + \sigma_{22}^{\nu} \end{pmatrix}\right]$$

and the correlation coefficient between the two variables can be written as

$$\rho_{II} = \frac{\sigma_{12}^{\alpha,\delta} + \sigma_{12}^{\epsilon,\nu}}{\sqrt{(\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon})(\sigma_{22}^{\delta} + \sigma_{22}^{\nu})}}$$

Case III: $y_{1it} = 0$ **and** $y_{2it} > 0$

This time the censored endogenous variable is the first one, so the situation is the opposite to the previous case. Since the simultaneous term drops from (3.5), this equation becomes its own reduced form. For the first variable,

$$\begin{aligned} y_{1it}^* &= \gamma_1 \left(\beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it}\right) + \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} + \epsilon_{1it} \\ &= \begin{pmatrix} \beta_{11} \\ \gamma_1 \beta_{21} \\ \beta_{12} + \gamma_1 \beta_{22} \end{pmatrix}' \begin{pmatrix} x_{1it} \\ x_{2it} \\ x_{it} \end{pmatrix} + (\alpha_{1i} + \gamma_1 \alpha_{2i}) + (\epsilon_{1it} + \gamma_1 \epsilon_{2it}) \leq 0 \end{aligned}$$

In other words, we can express the system as

$$y_{1it}^* = \Pi_1' X_{it} + \delta_{1i} + \nu_{1it} \leq 0 \tag{3.10}$$

$$y_{2it}^* = \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it} \tag{3.11}$$

As in case II, the statistical properties of the error components remain unchanged and it is only needed to derive the covariance between them.

$$\mathrm{E}(\delta_{1i}\alpha_{2j}) = \sigma_{12}^{\delta,\alpha} = \begin{cases} \sigma_{12}^{\alpha} + \gamma_1 \sigma_{22}^{\alpha} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{E}(\nu_{1it}\epsilon_{2js}) = \sigma_{12}^{\nu,\epsilon} = \begin{cases} \sigma_{12}^{\epsilon} + \gamma_1 \sigma_{22}^{\epsilon} & \text{if } i = j,\, t = s \\ 0 & \text{otherwise} \end{cases}$$

As a result, the distributions can be expressed in the following way

$$\begin{pmatrix} \delta_{1i} \\ \alpha_{2i} \\ \nu_{1it} \\ \epsilon_{2it} \end{pmatrix} \sim \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\delta} & \sigma_{12}^{\delta,\alpha} & 0 & 0 \\ \sigma_{12}^{\delta,\alpha} & \sigma_{22}^{\alpha} & 0 & 0 \\ 0 & 0 & \sigma_{11}^{\nu} & \sigma_{12}^{\nu,\epsilon} \\ 0 & 0 & \sigma_{12}^{\nu,\epsilon} & \sigma_{22}^{\epsilon} \end{pmatrix} \right]$$

Finally, the joint distribution of the endogenous variables becomes

$$\begin{pmatrix} y_{1it}^* \\ y_{2it}^* \end{pmatrix} \sim \mathbb{N} \left[\begin{pmatrix} \Pi_1' X_{it} \\ \beta_{21}' x_{2it} + \beta_{22}' x_{it} \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\delta} + \sigma_{11}^{\nu} & \sigma_{12}^{\delta,\alpha} + \sigma_{12}^{\nu,\epsilon} \\ \sigma_{12}^{\delta,\alpha} + \sigma_{12}^{\nu,\epsilon} & \sigma_{22}^{\alpha} + \sigma_{22}^{\epsilon} \end{pmatrix} \right]$$

and the correlation coefficient between them

$$\rho_{III} = \frac{\sigma_{12}^{\delta,\alpha} + \sigma_{12}^{\nu,\epsilon}}{\sqrt{(\sigma_{11}^{\delta} + \sigma_{11}^{\nu})(\sigma_{22}^{\alpha} + \sigma_{22}^{\epsilon})}}$$

Case IV: $y_{1it} = 0$ **and** $y_{2it} = 0$

The final case, when both variables are censored, is also the simplest. Since the simultaneous terms drop from both equations, (3.4) and (3.5) become their own reduced forms. In other words, the model can be written as the following censored SUR model.

$$y^*_{1it} = \beta'_{11}x_{1it} + \beta'_{12}x_{it} + \alpha_{1i} + \epsilon_{1it} \leq 0 \tag{3.12}$$

$$y^*_{2it} = \beta'_{21}x_{2it} + \beta'_{22}x_{it} + \alpha_{2i} + \epsilon_{2it} \leq 0 \tag{3.13}$$

The statistical properties of the error components are described in assumption (i) above. By using these properties, it can be concluded that the joint density of the endogenous variables is given by

$$\begin{pmatrix} y^*_{1it} \\ y^*_{2it} \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} \beta'_{11}x_{1it} + \beta'_{12}x_{it} \\ \beta'_{21}x_{2it} + \beta'_{22}x_{it} \end{pmatrix}, \begin{pmatrix} \sigma^\alpha_{11} + \sigma^\epsilon_{11} & \sigma^\alpha_{12} + \sigma^\epsilon_{12} \\ \sigma^\alpha_{12} + \sigma^\epsilon_{12} & \sigma^\alpha_{22} + \sigma^\epsilon_{22} \end{pmatrix}\right]$$

The correlation coefficient between them is given by the following expression

$$\rho_{IV} = \frac{\sigma^\alpha_{12} + \sigma^\epsilon_{12}}{\sqrt{(\sigma^\alpha_{11} + \sigma^\epsilon_{11})(\sigma^\alpha_{22} + \sigma^\epsilon_{22})}}$$

3.1.2 Constructing the likelihood function

The likelihood function for this model presents a number of difficulties. First of all, because of the limited character of the endogenous variables, the function is not continuous: since the variables are censored at a certain value, their distributions become truncated multivariate normal densities. This will introduce integrals into the likelihood and it will complicate its calculation. Second, each observation can belong to one of the four cases described above. As a result, there are four possible likelihood functions for each point, a further source of discontinuity. Third, two observations from the same individual i are correlated through the error component α_i, thus invalidating the independence property commonly assumed in linear models.

In order to explicit the likelihood function, it is necessary to take into account that each observation belongs to one of the four cases listed above, depending on the censored character of its exogenous variables. Because observations from the same individual are correlated through the individual error component, they cannot be considered independently. However, one can notice that, conditioned on the individual effects, the observations become independent. This will be studied case by case.

In equations (3.6) and (3.7) from the first case, the variables

$$\nu_{1it}|\delta_{1i} = \Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}$$
$$\nu_{2it}|\delta_{2i} = \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}$$

are independent across both individuals and time periods. Let $f^{\mathrm{I}}(\cdot,\cdot)$ be the density function of the error terms for an observation from case I; that is, when both variables are greater than zero. In this situation,

$$\begin{aligned} f^{\mathrm{I}}(\nu_{1it},\nu_{2it}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{\mathrm{I}}(\nu_{1it},\nu_{2it},\delta_{1i},\delta_{2i})\,\mathrm{d}\delta_{1i}\,\mathrm{d}\delta_{2i} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{\mathrm{I}}(\nu_{1it},\nu_{2it}|\delta_{1i},\delta_{2i}) g^{\mathrm{I}}(\delta_{1i},\delta_{2i})\,\mathrm{d}\delta_{1i}\,\mathrm{d}\delta_{2i} \end{aligned}$$

where $g^{\mathrm{I}}(\delta_{1i},\delta_{2i})$ is the marginal distribution of the individual error components from Case I. From assumption (i) above, the error terms (ν_{1it},ν_{2it}) are independent across individuals and time periods.

For an observation from case II, from (3.8) and (3.9) the conditional error terms can be written in the following manner.

$$\epsilon_{1it}|\alpha_{1i} = y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}$$
$$\nu_{2it}|\delta_{2i} \leq -\Pi_2' X_{it} - \delta_{2i}$$

In this case, the second dependent variable is censored. The observed variable y_{2it} equals 0, implying that the latent variable y_{2it}^* is less than or equal to zero. This explains the inequality sign in the second error term. Because of this, the distribution of the error terms becomes

$$f^{\mathrm{II}}(\epsilon_{1it}, \nu_{2it} \leq -\Pi_2' X_{it} - \delta_{2i}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}(\epsilon_{1it}, \nu_{2it}|\alpha_{1i}, \delta_{2i})\,\mathrm{d}\nu_{2it}\right) \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i})\,\mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i}$$

where $g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i})$ is the marginal distribution of the individual error components for an observation from Case II. The error terms $(\epsilon_{1it}, \nu_{2it})$ are independent for all i and t.

In a similar way, one of the error terms from an observation from Case III will be an inequality. The next expressions can be obtained from equations (3.10) and (3.11).

$$\nu_{1it}|\delta_{1i} \leq -\Pi_1' x_{it} - \delta_{1i}$$

$$\epsilon_{2it}|\alpha_{2i} = y_{2it} - \beta_2' x_{2it} - \alpha_{2i}$$

Conditioned on the individual effects, the two variables are independent for all individuals and time periods. Their marginal density function can be written as follows.

$$f^{\mathrm{III}}(\nu_{1it} \leq -\Pi_1' X_{it} - \delta_{1i}, \epsilon_{2it}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, \epsilon_{2it}|\delta_{1i}, \alpha_{2i})\,\mathrm{d}\nu_{1it}\right) \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i})\,\mathrm{d}\delta_{1i}\,\mathrm{d}\alpha_{2i}$$

where $g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i})$ is the marginal distribution of the individual error components from this case.

In the final case, both (3.12) and (3.13) are censored, so the error terms can be expressed in the form of the following inequalities.

$$\epsilon_{1it}|\alpha_{1i} \leq -\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}$$

$$\epsilon_{2it}|\alpha_{2i} \leq -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}$$

As a consequence, the joint density of the error terms becomes a cumulative bivariate normal distribution.

$$f^{\mathrm{IV}}(\epsilon_{1it} \leq -\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \epsilon_{2it} \leq -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i})$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{-\beta_{12}' x_{1it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{2it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it} \right)$$
$$\cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}$$

where, as before, g^{IV} is the marginal distribution of the individual error components for an observation from Case IV.

All the $f(\cdot)$ and $g(\cdot)$ functions are bivariate normal distributions with the parameters given above for the different cases. With all these elements and taken into account that two observations from the same individual become independent when the error terms are conditioned on the individual effects, it is now possible to elicit the likelihood function. In order to write the function more compactly, define the following four sets.

$\mathbb{D}_i^{\mathrm{I}}$: observations from individual i that belong to case I

$\mathbb{D}_i^{\mathrm{II}}$: observations from individual i that belong to case II

$\mathbb{D}_i^{\mathrm{III}}$: observations from individual i that belong to case III

$\mathbb{D}_i^{\mathrm{IV}}$: observations from individual i that belong to case IV

The likelihood function for individual i can then be written as follows.

$$
\begin{aligned}
L_i = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i} | \delta_{1i}, \delta_{2i}) \\
& \quad \cdot g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}\left(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it} | \alpha_{1i}, \delta_{2i}\right) \, \mathrm{d}\nu_{2it} \\
& \quad \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i} | \delta_{1i}, \alpha_{2i}\right) \, \mathrm{d}\nu_{1it} \\
& \quad \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \, \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it} \\
& \quad \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}
\tag{3.14}
$$

Since the observations from different individuals are independent, they can be aggregated directly. Consequently, the complete likelihood function can be written as

$$
\begin{aligned}
L = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{I}}} & f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i} | \delta_{1i}, \delta_{2i}) \\
& \cdot g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{II}}} & \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \, \mathrm{d}\nu_{2it} \\
& \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{III}}} & \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i} | \delta_{1i}, \alpha_{2i} \right) \, \mathrm{d}\nu_{1it} \\
& \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{IV}}} & \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i} \right) \, \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it} \\
& \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

and its logarithm by

$$
\begin{aligned}
\log L = \sum_{i=1}^{N} \Bigg\{ \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{I}}} & f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i} | \delta_{1i}, \delta_{2i}) \\
& \cdot g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
+ \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{II}}} & \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \, \mathrm{d}\nu_{2it} \\
& \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
+ \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{III}}} & \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i} | \delta_{1i}, \alpha_{2i} \right) \, \mathrm{d}\nu_{1it} \\
& \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
+ \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{IV}}} & \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i} \right) \, \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it} \\
& \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \Bigg\}
\end{aligned}
\tag{3.15}
$$

Maximizing this function with respect to the unknown parameters will produce consistent and asymptotically normally distributed estimates when either N or T tend to infinity (see Hsiao (2003), chapter 8). However, the likelihood presents multiple integrals, discontinuities and probably multiple modes because it is formed by four parts that correspond to the four possible cases described before. Consequently, it is not possible to find an analytical solution to the likelihood maximization problem and traditional optimization methods may not be able to converge and find an adequate solution. Chapter 4 will describe the different methods that were attempted to optimize equation (3.15) and the one that was finally chosen.

3.1.3 Gradient vector

The gradient of a (log) likelihood function is a column vector that describes its shape with respect to infinitesimal changes in the parameters. It is an essential ingredient of the most widely used optimization method in econometrics: the Newton method and its variants. These gradient-based methods will be described in the following chapter. Here we will calculate the gradient for the logarithm of the likelihood (3.15).

$$\bigtriangledown \log L = \Big[\frac{\partial \log L}{\partial \beta_{11}} \quad \frac{\partial \log L}{\partial \beta_{12}} \quad \frac{\partial \log L}{\partial \beta_{21}} \quad \frac{\partial \log L}{\partial \beta_{22}} \quad \frac{\partial \log L}{\partial \gamma_{1}} \quad \frac{\partial \log L}{\partial \gamma_{2}} \\ \frac{\partial \log L}{\partial \sigma^{\epsilon}_{11}} \quad \frac{\partial \log L}{\partial \sigma^{\epsilon}_{12}} \quad \frac{\partial \log L}{\partial \sigma^{\epsilon}_{22}} \quad \frac{\partial \log L}{\partial \sigma^{\alpha}_{11}} \quad \frac{\partial \log L}{\partial \sigma^{\alpha}_{12}} \quad \frac{\partial \log L}{\partial \sigma^{\alpha}_{22}} \Big] \tag{3.16}$$

Since the calculations are very lengthy, they are left for Appendix B; the vector above can be filled with equations (B.1)-(B.12). Note that in equations (3.4b) and (3.5b) the exogenous variables, and hence the corresponding coefficients, may be vector quantities. However, the expression above treats them as scalars. This does not imply a loss of generality: in case one or more exogenous variables have a dimension greater than one, we would substitute the corresponding derivative in the gradient by the derivatives with respect to each element of the vector coefficient in question.

The calculation of the analytical Hessian matrix could be obtained by taking the derivative of the gradient vector (3.16) with respect to all the parameters in the model (in other words, the second and cross derivatives of the log likelihood with respect to all the parameters). However, the calculations are lengthy and its applicability limited due to its complicated form. Indeed, it would be composed of numerous products and sums of single or double integrals that would be extremely costly to compute. For this reason, we do not pursue this calculation in this document.

3.1.4 Marginal effects of continuous exogenous variables

Since this is a nonlinear model, the coefficients estimated through maximization of (3.15) are not meaningful for a direct interpretation. Instead, the marginal effects of the exogenous regressors on the expectation of the observed endogenous variables should be examined. For example, if after estimation of equation (3.4b) one is interested in measuring the effect on y_{1it} of a change in the exogenous regressor x_{1it}, it would be misguided to consider the coefficient vector β_{11} since this coefficient measures the effect of x_{1it} on the *latent* variable y^*_{1it} and not on its *observed* counterpart. Instead, one should calculate the marginal effect on the expected value of the observed endogenous variable; i.e.,

$$\frac{\partial \mathrm{E}(y_{1it})}{\partial x_{1it}}$$

This is called the full marginal effect of x_{1it} on y_{1it}. This section presents the derivation of these marginal effects for both endogenous variables. As in the derivation of the gradient in the preceding section, note that the exogenous variables may be vector quantities. Without loss of generality, here we will assume that they are scalars; in case one or more of them have a dimension greater than one, it would suffice to calculate the derivative with respect to each of the elements of the vector variable in question.

We will start with the first endogenous variable. The first step is the calculation of the expected value $\mathrm{E}(y_{1it})$. Following the four cases outlined above, this expectation can then be written as follows.

$$\begin{aligned}\mathrm{E}(y_{1it}) = {} & \mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}>0)\mathrm{P}(y_{1it}>0, y_{2it}>0)\\ & + \mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}=0)\mathrm{P}(y_{1it}>0, y_{2it}=0)\\ & + \mathrm{E}(y_{1it}|y_{1it}=0, y_{2it}>0)\mathrm{P}(y_{1it}=0, y_{2it}>0)\\ & + \mathrm{E}(y_{1it}|y_{1it}=0, y_{2it}=0)\mathrm{P}(y_{1it}=0, y_{2it}=0)\end{aligned} \tag{3.17}$$

The last two summands drop since the conditional expected value of y_{1it} is equal to zero in those cases. Using the properties of the truncated normal distribution detailed in Appendix A and in particular results (A.5) and (A.6), the components of the expectation (3.17) are the following, according to the results from Case I and Case II described before.

$$\begin{aligned}\mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}>0) = {} & \frac{1}{\Gamma}\Pi_1' X_{it} + \frac{\frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}{F_{it}^{I}}\cdot\\ & \left\{\phi\left(\frac{-\Pi_1' X_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right)\left[1-\Phi\left(\frac{\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_I\frac{-\Pi_1' X_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right]\right.\\ & \left. + \rho_I\phi\left(\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\left[1-\Phi\left(\frac{\frac{-\Pi_1' X_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_I\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right]\right\}\end{aligned} \tag{3.18}$$

$$\begin{aligned}\mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}=0) = {} & \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \frac{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}{F_{it}^{II}}\cdot\\ & \left\{\phi\left(-\frac{\beta_{11}' x_{1it}+\beta_{12}' x_{it}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}\right)\Phi\left(\frac{\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_{II}\frac{-(\beta_{11}' x_{1it}+\beta_{12}' x_{it})}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}}{\sqrt{1-\rho_{II}^2}}\right)\right.\\ & \left. - \rho_{II}\phi\left(\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\left[1-\Phi\left(\frac{\frac{-(\beta_{11}' x_{1it}+\beta_{12}' x_{it})}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}-\rho_{II}\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right\}\end{aligned} \tag{3.19}$$

with $F_{it}^{I} = \mathrm{P}(y_{1it} > 0, y_{2it} > 0)$ and $F_{it}^{II} = \mathrm{P}(y_{1it} > 0, y_{2it} = 0)$. Using these results, the full expected value becomes

$$
\begin{aligned}
\mathrm{E}(y_{1it}) = & \left(\frac{1}{\Gamma}\Pi_1' X_{it}\right) F_{it}^{I} + \frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}} \cdot \\
& \left\{ \phi\left(\frac{-\Pi_1' X_{it}}{\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}}}\right) \left[1 - \Phi\left(\frac{\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}} - \rho_I \frac{-\Pi_1' X_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right] \right. \\
& \left. + \rho_I \phi\left(\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}}\right) \left[1 - \Phi\left(\frac{\frac{-\Pi_1' X_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}} - \rho_I \frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right] \right\} \\
& + \left(\beta_{11}' x_{1it} + \beta_{12}' x_{it}\right) F_{it}^{II} + \sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}} \cdot \\
& \left\{ \phi\left(-\frac{\beta_{11}' x_{1it} + \beta_{12}' x_{it}}{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}}\right) \Phi\left(\frac{\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}} - \rho_{II} \frac{-(\beta_{11}' x_{1it}+\beta_{12}' x_{it})}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}}{\sqrt{1-\rho_{II}^2}}\right) \right. \\
& \left. - \rho_{II} \phi\left(\frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}}\right) \left[1 - \Phi\left(\frac{\frac{-(\beta_{11}' x_{1it}+\beta_{12}' x_{it})}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}} - \rho_{II} \frac{-\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_{II}^2}}\right)\right] \right\}
\end{aligned}
$$

This expression is a weighted average of the expected values for the two cases in which the variable y_{1it} is not censored (cases I and II) plus correction factors arising from the simultaneity in the system. In order to abbreviate the preceding expressions, let

$$
\begin{aligned}
a_{I1} &= -\frac{\Pi_1' X_{it}}{\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}}} \\
a_{I2} &= -\frac{\Pi_2' X_{it}}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}} \\
a_{II1} &= -\frac{\beta_{11}' x_{1it} + \beta_{12}' x_{it}}{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}} \\
a_{III2} &= -\frac{\beta_{21}' x_{2it} + \beta_{22}' x_{it}}{\sqrt{\sigma_{22}^{\alpha} + \sigma_{22}^{\epsilon}}}
\end{aligned}
$$

With these values, we can write the previous expression as

$$\begin{aligned}
\mathrm{E}(y_{1it}) = &\left(\frac{1}{\Gamma}\Pi_1' X_{it}\right) F_{it}^{I} + \frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}\left\{\phi(a_{I1})\left[1-\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\
&\left. + \rho_I\phi(a_{I2})\left[1-\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\
&+ \left(\beta_{11}' x_{1it} + \beta_{12}' x_{it}\right) F_{it}^{II} + \sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}\left\{\phi(a_{II1})\Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)\right. \\
&\left. - \rho_{II}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right\}
\end{aligned} \tag{3.20}$$

Taking derivatives of (3.20) with respect to each of the exogenous regressors gives the marginal effects.

For the first regressor x_{1it} we have the following marginal effect.

$$\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{1it}} = &\frac{1}{\Gamma}\frac{\partial(\Pi_1' X_{it})}{\partial x_{1it}} F_{it}^{I} + \left(\frac{1}{\Gamma}\Pi_1' X_{it}\right)\frac{\partial F_{it}^{I}}{\partial x_{1it}} \\
&+ \frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}\left\{\frac{\partial\phi(a_{I1})}{\partial x_{1it}}\left[1-\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right] - \phi(a_{I1})\left(\frac{\partial\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{1it}}\right)\right. \\
&\left. + \rho_I\left[\frac{\partial\phi(a_{I2})}{\partial x_{1it}}\left[1-\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right] - \phi(a_{I2})\left(\frac{\partial\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{1it}}\right)\right]\right\} \\
&+ \frac{\partial(\beta_{11}' x_{1it} + \beta_{12}' x_{it})}{\partial x_{1it}} F_{it}^{II} + \left(\beta_{11}' x_{1it} + \beta_{12}' x_{it}\right)\frac{\partial F_{it}^{II}}{\partial x_{1it}} \\
&+ \sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}\left\{\frac{\partial\phi(a_{II1})}{\partial x_{1it}}\Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right) + \phi(a_{II1})\left(\frac{\partial\Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{1it}}\right)\right. \\
&\left. - \rho_{II}\left[\frac{\partial\phi(a_{I2})}{\partial x_{1it}}\left[1-\Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right] - \phi(a_{I2})\left(\frac{\partial\Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{1it}}\right)\right]\right\}
\end{aligned}$$

We will calculate the derivatives that make up the previous expression separately. For this, we use the results (A.7) and (A.8) of Appendix A.

- $\dfrac{\partial F_{it}^{I}}{\partial x_{1it}} = \dfrac{\beta_{11}}{\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}}}\phi(a_{I1})\left[1 - \Phi\left(\dfrac{a_{I2} - \rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]$
- $\dfrac{\partial(\Pi_I' X_{it})}{\partial x_{1it}} = \beta_{11}$
- $\dfrac{\partial \phi(a_{I1})}{\partial x_{1it}} = a_{I1}\phi(a_{I1})\dfrac{\beta_{11}}{\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}}}$
- $\dfrac{\partial \Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{1it}} = -\dfrac{\beta_{11}}{\sqrt{1-\rho_I^2}}\phi\left(\dfrac{a_{I2} - \rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\left(\dfrac{\gamma_2}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}} - \dfrac{\rho_I}{\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}}}\right)$
- $\dfrac{\partial \phi(a_{I2})}{\partial x_{1it}} = a_{I2}\phi(a_{I2})\dfrac{\gamma_2\beta_{11}}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}}$
- $\dfrac{\partial \Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{1it}} = -\dfrac{\beta_{11}}{\sqrt{1-\rho_I^2}}\phi\left(\dfrac{a_{I1} - \rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\left(\dfrac{1}{\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}}} - \dfrac{\rho_I\gamma_2}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}}\right)$
- $\dfrac{\partial F_{it}^{II}}{\partial x_{1it}} = \dfrac{\beta_{11}}{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}}\phi(a_{II1})\Phi\left(\dfrac{a_{I2} - \rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)$
- $\dfrac{\partial(\beta_{11}' x_{1it} + \beta_{12}' x_{it})}{\partial x_{1it}} = \beta_{11}$
- $\dfrac{\partial \phi(a_{II1})}{\partial x_{1it}} = a_{II1}\phi(a_{II1})\dfrac{\beta_{11}}{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}}$
- $\dfrac{\partial \Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{1it}} = -\dfrac{\beta_{11}}{\sqrt{1-\rho_{II}^2}}\phi\left(\dfrac{a_{I2} - \rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)\left(\dfrac{\gamma_2}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}} - \dfrac{\rho_{II}}{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}}\right)$
- $\dfrac{\partial \Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{1it}} = -\dfrac{\beta_{11}}{\sqrt{1-\rho_{II}^2}}\phi\left(\dfrac{a_{II1} - \rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\left(\dfrac{1}{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}} - \dfrac{\rho_{II}\gamma_2}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}}\right)$

Putting all these elements together, the marginal effect becomes

$$\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{1it}} &= \frac{1}{\Gamma}\beta_{11} F_{it}^{I} + \left(\frac{1}{\Gamma}\Pi_1' X_{it}\right) \frac{\beta_{11}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right] \\
&+ \frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}\left\{\frac{\beta_{11}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}} a_{I1}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\
&+ \frac{\beta_{11}}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\left(\frac{\gamma_2}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\frac{\rho_I}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right) \\
&+ \rho_I \frac{\gamma_2\beta_{11}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}} a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right] \\
&\left.+\rho_I \frac{\beta_{11}}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\left(\frac{1}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\frac{\rho_I\gamma_2}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\right\} \\
&+ \beta_{11} F_{it}^{II} + \left(\beta_{11}' x_{1it} + \beta_{12}' x_{it}\right)\frac{\beta_{11}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}\phi(a_{II1})\Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right) \\
&+ \sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}\left\{\frac{\beta_{11}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}} a_{II1}\phi(a_{II1})\Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)\right. \\
&- \frac{\beta_{11}}{\sqrt{1-\rho_{II}^2}}\phi(a_{II1})\phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)\left(\frac{\gamma_2}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\frac{\rho_{II}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}\right) \\
&- \rho_{II}\frac{\gamma_2\beta_{11}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}} a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right] \\
&\left.-\rho_{II}\frac{\beta_{11}}{\sqrt{1-\rho_{II}^2}}\phi(a_{I2})\phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\left(\frac{1}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}-\frac{\rho_{II}\gamma_2}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\right\}
\end{aligned}$$

Rearranging terms,

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{1it}} = \frac{1}{\Gamma}\beta_{11} & \left\{ F_{it}^{I} + \frac{\rho_I}{\sqrt{1-\rho_I^2}} \left[\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right) - \phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\
& + \frac{1}{\Gamma}\gamma_2\beta_{11}\frac{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\left\{\rho_I a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\
& \quad \left. + \frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right) - \frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right\} \\
& + \beta_{11}\left\{F_{it}^{II} - \frac{\rho_{II}}{\sqrt{1-\rho_{II}^2}}\left[\phi(a_{I2})\phi\left(\frac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right) - \phi(a_{II1})\phi\left(\frac{a_{I2}-\rho_{II}a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right\} \\
& - \gamma_2\beta_{11}\frac{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\left\{\rho_{II}a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right. \\
& \quad \left. + \frac{1}{\sqrt{1-\rho_{II}^2}}\phi(a_{II1})\phi\left(\frac{a_{I2}-\rho_{II}a_{II1}}{\sqrt{1-\rho_{II}^2}}\right) - \frac{\rho_{II}^2}{\sqrt{1-\rho_{II}^2}}\phi(a_{I2})\phi\left(\frac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right\}
\end{aligned}
\tag{3.21}
$$

Even though this expression may seem complicated, in fact it is only a combination of the corrections due to the simultaneity and the censorship of the data. The different terms account for the expected effect of x_{1it} on y_{1it} directly or indirectly through the simultaneous structure of the system, taking into consideration that the variables may be censored.

In case there is no censorship in the first dependent variable (i.e., all data for this variable is observed), then both a_{I1} and a_{II1} would approach minus infinity. This occurs because the threshold c introduced in Appendix A would tend to minus infinity in order to have a normal variable observed over all its domain. In this case, the marginal effect (3.21) reduces to

$$\frac{\partial \mathrm{E}(y_{1it}|\text{no censorship in } y_1)}{\partial x_{1it}} = \frac{1}{\Gamma}\beta_{11}F_{it}^{I} + \beta_{11}F_{it}^{II}$$
$$+ \frac{\gamma_2}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}a_{I2}\phi(a_{I2})\left[\frac{1}{\Gamma}\beta_{11}\rho_I\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}} - \beta_{11}\rho_{II}\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}\right]$$

This is an average of the marginal effects for Case I ($\frac{1}{\Gamma}\beta_{11}$) and Case II ($\beta_{11}$), weighted by the probability of each case, plus a correction due to the simultaneity on the data. If the second endogenous variable is not censored either, then the probability to be in Case II would be zero ($F_{it}^{II} = 0$) while the probability to be in Case I would equal one ($F_{it}^{I} = 1$). Under this situation the marginal effect would be reduced to

$$\frac{\partial \mathrm{E}(y_{1it}|\text{no censorship in } y_1 \text{ and } y_2)}{\partial x_{1it}} = \frac{1}{\Gamma}\beta_{11}$$

the marginal effect for a simple bivariate SEM. On the other hand, if the model has censored observations but there is no simultaneity in equation (3.4b) (i.e., $\gamma_1 = 0$), then $\Gamma = 1$, $a_{I1} = a_{II1}$ and $\rho_I = \rho_{II}$ so that the marginal effect (3.21) becomes

$$\frac{\partial \mathrm{E}(y_{1it}|\gamma_1 = 0)}{\partial x_{1it}} = \beta_{11}\left(F_{it}^{I} + F_{it}^{II}\right)$$

So the marginal effect reduces to that of the traditional Tobit model when there is no simultaneity term in the first equation. Finally, we have the case in which there is neither censorship nor simultaneity. Starting from either (3.1.4) and setting $\Gamma = 1$ or from (3.1.4) and setting $F_{it}^{I} = 1$ and $F_{it}^{II} = 0$, we obtain that

$$\frac{\partial \mathrm{E}(y_{1it}|\gamma_1 = 0, \text{no censorship in } y_1)}{\partial x_{1it}} = \beta_{11}$$

This is the marginal effect from a simple linear model.

Now we study the marginal effect of x_{2it} on y_{1it}. Intuitively, since this regressor is not directly included in the first equation and it only affects y_{1it} through the simultaneity of the model, its effect should be constrained in that sense. To calculate it we take the derivative of (3.20) with respect to x_{2it} to obtain

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{2it}} &= \frac{1}{\Gamma}\frac{\partial(\Pi_1' X_{it})}{\partial x_{2it}} F_{it}^{I} + \left(\frac{1}{\Gamma}\Pi_1' X_{it}\right)\frac{\partial F_{it}^{I}}{\partial x_{2it}} \\
&+ \frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}} \left\{ \frac{\partial \phi(a_{I1})}{\partial x_{2it}} \left[1 - \Phi\left(\frac{a_{I2} - \rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right] - \phi(a_{I1}) \left(\frac{\partial \Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{2it}}\right)\right. \\
&\left. + \rho_I \left[\frac{\partial \phi(a_{I2})}{\partial x_{2it}} \left[1 - \Phi\left(\frac{a_{I1} - \rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right] - \phi(a_{I2}) \left(\frac{\partial \Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{2it}}\right)\right]\right\} \\
&+ \frac{\partial(\beta_{11}' x_{1it} + \beta_{12}' x_{it})}{\partial x_{2it}} F_{it}^{II} + (\beta_{11}' x_{1it} + \beta_{12}' x_{it}) \frac{\partial F_{it}^{II}}{\partial x_{2it}} \\
&+ \sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}} \left\{ \frac{\partial \phi(a_{II1})}{\partial x_{2it}} \Phi\left(\frac{a_{I2} - \rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right) + \phi(a_{II1}) \left(\frac{\partial \Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{2it}}\right)\right. \\
&\left. - \rho_{II} \left[\frac{\partial \phi(a_{I2})}{\partial x_{2it}} \left[1 - \Phi\left(\frac{a_{II1} - \rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right] - \phi(a_{I2}) \left(\frac{\partial \Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{2it}}\right)\right]\right\}
\end{aligned}
$$

As before, we calculate each of these derivatives separately.

- $\dfrac{\partial F_{it}^{I}}{\partial x_{2it}} = \dfrac{\gamma_1\beta_{21}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\phi(a_{I1})\left[1-\Phi\left(\dfrac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]$
- $\dfrac{\partial(\Pi_I' X_{it})}{\partial x_{2it}} = \gamma_1\beta_{21}$
- $\dfrac{\partial\phi(a_{I1})}{\partial x_{2it}} = a_{I1}\phi(a_{I1})\dfrac{\gamma_1\beta_{21}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}$
- $\dfrac{\partial\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{2it}} = -\dfrac{\beta_{21}}{\sqrt{1-\rho_I^2}}\phi\left(\dfrac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\left(\dfrac{1}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}} - \dfrac{\rho_I\gamma_2}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right)$
- $\dfrac{\partial\phi(a_{I2})}{\partial x_{2it}} = a_{I2}\phi(a_{I2})\dfrac{\beta_{21}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}$
- $\dfrac{\partial\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{2it}} = -\dfrac{\beta_{21}}{\sqrt{1-\rho_I^2}}\phi\left(\dfrac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\left(\dfrac{\gamma_2}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}} - \dfrac{\rho_I}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)$
- $\dfrac{\partial F_{it}^{II}}{\partial x_{2it}} = 0$
- $\dfrac{\partial(\beta_{11}' x_{1it}+\beta_{12}' x_{it})}{\partial x_{2it}} = 0$
- $\dfrac{\partial\phi(a_{II1})}{\partial x_{2it}} = 0$
- $\dfrac{\partial\Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{2it}} = -\dfrac{\beta_{21}}{\sqrt{1-\rho_{II}^2}}\dfrac{1}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}$
- $\dfrac{\partial\Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{2it}} = \dfrac{\rho_{II}\beta_{21}}{\sqrt{1-\rho_{II}^2}}\dfrac{1}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}$

Using these results for the derivatives and rearranging terms, the marginal effect can be written in the following manner.

$$\begin{aligned}\frac{\partial \mathrm{E} y_{1it}}{\partial x_{2it}} = &\frac{1}{\Gamma}\gamma_1\beta_{21}\left\{F_{it}^I + \frac{\rho_I}{\sqrt{1-\rho_I^2}}\left[\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right) - \phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\ &+ \frac{1}{\Gamma}\beta_{21}\frac{\sqrt{\sigma_{11}^\delta+\sigma_{11}^\nu}}{\sqrt{\sigma_{22}^\delta+\sigma_{22}^\nu}}\left\{\rho_I a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\ &\left. + \frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right) - \frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right\} \\ &- \beta_{21}\frac{\sqrt{\sigma_{11}^\alpha+\sigma_{11}^\epsilon}}{\sqrt{\sigma_{22}^\delta+\sigma_{22}^\nu}}\left\{\rho_{II} a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right. \\ &\left. + \frac{1}{\sqrt{1-\rho_{II}^2}}\phi(a_{II1})\phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right) - \frac{\rho_{II}^2}{\sqrt{1-\rho_{II}^2}}\phi(a_{I2})\phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right\}\end{aligned} \tag{3.22}$$

This expression is also a mixture of corrections for both the censorship and the simultaneity in the model.

For a situation with no censored data in y_{1it}, the marginal effect would compress to

$$\begin{aligned}\frac{\partial \mathrm{E}(y_{1it}|\text{no censorship in } y_1)}{\partial x_{2it}} = &\frac{1}{\Gamma}\gamma_1\beta_{21}F_{it}^I \\ &+ \frac{a_{I2}\phi(a_{I2})}{\sqrt{\sigma_{22}^\delta+\sigma_{22}^\nu}}\left[\frac{1}{\Gamma}\beta_{21}\rho_I\sqrt{\sigma_{11}^\delta+\sigma_{11}^\nu} - \beta_{21}\rho_{II}\sqrt{\sigma_{11}^\alpha+\sigma_{11}^\epsilon}\right]\end{aligned}$$

If, in addition, the second endogenous variable does not have any censored observations either, it can be seen that the marginal effect further reduces to the marginal effect from a SEM.

$$\frac{\partial \mathrm{E}(y_{1it}|\text{no censorship in } y_1 \text{ and } y_2)}{\partial x_{2it}} = \frac{1}{\Gamma}\gamma_1\beta_{21}$$

On the other hand, when there is no simultaneity in the first equation (3.4b) (i.e., $\gamma_1 = 0$), the marginal effect equals to zero. This can be seen in the following development.

$$\begin{aligned}\frac{\partial \mathrm{E}(y_{1it}|\gamma_1 = 0)}{\partial x_{2it}} &= 0 + \beta_{21}\frac{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}}\left\{\rho_I a_{I2}\phi(a_{I2})\left[1 - \Phi\left(\frac{a_{I1} - \rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right.\\
&\quad \left. + \frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2} - \rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right) - \frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1} - \rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right\}\\
&\quad - \beta_{21}\frac{\sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}}{\sqrt{\sigma_{22}^{\delta} + \sigma_{22}^{\nu}}}\left\{\rho_I a_{I2}\phi(a_{I2})\left[1 - \Phi\left(\frac{a_{I1} - \rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right.\\
&\quad \left. + \frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2} - \rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right) - \frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1} - \rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right\} = 0\end{aligned}$$

So, as expected, x_{2it} has no impact on y_{1it} if there is no simultaneity included in the equation for this endogenous variable. Evidently, if there is neither censorship nor simultaneity in the model, the marginal effect of x_{2it} on y_{1it} would also be null.

Finally, we will calculate the marginal effect on y_{1it} of the variable present in both equations, x_{it}. For this purpose, we take the derivative of (3.20) with respect to this variable.

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{it}} &= \frac{1}{\Gamma}\frac{\partial(\Pi_1' X_{it})}{\partial x_{it}} F_{it}^{I} + \left(\frac{1}{\Gamma}\Pi_1' X_{it}\right)\frac{\partial F_{it}^{I}}{\partial x_{it}} \\
&\quad + \frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta} + \sigma_{11}^{\nu}}\left\{\frac{\partial \phi(a_{I1})}{\partial x_{it}}\left[1 - \Phi\left(\frac{a_{I2} - \rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right] - \phi(a_{I1})\left(\frac{\partial \Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{it}}\right)\right. \\
&\qquad \left. +\ \rho_I\left[\frac{\partial \phi(a_{I2})}{\partial x_{it}}\left[1 - \Phi\left(\frac{a_{I1} - \rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right] - \phi(a_{I2})\left(\frac{\partial \Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{it}}\right)\right]\right\} \\
&\quad + \frac{\partial(\beta_{11}' x_{1it} + \beta_{12}' x_{it})}{\partial x_{it}} F_{it}^{II} + (\beta_{11}' x_{1it} + \beta_{12}' x_{it})\frac{\partial F_{it}^{II}}{\partial x_{it}} \\
&\quad + \sqrt{\sigma_{11}^{\alpha} + \sigma_{11}^{\epsilon}}\left\{\frac{\partial \phi(a_{II1})}{\partial x_{it}}\Phi\left(\frac{a_{I2} - \rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right) + \phi(a_{II1})\left(\frac{\partial \Phi\left(\frac{a_{I2}-\rho_{II} a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{it}}\right)\right. \\
&\qquad \left. -\ \rho_{II}\left[\frac{\partial \phi(a_{I2})}{\partial x_{it}}\left[1 - \Phi\left(\frac{a_{II1} - \rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right] - \phi(a_{I2})\left(\frac{\partial \Phi\left(\frac{a_{II1}-\rho_{II} a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{it}}\right)\right]\right\}
\end{aligned}
$$

The various elements that compose this derivative are the following.

- $\dfrac{\partial F_{it}^{I}}{\partial x_{it}} = \dfrac{\beta_{12}+\gamma_1\beta_{22}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\phi(a_{I1})\left[1-\Phi\left(\dfrac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]$
- $\dfrac{\partial(\Pi_I' X_{it})}{\partial x_{it}} = \beta_{12}+\gamma_1\beta_{22}$
- $\dfrac{\partial\phi(a_{I1})}{\partial x_{it}} = \dfrac{\beta_{12}+\gamma_1\beta_{22}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}a_{I1}\phi(a_{I1})$
- $\dfrac{\partial\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{it}} = -\dfrac{1}{\sqrt{1-\rho_I^2}}\phi\left(\dfrac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\left(\dfrac{\gamma_2\beta_{12}+\beta_{22}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_I\dfrac{\beta_{12}+\gamma_1\beta_{22}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right)$
- $\dfrac{\partial\phi(a_{I2})}{\partial x_{it}} = \dfrac{\gamma_2\beta_{12}+\beta_{22}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}a_{I2}\phi(a_{I2})$
- $\dfrac{\partial\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)}{\partial x_{it}} = -\dfrac{1}{\sqrt{1-\rho_I^2}}\phi\left(\dfrac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\left(\dfrac{\beta_{12}+\gamma_1\beta_{22}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_I\dfrac{\gamma_2\beta_{12}+\beta_{22}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)$
- $\dfrac{\partial F_{it}^{II}}{\partial x_{it}} = \dfrac{\beta_{12}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}\phi(a_{II1})\Phi\left(\dfrac{a_{I2}-\rho_{II}a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)$
- $\dfrac{\partial(\beta_{11}' x_{1it}+\beta_{12}' x_{it})}{\partial x_{it}} = \beta_{12}$
- $\dfrac{\partial\phi(a_{II1})}{\partial x_{it}} = \dfrac{\beta_{12}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}a_{II1}\phi(a_{II1})$
- $\dfrac{\partial\Phi\left(\frac{a_{I2}-\rho_{II}a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{it}} = -\dfrac{1}{\sqrt{1-\rho_{II}^2}}\phi\left(\dfrac{a_{I2}-\rho_{II}a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)\left(\dfrac{\gamma_2\beta_{12}+\beta_{22}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_{II}\dfrac{\beta_{12}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}\right)$
- $\dfrac{\partial\Phi\left(\frac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)}{\partial x_{it}} = -\dfrac{1}{\sqrt{1-\rho_{II}^2}}\phi\left(\dfrac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\left(\dfrac{\beta_{12}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}-\rho_{II}\dfrac{\gamma_2\beta_{12}+\beta_{22}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)$

Inserting these expressions and rearranging terms, the marginal effect becomes

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{it}} &= \frac{1}{\Gamma}(\beta_{12}+\gamma_1\beta_{22})\left\{F_{it}^{I}+\frac{\rho_I}{\sqrt{1-\rho_I^2}}\left[\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)-\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\
&+\frac{1}{\Gamma}(\gamma_2\beta_{12}+\beta_{22})\frac{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\left\{\rho_I a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\
&\left.+\frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)-\frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right\} \\
&+\beta_{12}\left\{F_{it}^{II}-\frac{\rho_{II}}{\sqrt{1-\rho_{II}^2}}\left[\phi(a_{I2})\phi\left(\frac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)-\phi(a_{II1})\phi\left(\frac{a_{I2}-\rho_{II}a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right\} \\
&-(\gamma_2\beta_{12}+\beta_{22})\frac{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\left\{\rho_{II}a_{I2}\phi(a_{I2})\left[1-\Phi\left(\frac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right. \\
&\left.+\frac{1}{\sqrt{1-\rho_{II}^2}}\phi(a_{II1})\phi\left(\frac{a_{I2}-\rho_{II}a_{II1}}{\sqrt{1-\rho_{II}^2}}\right)-\frac{\rho_{II}^2}{\sqrt{1-\rho_{II}^2}}\phi(a_{I2})\phi\left(\frac{a_{II1}-\rho_{II}a_{I2}}{\sqrt{1-\rho_{II}^2}}\right)\right\}
\end{aligned}
\tag{3.23}
$$

In a situation where there is no censorship of the first endogenous variable, the marginal effect of the common variable x_{it} on y_{1it} simplifies to

$$
\begin{aligned}
\frac{\partial \mathrm{E}(y_{1it}|\text{no censorship in } y_1)}{\partial x_{it}} &= \frac{1}{\Gamma}(\beta_{12}+\gamma_1\beta_{22})F_{it}^{I}+\beta_{12}F_{it}^{II} \\
&+\frac{a_{I2}\phi(a_{I2})}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\left[\frac{1}{\Gamma}(\gamma_2\beta_{12}+\beta_{22})\rho_I\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}-(\gamma_2\beta_{12}+\beta_{22})\rho_{II}\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}\right]
\end{aligned}
$$

If the second endogenous variable does not present any censorship either, it can be seen that the marginal effect equals that of a classical SEM.

$$\frac{\partial \mathrm{E}(y_{1it}|\text{no censorship in } y_1 \text{ and } y_2)}{\partial x_{it}} = \frac{1}{\Gamma}(\beta_{12}+\gamma_1\beta_{22})$$

If we find ourselves in a case where y_{1it} is censored but there is no simultaneity in the equation, the marginal effect (3.23) reduces to

$$\frac{\partial \mathrm{E}(y_{1it}|\gamma_1=0)}{\partial x_{it}} = \beta_{12}\left(F_{it}^{I}+F_{it}^{II}\right)$$

This is the marginal effect from a classical Tobit model. Finally, proceeding from any of the two preceding equations, it can be seen that when both simplifications are present (in other words, there is neither censorship nor simultaneity in the model), the marginal effect is simply β_{12}, as in a simple linear model.

The marginal effects of each of the three exogenous variables x_{1it}, x_{2it} and x_{it} on the second endogenous variables y_{2it} can be obtained with a parallel development. The final results are given below.

$$\begin{aligned}
\frac{\partial \mathrm{E} y_{2it}}{\partial x_{1it}} &= \frac{1}{\Gamma}\gamma_2\beta_{11}\left\{F_{it}^{I} + \frac{\rho_I}{\sqrt{1-\rho_I^2}}\left[\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right) - \phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}\beta_{11}\frac{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\left\{\rho_I a_{I1}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right) - \frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right\} \\
&- \beta_{11}\frac{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\left\{\rho_{III} a_{I1}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{III2}-\rho_{III} a_{I1}}{\sqrt{1-\rho_{III}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-\rho_{III}^2}}\phi(a_{III2})\phi\left(\frac{a_{I1}-\rho_{III} a_{III2}}{\sqrt{1-\rho_{III}^2}}\right) - \frac{\rho_{III}^2}{\sqrt{1-\rho_{III}^2}}\phi(a_{I1})\phi\left(\frac{a_{III2}-\rho_{III} a_{I1}}{\sqrt{1-\rho_{III}^2}}\right)\right\}
\end{aligned} \tag{3.24}$$

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{2it}}{\partial x_{2it}} &= \frac{1}{\Gamma}\beta_{21}\left\{F_{it}^{I} + \frac{\rho_I}{\sqrt{1-\rho_I^2}}\left[\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right) - \phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}\gamma_1\beta_{21}\frac{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\left\{\rho_I a_{I1}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right) - \frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right\} \\
&+ \beta_{21}\left\{F_{it}^{III} - \frac{\rho_{III}}{\sqrt{1-\rho_{III}^2}}\left[\phi(a_{I1})\phi\left(\frac{a_{III2}-\rho_{III}a_{I1}}{\sqrt{1-\rho_{III}^2}}\right) - \phi(a_{III2})\phi\left(\frac{a_{I1}-\rho_{III}a_{III2}}{\sqrt{1-\rho_{III}^2}}\right)\right]\right\} \\
&- \gamma_1\beta_{21}\frac{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\left\{\rho_{III}a_{I1}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{III2}-\rho_{III}a_{I1}}{\sqrt{1-\rho_{III}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-\rho_{III}^2}}\phi(a_{III2})\phi\left(\frac{a_{I1}-\rho_{III}a_{III2}}{\sqrt{1-\rho_{III}^2}}\right) - \frac{\rho_{III}^2}{\sqrt{1-\rho_{III}^2}}\phi(a_{I1})\phi\left(\frac{a_{III2}-\rho_{III}a_{I1}}{\sqrt{1-\rho_{III}^2}}\right)\right\}
\end{aligned}
\tag{3.25}
$$

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{2it}}{\partial x_{it}} &= \frac{1}{\Gamma}(\gamma_2\beta_{12}+\beta_{22})\left\{F_{it}^{I} + \frac{\rho_I}{\sqrt{1-\rho_I^2}}\left[\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right) - \phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}(\beta_{12}+\gamma_1\beta_{22})\frac{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\left\{\rho_I a_{I1}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-\rho_I^2}}\phi(a_{I2})\phi\left(\frac{a_{I1}-\rho_I a_{I2}}{\sqrt{1-\rho_I^2}}\right) - \frac{\rho_I^2}{\sqrt{1-\rho_I^2}}\phi(a_{I1})\phi\left(\frac{a_{I2}-\rho_I a_{I1}}{\sqrt{1-\rho_I^2}}\right)\right\} \\
&+ \beta_{22}\left\{F_{it}^{III} - \frac{\rho_{III}}{\sqrt{1-\rho_{III}^2}}\left[\phi(a_{I1})\phi\left(\frac{a_{III2}-\rho_{III}a_{I1}}{\sqrt{1-\rho_{III}^2}}\right) - \phi(a_{III2})\phi\left(\frac{a_{I1}-\rho_{III}a_{III2}}{\sqrt{1-\rho_{III}^2}}\right)\right]\right\} \\
&- (\beta_{12}+\gamma_1\beta_{22})\frac{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\left\{\rho_{III}a_{I1}\phi(a_{I1})\left[1-\Phi\left(\frac{a_{III2}-\rho_{III}a_{I1}}{\sqrt{1-\rho_{III}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-\rho_{III}^2}}\phi(a_{III2})\phi\left(\frac{a_{I1}-\rho_{III}a_{III2}}{\sqrt{1-\rho_{III}^2}}\right) - \frac{\rho_{III}^2}{\sqrt{1-\rho_{III}^2}}\phi(a_{I1})\phi\left(\frac{a_{III2}-\rho_{III}a_{I1}}{\sqrt{1-\rho_{III}^2}}\right)\right\}
\end{aligned}
\tag{3.26}
$$

3.1.5 Effects of dichotomous exogenous variables

In the previous section we presented the marginal effects, changes in the endogenous variables brought about by infinitesimal changes in the exogenous variables. However, these calculations apply only for continuous variables. A discrete variable can not change "infinitesimally" and consequently its marginal effect does not exist. However, an analogous idea can lead us to the effect on an endogenous variable of a discrete change in the exogenous variable. In this section we will develop this idea for dichotomous, or dummy, exogenous variables.

For the purpose of this development, suppose x_2, the exogenous variable included only in the second equation, is a realization of a Bernoulli distribution. We are interested in the change of the expected values of y_1 and y_2 when x_2 changes from one state to the other. We start with the first endogenous variable.

$$
\begin{aligned}
\frac{\Delta \mathrm{E}(y_{1it})}{\Delta x_{2it}} &= \mathrm{E}(y_{1it}|x_{2it}=1) - \mathrm{E}(y_{1it}|x_{2it}=0) \\
&= [\mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}>0, x_{2it}=1) - \mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}>0, x_{2it}=0)]\, F_{it}^{I} \\
&+ [\mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}=0, x_{2it}=1) - \mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}=0, x_{2it}=0)]\, F_{it}^{II} \\
&+ [\mathrm{E}(y_{1it}|y_{1it}=0, y_{2it}>0, x_{2it}=1) - \mathrm{E}(y_{1it}|y_{1it}=0, y_{2it}>0, x_{2it}=0)]\, F_{it}^{III} \\
&+ [\mathrm{E}(y_{1it}|y_{1it}=0, y_{2it}=0, x_{2it}=1) - \mathrm{E}(y_{1it}|y_{1it}=0, y_{2it}=0, x_{2it}=0)]\, F_{it}^{IV}
\end{aligned}
$$

The last two lines drop because the conditional expectations are zero. For the expected value from case I, we use result (3.18) to obtain that

$$
\begin{aligned}
&\mathrm{E}(y_{1it}|y_{1it}>0,y_{2it}>0,x_{2it}=1)F_{it}^{I}=\frac{1}{\Gamma}\left(\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}\right)F_{it}^{I}\\
&+\frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}\left\{\phi\left(-\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right)\right.\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_I\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right]\\
&+\rho_I\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\\
&\left.\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_I\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right]\right\}\\
&\mathrm{E}(y_{1it}|y_{1it}>0,y_{2it}>0,x_{2it}=0)F_{it}^{I}=\frac{1}{\Gamma}\left(\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}\right)F_{it}^{I}\\
&+\frac{1}{\Gamma}\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}\left\{\phi\left(-\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right)\right.\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_I\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right]\\
&+\rho_I\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\\
&\left.\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_I\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right]\right\}
\end{aligned}
$$

The expected value for an observation from case II given the two states of the dichotomous variable can be derived from equation (3.19).

$$
\begin{aligned}
&\mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}=0, x_{2it}=1)F_{it}^{II} = (\beta_{11}x_{1it}+\beta_{12}x_{it})\,F_{it}^{II} \\
&\quad + \sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}\left\{\phi\left(-\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}\right)\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_{II}\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}}{\sqrt{1-\rho_{II}^2}}\right)\right. \\
&\quad - \rho_{II}\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right) \\
&\quad \left.\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}-\rho_{II}\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}(y_{1it}|y_{1it}>0, y_{2it}=0, x_{2it}=0)F_{it}^{II} = (\beta_{11}x_{1it}+\beta_{12}x_{it})\,F_{it}^{II} \\
&\quad + \sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}\left\{\phi\left(-\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}\right)\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_{II}\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}}{\sqrt{1-\rho_{II}^2}}\right)\right. \\
&\quad \left.-\rho_{II}\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma_{11}^{\alpha}+\sigma_{11}^{\epsilon}}}-\rho_{II}\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_{II}^2}}\right)\right]\right\}
\end{aligned}
$$

Combining these elements and rearranging terms, we obtain the full effect of a change of state of the dummy variable x_{2it}.

$$
\begin{aligned}
\frac{\Delta \mathrm{E}(y_{1it})}{\Delta x_{2it}} = {} & \frac{1}{\Gamma}\gamma_1\beta_{21}F^I_{it} + \frac{1}{\Gamma}\sqrt{\sigma^\delta_{11}+\sigma^\nu_{11}} \Bigg\{ \phi\left(-\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{11}+\sigma^\nu_{11}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}} - \rho_I\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{11}+\sigma^\nu_{11}}}}{\sqrt{1-\rho_I^2}}\right)\right] \\
& - \phi\left(-\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{11}+\sigma^\nu_{11}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}} - \rho_I\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{11}+\sigma^\nu_{11}}}}{\sqrt{1-\rho_I^2}}\right)\right] \\
& + \rho_I\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{11}+\sigma^\nu_{11}}} - \rho_I\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}}{\sqrt{1-\rho_I^2}}\right)\right] \\
& - \rho_I\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{11}+\sigma^\nu_{11}}} - \rho_I\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}}{\sqrt{1-\rho_I^2}}\right)\right]\Bigg\} \\
& + \sqrt{\sigma^\alpha_{11}+\sigma^\epsilon_{11}}\Bigg\{\phi\left(-\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma^\alpha_{11}+\sigma^\epsilon_{11}}}\right)\Bigg[\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}} - \rho_{II}\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma^\alpha_{11}+\sigma^\epsilon_{11}}}}{\sqrt{1-\rho_{II}^2}}\right) \\
& - \Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}} - \rho_{II}\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma^\alpha_{11}+\sigma^\epsilon_{11}}}}{\sqrt{1-\rho_{II}^2}}\right)\Bigg] \\
& - \rho_{II}\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma^\alpha_{11}+\sigma^\epsilon_{11}}} - \rho_{II}\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}}{\sqrt{1-\rho_{II}^2}}\right)\right] \\
& - \rho_{II}\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}\right)\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}}{\sqrt{\sigma^\alpha_{11}+\sigma^\epsilon_{11}}} - \rho_{II}\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma^\delta_{22}+\sigma^\nu_{22}}}}{\sqrt{1-\rho_{II}^2}}\right)\right]\Bigg\}
\end{aligned}
\tag{3.27}
$$

A parallel development leads us to the full effect of the dichotomous variable on the second endogenous variable.

$$
\begin{aligned}
\frac{\Delta \mathrm{E}(y_{2it})}{\Delta x_{2it}} &= \frac{1}{\Gamma}\beta_{21}F_{it}^{I} + \frac{1}{\Gamma}\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}\left\{\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right)\right. \\
&\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_I\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right] \\
&-\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}\right) \\
&\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_I\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right] \\
&+\rho_I\phi\left(-\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right) \\
&\cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_I\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right] \\
&\left.-\rho_I\phi\left(-\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right)\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}}{\sqrt{\sigma_{22}^{\delta}+\sigma_{22}^{\nu}}}-\rho_I\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_I^2}}\right)\right]\right\} \\
&+\beta_{21}F_{it}^{III}+\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}\left\{\phi\left(-\frac{\beta_{21}x_{2it}+\beta_{22}x_{it}}{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}\right)\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_{III}\frac{\beta_{21}x_{2it}+\beta_{22}x_{it}}{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}}{\sqrt{1-\rho_{III}^2}}\right)\right. \\
&-\phi\left(-\frac{\beta_{22}x_{it}}{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}\right)\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}-\rho_{III}\frac{\beta_{22}x_{it}}{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}}{\sqrt{1-\rho_{III}^2}}\right) \\
&-\rho_{III}\phi\left(-\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right) \\
&\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{21}x_{2it}+\beta_{22}x_{it}}{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}-\rho_{III}\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_{III}^2}}\right)\right] \\
&\left.-\rho_{III}\phi\left(-\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}\right)\left[1-\Phi\left(-\frac{\frac{\beta_{22}x_{it}}{\sqrt{\sigma_{22}^{\alpha}+\sigma_{22}^{\epsilon}}}-\rho_{III}\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}}{\sqrt{\sigma_{11}^{\delta}+\sigma_{11}^{\nu}}}}{\sqrt{1-\rho_{III}^2}}\right)\right]\right\}
\end{aligned}
\tag{3.28}
$$

3.2 Panel with individual fixed effects

The random effects model presented in the preceding section makes the strong assumption that the individual effects are uncorrelated with the exogenous variables. This assumption may well not hold and this would introduce a bias in the estimation of the coefficients. One solution is the use of a fixed effects model. In applications, the fixed effects are used to control for the part of the error term that is correlated with the exogenous variables and, as such, they are treated as nuisance parameters that are frequently eliminated through a transformation (such as first differences or the within transformation) when performing the estimation of the rest of the parameters of the model. In our case, however, we cannot drop them directly by using the usual transformations because of the nonlinearity introduced by the limited dependent character of the endogenous variables.

Nonetheless, treating the fixed effects as constants and then proceed to estimate the model cannot be proposed as a solution. This is because of the incidental parameters problem, which arises in this context because the number of coefficients to be estimated (the individual-specific constants) increases as the sample size grows. More precisely, if there are individual fixed effects in the model, this problem will appear when T is fixed and N, the number of individuals in the sample, increases. As a result, the estimation of the coefficients would be inconsistent if T is fixed.

In this section we will present the development of the fixed effect model and its marginal effects, as well as the likelihood function and the gradient vector. For this we would initially keep the fixed effects since, as discussed, the usual transformations would not succeed in eliminating them. Later in the document, we will describe how the estimation methodology can be used in a way that the incidental parameter problem is avoided.

3.2.1 Construction of the likelihood

As it was done previously, the bivariate observations have to be classified into four cases in order to compute the reduced form and the likelihood of the model. These cases will be detailed in the following paragraphs, although in a more concise manner than in the preceding section since the calculations are parallel.

For an observation from case I, both endogenous variables are observed and it is thus possible to compute the reduced form by cross-substitution. The resulting system will be the parallel to (3.6) and (3.7) in the previous section. The error terms

$$\nu_{1it} = \epsilon_{1it} + \gamma_1 \epsilon_{2it}$$
$$\nu_{2it} = \epsilon_{2it} + \gamma_2 \epsilon_{1it}$$

will be distributed, as before, according to a bivariate normal function with the following parameters

$$\begin{pmatrix} \nu_{1it} \\ \nu_{2it} \end{pmatrix} \sim \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\nu} & \sigma_{12}^{\nu} \\ \sigma_{12}^{\nu} & \sigma_{22}^{\nu} \end{pmatrix} \right]$$

As a consequence of the previous result, the following distribution for the bivariate observation (y_{1it}, y_{2it}) can be derived.

$$\begin{pmatrix} y_{1it}^* \\ y_{2it}^* \end{pmatrix} \sim \mathbb{N} \left[\frac{1}{\Gamma} \begin{pmatrix} \Pi_1' X_{it} + \delta_{1i} \\ \Pi_2' X_{it} + \delta_{2i} \end{pmatrix}, \frac{1}{\Gamma^2} \begin{pmatrix} \sigma_{11}^{\nu} & \sigma_{12}^{\nu} \\ \sigma_{12}^{\nu} & \sigma_{22}^{\nu} \end{pmatrix} \right]$$

with a correlation coefficient

$$r_I = \frac{\sigma_{12}^{\nu}}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu}}}$$

The likelihood for an observation from this case will then be written as

$$L_{it}^{\mathrm{I}} = f^{\mathrm{I}}(\nu_{1it}, \nu_{2it}) = f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}) \tag{3.29}$$

For a bivariate observation from case II, the reduced form will be parallel to equations (3.8) and (3.9). The error terms follow the same joint distribution.

$$\begin{pmatrix} \epsilon_{1it} \\ \nu_{2it} \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\epsilon} & \sigma_{12}^{\epsilon,\nu} \\ \sigma_{12}^{\epsilon,\nu} & \sigma_{22}^{\nu} \end{pmatrix}\right]$$

The joint distribution of the endogenous variables is

$$\begin{pmatrix} y_{1it}^* \\ y_{2it}^* \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} \\ \Pi_2' X_{it} + \delta_{2i} \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\epsilon} & \sigma_{12}^{\epsilon,\nu} \\ \sigma_{12}^{\epsilon,\nu} & \sigma_{22}^{\nu} \end{pmatrix}\right]$$

with a correlation coefficient

$$r_{II} = \frac{\sigma_{12}^{\epsilon,\nu}}{\sqrt{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}}}$$

The likelihood function becomes

$$\begin{aligned} L_{it}^{\mathrm{II}} &= f^{\mathrm{II}}(\epsilon_{1it}, \nu_{2it} \leq -\Pi_2' X_{it} - \delta_{2i}) \\ &= \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it})\, \mathrm{d}\nu_{2it} \end{aligned} \tag{3.30}$$

In the same manner, an observation from case III could be modeled in a similar way to the reduced form (3.10) and (3.11), where the error terms are distributed according to the following normal distribution.

$$\begin{pmatrix} \nu_{1it} \\ \epsilon_{2it} \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\nu} & \sigma_{12}^{\nu,\epsilon} \\ \sigma_{12}^{\nu,\epsilon} & \sigma_{22}^{\epsilon} \end{pmatrix}\right]$$

and the following joint distribution for the endogenous variables is obtained

$$\begin{pmatrix} y^*_{1it} \\ y^*_{2it} \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} \Pi_1' X_{it} + \delta_{1i} \\ \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\nu} & \sigma_{12}^{\nu,\epsilon} \\ \sigma_{12}^{\nu,\epsilon} & \sigma_{22}^{\epsilon} \end{pmatrix}\right]$$

where the correlation coefficient between the two variables is given by

$$r_{III} = \frac{\sigma_{12}^{\nu,\epsilon}}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}}}$$

The likelihood function for an observation from this case is

$$\begin{aligned} L_{it}^{\mathrm{III}} &= f^{\mathrm{III}}(\nu_{1it} \leq -\Pi_1' X_{it} - \delta_{1i}, \epsilon_{2it}) \\ &= \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i})\, \mathrm{d}\nu_{1it} \end{aligned} \tag{3.31}$$

Finally, for case IV the reduced form of the system will parallel (3.12) and (3.13). The distribution of the error terms is reduced to

$$\begin{pmatrix} \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\epsilon} & \sigma_{12}^{\epsilon} \\ \sigma_{12}^{\epsilon} & \sigma_{22}^{\epsilon} \end{pmatrix}\right]$$

which leads to the following distribution of the latent variables

$$\begin{pmatrix} y^*_{1it} \\ y^*_{2it} \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} \\ \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\epsilon} & \sigma_{12}^{\epsilon} \\ \sigma_{12}^{\epsilon} & \sigma_{22}^{\epsilon} \end{pmatrix}\right]$$

with a correlation coefficient

$$r_{IV} = \frac{\sigma_{12}^{\epsilon}}{\sqrt{\sigma_{11}^{\epsilon}\sigma_{22}^{\epsilon}}}$$

In this situation both observations are censored and the likelihood can be written as a cumulative bivariate normal distribution:

$$\begin{aligned} L_{it}^{\mathrm{IV}} &= f^{\mathrm{IV}}(\epsilon_{1it} \leq -\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \epsilon_{2it} \leq -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}) \\ &= \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})\, \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \end{aligned} \tag{3.32}$$

By combining equations (3.29)-(3.32) for all individuals and periods, we obtain the likelihood and its logarithm for the model with fixed effects.

$$
\begin{aligned}
L = \prod_{i=1}^{N} \Bigg\{ & \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}) \\
& \prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it})\, \mathrm{d}\nu_{2it} \\
& \prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i})\, \mathrm{d}\nu_{1it} \\
& \prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})\, \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\log L = \sum_{i=1}^{N} \Bigg\{ & \sum_{\mathbb{D}_i^{\mathrm{I}}} \log f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}) \\
& + \sum_{\mathbb{D}_i^{\mathrm{II}}} \log \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it})\, \mathrm{d}\nu_{2it} \\
& + \sum_{\mathbb{D}_i^{\mathrm{III}}} log \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i})\, \mathrm{d}\nu_{1it} \qquad (3.33) \\
& + \sum_{\mathbb{D}_i^{\mathrm{IV}}} \log \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})\, \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \Bigg\}
\end{aligned}
$$

Even though this function is somewhat simpler than (3.15), it is still not easy to handle since it also has multiple integrals and discontinuities. But even if the MLE was found by maximizing (3.33), it would be inconsistent when one of the dimensions of the panel grows while the other remains fixed. This arises because there is a limited number of observations available on the fixed dimension to estimate the effects on the other large dimension of the panel. This is the incidental parameters problem (see Hsiao (2003), chapter

7). Since in this case we consider only individual effects, this problem arises with large N and T fixed (which is the typical case of a panel)[6]. As argued before, the obvious solution, to treat the individual effects as nuisance parameters and transform the data to eliminate them, cannot be applied here because of the nonlinearities introduced by the limited character of the endogenous variables. Furthermore, Honoré's trimmed-sample GMM solution cannot be used since it is not directly generalized to multivariate systems. The following section will present in detail the methods used to maximize this likelihood function and how they work around the incidental parameter problem.

3.2.2 Gradient vector

The gradient of the logarithm of likelihood function obtained in the preceding section is given by the following vector.

$$\nabla \log L = \Big[\tfrac{\partial \log L}{\partial \beta_{11}} \quad \tfrac{\partial \log L}{\partial \beta_{12}} \quad \tfrac{\partial \log L}{\partial \beta_{21}} \quad \tfrac{\partial \log L}{\partial \beta_{22}} \quad \tfrac{\partial \log L}{\partial \gamma_{1}} \quad \tfrac{\partial \log L}{\partial \gamma_{2}} \quad \tfrac{\partial \log L}{\partial \alpha_{1i}} \quad \tfrac{\partial \log L}{\partial \alpha_{2i}} \quad \tfrac{\partial \log L}{\partial \sigma^{\epsilon}_{11}} \quad \tfrac{\partial \log L}{\partial \sigma^{\epsilon}_{12}} \quad \tfrac{\partial \log L}{\partial \sigma^{\epsilon}_{22}} \Big] \quad (3.34)$$

The calculations of the quantities that compose the gradient vector are summarized in Appendix B, specifically in equations (B.13)-(B.23). In case one of the coefficients is a vector quantity, we would have to include the derivative of the likelihood with respect to each of the elements of this vector. The Hessian matrix could be calculated by taking the derivative of the gradient vector (3.34) with respect to all the coefficients of the model. However, this would lead to lengthy and computationally complicated expressions that would take a lot of resources to calculate. For this reason,

[6]The MLEs will only be consistent if N is fixed and $T \to \infty$.

we do not consider it in this document.

3.2.3 Marginal effects of exogenous continuous variables

The marginal effects for the model with individual fixed effects can be calculated with a parallel procedure than that obtained to obtain the marginal effects for the random effects model, (3.21)-(3.26), only taking into consideration than the individual effects are not taken as realizations from an error component. In order to save space, only the final expressions will be included here. Define the following quantities

$$
\begin{aligned}
b_{I1} &= -\frac{\Pi_1' X_{it} + \delta_{1i}}{\sqrt{\sigma_{11}^{\nu}}} \\
b_{I2} &= -\frac{\Pi_2' X_{it} + \delta_{2i}}{\sqrt{\sigma_{22}^{\nu}}} \\
b_{II1} &= -\frac{\beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i}}{\sqrt{\sigma_{11}^{\epsilon}}} \\
b_{III2} &= -\frac{\beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i}}{\sqrt{\sigma_{22}^{\epsilon}}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{1it}} &= \frac{1}{\Gamma}\beta_{11}\left\{F_{it}^{I} + \frac{r_I}{\sqrt{1-r_I^2}}\left[\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right) - \phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}\gamma_2\beta_{11}\sqrt{\frac{\sigma_{11}^{\nu}}{\sigma_{22}^{\nu}}}\left\{r_I b_{I2}\phi(b_{I2})\left[1-\Phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_I^2}}\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right) - \frac{r_I^2}{\sqrt{1-r_I^2}}\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right\} \\
&+ \beta_{11}\left\{F_{it}^{II} - \frac{r_{II}}{\sqrt{1-r_{II}^2}}\left[\phi(b_{I2})\phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right) - \phi(b_{II1})\phi\left(\frac{b_{I2}-r_{II} b_{II1}}{\sqrt{1-r_{II}^2}}\right)\right]\right\} \\
&- \gamma_2\beta_{11}\sqrt{\frac{\sigma_{11}^{\epsilon}}{\sigma_{22}^{\nu}}}\left\{r_{II} b_{I2}\phi(b_{I2})\left[1-\Phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_{II}^2}}\phi(b_{II1})\phi\left(\frac{b_{I2}-r_{II} b_{II1}}{\sqrt{1-r_{II}^2}}\right) - \frac{r_{II}^2}{\sqrt{1-r_{II}^2}}\phi(b_{I2})\phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right)\right\}
\end{aligned}
\tag{3.35}
$$

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{2it}} &= \frac{1}{\Gamma}\gamma_1\beta_{21}\left\{F_{it}^{I} + \frac{r_I}{\sqrt{1-r_I^2}}\left[\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right) - \phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}\beta_{21}\sqrt{\frac{\sigma_{11}^{\nu}}{\sigma_{22}^{\nu}}}\left\{r_I b_{I2}\phi(b_{I2})\left[1-\Phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_I^2}}\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right) - \frac{r_I^2}{\sqrt{1-r_I^2}}\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right\} \\
&- \beta_{21}\sqrt{\frac{\sigma_{11}^{\epsilon}}{\sigma_{22}^{\nu}}}\left\{r_{II} b_{I2}\phi(b_{I2})\left[1-\Phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_{II}^2}}\phi(b_{II1})\phi\left(\frac{b_{I2}-r_{II} b_{II1}}{\sqrt{1-r_{II}^2}}\right) - \frac{r_{II}^2}{\sqrt{1-r_{II}^2}}\phi(b_{I2})\phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right)\right\}
\end{aligned}
\tag{3.36}
$$

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{1it}}{\partial x_{it}} &= \frac{1}{\Gamma}(\beta_{12}+\gamma_1\beta_{22})\left\{F_{it}^{I} + \frac{r_I}{\sqrt{1-r_I^2}}\left[\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right) - \phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}(\gamma_2\beta_{12}+\beta_{22})\sqrt{\frac{\sigma_{11}^{\nu}}{\sigma_{22}^{\nu}}}\left\{r_I b_{I2}\phi(b_{I2})\left[1-\Phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_I^2}}\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right) - \frac{r_I^2}{\sqrt{1-r_I^2}}\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right\} \\
&+ \beta_{12}\left\{F_{it}^{II} - \frac{r_{II}}{\sqrt{1-r_{II}^2}}\left[\phi(b_{I2})\phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right) - \phi(b_{II1})\phi\left(\frac{b_{I2}-r_{II} b_{II1}}{\sqrt{1-r_{II}^2}}\right)\right]\right\} \\
&- (\gamma_2\beta_{12}+\beta_{22})\sqrt{\frac{\sigma_{11}^{\epsilon}}{\sigma_{22}^{\nu}}}\left\{r_{II} b_{I2}\phi(b_{I2})\left[1-\Phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_{II}^2}}\phi(b_{II1})\phi\left(\frac{b_{I2}-r_{II} b_{II1}}{\sqrt{1-r_{II}^2}}\right) - \frac{r_{II}^2}{\sqrt{1-r_{II}^2}}\phi(b_{I2})\phi\left(\frac{b_{II1}-r_{II} b_{I2}}{\sqrt{1-r_{II}^2}}\right)\right\}
\end{aligned}
\tag{3.37}
$$

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{2it}}{\partial x_{1it}} &= \frac{1}{\Gamma}\gamma_2\beta_{11}\left\{F_{it}^{I} + \frac{r_I}{\sqrt{1-r_I^2}}\left[\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right) - \phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}\beta_{11}\sqrt{\frac{\sigma_{22}^{\nu}}{\sigma_{11}^{\nu}}}\left\{r_I b_{I1}\phi(b_{I1})\left[1-\Phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_I^2}}\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right) - \frac{r_I^2}{\sqrt{1-r_I^2}}\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right\} \\
&- \beta_{11}\sqrt{\frac{\sigma_{22}^{\epsilon}}{\sigma_{11}^{\nu}}}\left\{r_{III} b_{I1}\phi(b_{I1})\left[1-\Phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_{III}^2}}\phi(b_{III2})\phi\left(\frac{b_{I1}-r_{III} b_{III2}}{\sqrt{1-r_{III}^2}}\right) - \frac{r_{III}^2}{\sqrt{1-r_{III}^2}}\phi(b_{I1})\phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right)\right\}
\end{aligned} \tag{3.38}
$$

$$
\begin{aligned}
\frac{\partial \mathrm{E} y_{2it}}{\partial x_{2it}} &= \frac{1}{\Gamma}\beta_{21}\left\{F_{it}^{I} + \frac{r_I}{\sqrt{1-r_I^2}}\left[\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right) - \phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}\gamma_1\beta_{21}\sqrt{\frac{\sigma_{22}^{\nu}}{\sigma_{11}^{\nu}}}\left\{r_I b_{I1}\phi(b_{I1})\left[1-\Phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_I^2}}\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right) - \frac{r_I^2}{\sqrt{1-r_I^2}}\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right\} \\
&+ \beta_{21}\left\{F_{it}^{III} - \frac{r_{III}}{\sqrt{1-r_{III}^2}}\left[\phi(b_{I1})\phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right) - \phi(b_{III2})\phi\left(\frac{b_{I1}-r_{III} b_{III2}}{\sqrt{1-r_{III}^2}}\right)\right]\right\} \\
&- \gamma_1\beta_{21}\sqrt{\frac{\sigma_{22}^{\epsilon}}{\sigma_{11}^{\nu}}}\left\{r_{III} b_{I1}\phi(b_{I1})\left[1-\Phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right)\right]\right. \\
&\left. + \frac{1}{\sqrt{1-r_{III}^2}}\phi(b_{III2})\phi\left(\frac{b_{I1}-r_{III} b_{III2}}{\sqrt{1-r_{III}^2}}\right) - \frac{r_{III}^2}{\sqrt{1-r_{III}^2}}\phi(b_{I1})\phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right)\right\}
\end{aligned} \tag{3.39}
$$

$$
\begin{aligned}
\frac{\partial \mathrm{E}y_{2it}}{\partial x_{it}} &= \frac{1}{\Gamma}(\gamma_2\beta_{12}+\beta_{22})\left\{F_{it}^{I} + \frac{r_I}{\sqrt{1-r_I^2}}\left[\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right) - \phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right)\right]\right\} \\
&+ \frac{1}{\Gamma}(\beta_{12}+\gamma_1\beta_{22})\sqrt{\frac{\sigma_{22}^{\nu}}{\sigma_{11}^{\nu}}}\left\{r_I b_{I1}\phi(b_{I1})\left[1-\Phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right]\right. \\
&+ \left.\frac{1}{\sqrt{1-r_I^2}}\phi(b_{I2})\phi\left(\frac{b_{I1}-r_I b_{I2}}{\sqrt{1-r_I^2}}\right) - \frac{r_I^2}{\sqrt{1-r_I^2}}\phi(b_{I1})\phi\left(\frac{b_{I2}-r_I b_{I1}}{\sqrt{1-r_I^2}}\right)\right\} \\
&+ \beta_{22}\left\{F_{it}^{III} - \frac{r_{III}}{\sqrt{1-r_{III}^2}}\left[\phi(b_{I1})\phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right) - \phi(b_{III2})\phi\left(\frac{b_{I1}-r_{III} b_{III2}}{\sqrt{1-r_{III}^2}}\right)\right]\right\} \\
&- (\beta_{12}+\gamma_1\beta_{22})\sqrt{\frac{\sigma_{22}^{\epsilon}}{\sigma_{11}^{\nu}}}\left\{r_{III} b_{I1}\phi(b_{I1})\left[1-\Phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right)\right]\right. \\
&+ \left.\frac{1}{\sqrt{1-r_{III}^2}}\phi(b_{III2})\phi\left(\frac{b_{I1}-r_{III} b_{III2}}{\sqrt{1-r_{III}^2}}\right) - \frac{r_{III}^2}{\sqrt{1-r_{III}^2}}\phi(b_{I1})\phi\left(\frac{b_{III2}-r_{III} b_{I1}}{\sqrt{1-r_{III}^2}}\right)\right\}
\end{aligned}
\tag{3.40}
$$

3.2.4 Effects of dichotomous exogenous variables

Expressions equivalent to (3.27) and (3.28) but adapted to the fixed-effects model would measure the effect on the endogenous variables of a change of state of a dummy variable. Here we will report on these quantities assuming that x_2 is a dichotomous exogenous variable.

$$
\begin{aligned}
\frac{\Delta \mathrm{E}(y_{1it})}{\Delta x_{2it}} &= \frac{1}{\Gamma}\gamma_1\beta_{21}F^I_{it} + \frac{1}{\Gamma}\sqrt{\sigma^\nu_{11}}\Bigg\{\phi\left(-\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^\nu_{11}}}\right)\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}-r_I\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^\nu_{11}}}}{\sqrt{1-r_I^2}}\right)\right]\\
&-\phi\left(-\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^\nu_{11}}}\right)\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}-r_I\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^\nu_{11}}}}{\sqrt{1-r_I^2}}\right)\right]\\
&+r_I\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}\right)\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^\nu_{11}}}-r_I\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}}{\sqrt{1-r_I^2}}\right)\right]\\
&-r_I\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}\right)\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^\nu_{11}}}-r_I\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}}{\sqrt{1-r_I^2}}\right)\right]\Bigg\}\\
&+\sqrt{\sigma^\epsilon_{11}}\Bigg\{\phi\left(-\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}+\alpha_{1i}}{\sqrt{\sigma^\epsilon_{11}}}\right)\left[\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}-r_{II}\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}+\alpha_{1i}}{\sqrt{\sigma^\epsilon_{11}}}}{\sqrt{1-r_{II}^2}}\right)\right.\\
&\left.-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}-r_{II}\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}+\alpha_{1i}}{\sqrt{\sigma^\epsilon_{11}}}}{\sqrt{1-r_{II}^2}}\right)\right]\\
&-r_{II}\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}\right)\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}+\alpha_{1i}}{\sqrt{\sigma^\epsilon_{11}}}-\rho_{II}\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}}{\sqrt{1-r_{II}^2}}\right)\right]\\
&-r_{II}\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^\nu_{22}}}\right)\\
&\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\beta_{12}x_{it}+\alpha_{1i}}{\sqrt{\sigma^\epsilon_{11}}}-r_{II}\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^\nu_{22}}}}{\sqrt{1-r_{II}^2}}\right)\right]\Bigg\}
\end{aligned}
\tag{3.41}
$$

$$
\begin{aligned}
\frac{\Delta \mathrm{E}(y_{2it})}{\Delta x_{2it}} = {} & \frac{1}{\Gamma}\beta_{21}F^{I}_{it} + \frac{1}{\Gamma}\sqrt{\sigma^{\nu}_{22}}\left\{\phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^{\nu}_{22}}}\right)\right. \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}} - r_I\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^{\nu}_{22}}}}{\sqrt{1-r_I^2}}\right)\right] \\
& - \phi\left(-\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^{\nu}_{22}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}} - r_I\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^{\nu}_{22}}}}{\sqrt{1-r_I^2}}\right)\right] \\
& + r_I\phi\left(-\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+\beta_{21}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^{\nu}_{22}}} - r_I\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}}{\sqrt{1-r_I^2}}\right)\right] \\
& - r_I\phi\left(-\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}\right) \\
& \left.\cdot\left[1-\Phi\left(-\frac{\frac{\gamma_2\beta_{11}x_{1it}+(\gamma_2\beta_{12}+\beta_{22})x_{it}+\delta_{2i}}{\sqrt{\sigma^{\nu}_{22}}} - r_I\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}}{\sqrt{1-r_I^2}}\right)\right]\right\} \\
& + \beta_{21}F^{III}_{it} + \sqrt{\sigma^{\epsilon}_{22}}\left\{\phi\left(-\frac{\beta_{21}x_{2it}+\beta_{22}x_{it}+\alpha_{2i}}{\sqrt{\sigma^{\epsilon}_{22}}}\right)\right. \\
& \cdot\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}} - r_{III}\frac{\beta_{21}x_{2it}+\beta_{22}x_{it}+\alpha_{2i}}{\sqrt{\sigma^{\epsilon}_{22}}}}{\sqrt{1-r_{III}^2}}\right) \\
& - \phi\left(-\frac{\beta_{22}x_{it}+\alpha_{2i}}{\sqrt{\sigma^{\epsilon}_{22}}}\right)\Phi\left(-\frac{\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}} - r_{III}\frac{\beta_{22}x_{it}+\alpha_{2i}}{\sqrt{\sigma^{\epsilon}_{22}}}}{\sqrt{1-r_{III}^2}}\right) \\
& - r_{III}\phi\left(-\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}\right) \\
& \cdot\left[1-\Phi\left(-\frac{\frac{\beta_{21}x_{2it}+\beta_{22}x_{it}+\alpha_{2i}}{\sqrt{\sigma^{\epsilon}_{22}}} - r_{III}\frac{\beta_{11}x_{1it}+\gamma_1\beta_{21}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}}{\sqrt{1-r_{III}^2}}\right)\right] \\
& - r_{III}\phi\left(-\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}\right) \\
& \left.\cdot\left[1-\Phi\left(-\frac{\frac{\beta_{22}x_{it}+\alpha_{2i}}{\sqrt{\sigma^{\epsilon}_{22}}} - r_{III}\frac{\beta_{11}x_{1it}+(\beta_{12}+\gamma_1\beta_{22})x_{it}+\delta_{1i}}{\sqrt{\sigma^{\nu}_{11}}}}{\sqrt{1-r_{III}^2}}\right)\right]\right\}
\end{aligned}
\tag{3.42}
$$

3.3 Consistency and identification constraints

3.3.1 Consistency

Each of the four cases described in the preceding sections produce a different reduced form of the system. Table 3.1 summarizes this information, which is identical for random and fixed effects.

Table 3.1: Summary of the reduced form of the model

	$y_{1it} > 0$	$y_{1it} = 0$
$y_{2it} > 0$	Case I $y^*_{1it} = 1/\Gamma\left(\Pi_1' X_{it} + \delta_{1i} + \nu_{1it}\right) > 0$ $y^*_{2it} = 1/\Gamma\left(\Pi_2' X_{it} + \delta_{2i} + \nu_{2it}\right) > 0$	Case III $y^*_{1it} = \Pi_1' X_{it} + \delta_{1i} + \nu_{1it} \leq 0$ $y^*_{2it} = \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it} > 0$
$y_{2it} = 0$	Case II $y^*_{1it} = \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} + \epsilon_{1it} > 0$ $y^*_{2it} = \Pi_2' X_{it} + \delta_{2i} + \nu_{2it} \leq 0$	Case IV $y^*_{1it} = \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} + \epsilon_{1it} \leq 0$ $y^*_{2it} = \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it} \leq 0$

Each bivariate observation (y_{1it}, y_{2it}) can only belong to one of the cells of Table 3.1 (for example, y_{1it} can either be positive or equal to zero, it can never belong to both cases). One way to obtain the consistency constraints for the reduced form of the model is to find the values of the parameters that violate this rule. Let us start with the first column, where $y_{1it} > 0$, corresponding to cases I and II. In case I, $y_{2it} > 0$ implies that

$$\frac{1}{\Gamma}\left(\Pi_2' X_{it} + \delta_{2i} + \nu_{2it}\right) > 0$$

For case II, $y_{2it} = 0$ implies that

$$\Pi_2' X_{it} + \delta_{2i} + \nu_{2it} \leq 0$$

Both conditions are mutually exclusive if and only if

$$\begin{aligned} \frac{1}{\Gamma} &> 0 \\ \Rightarrow \frac{1}{1-\gamma_1\gamma_2} &> 0 \\ \Rightarrow 1-\gamma_1\gamma_2 &> 0 \\ \Rightarrow \gamma_1\gamma_2 &< 1 \end{aligned} \tag{3.43}$$

In a similar way, if we now consider the first row of Table 3.1, we have that $y_{2it} > 0$. This situation corresponds to cases I and III described above. We have that, for the former

$$\frac{1}{\Gamma}\left(\Pi_1' X_{it} + \delta_{1i} + \nu_{1it}\right) > 0$$

and, for the latter,

$$\Pi_1' X_{it} + \delta_{1i} + \nu_{1it} \leq 0$$

Just like before, both inequalities are mutually exclusive if and only if

$$\gamma_1\gamma_2 < 1$$

the same condition as (3.43). As a consequence, this is the logical consistency condition to define the dependent variables uniquely[7].

Another way to obtain this result is to follow Amemiya (1974) and write the simultaneous equations in the following form

$$Ay_{it} = Bx_{it} + \epsilon_{it}$$

For our model, we have that

$$\begin{aligned} y_{1it} - \gamma_1 y_{2it} &= \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} + \epsilon_{1it} \\ y_{2it} - \gamma_2 y_{1it} &= \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it} \end{aligned}$$

[7]The other two settings (i.e., by studying either the second column or the second row of Table 3.1) produces trivial situations.

and the matrix A becomes

$$A = \begin{pmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{pmatrix}$$

According to Theorem 3 of Amemiya (1974), y_{it} is uniquely defined through its reduced form if and only if every principal minor of A is positive. In our case, A is of dimension (2×2) so there are two principal minors:

$$[A]_{11} = \begin{vmatrix} 1 \end{vmatrix} = 1$$

$$[A]_{22} = \begin{vmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{vmatrix} = 1 - \gamma_1\gamma_2$$

The first principal minor is always positive and the second is positive only if $\gamma_1\gamma_2 < 1$, the same condition as (3.43).

3.3.2 Identification

The usual order and rank conditions for identifiability apply to our model. To exemplify how this is performed in this particular situation, take the reduced form of case I. In a previous subsection the $(K \times 1)$ vector of all exogenous variables of the system was defined as

$$X_{it} = \begin{bmatrix} x_{1it} \\ x_{2it} \\ x_{it} \end{bmatrix}$$

where x_{1it} is the $(k_1 \times 1)$ vector of exogenous variables included only in the first equation, x_{2it} is the $(k_2 \times 1)$ vector of exogenous variables included only in the second equation, and x_{it} is the $(k \times 1)$ vector of exogenous variables common to both equations. It can be seen that $K = k_1 + k_2 + k$. The order condition for identification indicates that the number of excluded exogenous variables in each equation must be greater than or equal to the number of included endogenous variables (in this case, one for each equation). This translates to the condition that $k_1 \geq 1$ and $k_2 \geq 1$. This means that at least one exogenous variable included in one equation is not included in the other.

The rank condition can be checked in the usual way as well. Take the structural form of the system from equations (3.4b) and (3.5b). It is possible to write both equations in the following form.

$$\begin{aligned} -y_{1it} + \gamma_1 y_{2it} + \beta'_{11} x_{1it} + \beta'_{12} x_{it} + \alpha_{1i} + \epsilon_{1it} &= 0 \\ \gamma_2 y_{1it} - \quad y_{2it} + \beta'_{21} x_{2it} + \beta'_{22} x_{it} + \alpha_{2i} + \epsilon_{2it} &= 0 \end{aligned}$$

Define the matrix C that compiles all the coefficients of the model

$$C = \begin{bmatrix} -1 & \gamma_1 & \beta'_{11} & \beta'_{12} & 0 \\ \gamma_2 & -1 & 0 & \beta'_{22} & \beta'_{21} \end{bmatrix}$$

The rank condition for a system with two equations like the one exposed here requires that, for each equation, the rank of the submatrix of excluded variables be at least one. This translates to rank$[\beta_{21}] \geq 1$ for the first equation and rank$[\beta_{11}] \geq 1$ for the second. This coincides with the order condition obtained in the previous paragraph.

In conclusion, the necessary and sufficient condition for identifiability in our system is to have $k_1 \geq 1$ and $k_2 \geq 1$. If $k_1 = k_2 = 1$, both equations are just-identified. On the other hand, if $k_1 > 1$ or $k_2 > 1$, the system is over-identified.

Chapter 4

Optimization methods

The traditional estimation procedure for a random-effect panel model is to incorporate the structure of the error components into the estimation of the variance-covariance matrix of the residuals and then apply Generalized Least Squares (GLS). If there are two or more simultaneous equations present, equation-by-equation (2SLS) and system (3SLS) estimators have been proposed, according to what was summarized in Chapter 2.

If the effects are assumed fixed, the traditional answer is to treat the individual effects as nuisance parameters and wipe them out of the model through a transformation. This is typically done by the within transformation (i.e., by transforming each variable to deviations from individual means); this is called the Least Squares Dummy Variables (LSDV) estimator. This transformation can also be applied when we are dealing with a system of simultaneous equations. On the other hand, if the data is censored, Honoré's trimmed-sample GMM estimator is available.

However, none of these solutions can be applied in our setting, when we have all three situations: a SEM with censored endogenous variables and data with a panel structure. In the preceding chapter we obtained the likelihood function (3.15) for the random-effects model and (3.33) for the fixed-effects, so that we can obtain the MLE by direct maximization of those functions. However, those functions are complex, since they include multiple integrals, discontinuities and probably multiple modes arising from their "case" structure. An analytical solution is impossible to find in this situation. This chapter will describe the main methodologies used to optimize the likelihood functions and the one that was finally selected. Since most methods are applications of standard procedures, they will only be succinctly described here.

It must be remarked that the objective of this chapter is to report on the different methodologies that were attempted to find an estimator that maximizes the likelihood function; these methods were sequentially applied until one achieved adequate results and was selected. For this reason, not all possible optimization alternatives are listed here. Some other procedures, like the quadrature method applied over a dimensional reduction formula of the likelihood function proposed by Huguenin et al. (2009), could also provide attractive optimization alternatives, but they will not be included here.

4.1 Standard gradient methods

Standard iterative algorithms based on gradients are commonplace in optimization problems. They are powerful methods that can be shown to converge to a (local) optimum under certain conditions. In addition, they are directly applicable to a wide range of situations. Because of this, they are the traditional solution of an optimization problem and they are pre-installed (commonly as the default algorithm) in most mathematical and statistical software. Different versions exist, depending if they work with the Hessian of the function or with an approximation of its gradient[8]. Some of the most common are the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm and the Berndt-Hall-Hall-Hausman (BHHH) algorithm.

However, in functions with a discontinuous gradient or other complexities, this kind of methods are not guaranteed to converge. Furthermore, they are never guaranteed to find the global optimum, a problem in a function with multiple local optima. Since the likelihoods (3.15) and (3.33) very likely present these problems, gradient methods are not expected to produce appropriate results.

The standard BFGS procedure was applied to find the maximum of the likelihood of the random-effects model, equation (3.15), in a simulation study. The algorithm was not able to find a solution. Even if the function was further simplified (by assuming that the variance-covariance matrices of the error components were known and by censoring only one of the endogenous variables) the algorithm had a difficult time converging and was very slow, because the approximation of the gradient for a function such as this one can be very computationally intensive. Moreover, when a solution was proposed, it was in general far from the true values of the coefficients and it was very unstable in the sense that small changes in the initial values produced widely distinct results (an indication that the algorithm was stopping at local optima). For these reasons, as it was hypothesized in the previous section, standard gradient methods are not suitable for functions

[8]The latter are commonly referred to as Quasi-Newton's methods.

such as the likelihoods studied here.

These optimization methods are computed by mathematical software through a numerical approximation of the gradient of the likelihood function. Another option is to provide the software directly with the analytical gradient (3.16) or (3.34). However, in our case this does not simplify the task (on the contrary) since, as shown in Appendix B, the gradients are complex functions of the data and the coefficients, involving many single and double integrals. These are computationally complicated to evaluate, possibly even more than the likelihood itself, and cannot provide any immediate help for the optimization.

4.2 Simulated maximum likelihood and simulated score method

The idea behind this method is to use simulation to reduce the complexity of the function that needs to be optimized. For example, the expected value of $f(x)$ is approximated through the average of R simulations in the following way

$$\mathrm{E}_g f(x) = \int f(x)g(x)\,\mathrm{d}x \approx \frac{1}{R}\sum_{r=1}^{R} f^{(r)}(x)$$

where $f^{(r)}(x)$ is the r-th simulation of the function.

Suppose we are dealing with a random-effects model with likelihood (3.15). Each of the four g densities are bivariate normal but with different parameters. In order to simplify the simulation by homogenizing these densities, a Whitening transformation was applied to each one of them. For instance, for case I, in the previous section it was described that the variables

$(\delta_{1i}, \delta_{2i})$ follow a bivariate normal density with the following parameters

$$\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix} \sim \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\delta} & \sigma_{12}^{\delta} \\ \sigma_{12}^{\delta} & \sigma_{22}^{\delta} \end{pmatrix}\right] = \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_g^{\mathrm{I}}\right]$$

Let $(\theta_{1i}, \theta_{2i})$ follow a bivariate standard normal density and define the factorization

$$C_g^{\mathrm{I}}(C_g^{\mathrm{I}})' = \Sigma_g^{\mathrm{I}}$$

Performing the change of variables

$$\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix} = C_g^{\mathrm{I}} \begin{pmatrix} \theta_{1i} \\ \theta_{2i} \end{pmatrix}$$

the likelihood for one observations from this case becomes

$$L_{it}^{\mathrm{I}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{|C_g^{\mathrm{I}}|} f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' x_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' x_{it} - \delta_{2i} | \theta_{1i}, \theta_{2i}) g(\theta_{1i}, \theta_{2i}) \, \mathrm{d}\theta_{1i} \, \mathrm{d}\theta_{2i}$$

Note that the superscript from g disappears because it is now a bivariate standard normal density. By performing the same transformation to the other cases, the likelihood functions takes this (slightly more simplified) form

$$\begin{aligned} L = \prod_{i=1}^{N} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Bigg\{ & \prod_{t\in\mathbb{D}^{\mathrm{I}}} \frac{1}{|C_g^{\mathrm{I}}|} f^{\mathrm{I}}(\Delta y_{1it} - \Pi_1' x_{it} - \delta_{1i}, \Delta y_{2it} - \Pi_2' x_{it} - \delta_{2i} | \theta_{1i}, \theta_{2i}) \\ & \prod_{t\in\mathbb{D}^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' x_{it} - \delta_{2i}} \frac{1}{|C_g^{\mathrm{II}}|} f^{\mathrm{II}}\left(y_{1it} - \beta_1' x_{1it} - \alpha_{1i}, \nu_{2it} | \theta_{1i}, \theta_{2i}\right) \mathrm{d}\nu_{2it} \\ & \prod_{t\in\mathbb{D}^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' x_{it} - \delta_{1i}} \frac{1}{|C_g^{\mathrm{III}}|} f^{\mathrm{III}}\left(\nu_{1it}, y_{2it} - \beta_2' x_{2it} - \alpha_{2i} | \theta_{1i}, \theta_{2i}\right) \mathrm{d}\nu_{1it} \\ & \prod_{t\in\mathbb{D}^{\mathrm{IV}}} \int_{-\infty}^{-\beta_2' x_{2it} - \alpha_{2i}} \int_{-\infty}^{-\beta_1' x_{1it} - \alpha_{1i}} \frac{1}{|C_g^{\mathrm{IV}}|} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \theta_{1i}, \theta_{2i}\right) \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it} \Bigg\} \\ & \cdot g(\theta_{1i}, \theta_{2i}) \, \mathrm{d}\theta_{1i} \, \mathrm{d}\theta_{2i} \end{aligned}$$

which can be approximated through R simulations of $(\theta_{1i}, \theta_{2i})$ in the following way

$$\widetilde{L}(R) = \prod_{i=1}^{N} \frac{1}{R} \sum_{r=1}^{R} \Bigg\{ \prod_{t \in \mathbb{D}^{\mathrm{I}}} \frac{1}{|C_g^{\mathrm{I}}|} f^{\mathrm{I}}(\Delta y_{1it} - \Pi_1' x_{it} - \delta_{1i}, \Delta y_{2it} - \Pi_2' x_{it} - \delta_{2i} | \theta_{1i}^{(r)}, \theta_{2i}^{(r)})$$

$$\prod_{t \in \mathbb{D}^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' x_{it} - \delta_{2i}} \frac{1}{|C_g^{\mathrm{II}}|} f^{\mathrm{II}} \left(y_{1it} - \beta_1' x_{1it} - \alpha_{1i}, \nu_{2it} | \theta_{1i}^{(r)}, \theta_{2i}^{(r)} \right) \mathrm{d}\nu_{2it}$$

$$\prod_{t \in \mathbb{D}^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' x_{it} - \delta_{1i}} \frac{1}{|C_g^{\mathrm{III}}|} f^{\mathrm{III}} \left(\nu_{1it}, y_{2it} - \beta_2' x_{2it} - \alpha_{2i} | \theta_{1i}^{(r)}, \theta_{2i}^{(r)} \right) \mathrm{d}\nu_{1it}$$

$$\prod_{t \in \mathbb{D}^{\mathrm{IV}}} \int_{-\infty}^{-\beta_2' x_{2it} - \alpha_{2i}} \int_{-\infty}^{-\beta_1' x_{1it} - \alpha_{1i}} \frac{1}{|C_g^{\mathrm{IV}}|} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \theta_{1i}^{(r)}, \theta_{2i}^{(r)} \right) \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it} \Bigg\}$$

It can be shown that (see Gouriéroux and Monfort, 1997)

$$\lim_{R \to \infty} \widetilde{L}(R) = L$$

However, this function is still too complex to implement. In a trial for our likelihood, $\widetilde{L}(R)$ was probably only a poor, highly variable approximation of L even for a large R, so the maximization algorithm could not converge. The method of simulated maximum likelihood is very costly in terms of computational effort and the function in question is too complex; as a result, no maximization result could be obtained.

Instead of simulating the likelihood function directly, it is also possible to simulate the score function. This function is the gradient of the log likelihood given by (3.16) for the random-effects model and (3.34) for the fixed-effects model. As is was explained before, the gradient is as complex a function as the likelihood itself. For this reason, its direct use in a gradient method does not make the optimization problem any easier. However, they can be approximated by simulation and then minimized to obtain computationally feasible estimators of the coefficients. Because of the close link between this method and the EM algorithm, further discussion will be delayed to the section that deals with this method and its simulated version.

4.3 Metaheuristic algorithms

When traditional optimization methods fail to be effective or efficient, a number of alternatives exist. One of them is the use of metaheuristic computational methods. This type of procedures iteratively proposes a candidate solution (frequently draw randomly according to a set of rules that defines the algorithm) and evaluates if it is an improvement according to a certain measure. Few or no assumptions are imposed on the objective function, so they can be applied to discontinuous or "ill-behaved" functions. In addition, they do not rely on gradients or Hessians, so the computational cost may decrease importantly in comparison with the standard gradient methods outlined previously. One of the disadvantages of these algorithms is that they are generally not guaranteed to reach an optimal or near-optimal solution; also, they depend on a number of parameters that are critical for their performance and for whose values little empirical guidance exists.

Many of such metaheuristic methodologies have been proposed. A number of reviews, such as that of Gilli and Winker (2008), presents summaries and comparisons among them. In this research we only applied the genetic algorithm and simulated annealing, two of the most common methods, to our likelihood maximization problem.

4.3.1 Genetic algorithm

This algorithm is inspired by the reproductive sequence of species in an evolutionary process. An initial population[9] P_0 is proposed, consisting of several solutions. They can be drawn from the neighborhood of an initial value (such as the estimators from a linear regression) or randomly drawn from the parameter space; the latter alternative is preferred since it allows to explore all sections of the likelihood function. Then, a new population is generated by crossover which translates to a random mixture of the "genes" of both parents (their values of the coefficients); random mutations are also applied to the new solutions in order to explore new sections of the parameter space. If a new solution inherits good properties (i.e., a high value of the likelihood function) from its parents or if it obtains them through mutation, it has a higher probability to survive and "reproduce". This is repeated until a pre-defined number of generations is reached or until another stopping criteria is met. As argued by Reeves and Rowe (2002) there is no general convergence proof for the genetic algorithm because it depends of the function to be optimized and the specific type of genetic algorithm applied (in fact, the term "genetic algorithm" refers to a wide family of evolutionary algorithms). However, its good empirical performance has made it a popular metaheuristic method. This sequence is presented in Algorithm 1 below, which is the specific version of the methodology applied in this thesis.

Even though it is reported to produce acceptable results in many optimization problems, it requires many evaluations of the objective function (one for each individual of the population at each generation). This is very

[9] The mention of a population in this algorithm is used in the sense of the available pool of solutions with which the genetic algorithm works and not in the statistical sense of the word.

Algorithm 1 Genetic algorithm

1. Initialize the algorithm by proposing the initial population P_0 of solutions obtained through random draws from the entire parameter space. Set $i = 0$.

2. Evaluate the likelihood L at all values of the population P_i and select surviving individuals at random (the higher L, the more likely a solution will survive). This will compose the surviving population P'_i

3. Repeat n times the following procedure (n is the number of children to be generated) to obtain a new generation of solutions P''_i

   (a) Randomly select two individuals a and b from P'_i.

   (b) Apply crossover from a and b to obtain a child c.

   (c) Apply a random mutation to child c that may randomly change one or more of the elements of the solution.

4. Combine the surviving population P'_i with the new generation P''_i. This will become the new population P_{i+1}.

5. If stopping criteria is not met, set $i = i + 1$ and go to 2.

6. If stopping criteria is met, P_{i+1} would be the population of genetic optimum. Optionally, take an average of these population as the initial value for a gradient or another method in order to search the vicinity of the optimum.

---

costly with a complicated function such as (3.15) or (3.33). In addition, this algorithm has many settings that need to be defined: the probability of survival, the number of individuals in the initial population P_0, the number n of new individuals at each generation, the probability of mutations and the stopping criteria. It is not clear how to define these values and they clearly have a significant impact on the performance of the algorithm. More troublesome, the algorithm may propose individuals that are not plausible (i.e., with values of the parameters that result in variance-covariance matrices that are not positive definite); it is not clear how the imposition of restrictions may affect the algorithm. A small Monte Carlo experiment on a simplified version of the likelihood (3.15) (with known variance-covariance matrices) produced relatively appropriate estimates, but it was extremely slow due to the high number of evaluations of such a complex objective function.

4.3.2 Simulated annealing

This metaheuristic procedure, closely related to the Metropolis algorithm, belongs to the class of threshold methods. Starting with an initial value π_0 and its corresponding likelihood L_0, a neighboring solution π' is proposed (with likelihood L'). This neighbor can be any value in the vicinity of the initial solution, but a common practice is to randomly change one of its elements. If the neighbor improves on the initial value (in other words, if $L' > L_0$), it is accepted. If it does not, it can still be accepted subject to a probabilistic threshold. In fact, it is accepted with probability $\mathrm{P}(L_0 - L', T_0)$. This probability is directly proportional to the difference between L_0 and L' (the closer L' lies with respect to L_0, the higher the probability to accept it) and to the temperature T_0 (the higher the temperature, the higher the probability to move to the proposed value)[10]. The temperature parameter T_i has a subscript because it changes as the algorithm progresses: it is gradually decreased. At the beginning of the procedure, the temperature takes high values so that neighbors are accepted with a high probability, thus allowing the exploration of the entire objective function. As the process advances, it is gradually "cooled" so that explo-

[10] Usually, this probability is obtained through an exponential density function with parameter T_i evaluated at $L_i - L'$, but it can take any other form.

ration concentrates in a section of the parameter space until, towards the end of the algorithm, neighbors that result in a lower likelihood will only be accepted with a very low probability. This sequence is repeated until a certain stopping criteria is reached. The simulated annealing algorithm, under appropriate but general cooling schedules, converges in probability to the set of optimal solutions as the number of iterations goes to infinity (see Aarts et al., 1997 and Winker, 2001). However, the rate of convergence is logarithmic leading to exponential complexity for arbitrarily close approximations of the optimal solution. Algorithm 2 below presents a summary of this method.

Algorithm 2 Simulated annealing

1. Initialize the algorithm by proposing an initial value π_0 for the parameters. Set $i = 0$.
2. Evaluate the likelihood at the current value of the parameters, $L_i = L(\pi_i)$.
3. Propose a neighbor π', usually π_i with one of its elements randomly changed, and evaluate the likelihood at this value, $L' = L(\pi')$. If $L' > L_i$, accept the new value. Otherwise, move to it only with probability $\mathrm{P}(L_i - L', T_i)$. If the neighbor is accepted, set $\pi_{i+1} = \pi'$, otherwise $\pi_{i+1} = \pi_i$.
4. If stopping criteria is not met, set $i = i + 1$ and go to 2.
5. If stopping criteria is met, π_{i+1} would be the optimum proposed by the algorithm. Optionally, take it as the initial value for a gradient or another method in order to search the vicinity of the optimum.

Even though it is a simple and effective algorithm, it has the same disadvantages listed above for the genetic algorithm. First, some settings such as the stopping criteria, the proposal of neighbors and, more importantly, the cooling schedule, need to be defined in advance and they are crucial to the performance of the algorithm. Also, although less severely than the genetic algorithm, it goes through the evaluation of the likelihood L many times, a costly requirement with a complex function such as our objective. Finally, the neighbor proposed can be an implausible value because it implies a variance-covariance matrix that is not positive definite; the effects of imposing restrictions on the algorithm are not clear. As before, a small Monte Carlo experiment was undertaken on a simplified version of the likelihood (3.15) (with known variance-covariance matrices of the error components). It produced acceptable estimators but, although faster than the genetic algorithm, it was still too slow to be recommended.

4.4 EM algorithm

The Expectation-Maximization (EM) algorithm is an iterative method constructed around the recursive implementation of an Expectation step and a Maximization step. If the unobserved or censored values are imputed with their expected values during the E-step, the likelihood becomes much simpler and easy to optimize during the M-step[11]. The idea behind the use of the EM algorithm is to replace the maximization of a complex likelihood function with a sequence of maximizations of easier functions through the latent structure of the data.

From Jensen's inequality, it can be proved that each iteration of the EM algorithm will produce the same or a higher value of the likelihood; however, this method is not guaranteed to converge to the global optimum. Although the performance of this algorithm is acceptable in many situations and its

[11] In our model, with the censored values "completed" in the E-step, the likelihood of the reduced form would become that of a regular multivariate regression model.

Algorithm 3 EM algorithm

1. Initialize the algorithm with a consistent initial estimator π_0. Set $i = 0$.
2. E-step: given the current parameter estimate π_i compute the statistic

$$J(\pi|\pi_i, Y) = \int \log f(\widetilde{Y}^*|\pi) f(\widetilde{Y}^*|Y, \pi_i)\, \mathrm{d}\widetilde{Y}^*$$

   with Y the observed data and $\widetilde{Y}^*$ the complete data. $J(\pi|\pi_i, Y)$ is the expected value of the complete data log-likelihood where the expectation is with respect to the conditional predictive distribution at the current parameter estimate π_i.
3. M-step: compute π_{i+1} as

$$\pi_{i+1} = \arg\max_{\pi} J(\pi|\pi_i, Y)$$

4. Set $i = i + 1$ and repeat steps 2-3 until convergence.

---

properties are well known, its application may not be simple. Moving from a univariate Tobit model to a bivariate SUR Tobit model (i.e., the reduced form of a two-equation simultaneous model) severely complicates the E-step, as it can be seen in Huang et al. (1987). Further moving to a panel setting and/or considering more than two equations turns the E-step into an intractable problem. For this reason, this algorithm cannot be applied for our maximization problem.

However, before we proceed to the next method, it would be useful to present the link between the J function used in the EM algorithm and the score function (i.e., the gradient of the log likelihood). Following the derivation in Hajivassiliou and Ruud (1994), we have the next result.

$$
\begin{aligned}
E_{\pi_i}\left[\triangledown \log f(\pi|Y^*)|Y\right] &= \int\limits_{\{Y^*|\psi(Y^*)=Y\}} \triangledown \log f(\pi|Y^*)\,\mathrm{d}F(\pi|Y^*,Y) \\
&= \int\limits_{\{Y^*|\psi(Y^*)=Y\}} \frac{\triangledown f(\pi|Y^*)}{f(\pi|Y^*)} f(\pi|Y^*,Y)\,\mathrm{d}Y^* \\
&= \int\limits_{\{Y^*|\psi(Y^*)=Y\}} \frac{\triangledown f(\pi|Y^*)}{f(\pi|Y^*)}\frac{f(\pi|Y^*)}{f(\pi|Y)}\,\mathrm{d}Y^* \\
&= \frac{1}{f(\pi|Y)} \int\limits_{\{Y^*|\psi(Y^*)=Y\}} \triangledown f(\pi|Y^*)\,\mathrm{d}Y^* \\
&= \frac{1}{f(\pi|Y)} \int\limits_{\{Y^*|\psi(Y^*)=Y\}} \triangledown \mathrm{d}F(\pi|Y^*) \\
&= \frac{\triangledown f(\pi|Y)}{f(\pi|Y)} = \triangledown \log f(\pi|Y)
\end{aligned}
$$

where $\psi(Y^*)$ is the rule (3.3) that defines the limited character of the dependent variable. Now, we have that the J function of the E-step of the EM algorithm is given by the following expression.

$$J(\pi|\pi_i, Y) = E_{\pi_i}\left[\log f(\pi|Y^*)|Y\right]$$

Ruud (1991) proves that if $\pi_{i+1} = \pi_i = \pi$ then

$$\triangledown J(\pi_{i+1}|\pi_i, Y) = \triangledown \log f(\pi|Y)$$

so the first order conditions of the optimization problem that defines the M-step of the EM algorithm and the normal equations of the log likelihood are closely related. This result will be used shortly, when we present a simulated version of the present algorithm.

4.5 Gibbs sampler

This procedure belongs to the class of Markov Chain Monte Carlo (MCMC) methods and it is used to generate random variates from a distribution indirectly, without having to calculate the density. Frequently it is the case that one can find a set of conditional densities that are much simpler to deal with than the joint density. The idea is to sample from these simpler conditional distributions in a sequential manner.

For example, let the objective function be a joint density of the parameters (τ_1, τ_2). If this joint density is cumbersome but it is possible to find simpler conditional distributions $f(\tau_1|\tau_2)$ and $f(\tau_2|\tau_1)$, the Gibbs sampler can be employed. If we let $\tau_j^{(i)}$ be the i-th draw from τ_j, the sampler can be applied sequentially as

$$\begin{aligned}
\tau_1^{(i)} &\sim f(\tau_1|\tau_2^{(i-1)}) \\
\tau_2^{(i)} &\sim f(\tau_2|\tau_1^{(i)}) \\
\tau_1^{(i+1)} &\sim f(\tau_1|\tau_2^{(i)}) \\
\tau_2^{(i+1)} &\sim f(\tau_2|\tau_1^{(i+1)}) \\
\tau_1^{(i+2)} &\sim f(\tau_1|\tau_2^{(i+1)}) \\
\tau_2^{(i+2)} &\sim f(\tau_2|\tau_1^{(i+2)}) \\
&\vdots
\end{aligned}$$

An initial burn-in sample of random variates is discarded in order to cancel the influence of the initial values. After that, it can be proved that the draws $(\tau_i^{(r)}, \tau_2^{(r)})$ belong to $f(\tau_1, \tau_2)$. With enough draws, characteristics of the joint and marginal densities can be derived. For instance, the shape can be approximated through kernel methods and the mean of the marginal density for τ_1 as

$$\mathrm{E}\left[f(\tau_1)\right]_R = \frac{1}{R}\sum_{r=1}^{R} f(\tau_1^{(r)})$$

for R draws or, for improved efficiency, through the "Rao-Blackwell" estimator of the mean

$$\widetilde{\mathrm{E}}\left[f(\tau_1)\right]_R = \frac{1}{R}\sum_{r=1}^{R} \mathrm{E}\left[f(\tau_1|\tau_2^{(r)})\right]$$

The conditional densities used to draw variates are in fact posterior densities, so a prior distribution is needed for each parameter. To remain in frequentist terrain, it is possible to take a diffuse prior, for which posterior densities have been found for many models. In this sense, the Gibbs sampler can be thought of as a stochastic version of the EM algorithm in which sampling replaces both the E-step and the M-step. The adequate performance of this method has been well documented; another advantage is that, since draws are always obtained from conditional distributions, impossible values (such as the negative definitive variance-covariance matrices that came up in the metaheuristic methods) are avoided. However, by construction, the random draws are autocorrelated, but this can easily be surmounted by only considering one of every p draws. Also, as with any sampling solution, a large R is required to fully characterize the densities of the parameters.

The idea to use sampling (or multiple imputation) in a latent variable model has been called "data augmentation" (see Tanner and Wong, 1987). Wei and Tanner (1990a, 1991) applied this methodology specifically to censored regression data. Gelfand and Smith (1990) then noted the close relationship between this method and the Gibbs sampler. Further details about the data augmentation solution and refinements thereof (such as marginal data augmentation to improve the efficiency of the MCMC) can be found in Van Dyk and Meng (2001).

For a model with censored dependent variables such as the one studied in this thesis, it is possible to define the sampling procedure into two conditional steps. These two steps are repeated sequentially to obtain the Gibbs sample. Algorithm 4 presents the pseudocode that can be followed in this case[12].

Algorithm 4 Gibbs sampler

1. Initialize the algorithm with an initial value of the parameter $\pi_{(0)}$. Define the size of the burn-in sample R' and the size of the sample that will be kept R. Set $r = 1$.
2. Given the current value of the parameter $\pi_{(r-1)}$, obtain a random draw from the conditional distribution of the censored data,
$$f\left(y^*_{t,(r)}|y_t = 0, \pi_{(r-1)}\right)$$
where $y^*_{t,(r)}$ represents the r-th draw from the distribution of the censored dependent variable at period t. By taking
$$\widetilde{Y}^*_{(r)} = \{y_t|y_t > 0\} \cup \{y^*_{t,(r)}|y_t = 0, \pi_{(r-1)}\}$$
we obtain the r-th draw of the "complete" data.
3. Given the complete data from the last step, generate a random draw from the conditional distribution of the parameters,
$$f\left(\pi_{(r)}|\widetilde{Y}_{(r)}\right)$$
4. Set $r = r + 1$ and repeat steps 2-3 $(R' + R)$ times.
5. Discard the first R' observations (burn-in sample) and use the rest to obtain the desired characteristics of the marginal distribution of the parameters. Since the draws are autocorrelated, take one in every p draws.

Central to the algorithm is the definition of the conditional distributions for the censored data and the parameters. To exemplify how these densities are obtained, consider the case of a standard univariate Tobit model, in which $y^*_i = x_i\beta + \epsilon_i$, where ϵ_i follows an *iid* $\mathbb{N}(0, \sigma^2)$ and the dependent variable is censored at zero. In this situation, under diffuse priors, it can be shown that (see Lancaster, 2004)

[12] This algorithm follows that exposed in Chib and Greenberg (1998).

$$f(y_i^*|y_i = 0, \beta, \sigma^2) \sim \mathrm{T}\mathbb{N}_{(-\infty,0)}\left(x_i\beta, \sigma^2\right)$$
$$f(\beta|Y^*, \sigma^2) \sim \mathbb{N}\left((X'X)^{-1}X'Y, \sigma^2(X'X)^{-1}\right)$$
$$f(\sigma^2|Y^*, \beta) \sim \Gamma^{-1}\left(\frac{N-k}{2}, \frac{(Y^* - X\beta)'(Y^* - X\beta)}{2}\right)$$

where N is the total number of observations and k is the number of regressors. T$\mathbb{N}$ represents a truncated normal distribution while Γ^{-1} is an inverted gamma distribution. By cycling through these conditional distributions according to Algorithm 4, it is possible to obtain a random sample from which the desired characteristics of β and σ^2 can be deducted. These distributions were calculated through diffuse priors; another set of priors would produce different conditional densities, which could have a significant effect on the results obtained.

In order to assess the Gibbs sampler for our (random-effects) model, we derived the conditional distributions of the complete data, the coefficients in the regression and the parameters in the variance-covariance matrices through diffuse prior in the same manner as the example above. We then applied the algorithm to simulated data. Good results were obtained since the estimators seemed to converge to the true values, the algorithm converged quickly (so that a small burn-in sample was sufficient) and the autocorrelation of the draws was rather low (about 0.2 for the coefficients in the equations and close to 0.4 for the parameters of the variance-covariance matrices), so taking one draw out of every four or five seemed enough to approximate an independent sample. However, a large sample is needed to fully characterize the parameters, so the algorithm quickly becomes so slow that a simulation study becomes unfeasible. Additionally, the initial assumptions about the prior distributions may not have a negligible effect on the results.

4.6 MCECM algorithm

Combining the two previous procedures may seem like a reasonable compromise. In this sense, we keep the feasibility and versatility of a simulation methodology but with the frequentist, likelihood-maximizing approach of the EM algorithm. The basic idea is to estimate the unfeasible E-step through simulation and then use the conditional maximization technique of Meng and Rubin (1993) to simplify the M-step. Since simulation only takes part in the E-step, this method will likely be faster than the Gibbs sampler. Wei and Tanner (1990b) first proposed this solution in a general framework (and without the conditional maximization step). It was applied to an univariate probit model by Chib and Greenberg (1998), to a SUR Tobit model by Huang (1999) and to a multinomial probit model by Zhou and Liu (2008). The basic pseudocode for the Monte Carlo Expectation Conditional Maximization (MCECM) algorithm is described below.

For the specific model under study in this research, the MCECM algorithm is applied in the following manner. The censored values of the endogenous variables are simulated R times according to their density functions (taking into account the simultaneity and the structure of the panel data), given the current values of the parameters. This Monte Carlo sample is used to approximate the expected value of the censored observations, thus rendering feasible the E-step of the EM algorithm. With the data completed through these expected values, it is now easier to maximize the

Algorithm 5 MCECM algorithm

1. Initialize the algorithm with a consistent initial estimator π_0. Set $i = 0$.
2. Monte Carlo E-step: given the current parameter estimate π_i compute the statistic
$$\widehat{J}(\pi|\pi_i, Y) = \frac{1}{R}\sum_{r=1}^{R} \log f(\widetilde{Y}^{*(r)}|\pi)$$
where $\widetilde{Y}^{*(r)}$ is the r-th simulation of the complete data. $\widetilde{Y}^{*(r)}$ is equal to Y for the observed values but the censored values are replaced by random draws from the conditional predictive distribution given the current value of the parameters, $f(\widetilde{Y}^*|Y, \pi_i)$. By repeating this simulation R times, $\widehat{J}(\pi|\pi_i, Y)$ becomes an estimation of the expected value of the complete data log-likelihood where the expectation is with respect to the conditional predictive distribution at the current parameter estimate π_i.
3. Conditional M-step: compute π_{i+1} as
$$\pi_{i+1} = \arg\max_{\pi} \widehat{J}(\pi|\pi_i, Y)$$
through conditional maximization methods.
4. Set $i = i + 1$ and repeat steps 2-3 until convergence.

(conditional) likelihood and obtain the MLEs or any desired estimator for the coefficients and the variance-covariance matrices[13]. These two steps are iterated until convergence or until a specified stopping rule is reached.

For the fixed effects model, we argued in Chapter 3 that traditional maximization procedures suffer from the incidental parameters problem, but that in this setting it is not possible to apply the usual solution (treat the individual effects as nuisance parameters, eliminate them through a

[13] In the conditional-maximization step, according to Meng and Rubin (1993), the likelihood is divided into several subgroups of parameters. The likelihood is maximized for each subgroup, fixing the rest at the latest optimum available. For our specific case, the likelihood of the coefficients is maximized given the latest value of the variance-covariance matrices; then, the optimal values for those matrices are calculated by fixing the coefficients at the maximum values from the previous step.

transformation, and concentrate the maximization procedure on the other parameters of the model). However, the MCECM solution proposed here presents the advantage that it allows to work around the incidental parameter problem even in a nonlinear setting. First, the censored observations are imputed through simulation according to the Monte Carlo E-step described above. In this way, we obtain a complete set of endogenous variables, either with observed values or with an estimation of their expected value. We are now working in a linear setting so we can then proceed to the traditional solution for these models: use a linear transformation (for example, the within transformation) to eliminate the fixed effects and then maximize the likelihood of the transformed model. In this way, we can use the MCECM to work around the incidental parameters problem in the estimation of the other parameters of the model.

As it was noted previously, the first order conditions in the M-step of the EM algorithm and the score function of the log likelihood are closely related. If the expectation in the E-step is analytically or numerically difficult to compute, we can approximate it through simulation as described above and then proceed to the maximization of the M-step. This procedure will consequently be closely related to the method of simulated score. This idea is already mentioned although not completely developed in Hajivassiliou and Ruud (1994), where the authors note that unbiased simulation of the J function implies a means for unbiased simulation of the score function.

Louis (1982) developed a method to find the asymptotic standard errors for estimators that were obtained through an EM-type algorithm. If the MLE is obtained at the conditional-maximization step of the MCECM algorithm, we can take advantage of a simulation version of Louis' method to obtain the standard errors of the estimators. For examples of the application of this method in a simulation context, see Huang (1999) and Zhou and Liu (2008).

Concerning the consistency restriction mentioned at the end of the previous chapter, we incorporate it into the methodology directly in the algorithm. In step 3 of the Algorithm 5 if the value of π_{i+1} proposed in the M-step violates this restriction, it is rejected and we go back to the Monte Carlo E-step to generate another sample of the complete data. This procedure gave good results in the simulation exercise that will be presented in the following chapter.

As it can be seen, the MCECM algorithm can be readily applied to estimate our model. However, there are several computational considerations that must be tackled previously. First, the algorithm may not converge at all because the simulation introduces a persistent Monte Carlo error that may invalidate the ascent property of the EM method. For this reason, a special version of the algorithm is applied, the ascent-based MCECM. Second, the simulations obtained for the Monte Carlo E-step of Algorithm 5 require many draws from univariate and bivariate truncated normal distributions; this must be performed as efficiently as possible to minimize the computational time. Finally, even though this solution is *a priori* faster than the MCMC algorithm described in the preceding section, it is still computationally intensive. A code that maximizes the resources available is required. These three computational issues are discussed in Appendix C.

As it will be lengthly explained in the next chapter, an initial simulation study showed that the MCECM algorithm finds adequate estimators, close to the true values of the parameters, and it is considerably faster than the

Gibbs sampler. Because of this combination of accurate results and feasible computational time, it was chosen to estimate our models. Chapter 5 will describe the specifics of the MCECM algorithm used in a complete simulation study and the main results obtained for both a random-effects and a fixed-effects model.

Even though in this thesis we illustrate the performance of the MCECM algorithm in situations where the endogenous variables are left-censored at zero, the procedure can be directly extended to other types of data limitations. For example, if the dependent variables are dichotomous and we are estimating a probit model, the latent variables can be "completed" through simulation following the same procedure described in Algorithm 5. The same method can also be used to simulate data and estimate a sample selection (or type-II Tobit) model. In fact, the idea of simulating data to simplify the Estimation-step of the algorithm can be readily applied to other type of data limitation. We present two examples of this in Chapter 6.

Chapter 5

Simulation study

With the objective of evaluating the performance of the estimator developed in Chapter 3, two simulation studies were undertaken. The idea was to verify its properties and compare it to other common estimators under different configurations of the model. Of all the optimization methods tested in Chapter 4, the MCECM algorithm was selected because it was successful in converging to an optimal estimator at a reasonable computational cost. The results will be presented in the remaining of this section, first for the random-effects model and then for the fixed-effects model.

5.1 Simulation study for the random-effects model

To study the properties of the solution proposed in this thesis and compare it to other estimators in a random-effects setting, we perform a simulation study on a simple system very similar to that described by equations (3.4b) and (3.5b). Specifically, we simulate from the following bivariate simultaneous system with one-way error components

$$y^*_{1it} = \gamma_1 y_{2it} + \beta_{11} x_{1it} + \beta_{12} x_{it} + \alpha_{1i} + \epsilon_{1it} \tag{5.1}$$

$$y^*_{2it} = \gamma_2 y_{1it} + \beta_{21} x_{2it} + \beta_{22} x_{it} + \alpha_{2i} + \epsilon_{2it} \tag{5.2}$$

where x_{1it} is an exogenous variable included only in the first equation, x_{2it} is an exogenous variable included only in the second equation and x_{it} is an exogenous variable common to both equations. In order to simplify the model and reduce the computational burden of the simulation, x_{1it} and x_{2it} include one variable each; x_{it} on the other hand includes only a constant term.

In the rest of this section we will describe the features of the base simulation that was developed. Variations on these base characteristics, such as different functional forms, densities of the error term or sample sizes, were also undertaken with the objective of evaluating the robustness of the estimation method to different Data Generating Processs (DGPs). The results will be summarized in what follows.

The parameters of the model were set so that they mimic an empirical application. A classic article in the literature of limited dependent variables is the study of attrition bias by Hausman and Wise (1979). As it was described in Chapter 2, in that article the authors estimate a model based on the Gary experiment of potential labor supply and earning responses to possible income maintenance plans. Specifically, they estimate an earnings equation with random individual effects taking into account attrition in the panel sample. The results are also reproduced in Section 8.2.3 and the first column of Table 8.1 in Hsiao (2003). Our intention is to design our simulation study so that it mimics the exogenous variables and the coefficients estimated by these authors.

The parameters of the exogenous variables were set at the following values.

$$\beta_{11} = 0.04$$
$$\beta_{21} = 0.21$$

In other words, we assign the estimated coefficient of the variable "experience" to the variable that appears exclusively in the first equation and the estimated coefficient of "education" to the exogenous variable included only in the second equation[14]. For the constant term, the estimated value reported on the article is so large that it would cause the dependent variables to be practically always observed (all observations would belong to case I). This is clearly not desirable if we want to test our methodology. We set the coefficients at a smaller value, $\beta_{12} = \beta_{22} = 0.1$. Since the original study only estimates one equation, there are no available values for the simultaneous parameters. We set them at $\gamma_1 = 0.4$ and $\gamma_2 = 0.2$ to have a stronger simultaneity factor in one equation and a weaker in the other.

The exogenous variables were generated through uniform distributions for the base simulation. x_{1it}, the variable that appears only in the first equation and whose coefficient was taken from the variable "experience" in the empirical study, was generated through iid draws from $U(0, 30)$. We generate the variable included only in the second equation, for which we assigned the coefficient that corresponds to "education", through the following density: $x_{2it} \sim U(0, 20)$.

The variance-covariance matrices of the error components also mimic the

[14] The coefficients from the original article were multiplied by 10 to obtain results that are easier to read.

empirical study. We set the variance of the error term and the individual effect for each equation at the values estimated in that article. There is no covariance because the authors estimate a one-equation model, so we set it at one fourth of the variance. This means that the error components were randomly drawn from the following density.

$$\begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.60 & 0.15 & 0 & 0 \\ 0.15 & 0.60 & 0 & 0 \\ 0 & 0 & 1.80 & 0.45 \\ 0 & 0 & 0.45 & 1.80 \end{pmatrix}\right] \tag{5.3}$$

Equations (5.1) and (5.2) are observed simultaneity and this creates a challenge in the design of the simulation study because in order to obtain the value of y^*_{1it} we need the observed value of y_{2it} and vice versa. One possible alternative to work around this problem is to use the reduced form: obtain the coefficients of the reduced form by transforming the structural-form coefficients as it is described in Chapter 3 and then use them to generate the endogenous variables. This could be applied to this specific simulation study since the model is exactly identified. However, with the intention of developing a more general procedure, we chose instead to generate them through an iterative procedure. We start by generating the first endogenous variable assuming that the second is not observed and we censor it if it smaller than zero. After that we generate y^*_{2it} using the value of y_{1it} obtained in the previous step and we apply the censorship rule. We then take the value of the second endogenous variable and we use it to obtain a random draw for the first one. We repeat this procedure until convergence, which usually happens after only a few iterations. With such a procedure we can generate both endogenous variables for a general model and by using only the coefficients of the structural form.

Another detail that needs to be clarified is the generation of the error components. Above we have the joint density function of the four components. However, we should have one $(\alpha_{1i}, \alpha_{2i})$ per individual, but we have one $(\epsilon_{1it}, \epsilon_{2it})$ for each individual and time period, so a random draw cannot be obtained directly. We dealt with this difficulty by drawing values from the following density

$$\begin{pmatrix} \alpha_{1it} \\ \alpha_{2it} \\ \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} T(0.60) & T(0.15) & 0 & 0 \\ T(0.15) & T(0.60) & 0 & 0 \\ 0 & 0 & 1.80 & 0.45 \\ 0 & 0 & 0.45 & 1.80 \end{pmatrix} \right]$$

and then calculating the "real" individual effects as

$$\alpha_{1i} = \frac{1}{T} \sum_{t=1}^{T} \alpha_{1it}$$
$$\alpha_{2i} = \frac{1}{T} \sum_{t=1}^{T} \alpha_{2it}$$

We can verify that that in this way we obtain random draws from the joint distribution (5.3).

Since the model in question represents the overlapping of three estimation difficulties (censored dependent variables, simultaneous equations and one-way error components), several estimators that deal with one or more of these problems were compared in the simulation study.

1. Method that does not correct any problem: Ordinary Least Squares (OLS).

2. Methods that correct only one problem: Generalized Least Squares

for a Random-Effects Model (GLS-RE), a SEM estimator and a Tobit Model (TOB) estimator.

3. Methods that correct two problems: Simultaneous Equation Model with Random Effects (SEM-RE) and Tobit Model with Random Effects (TOB-RE) estimators.

4. Method that corrects all three problems: the procedure proposed in this research, estimated through the MCECM algorithm.

The idea is to compare their performance and identify if our proposed estimator achieves the best results. In the remainder of this section, the methodology we propose here will be called "the MCECM estimator" (although strictly speaking this only refers to the optimization algorithm).

As a recapitulation of the MCECM methodology described in Chapter 4, the censored endogenous variables are imputed with an estimation of their expected value; this is accomplished by simulation under the current estimates of the coefficients. Once the dependent variables are completed in this way, the model becomes a linear model and it is straightforward to obtain the maximum likelihood estimates for the coefficients of the model. This process is iterated until the stopping rule, according the ascent-version of the algorithm described in Appendix C, is reached.

We now give a few specifications of the rest of the estimators included above. The GLS-RE is estimated through Generalized Least Squares where the error-component structure is modeled through a block-diagonal variance-covariance matrix; a feasible version is obtained by estimating the variances through the Within-type residuals (see Baltagi, 2005). The SEM estimator is obtained through a standard FIML for the system and TOB is the classical MLE for an univariate Tobit model. Since in this simulation study we deal with a just-identified simultaneous system, we first obtained the reduced form of the system and then proceeded to estimate it through GLS for a SUR model with random effects (see Baltagi (2005), chapter 6): this is the SEM-RE estimator (which treats the simultaneity in the model as well as the presence of random individual effects) performed on the reduced form. If we were dealing with more general (over-identified) simultaneous systems, any of the estimators described in Section 2.2 of Chapter 2 could be employed. Finally, the TOB-RE estimator deals with a model with censored dependent variables and data in a panel structure. This estimator posed a problem since the "classical" way to obtain it is through MLE; however, its calculation suffered from the same kind of difficulties described in Section 3. For this reason, it was preferred to employ the MCECM algorithm here as well: the data was "completed" through simulation and then an univariate GLS-RE estimator was applied to each equation separately.

The panel included $N = 30$ individuals and $T = 7$ periods for the base simulation. The study consisted in drawing $H = 1000$ values simultaneously from equations (5.1) and (5.2) according to the parameters described above. With these characteristics, we obtain an average censorship rate of 20.5% and 12.3% of the observations for y_{1it} and y_{2it}, respectively. For each draw we estimated the model using the proposed procedure and relying on the parameters of the ascent-based version of the MCECM algorithm listed in Table C.1. Once the estimation was carried out, it was possible to evaluate the properties of the estimator and compare it to other alternatives. This was done through the bias, the variance and the Mean Squared Error (MSE)[15]. The average computational time for the experiment was 63803 seconds (17.7 hours).

A final word about the quantities in which the comparison is based on. The presence of censorship in the endogenous variables introduces nonlinearities in the model. Because of this, the coefficients of (5.1) and (5.2) cannot be used directly if one wants to study the effect of the exogenous variables on each equation and we will only present them for the base simulation. More meaningful conclusion can be obtained by using the marginal effects derived in Chapter 3 and these will be used for the other types of DGP.

As it is shown in Table 5.1, the results from the simulation study for the estimator described in this thesis are promising. In terms of the coefficients of the exogenous variables (β_{11}, β_{12}, β_{21}, β_{22}), the OLS estimator (column 1) and those that correct only one of the problems of the data (columns 2, 3, 4 and 5) perform worse in terms of estimation bias. This is because the censored character of the dependent variables and the simultaneity in the model, if neglected, will generate a bias in the estimation. In all cases except β_{12}, the smallest bias corresponds to the MCECM estimator (column 7), and for the latter it is second to best following closely the GLS-RE estimator.

Now for the estimation variance, the best results correspond to either the GLS-RE (column 2) or the OLS (column 1), estimator that show a low variability around a biased estimator. Even though the MCECM estimator has a higher variance, this is the cost of employing a simulation-based

[15]A natural object to evaluate the performance of different estimators is the average value of likelihood function at the optimum identified by each method. However, the likelihood (3.15) is so complex that its evaluation requires a lot of computational resources. As a consequence, its use in this simulation exercise would be extremely costly. The comparison among estimators is thus based only in the MSE.

Table 5.1: Comparison of the estimation results for the simulation study of the random-effects model, base simulation

		(1) OLS	**(2)** GLS-RE	**(3)** SEM	**(4)** TOB	**(5)** SEM-RE	**(6)** TOB-RE	**(7)** MCECM
β_{11}	Bias	-0.0093	-0.0089	-0.0075	-0.0018	-0.0073	-0.0015	-0.0008
	Variance	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	MSE	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
β_{12}	Bias	-0.2934	-0.1639	0.4841	-0.8958	0.4772	-0.7496	0.1831
	Variance	0.0329	0.0369	0.0747	0.0773	0.0654	0.0763	0.0946
	MSE	0.1189	0.0637	0.3091	0.8799	0.2931	0.6382	0.1282
β_{21}	Bias	-0.0574	-0.0554	-0.0139	-0.0417	-0.0174	-0.0435	0.0006
	Variance	0.0002	0.0002	0.0016	0.0002	0.0009	0.0002	0.0011
	MSE	0.0035	0.0033	0.0018	0.0020	0.0012	0.0021	0.0011
β_{22}	Bias	-0.3359	-0.3029	0.4923	-0.6814	0.4202	-0.4310	0.1802
	Variance	0.0214	0.0214	0.4494	0.0416	0.2585	0.0395	0.1750
	MSE	0.1342	0.1131	0.6917	0.5059	0.4351	0.2253	0.2075
γ_1	Bias	0.2351	0.1850	-0.0621	0.3389	-0.0604	0.3042	-0.0392
	Variance	0.0020	0.0022	0.0062	0.0031	0.0052	0.0029	0.0061
	MSE	0.0573	0.0364	0.0101	0.1179	0.0089	0.0955	0.0076
γ_2	Bias	0.5138	0.4869	-0.1232	0.5575	-0.0691	0.4981	-0.0792
	Variance	0.0024	0.0025	0.2347	0.0029	0.1318	0.0024	0.1072
	MSE	0.2664	0.2395	0.2499	0.3137	0.1366	0.2505	0.1135
σ_{11}^{ϵ}	Bias	-0.4055	-0.5964	0.1618	0.0908	-0.4840	-0.2781	-0.2272
	Variance	0.0257	0.0208	0.3275	0.0506	0.1123	0.0310	0.1515
	MSE	0.1901	0.3765	0.3537	0.0589	0.3465	0.1083	0.2031
σ_{12}^{ϵ}	Bias			0.4177		-0.1402		-0.0155
	Variance			0.4912		0.1908		0.2260
	MSE			0.6657		0.2104		0.2263
σ_{22}^{ϵ}	Bias	-0.6436	-0.7111	0.8450	-0.4353	-0.1692	-0.6055	0.0041
	Variance	0.0131	0.0136	1.4479	0.0206	0.1769	0.0185	0.2172
	MSE	0.4273	0.5194	2.1618	0.2101	0.2055	0.3852	0.2172
σ_{11}^{α}	Bias		-0.3412			-0.1114	-0.4042	0.0326
	Variance		0.0193			0.0540	0.0110	0.0854
	MSE		0.1357			0.0664	0.1743	0.0865
σ_{11}^{α}	Bias					0.3331		0.4579
	Variance					0.0447		0.0644
	MSE					0.1556		0.2741
σ_{11}^{α}	Bias		-0.4795			0.0301	-0.5469	0.1463
	Variance		0.0066			0.1899	0.0026	0.2536
	MSE		0.2365			0.1908	0.3017	0.2750

algorithm to obtain the estimators. Nonetheless, the variance arising from the MCECM algorithm is surprisingly small and our proposed estimator does not perform much worse (and it is, in fact, frequently better) than its competitors. The worst performers in terms of efficiency are either the Tobit-type solutions (columns 4 and 6) or the estimators of simultaneous models (columns 3 and 5). As a summary of both criteria, the MSE clearly indicates the good performance of the MCECM estimator for the coefficients of the included exogenous variables (β_{11} and β_{21}) and, although not the best, its acceptable performance for the constant terms (β_{12} and β_{22}). To show the results graphically, Figure 5.1 presents boxplots for the estimation of these coefficients under the different methodologies. The true values of the coefficients are marked with a dashed line. These charts confirm that the MCECM solution has small estimation bias, since the median of the estimators are close to the true values of the coefficients. This does not come with a high cost in terms of loss of efficiency, since the estimation variance is in general comparable to that of the other estimators.

For the coefficients that determine the simultaneity, γ_1 and γ_2, the worst performer is the solution that corrects only the censorship in the endogenous variables: the TOB estimator (column 4). On the other hand, the MCECM estimator outperforms the rest in terms of bias; even though it has a higher variance due to its computation through simulation, this is moderate and this estimator also has the smallest MSE for both coefficients. These results are also included in Table 5.1 and illustrated through the boxplots of Figure 5.2, where the center of the distribution of the estimators lies closer to the true values of the coefficients (marked with a dashed line).

Concerning the variance-covariance matrix of the general error of the

Figure 5.1: Comparison of the estimation results of β_{11}, β_{12}, β_{21} and β_{22} for the simulation study of the random-effects model, base simulation

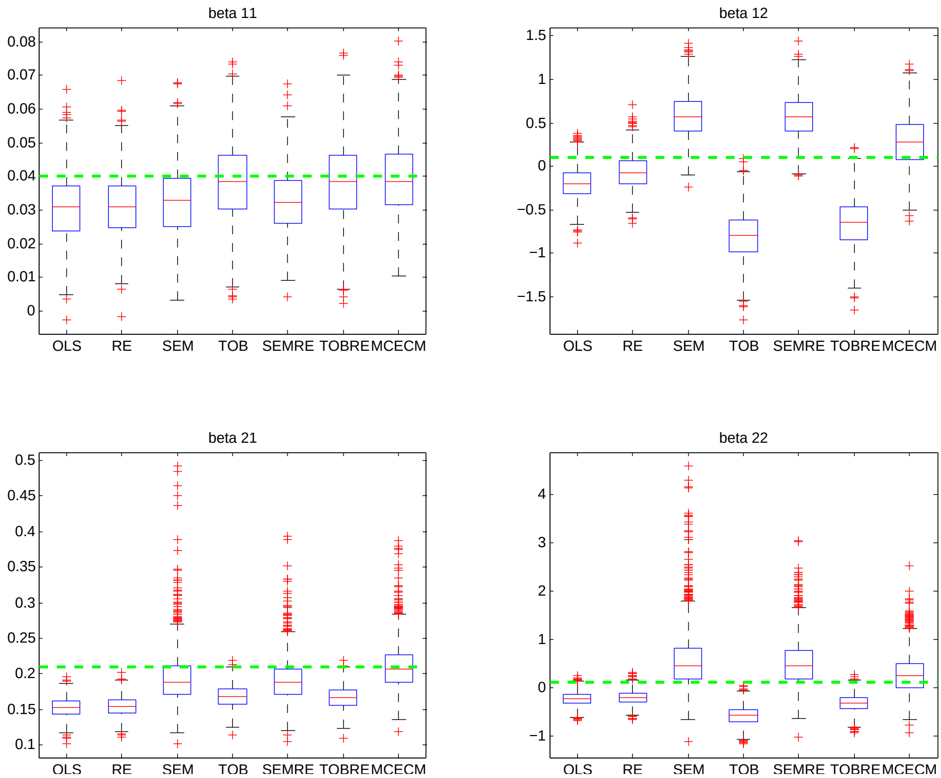
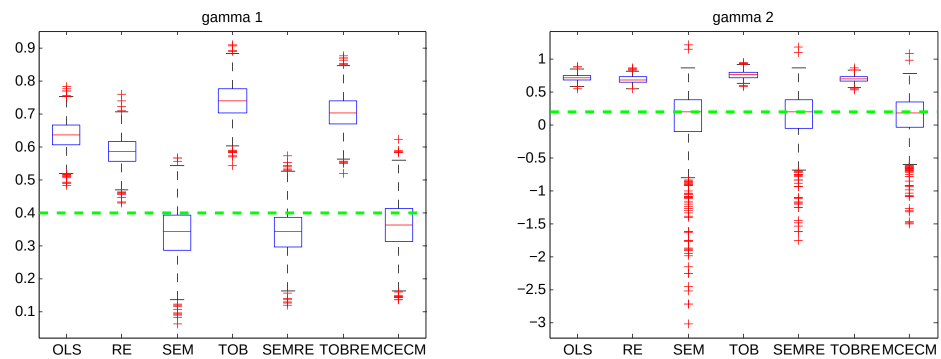


Figure 5.2: Comparison of the estimation results of γ_1 and γ_2 for the simulation study of the random-effects model, base simulation

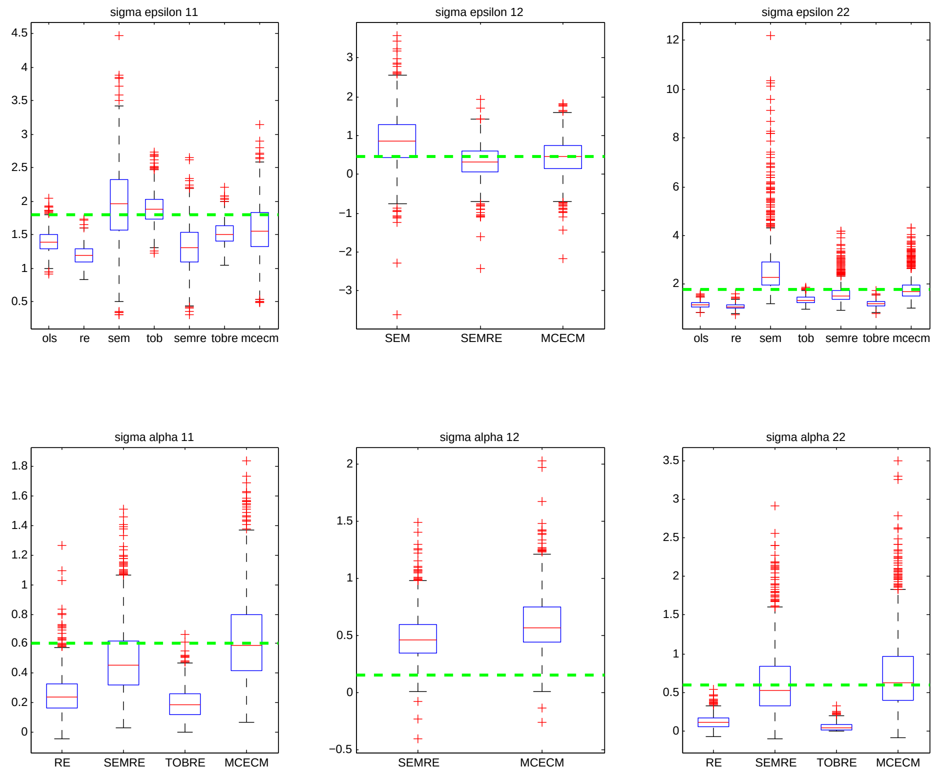


model (coefficients σ_{11}^{ϵ}, σ_{12}^{ϵ} and σ_{22}^{ϵ}), the most biased estimators for the variance terms are the GLS-RE and the SEM estimators (columns 2 and 3 of Table 5.1); the covariance is only estimated in those models that consider the simultaneous character of the model and, of these, the worst bias is found in the SEM estimator. The MCECM estimator has the best performance in terms of estimation bias for σ_{12}^{ϵ} and σ_{22}^{ϵ} and, although it does not lead to the smallest bias, it still has acceptable results for σ_{11}^{ϵ}. However, for the same reason stated previously, our proposed estimator has a higher variability which makes it less efficient (although not by a wide margin) than the competition. In terms of the MSE, the MCECM estimator does not outperform those of the best estimators (TOB for the first variance and SEM-RE for the second variance and the covariance) but it still produces good results considering its simulation nature.

The variance-covariance matrix of the individual random effect (coefficients σ_{11}^{α}, σ_{12}^{α} and σ_{22}^{α}) is only estimated through those methodologies that take into account the panel structure of the data. In this case, the MCECM estimator outperforms the others only for one coefficient (σ_{11}^{α}), that variable where the censorship is more severe. For the others, the best method is the SEM-RE estimator. This is likely due to the small sample size used in the experiment (the terms of the variance-covariance matrix of the individual components have to be estimated only through the $T = 7$ observations for each individual). The variability generated when drawing samples from this error component would explain its worse results.

As it was mentioned before, the limited character of the endogenous variables introduces nonlinearities in the model. For this reason, the coefficients calculated above do not reflect the effect that a change in the

Figure 5.3: Comparison of the estimation results of elements of the variance-covariance matrices of the error components for the simulation study of the random-effects model, base simulation



exogenous variable would have on the dependent variables and they were presented only as an initial evaluation of the performance of the estimation methodologies. Now we turn the comparison to the marginal effects derived in Chapter 3 and of more interest for this type of models. The results are summarized in Table 5.2 and Figure 5.4. We include the marginal effect of x_1 and x_2 on both endogenous variables; we could also calculate the effect of the common exogenous variable x but in this experiment we only include a constant term which, buy definition, does not change. Since the marginal effects also depend on the observed value of the exogenous and endogenous variables, they are different for each realization of the simulation experiment. Instead of calculating the marginal effects at the average value of the variables, we preferred to obtain the average marginal effects over all realizations. This is by far more resource consuming, but produces a wealth of information; in fact, we can obtain an empirical distribution of each marginal effect which could then be used for inference or to construct any statistic of interest. The average marginal effects evaluated at the true values of the coefficients are the following.

$$\text{Average marginal effect of } x_1 \text{ on } y_1 = 0.0423$$

$$\text{Average marginal effect of } x_2 \text{ on } y_1 = 0.0779$$

$$\text{Average marginal effect of } x_1 \text{ on } y_2 = 0.0065$$

$$\text{Average marginal effect of } x_2 \text{ on } y_2 = 0.2353$$

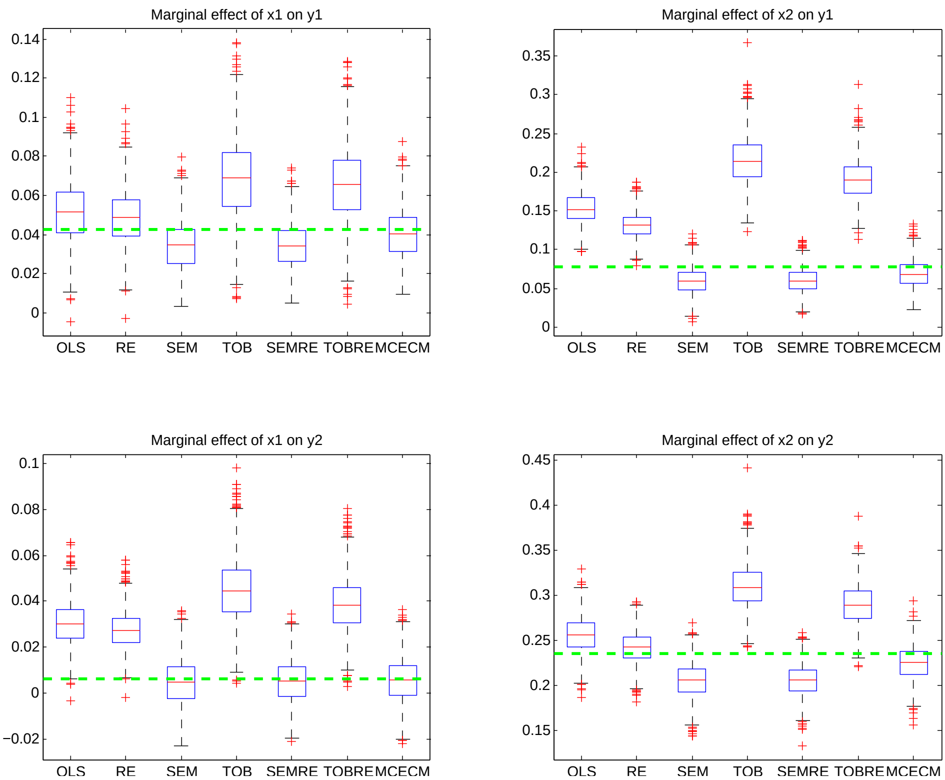
The included exogenous variables have an affect on their own equation very similar the coefficients they were assigned, but with corrections for the censorship in the dependent variables and the indirect effect through the other endogenous variable. The cross-effects (i.e., x_2 on y_1 and x_1 on y_2) are explained by the simultaneity between the two equations.

Table 5.2: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, base simulation

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	0.9167	0.6252	-0.8050	2.6427	-0.7718	2.3218	-0.1866
of x_1 on y_1	Var	0.0239	0.0196	0.0161	0.0412	0.0128	0.0369	0.0163
	MSE	0.0323	0.0235	0.0226	0.1110	0.0188	0.0908	0.0166
Marg. effect	Bias	7.5566	5.3669	-1.8034	13.7656	-1.7511	11.3261	-0.8940
of x_2 on y_1	Var	0.0406	0.0294	0.0304	0.0927	0.0247	0.0642	0.0322
	MSE	0.6116	0.3175	0.0629	1.9876	0.0553	1.3470	0.0402
Marg. effect	Bias	2.4070	2.1030	-0.1487	3.8441	-0.1231	3.2111	-0.0932
of x_1 on y_2	Var	0.0091	0.0068	0.0099	0.0203	0.0081	0.0143	0.0089
	MSE	0.0670	0.0510	0.0101	0.1681	0.0082	0.1174	0.0090
Marg. effect	Bias	2.0653	0.6706	-2.9689	7.4509	-2.9951	5.4299	-1.0364
of x_2 on y_2	Var	0.0401	0.0320	0.0360	0.0659	0.0293	0.0501	0.0343
	MSE	0.0827	0.0365	0.1242	0.6210	0.1190	0.3449	0.0451

In terms of bias, the classical estimator for censored dependent variables, TOB and TOB-RE (columns 4 and 6) present the worst performance for all the marginal effects since they usually overestimate the effect that changes in the exogenous factors would cause in the dependent variables. For all but the last one, the MCECM estimator has the smallest bias of all competing solutions. Although this comes at a cost of a higher variability, it can be seen that this loss of efficiency is limited and generally comparable to the most efficient estimators (and always much better than the worst performers in terms of efficiency, also TOB and TOB-RE). Using the MSE as a summary of estimation performance, our proposed solution produces the best results for the first endogenous variable and comes at a close second place for the second variable (where the censorship is less pronounced). These same results are presented graphically in the boxplots of Figure 5.4. In this graph, the dashed line correspond to the average marginal effects calculated at the true value of the parameters. Here it is straightforward to verify the good performance of the MCECM estimator.

Figure 5.4: Comparison of the average marginal effects for the simulation study of the random-effects model, base simulation



In conclusion, our proposed methodology has a good performance when compared to the competing solutions. Even though it is an estimator based on simulation, with an intrinsic variability explained by the random drawing from probabilistic distributions, it remains remarkably efficient and it converges very close to the true parameters, generally ranking as the best estimator in terms of the MSE criterion both when studying the separate coefficients of the model and the average marginal effects.

Even though this experiment was designed to mimic the results of an empirical study, namely a landmark estimation in the limited-dependent variables literature, there were still many characteristics that were chosen by the researcher; for example, the sample size, the distribution used to draw the exogenous variables, the density of the error term, etc. To evaluate the robustness of the different estimator to different DGPs, the following section will present variations to the base model presented above. To save space, comparisons will be based only on the marginal effects, since these would be the quantities of interest in any applied study. The results from each of the different simulation exercises will be briefly presented in what follows, before a concluding section summarizes the main findings.

5.1.1 Robustness to different distributions of the exogenous variables

The exogenous variables x_1 and x_2 of the base simulation were generated through a uniform distribution. We will now explore the performance of the estimators when other densities are used to generate these variables. We will start with variables generated through a normal distribution. In comparison to the uniform distribution, in this case values are more concentrated around a central value, but there is also the possibility to obtain extreme observations. In particular, we generate x_1 from a normal distribution with mean 15 and standard deviation 6; this means that 93.9% of the draws will theoretically fall in the interval $[0, 30]$. For the second exogenous variables we use an independent normal distribution with mean 10 and standard deviation 5, so that 95.4% of the draws will be between zero and 20. The rest of the configuration is unchanged with respect to the original experiment. In this case, since the mean of the exogenous variables does not change and they also follow a symmetric distribution, the percentage of censorship remains close to that observed in the base simulation. The results are summarized in Table 5.3 and Figure 5.5.

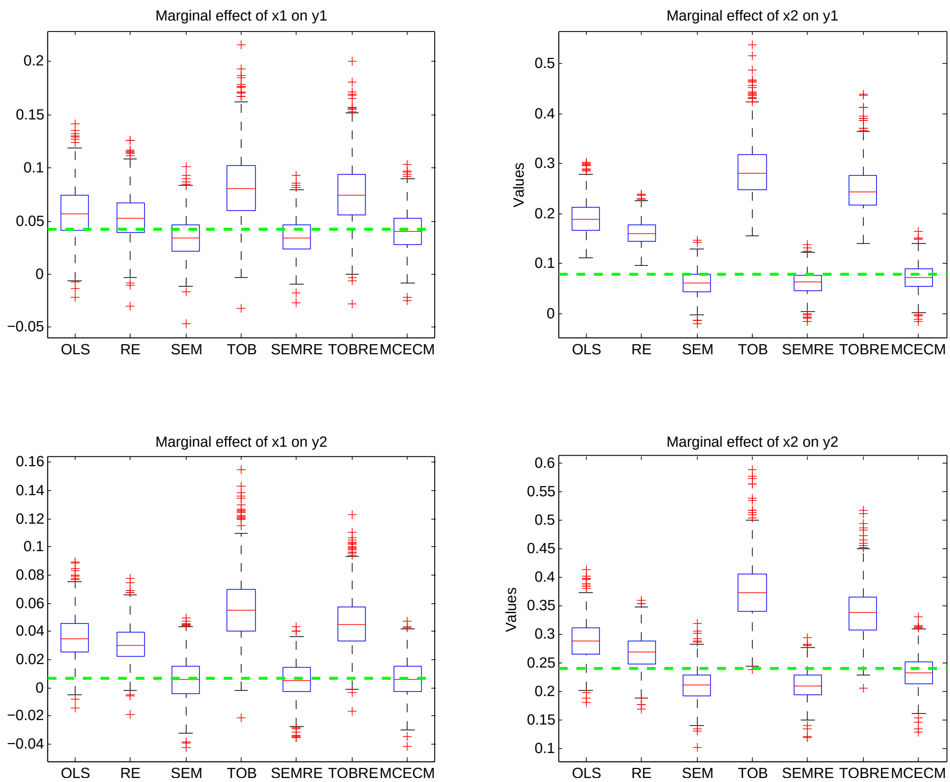
Table 5.3: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, exogenous variables generated with a normal distribution

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	1.5356	1.0782	-0.7318	3.9033	-0.7106	3.3076	-0.1269
of x_1 on y_1	Var	0.0589	0.0474	0.0322	0.1137	0.0252	0.0967	0.0323
	MSE	0.0825	0.0590	0.0376	0.2660	0.0302	0.2061	0.0325
Marg. effect	Bias	11.2749	8.2953	-1.6153	20.7076	-1.5667	16.9836	-0.6325
of x_2 on y_1	Var	0.1006	0.0617	0.0644	0.2828	0.0537	0.1892	0.0690
	MSE	1.3718	0.7498	0.0905	4.5708	0.0783	3.0736	0.0730
Marg. effect	Bias	2.9066	2.4537	-0.0895	4.9491	-0.0698	3.9346	-0.0273
of x_1 on y_2	Var	0.0233	0.0169	0.0214	0.0594	0.0172	0.0380	0.0182
	MSE	0.1078	0.0771	0.0215	0.3043	0.0173	0.1928	0.0182
Marg. effect	Bias	5.0007	2.8209	-2.8885	13.6661	-2.9035	9.8836	-0.7269
of x_2 on y_2	Var	0.1174	0.0865	0.0778	0.2513	0.0675	0.1779	0.0826
	MSE	0.3675	0.1661	0.1613	2.1189	0.1518	1.1548	0.0879

The results are very similar to those of the base simulation. The TOB and TOB-RE are still the worse estimators in terms of bias and variance. The classic random-effects estimator performs marginally worse, while the SEM-RE is improved in this simulation. Our estimator beats the rest in terms of bias for the four marginal effect and, in spite of showing a relatively higher variance, it is still the minimum MSE estimator of the seven for two effects and a close second best performer for the other two.

We now generate the exogenous variable through a log-normal distribution, with the objective of studying the performance on the estimation methods when the variables have a long right-hand tail, similar to what is observed in variables such as income. x_1 was generated with a log-distribution with parameters 2.7 (equivalent to a mean of 14.9 on a normal distribution) and 0.6, while the second variable was obtained from the same

Figure 5.5: Comparison of the average marginal effects for the simulation study of the random-effects model, exogenous variables generated with a normal distribution



distribution but with parameters 2.3 (which leads to a mean of 10.0 in a normal distribution) and 0.4. The rest of the parameters and the characteristics of the experiment was the same as in the base simulation. Under this configuration, the average censorship fell to 18.4% and 8.2% of the first and second dependent variables, respectively.

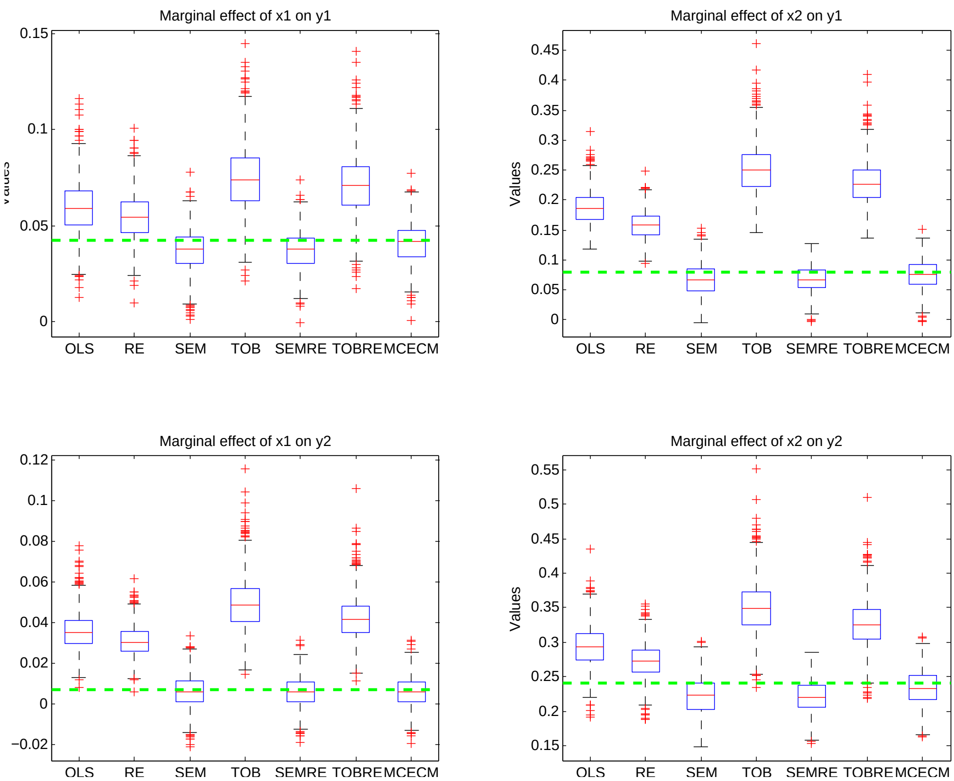
Table 5.4: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, exogenous variables generated with a log-normal distribution

		(1) OLS	(2) GLSRE	(3) SEM	(4) TOB	(5) SEMRE	(6) TOBRE	(7) MCECM
Marg. effect	Bias	1.6902	1.2014	-0.5197	3.2031	-0.5140	2.8352	-0.1527
of x_1 on y_1	Var	0.0193	0.0145	0.0116	0.0292	0.0092	0.0261	0.0100
	MSE	0.0479	0.0290	0.0143	0.1318	0.0118	0.1065	0.0102
Marg. effect	Bias	10.7396	7.8601	-1.2005	17.2282	-1.1653	14.9638	-0.3995
of x_2 on y_1	Var	0.0770	0.0477	0.0644	0.1652	0.0487	0.1259	0.0568
	MSE	1.2304	0.6655	0.0788	3.1333	0.0622	2.3650	0.0584
Marg. effect	Bias	2.9155	2.4115	-0.0957	4.2574	-0.0931	3.5461	-0.0789
of x_1 on y_2	Var	0.0084	0.0053	0.0066	0.0169	0.0053	0.0118	0.0055
	MSE	0.0934	0.0635	0.0067	0.1982	0.0054	0.1375	0.0055
Marg. effect	Bias	5.2739	3.1550	-1.8983	10.9127	-1.9457	8.4775	-0.6797
of x_2 on y_2	Var	0.0931	0.0654	0.0727	0.1505	0.0550	0.1237	0.0600
	MSE	0.3712	0.1650	0.1087	1.3414	0.0929	0.8424	0.0646

The results in Table 5.4 show that the performance of estimator remains relatively stable under this DGP. The best solution in terms of bias is the MCECM estimator for all four marginal effects. Even taking into account its higher variability, it remains the best estimator in terms of MSE with the exception of the effect of x_1 on y_2, where it is larger than the best performer by only 2.8%. The second best solution are the SEM-RE and the SEM estimator. GLS-RE, OLS and particularly the Tobit estimators do not lead to a good estimation. For both this and the previous simulation, the average marginal effect evaluated at the true values of the coefficients remains practically unchanged from those of the base simulation.

As a third and final exploration of the effect of different DGP of the

Figure 5.6: Comparison of the average marginal effects for the simulation study of the random-effects model, exogenous variables generated with a log-normal distribution



exogenous variables, we study a situation where one of the exogenous variables is dichotomous. In particular, the exogenous variable included only in the second variable is drawn from a Bernoulli distribution with parameter 0.5 (i.e, the probability to observe an outcome of 1 is equal to 50%). The parameters for this equation had to be adjusted accordingly: $\beta_{21} = 2.16$, the coefficient of the variable "union" of the same study we are mimicking, while the constant term β_{22} was set to 1. With these values, we obtain a percentage of observed endogenous variables equal to zero of 20.8% and 12.7%, very similar to those of the base simulation.

The marginal effects summarized in Table 5.5 had to be modified to accommodate the fact that one of the exogenous variables is dichotomous. In fact, the marginal effect is not defined for a discrete variable. However, as it was described in Chapter 3, an equivalent measure is the change in the dependent variable when the dummy variable changes from one state to the other; the relevant formulas are (3.27) and (3.28). The average effects evaluated at the true value of the coefficients are given by

$$\begin{aligned} \text{Average marginal effect of } x_1 \text{ on } y_1 &= 0.0422 \\ \text{Average effect of } x_2 \text{ on } y_1 &= 0.6061 \\ \text{Average marginal effect of } x_1 \text{ on } y_2 &= 0.0065 \\ \text{Average effect of } x_2 \text{ on } y_2 &= 2.0133 \end{aligned}$$

The introduction of a dummy exogenous variable does not alter the order of performance of the estimators. For all cases, our proposed estimator exhibits the minimal estimation bias and a variability that, although higher

Table 5.5: Comparison of the average effects for the simulation study of the random-effects model, values multiplied by 10^2, one dummy exogenous variable

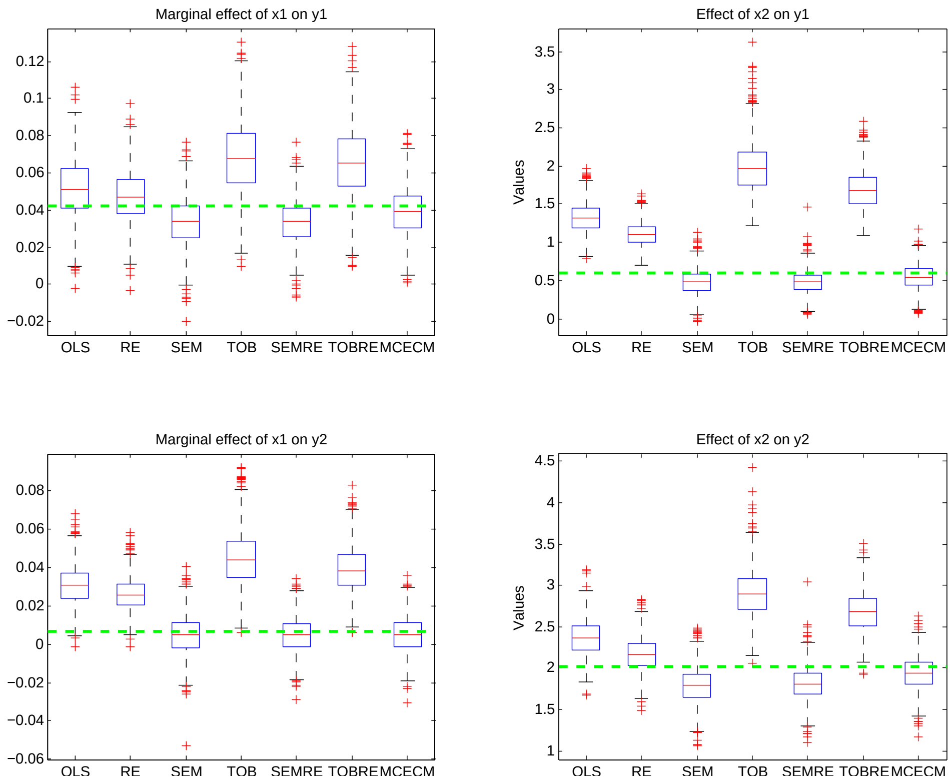
		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	0.9359	0.4919	-0.8566	2.6049	-0.8660	2.3288	-0.3066
of x_1 on y_1	Var	0.0248	0.0195	0.0170	0.0402	0.0136	0.0362	0.0171
	MSE	0.0335	0.0220	0.0243	0.1080	0.0211	0.0905	0.0181
Marg. effect	Bias	72.0528	50.2288	-12.8383	137.8808	-12.0927	108.5027	-6.2406
of x_2 on y_1	Var	3.4900	2.3872	2.6845	10.5316	2.2022	5.9076	2.5041
	MSE	55.4060	27.6164	4.3327	200.6427	3.6645	123.6359	2.8935
Marg. effect	Bias	2.4415	1.9558	-0.1808	3.8680	-0.1855	3.2535	-0.1560
of x_1 on y_2	Var	0.0097	0.0067	0.0103	0.0205	0.0086	0.0144	0.0095
	MSE	0.0693	0.0449	0.0106	0.1701	0.0090	0.1203	0.0098
Marg. effect	Bias	34.6573	14.3597	-22.1525	89.4279	-20.4080	67.0748	-8.3034
of x_2 on y_2	Var	4.7468	4.1694	4.5702	8.1101	4.0097	5.5679	3.9180
	MSE	16.7581	6.2314	9.4775	88.0835	8.1746	50.5582	4.6074

because of its calculation through simulation, is comparable to that of the best performer. Using the MSE as a summary comparison measure, the MCECM estimator is the best except for the average marginal effect of x_1 on y_2, where it it is outperformed only by the SEM-RE estimator. The rest of the solutions, particularly TOB and TOB-RE, show poorer results. The good results from the estimator derived in this thesis can also be verified in Figure 5.7.

5.1.2 Robustness to different distributions of the error term

The error terms ϵ_1 and ϵ_2 in the original configuration were generated through a normal distribution with a certain variance-covariance matrix, according to (5.3). The likelihood function was constructed around this distribution and it would be expected that any deviation from this would undermine the performance of the MCECM estimator (but also of the competing solutions). To explore this, we will run the simulation exercise under two alternative densities of the error term.

Figure 5.7: Comparison of the average effects for the simulation study of the random-effects model, one dummy exogenous variable



First, we change the density of the errors to a mixture of normals. Specifically, the marginal density of the errors is the following.

$$\begin{pmatrix} \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \overset{iid}{\sim} (1-\omega)\cdot \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.80 & 0.45 \\ 0.45 & 1.80 \end{pmatrix}\right] + \omega \cdot \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 18.0 & 4.5 \\ 4.5 & 18.0 \end{pmatrix}\right]$$

with $\omega = 0.05$. In other words, the original distribution is "contaminated" with a normal density with higher variance (10 times that of the original density).

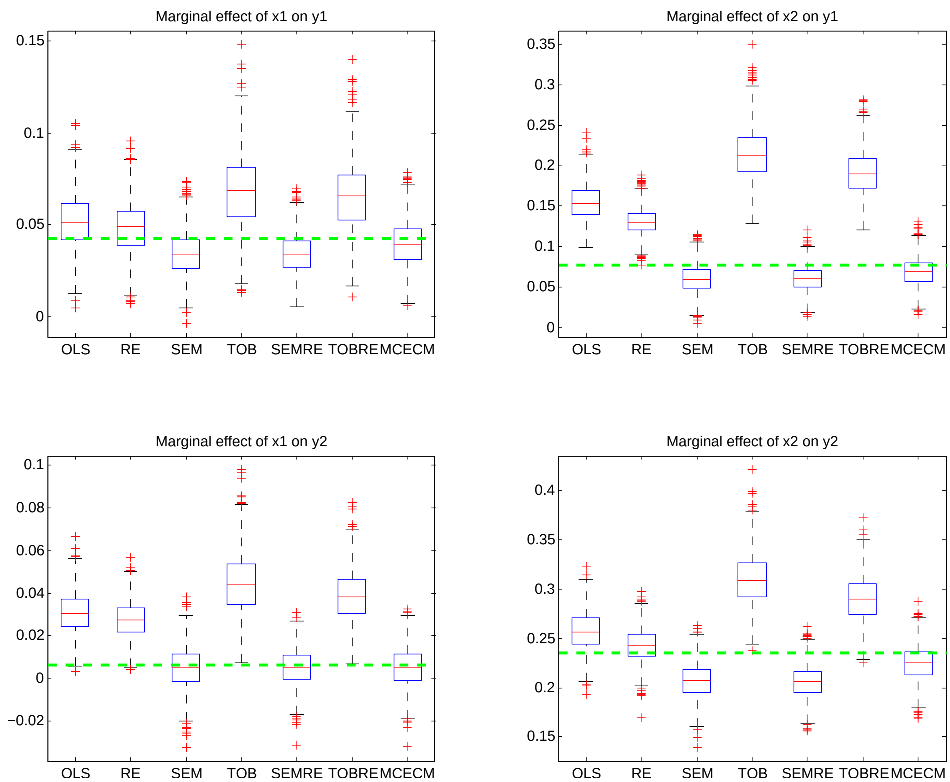
Table 5.6: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, error terms drawn from a mixture of normal distributions

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	0.9302	0.5918	-0.8006	2.5988	-0.8103	2.2802	-0.2582
of x_1 on y_1	Var	0.0227	0.0186	0.0150	0.0391	0.0119	0.0353	0.0150
	MSE	0.0314	0.0221	0.0214	0.1066	0.0184	0.0872	0.0157
Marg. effect	Bias	7.6985	5.3415	-1.7362	13.7217	-1.7194	11.3334	-0.8767
of x_2 on y_1	Var	0.0444	0.0290	0.0297	0.1010	0.0242	0.0696	0.0305
	MSE	0.6370	0.3143	0.0599	1.9839	0.0538	1.3541	0.0382
Marg. effect	Bias	2.4293	2.0972	-0.1363	3.8071	-0.1428	3.1963	-0.1125
of x_1 on y_2	Var	0.0089	0.0066	0.0093	0.0195	0.0073	0.0139	0.0082
	MSE	0.0680	0.0506	0.0094	0.1645	0.0075	0.1161	0.0083
Marg. effect	Bias	2.2371	0.7330	-2.8581	7.4446	-2.9140	5.4572	-1.0270
of x_2 on y_2	Var	0.0387	0.0287	0.0322	0.0664	0.0263	0.0500	0.0304
	MSE	0.0888	0.0340	0.1139	0.6207	0.1112	0.3478	0.0409

Table 5.6 and Figure 5.8 shows that the marginal effects remain stable even if the errors are generated with the contaminated distribution. As it was shown in Chapter 3, these expressions are complex, nonlinear functions of all the parameters in the model. The variance of the error terms appears in practically all the factors that compose the marginal effects, in the numerator and the denominator, with negative and positive sign; they frequently affect the expression through nonlinear expressions (like through a normal density function or a normal cumulative function). Because of this, the expected effect was ambiguous a priori. The previous table suggest that

the effect of a higher variance is at least partially mitigated in the calculation of the marginal effects and the results are not significantly affected. As in the base simulation, the best estimator in terms of bias is the MCECM solution for all but the last marginal effect; in terms of MSE it outperforms the others or it is in a close second position. As before, the worst results come from the TOB-RE and especially the TOB methodologies.

Figure 5.8: Comparison of the average marginal effects for the simulation study of the random-effects model, error terms drawn from a mixture of normal distributions



Even though they are of limited interest in an empirical application, in Table 5.7 we show the estimation results for the coefficients of equations (5.1) and (5.2). Even if the true value of the terms of the variance-covariance matrix of the error terms change under the contaminated distribution, we do not change it for the purpose of the calculation of the bias and the MSE in this table, to approximate a real-life situation where a researcher faces a contaminated distribution but has interest in estimating the parameters of the "true", uncontaminated distribution. We can conclude that the contamination has an detrimental effect in the estimation through a higher estimation variance; however, this affects all estimators equally and the ranking of results is in general not affected. The effect is more evident in the estimation of σ^{ϵ}_{11}, σ^{ϵ}_{12} and σ^{ϵ}_{22} where both the estimation bias and variance are larger since the true parameters are affected under the mixture or normals.

Second, we generate the error terms through an asymmetric distribution to increase the probability of having extreme values on one side of the distribution and study its effect on the estimation. For this, we draw the error terms from a bivariate skew-normal distribution following the procedure suggested by Azzalini and Capitanio (1999). We want to generate draws from a bivariate skew-normal distribution based on the marginal density of the error terms in (5.3) but with skew parameters s_1 and s_2. With this in mind, we draw random values from the following multivariate normal distribution

$$\begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & s_1\sqrt{1.80} & s_2\sqrt{1.80} \\ s_1\sqrt{1.80} & 1.80 & 0.45 \\ s_2\sqrt{1.80} & 0.45 & 1.80 \end{pmatrix}\right]$$

Table 5.7: Comparison of the estimation results for the simulation study of the random-effects model, error terms drawn from a mixture of normal distributions

		(1) OLS	**(2)** GLS-RE	**(3)** SEM	**(4)** TOB	**(5)** SEM-RE	**(6)** TOB-RE	**(7)** MCECM
	Bias	-0.0093	-0.0091	-0.0074	-0.0021	-0.0075	-0.0018	-0.0014
β_{11}	Variance	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	MSE	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
	Bias	-0.3042	-0.1629	0.4662	-0.8802	0.4683	-0.7410	0.1879
β_{12}	Variance	0.0296	0.0323	0.0684	0.0714	0.0605	0.0686	0.0863
	MSE	0.1221	0.0588	0.2857	0.8461	0.2798	0.6177	0.1216
	Bias	-0.0568	-0.0548	-0.0135	-0.0416	-0.0159	-0.0434	0.0013
β_{21}	Variance	0.0002	0.0002	0.0018	0.0002	0.0010	0.0002	0.0011
	MSE	0.0034	0.0032	0.0019	0.0019	0.0013	0.0021	0.0011
	Bias	-0.3415	-0.3081	0.4577	-0.6708	0.4202	-0.4315	0.1885
β_{22}	Variance	0.0192	0.0183	0.5121	0.0396	0.2940	0.0367	0.2006
	MSE	0.1358	0.1132	0.7216	0.4895	0.4706	0.2229	0.2361
	Bias	0.2342	0.1806	-0.0615	0.3344	-0.0613	0.3010	-0.0400
γ_1	Variance	0.0021	0.0021	0.0060	0.0031	0.0051	0.0029	0.0057
	MSE	0.0569	0.0347	0.0098	0.1149	0.0089	0.0935	0.0073
	Bias	0.5164	0.4894	-0.1066	0.5582	-0.0773	0.5004	-0.0859
γ_1	Variance	0.0023	0.0024	0.2817	0.0027	0.1538	0.0022	0.1279
	MSE	0.2689	0.2419	0.2931	0.3143	0.1598	0.2526	0.1352
	Bias	-0.4753	-0.6673	0.0749	-0.0214	-0.5708	-0.3717	-0.3375
σ^{ϵ}_{11}	Variance	0.0228	0.0186	0.2810	0.0456	0.0972	0.0268	0.1362
	MSE	0.2487	0.4639	0.2866	0.0460	0.4229	0.1649	0.2501
	Bias			0.3683		-0.1469		-0.0385
σ^{ϵ}_{12}	Variance			0.4894		0.1615		0.1919
	MSE			0.6251		0.1831		0.1934
	Bias	-0.7180	-0.7787	0.7250	-0.5316	-0.2595	-0.6823	-0.1071
σ^{ϵ}_{22}	Variance	0.0126	0.0131	4.4856	0.0184	0.2383	0.0178	0.2750
	MSE	0.5281	0.6195	5.0112	0.3010	0.3057	0.4833	0.2865
	Bias		-0.3410			-0.1150	-0.4063	0.0240
σ^{α}_{11}	Variance		0.0189			0.0550	0.0103	0.0886
	MSE		0.1352			0.0682	0.1754	0.0892
	Bias					0.3342		0.4510
σ^{α}_{12}	Variance					0.0405		0.0580
	MSE					0.1522		0.2614
	Bias		-0.4878			0.0281	-0.5546	0.1346
σ^{α}_{22}	Variance		0.0060			0.2398	0.0019	0.3017
	MSE		0.2440			0.2406	0.3095	0.3198

If we apply the following decision rule

$$\begin{pmatrix} \tilde{\epsilon}_{1it} \\ \tilde{\epsilon}_{2it} \end{pmatrix} = \begin{cases} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} & \text{if } w_0 \geq 0 \\ \begin{pmatrix} -w_1 \\ -w_2 \end{pmatrix} & \text{if } w_0 < 0 \end{cases}$$

$(\tilde{\epsilon}_{1it}, \tilde{\epsilon}_{2it})$ is drawn from a skew-normal distribution with the given variance-covariance matrix and skew parameters s_1 and s_2. We choose $s_1 = 0.75$ and $s_2 = 0.80$ to have different degrees of skewness for each of the error terms.

Table 5.8: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, error terms drawn from a skew-normal distribution

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	1.1366	0.8431	-0.7730	2.9235	-0.7780	2.5398	-0.1736
of x_1 on y_1	Var	0.0269	0.0223	0.0174	0.0448	0.0146	0.0400	0.0180
	MSE	0.0398	0.0294	0.0233	0.1303	0.0207	0.1045	0.0183
Marg. effect	Bias	8.4493	6.5360	-1.8420	14.6388	-1.7976	11.8543	-0.7953
of x_2 on y_1	Var	0.0284	0.0149	0.0330	0.0687	0.0268	0.0452	0.0330
	MSE	0.7423	0.4421	0.0670	2.2117	0.0591	1.4505	0.0393
Marg. effect	Bias	2.6349	2.3440	-0.1579	4.1265	-0.1608	3.4158	-0.1143
of x_1 on y_2	Var	0.0102	0.0076	0.0106	0.0216	0.0093	0.0150	0.0101
	MSE	0.0796	0.0625	0.0108	0.1919	0.0095	0.1317	0.0102
Marg. effect	Bias	2.9275	1.6628	-2.7883	8.4052	-2.8291	6.0756	-0.6903
of x_2 on y_2	Var	0.0327	0.0245	0.0360	0.0531	0.0304	0.0421	0.0342
	MSE	0.1184	0.0521	0.1138	0.7596	0.1104	0.4112	0.0389

Similar to what we verified in the preceding exercise, the introduction of a skew distribution to the error terms has little consequences on the marginal effects, as evidenced by Table 5.8 and Figure 5.9. The MCECM estimator has the minimal bias and MSE among the seven alternatives for three of the average marginal effects; for the effect of x_1 on the second endogenous variable, this estimator has the smallest bias but its higher variance makes it slightly worse than the SEM-RE estimator. For all the cases, the classical Tobit and the TOB-RE produce the worst estimation results. To explore the effect of the skew distribution in the estimation, we have to study the results of the original coefficients of the model. These are presented in Table 5.9. As we did before, we do not modify the true value of the variance-covariance matrix of (ϵ_1, ϵ_2) to consider a case in which the researcher deals with data from a skewed distribution but has interest in the "core" observations. We can observe a deterioration in the estimation, in terms of loss of efficiency, particularly for the parameters of the second equation (that with the highest skewness); compare for example the variance of β_{22} and γ_2 with the results from the base simulation in Table 5.1. This affects all the estimation methodologies. The same result can be verified for the other parameters of the second equation, σ^{ϵ}_{22} and σ^{α}_{22}, and to a lesser degree in the covariance terms.

5.1.3 Robustness to different functional forms

We tested the performance of the estimation methods to different distributions generating the exogenous or the error terms, now we maintain the configuration of the base simulation but we change the functional form of the model. First, we study the case where the functional form is quadratic in the exogenous variables. So, instead of (5.1) and (5.2), we now have the following model.

$$y^*_{1it} = \gamma_1 y_{2it} + \beta_{11} x^2_{1it} + \beta_{12} x^2_{it} + \alpha_{1i} + \epsilon_{1it}$$
$$y^*_{2it} = \gamma_2 y_{1it} + \beta_{21} x^2_{2it} + \beta_{22} x^2_{it} + \alpha_{2i} + \epsilon_{2it}$$

The two exogenous variables x_1 and x_2 were generated as in the base simulation from uniform distributions, but were divided by 4 in order to maintain a degree of censorship similar to the base simulation and make

Figure 5.9: Comparison of the average marginal effects for the simulation study of the random-effects model, error terms drawn from a skew-normal distribution

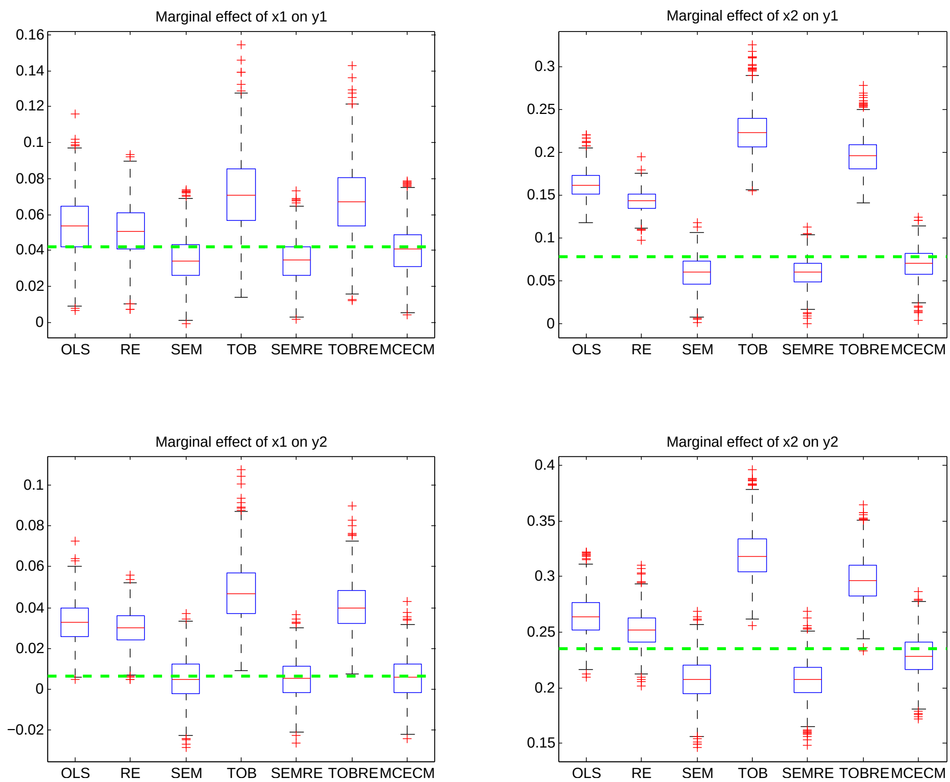


Table 5.9: Comparison of the estimation results for the simulation study of the random-effects model, error terms drawn from a skew-normal distribution

		(1) OLS	**(2)** GLS-RE	**(3)** SEM	**(4)** TOB	**(5)** SEM-RE	**(6)** TOB-RE	**(7)** MCECM
β_{11}	Bias	-0.0089	-0.0087	-0.0069	-0.0007	-0.0070	-0.0005	-0.0004
	Variance	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	MSE	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
β_{12}	Bias	-0.3416	-0.2388	0.4748	-0.9298	0.4723	-0.7780	0.1671
	Variance	0.0278	0.0284	0.0751	0.0635	0.0638	0.0602	0.0873
	MSE	0.1445	0.0854	0.3005	0.9280	0.2868	0.6654	0.1152
β_{21}	Bias	-0.0564	-0.0548	-0.0089	-0.0384	-0.0106	-0.0409	0.0076
	Variance	0.0002	0.0002	0.0037	0.0002	0.0019	0.0002	0.0015
	MSE	0.0034	0.0032	0.0038	0.0017	0.0021	0.0019	0.0016
β_{22}	Bias	-0.3916	-0.3652	0.4856	-0.7522	0.4587	-0.4912	0.1494
	Variance	0.0172	0.0159	1.0440	0.0346	0.5404	0.0321	0.2263
	MSE	0.1706	0.1493	1.2798	0.6004	0.7509	0.2733	0.2486
γ_1	Bias	0.2480	0.2089	-0.0669	0.3433	-0.0654	0.3073	-0.0391
	Variance	0.0014	0.0011	0.0069	0.0021	0.0057	0.0020	0.0061
	MSE	0.0629	0.0447	0.0114	0.1200	0.0100	0.0964	0.0077
γ_2	Bias	0.5341	0.5128	-0.1544	0.5734	-0.1338	0.5108	-0.1132
	Variance	0.0011	0.0008	0.6757	0.0014	0.3396	0.0012	0.1727
	MSE	0.2864	0.2638	0.6995	0.3303	0.3575	0.2621	0.1855
σ^{ϵ}_{11}	Bias	-0.3907	-0.5585	0.2259	0.1131	-0.4234	-0.2463	-0.1844
	Variance	0.0142	0.0104	0.3525	0.0263	0.1418	0.0122	0.1749
	MSE	0.1669	0.3223	0.4036	0.0391	0.3211	0.0728	0.2089
σ^{ϵ}_{12}	Bias			0.5404		-0.0052		0.0649
	Variance			0.6030		0.2074		0.2381
	MSE			0.8950		0.2074		0.2423
σ^{ϵ}_{22}	Bias	-0.6362	-0.6917	1.2112	-0.4403	0.0554	-0.5643	0.1869
	Variance	0.0035	0.0047	10.5216	0.0049	0.6071	0.0051	0.4527
	MSE	0.4082	0.4832	11.9885	0.1988	0.6102	0.3235	0.4876
σ^{α}_{11}	Bias		-0.3830			-0.1373	-0.4150	0.0114
	Variance		0.0124			0.0504	0.0112	0.0858
	MSE		0.1591			0.0692	0.1834	0.0859
σ^{α}_{12}	Bias					0.3335		0.4645
	Variance					0.0453		0.0661
	MSE					0.1566		0.2819
σ^{α}_{22}	Bias		-0.4799			0.0796	-0.5509	0.1974
	Variance		0.0067			0.3206	0.0024	0.3469
	MSE		0.2371			0.3270	0.3060	0.3859

it easier to do the comparison. The average incidence of censorship of the first and second dependent variable are now 20.8% and 18.5%, respectively. The results from the simulation exercise are presented in Table 5.10 and Figure 5.10.

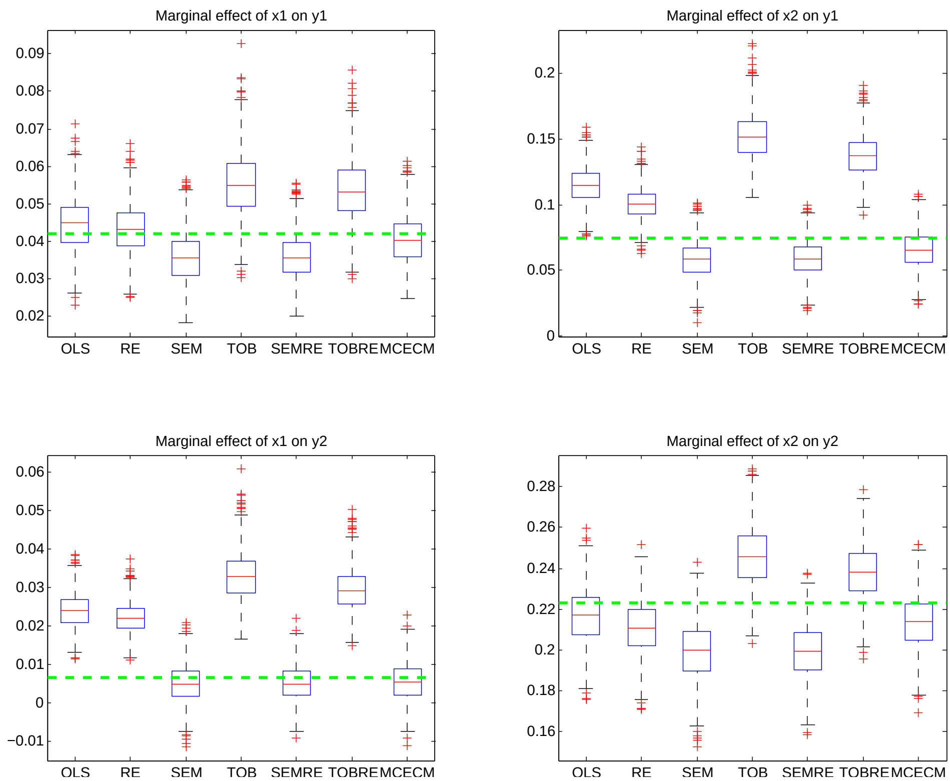
Table 5.10: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, functional form quadratic in the exogenous variables

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	0.2586	0.1193	-0.6606	1.3006	-0.6450	1.1433	-0.1604
of x_1 on y_1	Var	0.0051	0.0044	0.0047	0.0073	0.0038	0.0067	0.0043
	MSE	0.0058	0.0045	0.0091	0.0242	0.0079	0.0198	0.0046
Marg. effect	Bias	4.0497	2.6797	-1.5613	7.7897	-1.5197	6.2833	-0.8424
of x_2 on y_1	Var	0.0174	0.0140	0.0209	0.0347	0.0171	0.0236	0.0198
	MSE	0.1814	0.0858	0.0452	0.6415	0.0402	0.4184	0.0269
Marg. effect	Bias	1.7432	1.5577	-0.1458	2.6440	-0.1337	2.2989	-0.0824
of x_1 on y_2	Var	0.0018	0.0015	0.0024	0.0038	0.0020	0.0028	0.0024
	MSE	0.0322	0.0257	0.0026	0.0737	0.0022	0.0556	0.0025
Marg. effect	Bias	-0.6466	-1.2684	-2.3850	2.2346	-2.3856	1.4713	-0.9585
of x_2 on y_2	Var	0.0182	0.0164	0.0217	0.0223	0.0183	0.0195	0.0181
	MSE	0.0224	0.0325	0.0786	0.0722	0.0752	0.0411	0.0273

As expected, the increase in the dispersion of the exogenous variables due to the transformation has the inverse effect in the variance of the estimators. This can be verified when comparing the boxplots below to those of the original simulation. However, the comparative performance of the estimator does not change significantly. The MCECM solution is the best for three of the marginal effects in terms of bias and for two of them in terms of the MSE; for the rest, it is always a close second best. Under this functional form, GLS-RE in particular ameliorates its performance, while TOB and TOB-RE are still the estimators with the worse performance.

As a second test of the effect of changing the functional form, we now estimate equations that are logarithmic in the exogenous variables, in the

Figure 5.10: Comparison of the average marginal effects for the simulation study of the random-effects model, functional form quadratic in the exogenous variables



following form

$$y_{1it}^* = \gamma_1 y_{2it} + \beta_{11} \log x_{1it} + \beta_{12} x_{it} + \alpha_{1i} + \epsilon_{1it}$$
$$y_{2it}^* = \gamma_2 y_{1it} + \beta_{21} \log x_{2it} + \beta_{22} x_{it} + \alpha_{2i} + \epsilon_{2it}$$

Since the common exogenous variable only includes the constant term in this simulation, we do not apply the logarithmic function to it. As a consequence of the logarithmic transformation, the proportion of dependent cases that are censored reduces slightly to 17.6% and 16.0%, respectively. Figure 5.11 shows that, since the logarithm reduces the dispersion of the exogenous variables, the variability of the estimation augments considerably for all estimation methods. In fact, in the calculation of the results for Table 5.11 the bias remain stable in comparison to the base configuration but the variance dominates the calculation of the MSE. This increased variability is amplified in the calculation of our estimator since it is based on simulation. As a result, while it is still the solution with the lowest bias for all four marginal effects, its higher variance punishes its MSE; the best estimator according to this criterion is now SEM-RE followed by either GLS-RE or SEM. In fact the only methods that are worse than the MCECM are the two Tobit estimators. Even a basic ordinary least squares outperforms our estimator in two cases. This shows the effect of a higher estimation variability and how it can be amplified by simulation.

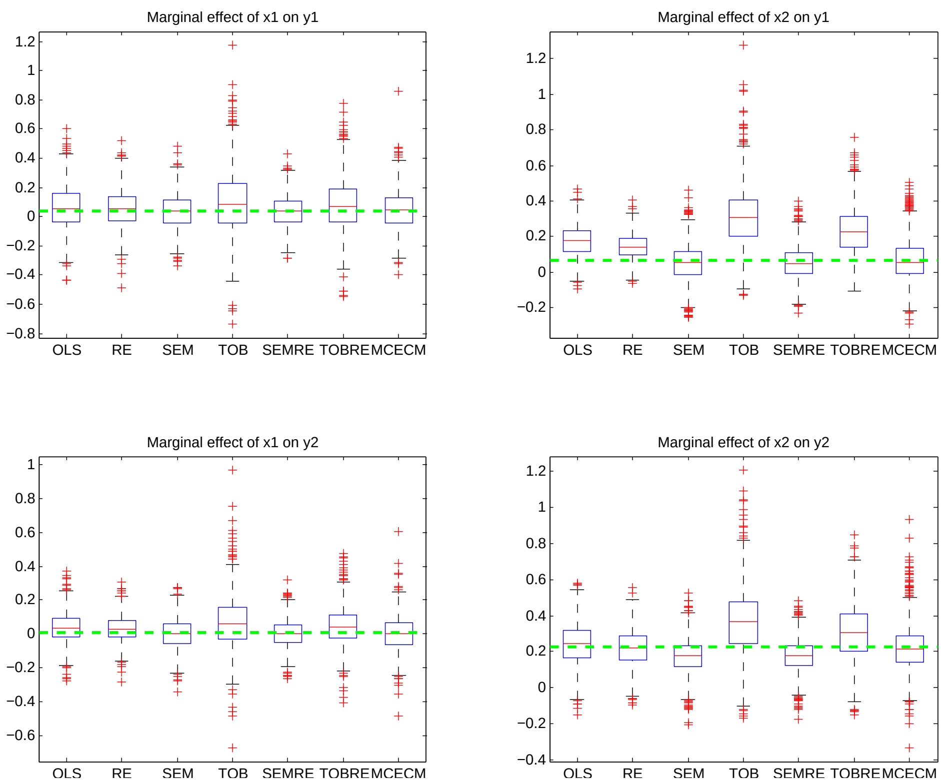
5.1.4 Results with different sample sizes

The computational cost of the simulation exercise is very high even for small samples. However, it would be interesting to try different sample sizes to study their impact on the evaluation and in the computational resources

Table 5.11: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, functional form logarithmic in the exogenous variables

		(1) OLS	(2) GLSRE	(3) SEM	(4) TOB	(5) SEMRE	(6) TOBRE	(7) MCECM
Marg. effect	Bias	1.9014	1.2230	-0.5543	4.8395	-0.6368	3.3361	0.3621
of x_1 on y_1	Var	1.9621	1.5729	1.3341	4.4252	1.1166	3.1043	1.7213
	MSE	1.9982	1.5879	1.3372	4.6594	1.1207	3.2156	1.7226
Marg. effect	Bias	10.8775	7.3465	-1.6063	24.4597	-1.4474	16.6739	-0.0497
of x_2 on y_1	Var	0.8037	0.5378	1.1619	2.9711	0.9113	1.8671	1.4338
	MSE	1.9868	1.0775	1.1877	8.9539	0.9322	4.6473	1.4338
Marg. effect	Bias	3.0658	2.4247	-0.5565	5.7606	-0.5830	3.8310	-0.4415
of x_1 on y_2	Var	0.7376	0.5095	0.8038	2.3424	0.6760	1.2497	0.9821
	MSE	0.8316	0.5683	0.8069	2.6742	0.6794	1.3964	0.9841
Marg. effect	Bias	1.6329	-0.7320	-5.0026	14.0915	-4.8001	8.4047	-0.2585
of x_2 on y_2	Var	1.3834	1.1536	0.9992	3.5332	0.8430	2.4374	1.7842
	MSE	1.4100	1.1589	1.2495	5.5189	1.0734	3.1438	1.7849

Figure 5.11: Comparison of the average marginal effects for the simulation study of the random-effects model, functional form logarithmic in the exogenous variables



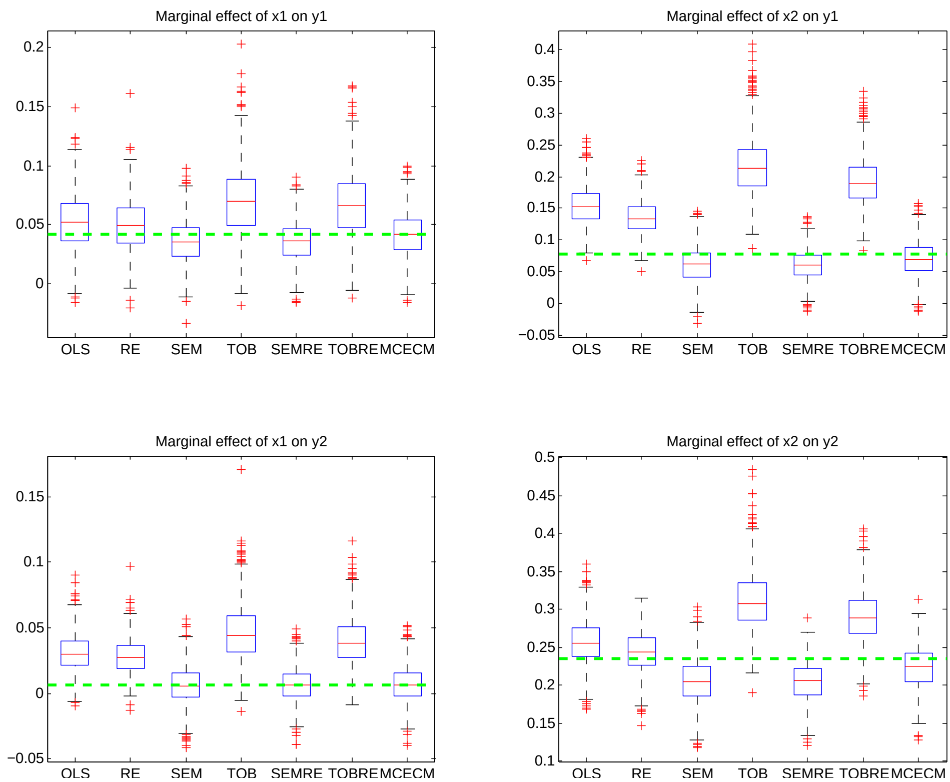
required. First we try with a smaller sample size: $N = 25$ and $T = 5$, which produces 125 observations, 40.4% fewer than in the base configuration. The estimation exercise takes 39659 seconds (11.0 hours), 37.8% less than in the original exercise, approximately the same reduction than that of the sample size. The average marginal effects are summarized by Table 5.12 and Figure 5.12. When comparing this figure with the results of the base simulation in Figure 5.4, we see that the bias of the estimation does not change significantly but, as expected, the variability increases for all the methodologies. Although not reproduced here to save space, this increase in variance is particular important for the estimation of the error components, and particulary for the individual random effects, since they have to be estimated with less information.

Table 5.12: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, sample size $N = 25$, $T = 5$

		(1) OLS	(2) GLSRE	(3) SEM	(4) TOB	(5) SEMRE	(6) TOBRE	(7) MCECM
Marg. effect	Bias	1.0158	0.7376	-0.6430	2.8105	-0.6310	2.4818	-0.0262
of x_1 on y_1	Var	0.0514	0.0452	0.0319	0.0884	0.0264	0.0795	0.0329
	MSE	0.0617	0.0506	0.0361	0.1674	0.0303	0.1411	0.0329
Marg. effect	Bias	7.5805	5.6531	-1.7020	13.8638	-1.7018	11.4220	-0.8514
of x_2 on y_1	Var	0.0862	0.0664	0.0752	0.2011	0.0574	0.1367	0.0731
	MSE	0.6608	0.3860	0.1041	2.1231	0.0863	1.4413	0.0804
Marg. effect	Bias	2.4742	2.1948	-0.0199	3.9813	-0.0095	3.3409	0.0380
of x_1 on y_2	Var	0.0203	0.0164	0.0204	0.0458	0.0161	0.0320	0.0179
	MSE	0.0815	0.0646	0.0204	0.2043	0.0161	0.1436	0.0179
Marg. effect	Bias	2.0551	0.8374	-3.0097	7.5400	-3.1007	5.5046	-1.1739
of x_2 on y_2	Var	0.0856	0.0749	0.0849	0.1440	0.0678	0.1074	0.0775
	MSE	0.1278	0.0820	0.1755	0.7125	0.1639	0.4104	0.0913

In terms of the comparative results included in the table, the results are similar to the base simulation (TOB and TOB-RE are largely the worst estimator, while the rest perform much better). However, the estimator proposed in this thesis loses some ground to the other methods. In fact, it now has the smallest bias for only two marginal effect and the smallest MSE only for one (it comes second for the rest). With the smaller sample size, the MCECM estimator needs to “complete” the vector of endogenous variables with less information and this reduces the precision of the simulation solution. However, even if the SEM-RE or GLS-RE outperform the MCECM, the latter still produces comparably acceptable results.

Figure 5.12: Comparison of the average marginal effects for the simulation study of the random-effects model, sample size $N = 25$, $T = 5$

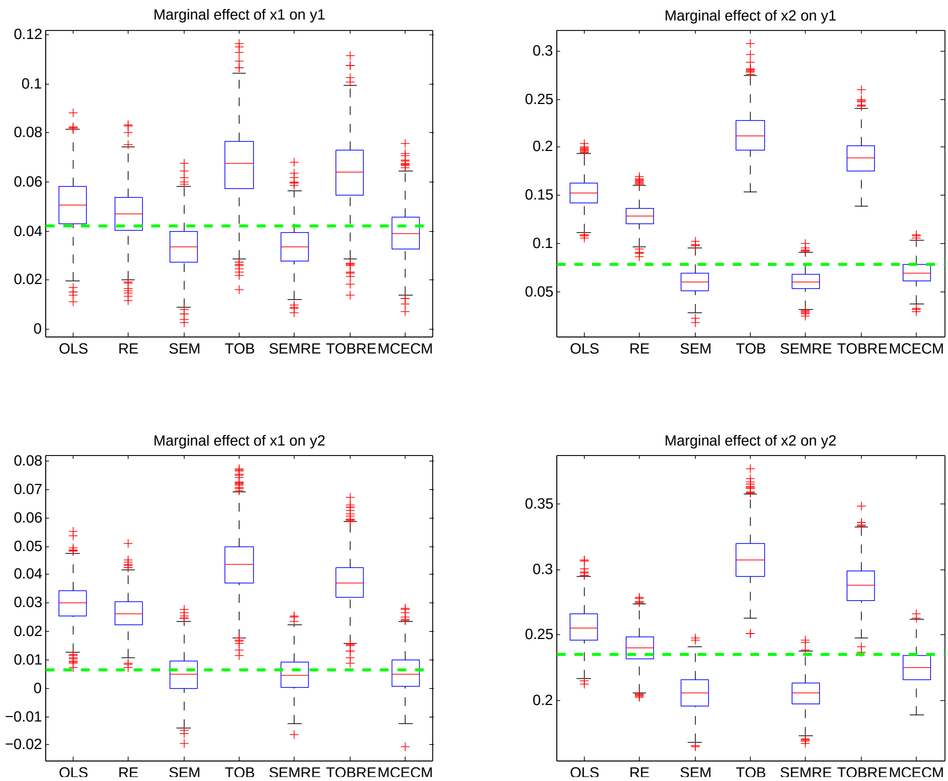


As a second test, we reproduce the exercise with a larger sample: $N = 40$ and $T = 10$, 400 observations. The complete simulation now takes 107556 seconds or 29.9 hours, 68.6% more than the original simulation (but less than the 90.5% augmentation in sample size). As a result of the increase of available information, all estimators are more efficient as it is shown in Figure 5.13. This is arguably the reason that the simulation becomes more computationally efficient (i.e., the increase in computation time is proportionally less than the rise in the number of observations): since the estimation is now more precise, the number of simulations required in the MCECM algorithm would also decrease. In comparative terms, this benefits other competing estimators more than the MCECM, as evidenced by the results of the marginal effects included in the table below. However, our proposed estimator still has the smallest bias for three of the effects and the minimum MSE for two of them (and, when not the best, it performs almost as well as the best estimator). The Tobit-type solutions are still the worst estimators.

Table 5.13: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, sample size $N = 40$, $T = 10$

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	0.8395	0.4878	-0.8648	2.4989	-0.8697	2.1748	-0.3019
of x_1 on y_1	Var	0.0132	0.0108	0.0090	0.0224	0.0073	0.0201	0.0095
	MSE	0.0203	0.0131	0.0165	0.0849	0.0149	0.0674	0.0104
Marg. effect	Bias	7.5010	5.0982	-1.8036	13.5729	-1.7479	11.1276	-0.8758
of x_2 on y_1	Var	0.0259	0.0163	0.0176	0.0572	0.0134	0.0391	0.0167
	MSE	0.5886	0.2762	0.0501	1.8995	0.0440	1.2773	0.0243
Marg. effect	Bias	2.3568	2.0043	-0.1624	3.7316	-0.1654	3.1069	-0.1317
of x_1 on y_2	Var	0.0050	0.0036	0.0052	0.0109	0.0044	0.0077	0.0049
	MSE	0.0605	0.0438	0.0055	0.1502	0.0046	0.1042	0.0051
Marg. effect	Bias	2.0081	0.4716	-2.9805	7.3031	-2.9820	5.2720	-1.0050
of x_2 on y_2	Var	0.0217	0.0158	0.0190	0.0359	0.0151	0.0268	0.0168
	MSE	0.0620	0.0181	0.1079	0.5692	0.1040	0.3048	0.0269

Figure 5.13: Comparison of the average marginal effects for the simulation study of the random-effects model, sample size $N = 40$, $T = 10$



5.1.5 Other simulations

Besides to the robustness exercises presented above, there are a few other possibilities that may be of interest when studying the performance of our estimator under different settings. We first evaluate the results with more parameters. In the exogenous variable common to both equation in (5.1) and (5.2), in addition to the constant term we include another exogenous variable. In terms of the empirical study we are trying to mimic, the new model is the following: as before, x_1 includes the variable "experience" with the same coefficient as in the base simulation; x_2 is the variable "education" with the same coefficient; finally, we add the common variables "nonlabor income" with the coefficient $\beta_{12} = \beta_{22} = -0.13$ and for the constant term we assign a coefficient of one for both equations. The distribution from which the exogenous variables "education" and "experience" are drawn is the same than before; for "nonlabor income" we use a $U(0, 15)$. This leads to an average censorship of 23.4% for the first endogenous variables and 14.2% for the second. The variance-covariance matrices of the error components and the rest of the configuration of the simulation is kept unchanged. The results are presented in Table 5.14 and Figure 5.14. The dashed lines in the graph represent to average marginal values evaluated at the true values of the parameters. They are equal to

$$\text{Average marginal effect of } x_1 \text{ on } y_1 = 0.0413$$

$$\text{Average marginal effect of } x_2 \text{ on } y_1 = 0.0767$$

$$\text{Average marginal effect of } x \text{ on } y_1 = -0.1819$$

$$\text{Average marginal effect of } x_1 \text{ on } y_2 = 0.0066$$

$$\text{Average marginal effect of } x_2 \text{ on } y_2 = 0.2333$$

$$\text{Average marginal effect of } x \text{ on } y_2 = -0.1659$$

As in the original simulation, for the exogenous variables included only

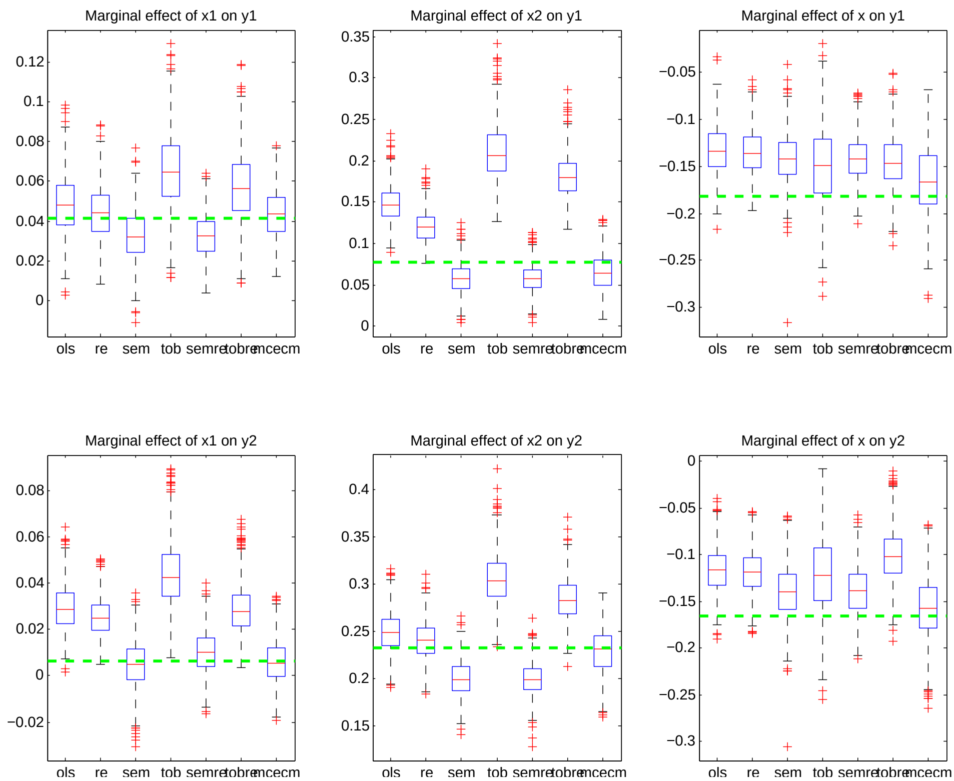
Table 5.14: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, one additional exogenous variable

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	0.7167	0.3218	-0.8953	2.4182	-0.8544	1.6118	0.2125
of x_1 on y_1	Var	0.0230	0.0181	0.0161	0.0388	0.0119	0.0319	0.0160
	MSE	0.0281	0.0191	0.0241	0.0973	0.0192	0.0578	0.0164
Marg. effect	Bias	7.0714	4.3346	-1.8879	13.3316	-1.9048	10.4666	-1.1641
of x_2 on y_1	Var	0.0449	0.0312	0.0313	0.1052	0.0258	0.0613	0.0480
	MSE	0.5449	0.2191	0.0670	1.8825	0.0621	1.1568	0.0616
Marg. effect	Bias	4.8904	4.6911	4.0089	3.2424	3.9919	3.6796	1.6441
of x on y_1	Var	0.0655	0.0529	0.0666	0.1761	0.0526	0.0768	0.1226
	MSE	0.3046	0.2730	0.2273	0.2812	0.2119	0.2122	0.1496
Marg. effect	Bias	2.2728	1.8737	-0.1796	3.7260	0.3607	2.1927	-0.0686
of x_1 on y_2	Var	0.0091	0.0065	0.0100	0.0200	0.0082	0.0104	0.0087
	MSE	0.0608	0.0416	0.0103	0.1589	0.0095	0.0585	0.0088
Marg. effect	Bias	1.6770	0.7558	-3.2349	7.2783	-3.3119	5.0635	-0.3191
of x_2 on y_2	Var	0.0389	0.0383	0.0360	0.0683	0.0309	0.0481	0.0546
	MSE	0.0670	0.0440	0.1406	0.5980	0.1406	0.3045	0.0557
Marg. effect	Bias	4.9828	4.7714	2.6216	4.5226	2.7302	6.4904	0.8005
of x on y_2	Var	0.0593	0.0512	0.0830	0.1644	0.0646	0.0823	0.1020
	MSE	0.3076	0.2788	0.1517	0.3690	0.1392	0.5036	0.1084

in one of the equations (x_1 and x_2) the MCECM estimator is the best in terms of bias; it is also the method that produces the smallest MSE for three marginal effects and a close second best for the fourth one. The SEM-RE, GLS-RE and SEM also produce acceptable results while, as before, TOB and TOB-RE are the worst performers. For the common exogenous variable (other than the constant term which, by definition, does not have a marginal effect), the TOB and TOB-RE estimators significantly improve their comparative performance; our proposed estimator is clearly better in terms of bias and MSE.

Another possibility worth exploring is the behavior of the estimation methodologies in presence of a weak instrument; this is, when one of the identifying variables is only weakly correlated with the endogenous variable it is instrumenting. To achieve this, we generate a simulation with the same

Figure 5.14: Comparison of the average marginal effects for the simulation study of the random-effects model, one additional exogenous variable



configuration as the original exercise, but we change the coefficient of the exogenous variable included only in the first equation to $\beta_{11} = 0$. Then, we proceed with the estimation as usual. Now the variable x_1 will not have any correlation with any of the dependent variables. As a result, the estimation variance of the coefficients increases, particularly for the simultaneity parameters γ_1 and γ_2 and the elements of the variance-covariance matrices of the error components. In Figure 5.15 we see the results in terms of the average marginal effects. All methods correctly estimate that the effect of x_1 on the endogenous variable is zero, but with different precision. The best estimation methods are SEM-RE and GLS-RE, while the worse are the two Tobit estimators. For the marginal effect of the second variable on the endogenous variables, if we compare the boxplots below with those of Figure 5.4 we see that their variability increases. The MCECM estimator is the best solution in terms of MSE for the effect on y_1 and third for the other dependent variable. In general, we see a relative improvement on the performance of the non-Tobit methods that do not take into account the simultaneity of the system (i.e., OLS and GLS-RE) for this underidentified model.

We now test the results under different degrees of censorship in the endogenous variables. First, we study the effect of reducing the incidence of censored cases. To achieve this, we take the base configuration but we increase the constant terms to $\beta_{12} = \beta_{22} = 0.5$. This change reduces the proportion of censored observations to 13.5% and 8.4% of the first and second dependent variable, respectively. The computational time is also decreased to 59373 (or 16.5 hours) because less censorship means that the MCECM method needs to “complete” less data and it is calculated faster.

Table 5.15: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, weak instrument

		(1) OLS	(2) GLSRE	(3) SEM	(4) TOB	(5) SEMRE	(6) TOBRE	(7) MCECM
Marg. effect	Bias	0.0825	0.0829	0.0312	0.0912	0.0333	0.0953	0.0327
of x_1 on y_1	Var	0.0209	0.0182	0.0144	0.0447	0.0116	0.0422	0.0160
	MSE	0.0210	0.0182	0.0144	0.0447	0.0116	0.0423	0.0160
Marg. effect	Bias	5.9577	4.0380	-2.2637	14.0948	-2.2769	11.6834	-0.9137
of x_2 on y_1	Var	0.0386	0.0304	0.0369	0.1102	0.0265	0.0743	0.0411
	MSE	0.3935	0.1934	0.0881	2.0969	0.0783	1.4393	0.0495
Marg. effect	Bias	0.0538	0.0524	-0.0450	0.0729	-0.0449	0.0637	-0.0444
of x_1 on y_2	Var	0.0086	0.0068	0.0106	0.0257	0.0079	0.0177	0.0089
	MSE	0.0086	0.0068	0.0106	0.0257	0.0080	0.0177	0.0089
Marg. effect	Bias	0.9190	-0.4072	-3.9331	8.6708	-3.9160	6.0011	-1.3166
of x_2 on y_2	Var	0.0357	0.0295	0.0400	0.0918	0.0334	0.0567	0.0450
	MSE	0.0442	0.0312	0.1947	0.8437	0.1868	0.4168	0.0623

Figure 5.15: Comparison of the average marginal effects for the simulation study of the random-effects model, weak instrument

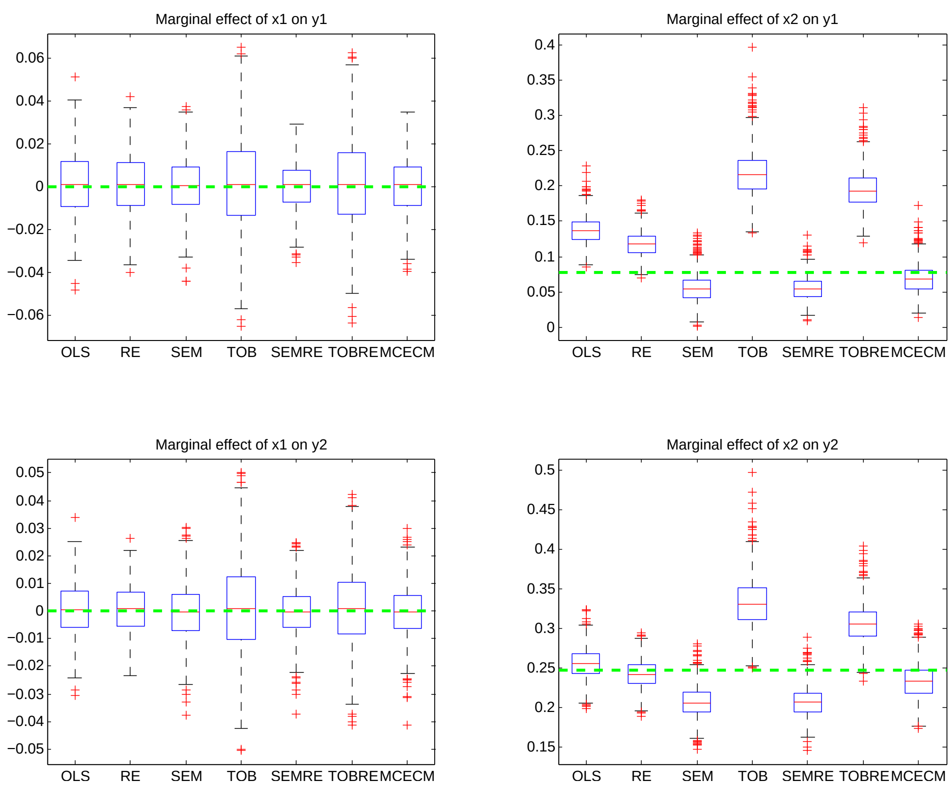


Table 5.16: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, less censorship

		(1) OLS	**(2)** GLSRE	**(3)** SEM	**(4)** TOB	**(5)** SEMRE	**(6)** TOBRE	**(7)** MCECM
Marg. effect	Bias	1.4725	1.0996	-0.5418	2.8006	-0.5427	2.4562	-0.1283
of x_1 on y_1	Var	0.0280	0.0234	0.0184	0.0420	0.0149	0.0373	0.0174
	MSE	0.0497	0.0355	0.0213	0.1204	0.0179	0.0976	0.0175
Marg. effect	Bias	9.2795	6.7224	-1.3114	13.9995	-1.2341	11.9498	-0.6129
of x_2 on y_1	Var	0.0448	0.0332	0.0355	0.0905	0.0279	0.0652	0.0333
	MSE	0.9059	0.4851	0.0527	2.0504	0.0431	1.4932	0.0371
Marg. effect	Bias	2.7394	2.3969	-0.0753	3.7575	-0.0674	3.2639	-0.0412
of x_1 on y_2	Var	0.0105	0.0081	0.0106	0.0184	0.0087	0.0139	0.0091
	MSE	0.0856	0.0655	0.0107	0.1596	0.0088	0.1204	0.0091
Marg. effect	Bias	3.9111	2.2697	-2.0088	7.9652	-2.0175	6.1916	-0.6971
of x_2 on y_2	Var	0.0423	0.0338	0.0362	0.0644	0.0296	0.0520	0.0328
	MSE	0.1953	0.0853	0.0765	0.6988	0.0703	0.4354	0.0376

As a result of these changes, we can verify in Figure 5.16 that variability is somewhat reduced and the bias of estimators that do not correct for the limited character of the dependent variable (OLS, GLS-RE, SEM and SEM-RE) is less pronounced. Even though in this case there is less to gain from our computational solution, it still performs comparatively well, with the minimal bias of all alternatives for the four marginal effects and the minimal MSE for all but one. This estimator is followed closely by the SEM-RE estimator, which is the second best. As before, the TOB and TOB-RE estimator are biased and less efficient.

Second, we increase the incidence of censorship of the dependent variables. For this, we change the constant term of both equations to $\beta_{12} = \beta_{22} = -0.5$. With this change, the proportion of censorship reaches 34.1% for the first variable and 20.3% for the second. The computational time increases marginally to 66533 seconds (18.5 hours). As expected, opposite to what was observed in the preceding simulation, in this case the estimators that do not correct for the censored character of the dependent variables perform relatively worse. The two Tobit estimators, although they do take

Figure 5.16: Comparison of the average marginal effects for the simulation study of the random-effects model, less censorship

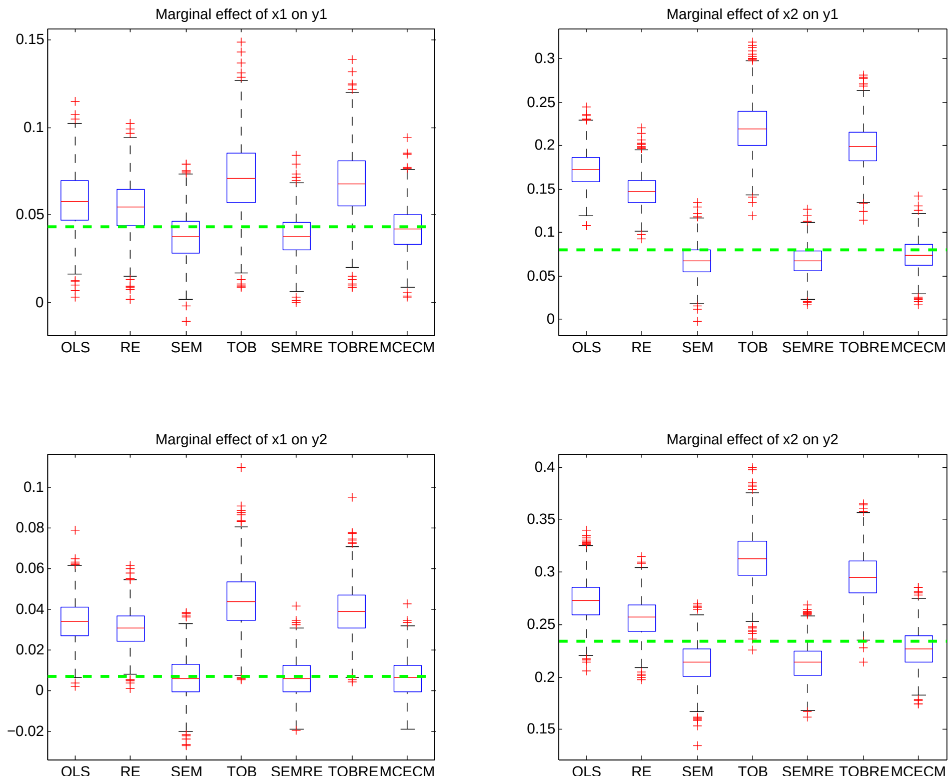
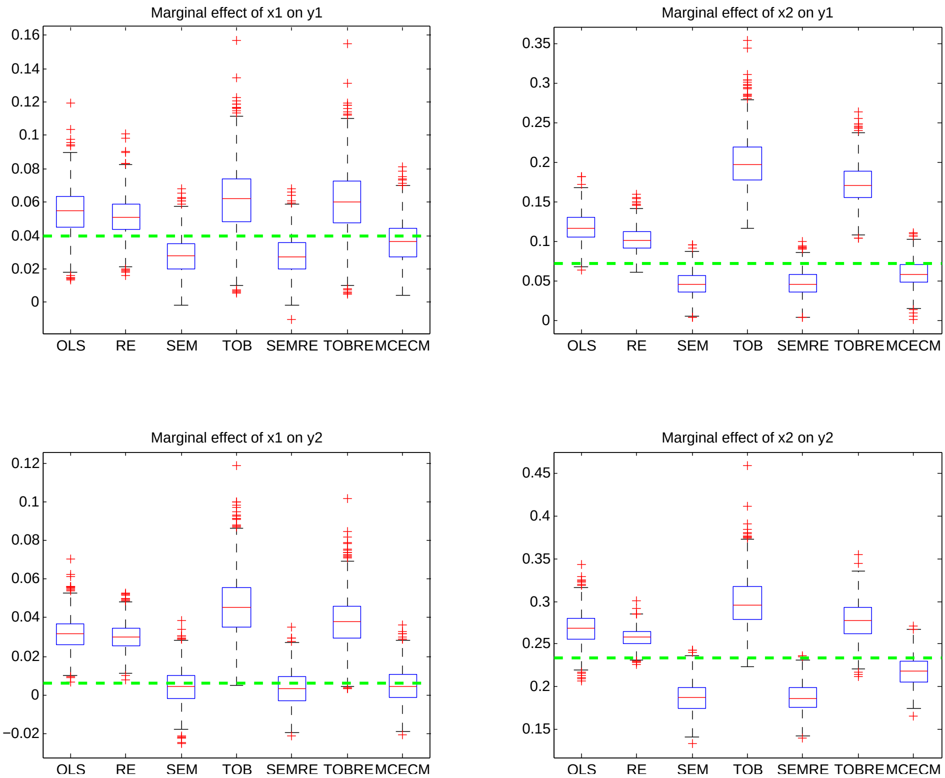


Table 5.17: Comparison of the average marginal effects for the simulation study of the random-effects model, values multiplied by 10^2, more censorship

		(1) OLS	(2) GLSRE	(3) SEM	(4) TOB	(5) SEMRE	(6) TOBRE	(7) MCECM
Marg. effect	Bias	1.5152	1.1806	-1.1415	2.2165	-1.1217	2.0715	-0.3158
of x_1 on y_1	Var	0.0192	0.0141	0.0132	0.0409	0.0130	0.0378	0.0153
	MSE	0.0422	0.0280	0.0263	0.0900	0.0256	0.0807	0.0163
Marg. effect	Bias	4.4841	2.9710	-2.6136	12.6764	-2.5635	9.9857	-1.3445
of x_2 on y_1	Var	0.0343	0.0239	0.0235	0.1075	0.0235	0.0620	0.0278
	MSE	0.2353	0.1122	0.0918	1.7145	0.0892	1.0591	0.0459
Marg. effect	Bias	2.5422	2.3828	-0.2110	3.9645	-0.2971	3.1632	-0.1505
of x_1 on y_2	Var	0.0071	0.0051	0.0088	0.0260	0.0077	0.0166	0.0082
	MSE	0.0717	0.0618	0.0092	0.1832	0.0086	0.1166	0.0084
Marg. effect	Bias	3.4499	2.4271	-4.6619	6.4473	-4.6664	4.3817	-1.5896
of x_2 on y_2	Var	0.0378	0.0112	0.0318	0.0835	0.0270	0.0484	0.0313
	MSE	0.1568	0.0701	0.2492	0.4991	0.2448	0.2404	0.0566

it into account, still underperform since they neglect the simultaneity in the model. The MCECM solution is not affected by the increase in censorship and it retains its good results: it is now the best estimator among the seven alternatives in terms of bias and MSE.

Figure 5.17: Comparison of the average marginal effects for the simulation study of the random-effects model, more censorship



5.1.6 Summary of the simulation study for the random-effects model

The preceding simulations were designed to test the comparative performance of the estimator proposed in this thesis. The simulation exercise was designed to mimic an empirical study: the coefficients of the model were taken from a well-known and influential study in the limited dependent variable literature and the exogenous variables were drawn from distributions "inspired" by the corresponding real-life variables. With the objective of exploring the performance of the estimator under different scenarios, the simulation was modified to accommodate different DGPs. Although the effect of this changes are generally more straightforward to verify in the estimated values of the coefficients themselves, these are not the quantities of interest in an empirical analysis because they do not reflect the effect that changes in the exogenous variables bring about to the dependent variables; for these reason, the analysis was mainly based on the average marginal effects. These are complex, nonlinear functions that depend on the observed values of the exogenous variables and all the estimated coefficients of the model.

Two main results can be derived from the simulation study. First, the average marginal effects remain remarkably stable under different DGPs. Even though the coefficients did present higher variability, for example, this effect was mitigated when in the calculation of the marginal effects. However, even if small, there were some changes in these quantities and in the performance of the different estimators to the changes in the parameters. Second, the MCECM estimator proposed in this thesis has a remarkable performance. It correctly takes into account the three problems present in the data (censorship, simultaneity and individual random effects) and leads to adequate estimation results. Most of the times it is the method with the smallest estimation bias. Even though it is a procedure based on simulation, with an inherent increase in variability, its loss of efficiency remains moderate in comparative terms and it frequently ends up as the best solution in terms of a minimal MSE among the alternatives. Although some DGPs may lead to less precise estimates, it affects all the methods and our estimator retains its good results. However, some situations like a small sample size, low dispersion in the exogenous variables (through a logarithmic transformation, for example) or the presence of a weak instrument increase the variance of the estimation and this is amplified by the simulation nature of the MCECM; in those cases, although it still leads to acceptable results, it is outperformed by other more robust estimators. In contrast, while other methods see their estimation bias increase when a higher proportion of the dependent variables are censored; the MCECM method retains its good properties. In general terms, the SEM-RE and SEM are the methodologies that follow the MCECM in terms of estimation performance (and in some cases show a better results than our estimator). Without exception, the two Tobit estimators TOB and TOB-RE lead to the worse results.

5.2 Simulation study for the fixed-effects model

A parallel simulation study was performed for a panel with fixed individual effects. For this experiment, we simulate from the following system

$$y^*_{1it} = \gamma_{11} y_{2it} + \beta_{11} x_{1it} + \beta_{12} x_{it} + \alpha_{1i} + \epsilon_{1it} \tag{5.4}$$

$$y^*_{2it} = \gamma_{21} y_{1it} + \beta_{21} x_{2it} + \beta_{22} x_{it} + \alpha_{2i} + \epsilon_{2it} \tag{5.5}$$

where x_{1it} is an exogenous variable included only in the first equation and x_{2it} is an exogenous variable present only in the second equations; x_{it} is a variable common to both equations. α_{1i} and α_{2i} are now fixed effects that will be treated as nuisance parameters (in other words, the objective of the exercise is to estimate the coefficients β_{11}, β_{12}, β_{21}, β_{22} γ_1, γ_2 and the variance-covariance matrix of the error terms ϵ_{1it} and ϵ_{2it}). In fact, we will see that the estimation methods apply a transformation to the variables in the model to eliminate the fixed individual effects and then estimate the rest of the coefficients. Through this procedure we avoid the incidental parameter problem in the estimation of the coefficients of interest (although not in any attempt to estimate the individual effects). Note that the constant term included in the variable x_{it} in equations (5.1) and (5.2) is excluded here to avoid a perfect linear relationship with the fixed effects.

The experiment consisted in simulating data from the two equations above and estimating the coefficients. This was repeated $H = 1000$ for a panel of size $N = 30$ and $T = 10$ for a total of 300 observations. Here as well we try to mimic the same empirical study followed in the simulation for the random-effects model, Hausman and Wise (1979). We associate x_1 with the variable "experience", x_2 with "education" and x_{it} with "non-labor income". They are randomly generated from independent densities $\mathbb{N}(15, 6^2)$, $\mathbb{N}(10, 4^2)$ and $\mathbb{N}(8, 3^2)$, respectively. The coefficients are assigned accordingly (but multiplied by 10 to make it easier to read the results).

$$\beta_{11} = 0.04$$

$$\beta_{12} = -0.13$$

$$\beta_{21} = 0.21$$

$$\beta_{22} = -0.13$$

As before, in the article that we use as source for the coefficients only one equation is estimated, so we cannot extract the simultaneity coefficients from there. We therefore assign the values $\gamma_1 = 0.4$ and $\gamma_1 = 0.2$ arbitrarily but in such a way that there is a stronger simultaneity factor in one equation and a weaker factor in the other. The variance of the error terms was equally taken from the empirical study. There is no covariance available from the source, so we decided to set it at one fourth of the variance. Consequently, the error terms are randomly drawn from the following bivariate distribution.

$$\begin{pmatrix} \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.80 & 0.45 \\ 0.45 & 1.80 \end{pmatrix}\right]$$

In modern econometric applications of panel data methodologies, the fixed-effects estimator is chosen not because the individual effects are taken as fixed quantities but because the restriction of no correlation between the exogenous variables and the individual effects fundamental to the random-effects model is considered too strong. For this reason, we will generate the individual fixed effects in such a way that, as assumed in empirical applications, they are correlated to the exogenous variables. For this, following Laisney and Lechner (2003), we generate the fixed effects according to the following formula.

$$\alpha_{1i} = \frac{1}{\sqrt{2(T+1)}}\left[\sum_{t=1}^{T}\left(\frac{x_{1it}-\mu_{x_1}}{\sigma_{x_1}} + \frac{\theta_{1it}}{\sqrt{T}}\right)\right] + 1 \tag{5.6}$$

$$\alpha_{2i} = \frac{1}{\sqrt{2(T+1)}}\left[\sum_{t=1}^{T}\left(\frac{x_{2it}-\mu_{x_2}}{\sigma_{x_2}} + \frac{\theta_{2it}}{\sqrt{T}}\right)\right] + 1 \tag{5.7}$$

where θ_{1it} and θ_{2it} are uncorrelated iid draws from a standard normal distribution (uncorrelated with the exogenous variables). It can be verified that, when generated in this way and when the exogenous variables are normal and uncorrelated over time, the correlations between the α_{1i} and the first exogenous variable, on one hand, and α_{2i} and the second exogenous variable, on the other, are greater than zero and equal to $1/\sqrt{T+1}$. This equivalent to a correlation of 0.3015 for our simulation exercise. The summand one at the end of each individual effect was included in order to translate the distribution of the dependent variables to the right[16]. This is indeed equivalent to

[16] In fact, it can be verified that, when generated through equations (5.6) and (5.7), and if the exogenous variables have univariate densities iid $\mathbb{N}(\mu_{x_1}, \sigma_2^{x_1})$ and $\mathbb{N}(\mu_{x_2}, \sigma_2^{x_2})$

adding a constant term to the equations, so the variances of the individual effects and their covariance with the exogenous variables are unaffected. If we do not include it and we use the coefficients extracted from the empirical study we would have so high an average rate of censorship that all estimation procedures would collapse. Under the present configuration, the average censorship was 22.3% for y_{1it} and 12.1% for the second variable.

Finally, it must also be kept in mind that both y_1^* and y_2^* must be generated simultaneously. In the same manner than in the simulation for the random-effects model, we achieved this through an iterative procedure. We start by assuming that the second dependent variables is zero; we generate y_1^* and we censor it if it is smaller then zero. We then take this observed value to generate y_2^* and we derive the observed variable y_2 censored at zero. We then use it to generate the first endogenous variable and we repeat until stabilization, which usually is achieved after only a few iterations.

As before, the parameters for the ascent-based version of the MCECM algorithm were the default values listed in the last column of Table C.1. With the chosen values, the average computational time for this experiment was 55942 seconds (15.5 hours).

As in the previous simulation study, here we compare the solution proposed in this document with other standard econometric methodologies.

---

respectively, the individual effects will be distributed according to a distribution

$$\begin{pmatrix}\alpha_{1i}\\ \alpha_{2i}\end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix}1\\ 1\end{pmatrix}, \begin{pmatrix}1/2 & 0\\ 0 & 1/2\end{pmatrix}\right]$$

with correlations

$$\operatorname{corr}(\alpha_{1i}, x_{1it}) = \operatorname{corr}(\alpha_{2i}, x_{2it}) = \frac{1}{\sqrt{T+1}}$$
$$\operatorname{corr}(\alpha_{1i}, x_{2it}) = \operatorname{corr}(\alpha_{2i}, x_{1it}) = 0$$

Specifically, the following estimators were calculated in this experiment.

1. Method that does not correct any problem: OLS.
2. Methods that correct only one problem: LSDV, SEM and TOB estimators.
3. Methods that correct two problems: Simultaneous Equation Model with Fixed Effects (SEM-FE) and Tobit Model with Fixed Effects (TOB-FE).
4. Method that corrects all three problems: the procedure proposed in this research, estimated through the MCECM algorithm (as before, this will be called the "MCECM estimator").

Their properties will be studied through a simulation exercise in order to decide how does the proposed solution compare to the rest.

The MCECM methodology described in Chapter 4 is applied to this fixed effects model in the following manner. First, the censored endogenous variables are imputed through a simulation procedure under the current estimates of the coefficients. Once the dependent variables are completed in this way, the model becomes a linear model and we can now apply a linear transformation to eliminate the fixed effects; we chose the within transformation. After this, we proceed to obtain the maximum likelihood estimates for the rest of the coefficients of the transformed model. This process is repeated until the stopping rule, according the ascent-version of the algorithm described in Appendix C, is attained.

We now make a few remarks about the specific methodologies used to calculate the rest of the estimators included in the list above. The LSDV estimator is the standard solution for panels with fixed effects, calculated by adding a dummy variable for each individual in the sample or alternatively by applying the within transformation[17] to all the variables in each equation and then estimating the parameters by OLS. In fact, by using this transformation the individual fixed effects drop from the equations (as well as any variable that is fixed over time for each individual). The SEM and TOB estimators are identical to those described in the previous simulation exercise: for the former, the FIML estimator of the system and, for the latter, the classic MLE for a univariate Tobit model. The SEM-FE estimator corrects for both the simultaneity present in the model and the panel structure (with fixed effects) of the data; this is achieved by first using a within transformation to eliminate the individual effects and then applying a FIML procedure to the transformed model. Finally, the TOB-FE takes into the account the censored character of the dependent variables plus the presence of fixed individual effects. Here, the within transformation cannot be applied directly because of the nonlinearity introduced by the censorship in the data. In order to calculate this estimator the simulation algorithm described in Chapter 4 had to be used: the data was "completed" using simulation and then a standard LSDV estimation was applied to each equation.

The comparison of the different estimation methodologies will be based on the marginal effects, which measure the effect on an endogenous variables of an infinitesimal change in one of the exogenous variables. These quantities are nonlinear functions of all the parameters of the model and the observed

[17]The within transformation of a variable is equivalent to subtracting the individual mean from each observation.

values of the exogenous variables. Since each repetition of the simulation generates a different marginal effect, there are two possibilities to report them: the marginal effects at the average or the average marginal effects. For the former, we calculate the average values of the exogenous variables and the estimated coefficients across the simulation exercise and evaluate the marginal effects at these averages; for the latter, we calculate the marginal effect at each repetition of the simulation and we calculate the average at the end. This last method is computationally very costly, but we prefer it because it provides more information that could be used to study the distribution of the marginal effects or to estimate any function based on them if this is needed. At the true value of the coefficients, the average marginal effects are given by the following values.

$$\text{Average marginal effect of } x_1 \text{ on } y_1 = 0.0405$$
$$\text{Average marginal effect of } x_2 \text{ on } y_1 = 0.0770$$
$$\text{Average marginal effect of } x \text{ on } y_1 = -0.1793$$
$$\text{Average marginal effect of } x_1 \text{ on } y_2 = 0.0068$$
$$\text{Average marginal effect of } x_2 \text{ on } y_2 = 0.2371$$
$$\text{Average marginal effect of } x \text{ on } y_2 = -0.1688$$

These are similar to the values given for the simulation exercise of the random-effects model with the same number of parameters described in Section 5.1.5. This is because the coefficients of the slope parameters and the variance-covariance matrix of the error term take the same values. The individual fixed effects are generated in a way that they have a mean of one, just like the value of the intercept in the random-effects model. The small difference arises from the fact that the individual effects have each a variance of 0.60 and a covariance of 0.15 in the previous exercise; here they are generated in a way that each has a variance of 0.50 and they are not correlated, as it was explained above. The performance of each methodology in the estimation of the average marginal effects will be presented through the bias (to the average marginal effects at the true values of the coefficients), the variance and the MSE. It would also be advantageous to compare the different solutions by means of the average value of the likelihood function; however, (3.33) is a costly function to evaluate and it becomes prohibitive in a simulation exercise. Because of this, we concentrate only on the performance measured through the marginal effects.

The results of the simulation study are summarized in Table 5.18 and Figure 5.18. These also include, for comparison purposes, the estimation results of the MCECM for a random-effects model, marked with the label "M-RE" (column 7 of the table); this estimation method should show a bias in the present context because its assumptions are violated by the correlation between the individual effects and the exogenous variables.

We start the discussion of the results with the marginal effects of the two variables directly included in their own equation: x_1 on the first endogenous variable and x_2 on the second endogenous variable. In both cases, the MCECM estimator of column 8 in the table has the smallest bias of the alternatives; as the boxplots show, the median of the average marginal effects coincides with the true effects. It is followed closely by the LSDV solution. The most biased estimator is TOB in the two effects, followed by OLS in the first one and the fixed-effects Tobit estimator in the second. Although it is better than other alternatives, the estimator for random ef-

Table 5.18: Comparison of the average marginal effects for the simulation study of the fixed-effects model, values multiplied by 10^2

		(1) OLS	(2) LSDV	(3) SEM	(4) TOB	(5) SEMFE	(6) TOBFE	(7) M-RE	(8) MCECM
Marginal	Bias	3.5850	0.3444	2.2884	5.9558	-0.6756	2.0719	2.1953	0.1581
effect of	Var	0.0232	0.0197	0.0189	0.0343	0.0154	0.0387	0.0369	0.0225
x_1 on y_1	MSE	0.1517	0.0209	0.0713	0.3891	0.0200	0.0816	0.0851	0.0227
Marginal	Bias	4.9919	3.9663	-1.2512	9.9254	-1.6612	9.3907	0.4901	-0.4263
effect of	Var	0.0354	0.0215	0.0355	0.0607	0.0340	0.0618	0.0733	0.0444
x_2 on y_1	MSE	0.2846	0.1789	0.0511	1.0459	0.0616	0.9437	0.0757	0.0463
Marginal	Bias	4.8898	4.6485	3.7135	0.4051	3.4615	-1.6076	-3.9061	0.2664
effect of	Var	0.0954	0.1010	0.1094	0.1794	0.1094	0.1828	0.2774	0.1381
x on y_1	MSE	0.3345	0.3171	0.2473	0.1810	0.2292	0.2087	0.4300	0.1388
Marginal	Bias	3.0383	1.6837	0.2913	4.6835	-0.1599	2.9460	0.1455	-0.0911
effect of	Var	0.0061	0.0059	0.0151	0.0134	0.0142	0.0140	0.0150	0.0162
x_1 on y_2	MSE	0.0984	0.0342	0.0160	0.2328	0.0145	0.1008	0.0152	0.0163
Marginal	Bias	3.2740	-1.0476	2.0977	8.2473	-2.9901	4.4261	2.0097	-0.1802
effect of	Var	0.0872	0.0561	0.0725	0.0921	0.0624	0.0818	0.0604	0.0646
x_2 on y_2	MSE	0.1943	0.0671	0.1165	0.7723	0.1518	0.2777	0.1008	0.0649
Marginal	Bias	4.6515	4.6067	2.1498	0.8163	2.0490	-0.8479	0.3547	0.3022
effect of	Var	0.0975	0.0855	0.1392	0.1480	0.1438	0.1616	0.1390	0.1572
x on y_2	MSE	0.3139	0.2978	0.1855	0.1547	0.1858	0.1688	0.1402	0.1581

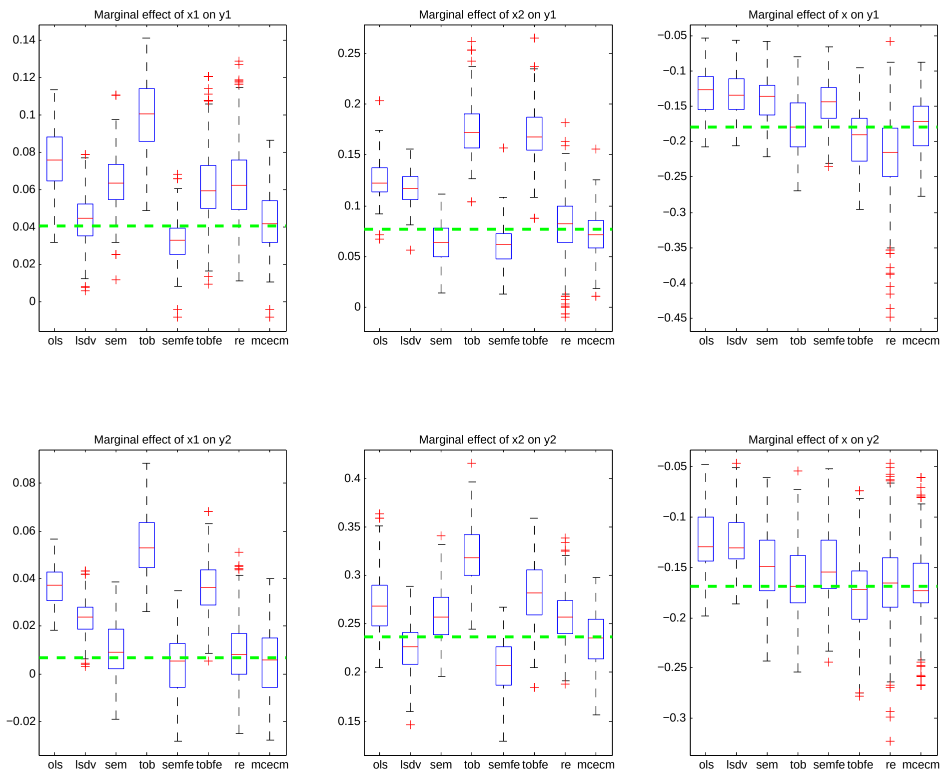
fects is affected by the correlation between the exogenous variables and the individual effect, with a resulting bias. In terms of estimation variance, as we have mentioned before, the cost of using an estimator based on simulation is a loss of efficiency, evident in these results. However, we see that this trade-off is only moderate, as the variance of the MCECM is not much larger than that of the most efficient estimator (SEM-FE for the first effect and LSDV for the second). Using the MSE as a summary measure, our proposed methodology is the best estimator for the marginal effect of x_2 on y_2 and, although not the best, still produces appropriate results for the marginal effect of x_1 on y_1.

We now analyze the results for the exogenous variables that are not included in each equation and that only have an indirect effect through the simultaneity in the system: x_2 on the first dependent variable and x_1 on the second one. Here, the estimator proposed in this thesis also has the best performance in terms of bias, followed by the SEM and SEM-FE alternatives. For both effects, the estimators with the highest bias are the two Tobit estimators and OLS. The correlation between the exogenous variables and the individual effects has little effect on the random-effects estimator in this cases; we can see in Figure 5.18 how it produces good results in terms of bias. This is because the individual effects are constructed in such a way that they are correlated with the included exogenous in each equation, so that it causes little problem in the estimation of the effects of the other exogenous variables. On the other hand, the loss of efficiency of the MCECM is more marked in these cases, especially for the effect of the first exogenous variable on y_2, where it is the estimator with the highest variance. Even if it still is the best alternative in terms of MSE for the marginal effect of x_2 on y_1, it is outperformed (although by a small margin) for the other marginal effect.

Finally, we study the effect that changes in x, the variable included in both equations, have on the endogenous variables. In the boxplots we can see that the estimation is much better for all estimators, particularly for the worse performers in the previous effects (the Tobit estimators). For the two effects, the MCECM solution has the minimal bias among the alternatives, closely followed by TOB and TOB-FE. Although its loss of efficiency brought about by its simulation nature has a negative effect, this is only moderate, resulting in it still being the best solution measured by MSE for the effect on the first endogenous variable, and third rank for the other one (but very close to the best solutions).

In summary, the MCECM estimator proposed in this document has a remarkable performance in terms of estimation bias: for the six average

Figure 5.18: Comparison of the average marginal effects for the simulation study of the fixed-effects model



marginal effects reported in this section, it is the best solution among the alternatives, sometimes by a wide margin. However, it is an estimator whose construction is based on simulation and this has a cost that manifests itself in a loss of efficiency. This increase in variance was more pronounced in the simulation for the fixed-effects than for the random-effects exercise of the previous section. Nonetheless, it still is the best estimator using the MSE as comparison measure for three effects and shows appropriate results, very close to the best performer, in the rest of the cases. The violation of the assumption of no correlation between the exogenous variables and the individual effects of the estimator for the random-effects model brings consequences that can be verified here: even for a moderate correlation as that used in this exercise, this estimator shows a significant bias in the estimation of the marginal effect of the exogenous variable that is correlated with the individual effect (x_1 for the first equation and x_2 for the second).

Chapter 6

Extensions

The methodology developed in the preceding chapters can be extended to other cases of limited dependent variables. The fundamental idea remains the same: in situations where the limited character of the endogenous variables derives into a complicated likelihood function, the MCECM algorithm presented in Chapter 4 can be employed as computational solution to work around the complexity of the optimization problem. This chapter will present two further examples of how can this methodology be applied. It would also present a final section with a short discussion about the effect of extending our model to a dynamic setting.

6.1 Probit-probit model

In this setting, the function (3.3) that defines the limited character of the endogenous variables is given by

$$y_{mit} = \psi(y^*_{mit}) = \begin{cases} 1 & \text{if } y^*_{mit} > 0 \\ 0 & \text{otherwise} \end{cases}$$

In other words, the observed dependent variables is equal to one if its latent counterpart takes a positive value and zero otherwise.

This section will develop a model with two equations and individual random effects.

$$y^*_{1it} = \gamma_1 y_{2it} + \beta'_{11} x_{1it} + \beta'_{12} x_{it} + \alpha_{1i} + \epsilon_{1it}$$

$$y^*_{2it} = \gamma_2 y_{1it} + \beta'_{21} x_{2it} + \beta'_{22} x_{it} + \alpha_{2i} + \epsilon_{2it}$$

$$y_{1it} = \begin{cases} 1 & \text{if } y^*_{1it} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$y_{2it} = \begin{cases} 1 & \text{if } y^*_{2it} > 0 \\ 0 & \text{otherwise} \end{cases}$$

with, for all i and t,

$$\begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{\alpha}_{11} & \sigma^{\alpha}_{12} & 0 & 0 \\ \sigma^{\alpha}_{12} & \sigma^{\alpha}_{22} & 0 & 0 \\ 0 & 0 & \sigma^{\epsilon}_{11} & \sigma^{\epsilon}_{12} \\ 0 & 0 & \sigma^{\epsilon}_{12} & \sigma^{\epsilon}_{22} \end{pmatrix} \right]$$

We will proceed, as in Chapter 3, by considering the four possible combinations of observed outcomes.

6.1.1 Four possible cases

Case I: $y_{1it} = 1$ **and** $y_{2it} = 1$

Both observed variables are equal to one and this implies that the two latent endogenous variables are positive. This can be used to construct the probability of falling under this case, given the exogenous variables and the values of the parameters.

$$y_{1it}^* = \gamma_1 + \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} + \epsilon_{1it} > 0$$

$$y_{2it}^* = \gamma_2 + \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it} > 0$$

This leads to the inequalities

$$\epsilon_{1it} > -\gamma_1 - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}$$

$$\epsilon_{2it} > -\gamma_2 - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}$$

Since all observations of the same individuals are correlated through the individual effects α_{1i} and α_{2i}, we cannot write the density of each one independently. However, as we did before, if we write the error term conditional on the realizations of the individual effects, we can get around this difficulty.

$$\begin{aligned}
&f(\epsilon_{1it} > -\gamma_1 - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i},\, \epsilon_{2it} > -\gamma_2 - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}) = \\
&\quad = f(\epsilon_{1it} < \gamma_1 + \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i},\, \epsilon_{2it} < \gamma_2 + \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i}) \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\epsilon_{1it} < \gamma_1 + \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i},\, \epsilon_{2it} < \gamma_2 + \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} | \alpha_{1i}, \alpha_{2i}) \\
&\quad\quad \cdot g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i} \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \int_{-\infty}^{\gamma_1 + \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i}} \int_{-\infty}^{\gamma_2 + \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})\, \mathrm{d}\epsilon_{2i}\, \mathrm{d}\epsilon_{1i} \\
&\quad\quad \cdot g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}
\end{aligned}$$

Case II: $y_{1it} = 1$ **and** $y_{2it} = 0$

The $\{1, 0\}$ combination means that the first latent endogenous variable is positive while the second takes a value smaller than zero.

$$y^*_{1it} = \beta'_{11}x_{1it} + \beta'_{12}x_{it} + \alpha_{1i} + \epsilon_{1it} > 0$$
$$y^*_{2it} = \gamma_2 + \beta'_{21}x_{2it} + \beta'_{22}x_{it} + \alpha_{2i} + \epsilon_{2it} < 0$$

or, written in another manner,

$$\epsilon_{1it} > -\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i}$$
$$\epsilon_{2it} < -\gamma_2 - \beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}$$

As before, with the construction of the likelihood function in mind, it would be necessary to write the probability of such a case independently for each individual and time period. This can be achieved by conditioning on the individual error components.

$$\begin{aligned}
&f(\epsilon_{1it} > -\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{2it} < -\gamma_2 - \beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}) = \\
&\quad = f(\epsilon_{1it} < \beta'_{11}x_{1it} + \beta'_{12}x_{it} + \alpha_{1i},\, \epsilon_{2it} < -\gamma_2 - \beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}) \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\epsilon_{1it} < \beta'_{11}x_{1it} + \beta'_{12}x_{it} + \alpha_{1i},\, \epsilon_{2it} < -\gamma_2 - \beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}|\alpha_{1i}, \alpha_{2i}) \\
&\quad\quad \cdot g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i} \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \int_{-\infty}^{\beta'_{11}x_{1it}+\beta'_{12}x_{it}+\alpha_{1i}} \int_{-\infty}^{-\gamma_2-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it}|\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\epsilon_{2i}\, \mathrm{d}\epsilon_{1i} \\
&\quad\quad \cdot g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}
\end{aligned}$$

Case III: $y_{1it} = 0$ **and** $y_{2it} = 1$

This case can be developed as in the preceding combination, but exchanging the first and second endogenous variables. The observed outcome is equivalent to say that the latent variable y^*_{1it} is negative while the second variable y^*_{2it} is positive. This leads to these inequalities for the error terms.

$$\epsilon_{1it} < -\gamma_1 - \beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i}$$
$$\epsilon_{2it} > -\beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}$$

So that the probability conditional on the individual effects is given by the following expression.

$$f(\epsilon_{1it} < -\gamma_1 - \beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{2it} > -\beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}) =$$
$$= f(\epsilon_{1it} < -\gamma_1 - \beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{2it} < \beta'_{21}x_{2it} + \beta'_{22}x_{it} + \alpha_{2i})$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\epsilon_{1it} < -\gamma_1 - \beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{2it} < \beta'_{21}x_{2it} + \beta'_{22}x_{it} + \alpha_{2i}|\alpha_{1i}, \alpha_{2i})$$
$$\cdot\, g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{-\gamma_1-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}}\int_{-\infty}^{\beta'_{21}x_{2it}+\beta'_{22}x_{it}+\alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it}|\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\epsilon_{2i}\, \mathrm{d}\epsilon_{1i}$$
$$\cdot\, g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}$$

Case IV: $y_{1it} = 0$ **and** $y_{2it} = 0$

Finally, in this situation we observe zero for both variables and this implies that their latent counterparts take negative values. Because of this, the error terms obey the following inequalities.

$$\epsilon_{1it} < -\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i}$$
$$\epsilon_{2it} < -\beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}$$

The probability that the individual i in time period t falls under this case is given by

$$f(\epsilon_{1it} < -\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{2it} < -\beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}) =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\epsilon_{1it} < -\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{2it} < -\beta'_{21}x_{2it} - \beta'_{22}x_{it} - \alpha_{2i}|\alpha_{1i}, \alpha_{2i})$$

$$\cdot\, g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \int_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} \int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it}|\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\epsilon_{2i}\, \mathrm{d}\epsilon_{1i}$$

$$\cdot\, g(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}$$

6.1.2 The likelihood function and its optimization

By writing the density of the error terms conditional on the individual random effects, all observations can be treated as independent and the likelihood becomes easier to elicit. In the tobit-tobit model developed in Chapter 3, the reduced forms have different parameters (linear combinations of the parameters of the structural form) and this leads to different densities for the error terms and the individual effects. The present situation is simpler in the sense that the four reduced forms have the same parameters (that is the reason why the functions $f(\cdot)$ and $g(\cdot)$ are not indexed according to case). Combining the four cases for all observations and taking logarithms, we can write the log likelihood for this model in the following way.

$$\log L = \sum_{i=1}^{N} \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Bigg\{ \prod_{\mathbb{D}_i^{\mathrm{I}}} \int_{-\infty}^{\gamma_1 + \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i}} \int_{-\infty}^{\gamma_2 + \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2i} \, \mathrm{d}\epsilon_{1i}$$
$$\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{\beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i}} \int_{-\infty}^{-\gamma_2 - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2i} \, \mathrm{d}\epsilon_{1i}$$
$$\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\gamma_1 - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \int_{-\infty}^{\beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2i} \, \mathrm{d}\epsilon_{1i} \quad (6.1)$$
$$\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} f(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2i} \, \mathrm{d}\epsilon_{1i}$$
$$g(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \Bigg\}$$

Although simpler than (3.15), this expression is still a difficult function to optimize since it is composed of four different section and features several double integrals. One way around this complexity is the computational solution applied in the previous chapters. This optimization algorithm can be applied directly for the probit-probit model. The idea remains the same: simulate the latent variables using the observed values of the endogenous variables (a sequence of ones and zeros) and the current values of the parameters, use them to obtain the conditional maximum likelihood estimates and iterate until convergence. The procedure presented in Algorithm 5 does not need any modification to deal with the probit-probit model.

6.2 Tobit type I-II model

Now we develop a situation that involves a mixture of two models: a Tobit type I and a Tobit type II. The first endogenous variable is censored when it is smaller than a constant threshold (Tobit type I). The second endogenous variable is censored when the value taken by a *third* variable does not reach a threshold (Tobit type II). Formally,

$$
\begin{aligned}
y_{1it}^* &= \gamma_1 y_{2it} + \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} + \epsilon_{1it} \\
y_{2it}^* &= \gamma_2 y_{1it} + \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it} \\
y_{1it} &= \begin{cases} y_{1it}^* & \text{if } y_{1it}^* > 0 \\ 0 & \text{otherwise} \end{cases} \\
y_{2it} &= \begin{cases} y_{2it}^* & \text{if } y_{3it}^* = \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i} + \epsilon_{3it} > 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

In other words, the first endogenous variable is censored at a fixed threshold (zero) while the second is censored at a stochastic threshold that depends on exogenous variables and an individual effect according to the linear equation defined by y_{3it} (the selection equation). An individual error component is present in all three equations.

The statistical properties of the error components are given by the following distribution.

$$
\begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \alpha_{3i} \\ \epsilon_{1it} \\ \epsilon_{2it} \\ \epsilon_{3it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\alpha} & \sigma_{12}^{\alpha} & \sigma_{13}^{\alpha} & 0 & 0 & 0 \\ \sigma_{12}^{\alpha} & \sigma_{22}^{\alpha} & \sigma_{23}^{\alpha} & 0 & 0 & 0 \\ \sigma_{13}^{\alpha} & \sigma_{23}^{\alpha} & \sigma_{33}^{\alpha} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{11}^{\epsilon} & \sigma_{12}^{\epsilon} & \sigma_{13}^{\epsilon} \\ 0 & 0 & 0 & \sigma_{12}^{\epsilon} & \sigma_{22}^{\epsilon} & \sigma_{23}^{\epsilon} \\ 0 & 0 & 0 & \sigma_{13}^{\epsilon} & \sigma_{23}^{\epsilon} & \sigma_{33}^{\epsilon} \end{pmatrix} \right] \tag{6.2}
$$

6.2.1 Four possible cases

Case I: $y_{1it} > 0$ **and** $y_{2it} > 0$

The fact that both endogenous variables are observed implies that both the latent variable y_{1it}^* and the selection variable y_{3it}^* take positive values. We calculate the reduced form of the system, using the same symbols for the reduced form parameters than we did in Chapter 3.

$$y_{1it} = \frac{1}{\Gamma}\left(\Pi_1' X_{it} + \delta_{1i} + \nu_{1it}\right)$$
$$y_{2it} = \frac{1}{\Gamma}\left(\Pi_2' X_{it} + \delta_{2i} + \nu_{2it}\right)$$

In other words,

$$\nu_{1it} = \Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}$$
$$\nu_{2it} = \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}$$

and we know, since the second variable is observed, that y^*_{3it} is positive, which implies that $\epsilon_{3it} > -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}$. How do we obtain the probability of observing these variables in such a situation? We use the Bayes rule.

$$\begin{aligned} P(\nu_{1it}, \nu_{2it}, y^*_{3it} > 0) &= P(\nu_{1it}, \nu_{2it}) P(y_{3it} > 0 | \nu_{1it}, \nu_{2it}) \\ &= P(\nu_{1it}, \nu_{2it}) P(\epsilon_{3it} > -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i} | \nu_{1it}, \nu_{2it}) \end{aligned}$$

The first probability is the same than in the model of Chapter 3 and it equals $f^{\mathrm{I}}(\cdot)$. For the second, we need to find the distribution of ϵ_{3it} conditioned on (ν_{1it}, ν_{2it}). For this, we start from the distributional assumptions (6.2) to obtain that

$$\begin{pmatrix} \epsilon_{3it} \\ \nu_{1it} \\ \nu_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{\epsilon}_{33} & \sigma^{\epsilon}_{13} + \gamma_1 \sigma^{\epsilon}_{23} & \sigma^{\epsilon}_{23} + \gamma_2 \sigma^{\epsilon}_{13} \\ \sigma^{\epsilon}_{13} + \gamma_1 \sigma^{\epsilon}_{23} & \sigma^{\nu}_{11} & \sigma^{\nu}_{12} \\ \sigma^{\epsilon}_{23} + \gamma_2 \sigma^{\epsilon}_{13} & \sigma^{\nu}_{12} & \sigma^{\nu}_{22} \end{pmatrix}\right]$$

Using this distribution, we can derive the conditional density

$$(\epsilon_{3it} | \nu_{1it}, \nu_{2it}) \overset{iid}{\sim} \mathbb{N}\left(\mu^{\mathrm{I}}_3, \Sigma^{\mathrm{I}}_3\right)$$

where

$$\mu_3^{\mathrm{I}} = \frac{[\sigma_{22}^{\nu}(\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon}) - \sigma_{12}^{\nu}(\sigma_{23}^{\epsilon} + \gamma_2 \sigma_{13}^{\epsilon})]\nu_{1it} + [\sigma_{11}^{\nu}(\sigma_{23}^{\epsilon} + \gamma_2 \sigma_{13}^{\epsilon}) - \sigma_{12}^{\nu}(\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon})]\nu_{2it}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2}$$
$$\Sigma_3^{\mathrm{I}} = \sigma_{33}^{\epsilon} - \frac{\sigma_{22}^{\nu}(\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon})^2 + \sigma_{11}^{\nu}(\sigma_{23}^{\epsilon} + \gamma_2 \sigma_{13}^{\epsilon})^2 - 2\sigma_{12}^{\nu}(\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon})(\sigma_{23}^{\epsilon} + \gamma_2 \sigma_{13}^{\epsilon})}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2}$$

We will call this density $f_3^{\mathrm{I}}(\epsilon_{3it}|\nu_{1it}, \nu_{2it})$.

As we have done in the preceding sections, we need to write all distribution conditioned on the individual error components, so that observations for all individuals and period can be written independently; this is a necessary step to be able to write the likelihood function.

$$\begin{aligned}
&f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}) f_3^{\mathrm{I}}(\epsilon_{3it} > -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}|\nu_{1it}, \nu_{2it}) = \\
&= f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}) f_3^{\mathrm{I}}(\epsilon_{3it} < \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i}|\nu_{1it}, \nu_{2it}) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}|\delta_{1i}, \delta_{2i}) \\
&\quad \cdot f_3^{\mathrm{I}}(\epsilon_{3it} < \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i}|\nu_{1it}, \nu_{2it}, \delta_{1i}, \delta_{2i}, \alpha_{3i}) \cdot g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}, \alpha_{3i})\, \mathrm{d}\delta_{1i}\, \mathrm{d}\delta_{2i}\, \mathrm{d}\alpha_{3i} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i}|\delta_{1i}, \delta_{2i}) \\
&\quad \cdot \int_{-\infty}^{\beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i}} f_3^{\mathrm{I}}(\epsilon_{3it}|\nu_{1it} = \Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i},\ \nu_{2it} = \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i},\ \delta_{1i}, \delta_{2i}, \alpha_{3i})\, \mathrm{d}\epsilon_{3i} \\
&\quad \cdot g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}, \alpha_{3i})\, \mathrm{d}\delta_{1i}\, \mathrm{d}\delta_{2i}\, \mathrm{d}\alpha_{3i}
\end{aligned}$$

where $g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}, \alpha_{2i})$ is the distribution of the individual components given by

$$\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \\ \alpha_{3i} \end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\delta} & \sigma_{12}^{\delta} & \sigma_{13}^{\alpha} + \gamma_1 \sigma_{23}^{\alpha} \\ \sigma_{12}^{\delta} & \sigma_{22}^{\delta} & \sigma_{23}^{\alpha} + \gamma_2 \sigma_{13}^{\alpha} \\ \sigma_{13}^{\alpha} + \gamma_1 \sigma_{23}^{\alpha} & \sigma_{23}^{\alpha} + \gamma_2 \sigma_{13}^{\alpha} & \sigma_{33}^{\alpha} \end{pmatrix}\right]$$

Case II: $y_{1it} > 0$ **and** $y_{2it} = 0$

In this case, we observe the first endogenous variable but not the second one. This implies that the first latent variable is positive, but we cannot say anything about the second latent variable. However, we know that the third variable in the selection equation is not positive. With this information, we can calculate the probability to fall in this case. Given that the second dependent variable is zero, the first one can be written as

$$y_{1it} = \beta_{11}' x_{1it} + \beta_{12}' x_{it} + \alpha_{1i} + \epsilon_{1it}$$

We can also derive the expressions

$$\begin{aligned} \epsilon_{1it} &= y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i} \\ \epsilon_{3it} &< -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i} \end{aligned}$$

In this case, we do not observe the second endogenous variable, so it does not enter the probability; we need to find the marginal distribution of the other two variables.

$$\begin{aligned} P(\epsilon_{1it},\, y_{3it}^* < 0) &= P(\epsilon_{1it},\, \epsilon_{3it} < -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}) \\ &= \int_{-\infty}^{\infty} P(\epsilon_{1it},\, \nu_{2it},\, \epsilon_{3it} < -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i})\, \mathrm{d}\nu_{2i} \\ &= \int_{-\infty}^{\infty} P(\epsilon_{1it},\, \nu_{2it}) P(\epsilon_{3it} < -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i} | \epsilon_{1it},\, \nu_{2it})\, \mathrm{d}\nu_{2i} \end{aligned}$$

We had obtained $P(\epsilon_{1it},\, \nu_{2it})$ previously, this is the density $f^{\mathrm{II}}(\cdot)$. We now derive the second part of the expression. For this, we need the joint density of the three error terms.

$$\begin{pmatrix} \epsilon_{3it} \\ \epsilon_{1it} \\ \nu_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{33}^{\epsilon} & \sigma_{13}^{\epsilon} & \sigma_{23}^{\epsilon} + \gamma_2 \sigma_{13}^{\epsilon} \\ \sigma_{13}^{\epsilon} & \sigma_{11}^{\epsilon} & \sigma_{12}^{\epsilon,\nu} \\ \sigma_{23}^{\epsilon} + \gamma_2 \sigma_{13}^{\epsilon} & \sigma_{12}^{\epsilon,\nu} & \sigma_{22}^{\nu} \end{pmatrix} \right]$$

so that the conditional density of the third error term given the other two, $f_3^{\mathrm{II}}(\epsilon_{3it}|\epsilon_{1it}, \nu_{2it})$, is given by

$$(\epsilon_{3it}|\epsilon_{1it}, \nu_{2it}) \overset{iid}{\sim} \mathbb{N}\left(\mu_3^{\mathrm{II}}, \Sigma_3^{\mathrm{II}}\right)$$

$$\mu_3^{\mathrm{II}} = \frac{[\sigma_{22}^{\nu}\sigma_{13}^{\epsilon} - \sigma_{12}^{\epsilon,\nu}(\sigma_{23}^{\epsilon} + \gamma_2\sigma_{13}^{\epsilon})]\epsilon_{1it} + [\sigma_{11}^{\epsilon}(\sigma_{23}^{\epsilon} + \gamma_2\sigma_{13}^{\epsilon}) - \sigma_{12}^{\epsilon,\nu}\sigma_{13}^{\epsilon}]\nu_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2}$$

$$\Sigma_3^{\mathrm{II}} = \sigma_{33}^{\epsilon} - \frac{\sigma_{22}^{\nu}(\sigma_{13}^{\epsilon})^2 + \sigma_{11}^{\epsilon}(\sigma_{23}^{\epsilon} + \gamma_2\sigma_{13}^{\epsilon})^2 - 2\sigma_{12}^{\epsilon,\nu}\sigma_{13}^{\epsilon}(\sigma_{23}^{\epsilon} + \gamma_2\sigma_{13}^{\epsilon})}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2}$$

With these elements we can write the probability of one observation from this case.

$$\int_{-\infty}^{\infty} f^{\mathrm{II}}(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i},\, \nu_{2it}) f_3^{\mathrm{II}}(\epsilon_{3it} < -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}|\epsilon_{1it},\, \nu_{2it})\,\mathrm{d}\nu_{2i} =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{\mathrm{II}}(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i},\, \nu_{2it}|\alpha_{1i}, \delta_{2i})$$
$$\cdot f_3^{\mathrm{II}}(\epsilon_{3it} < -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}|\epsilon_{1it}, \nu_{2it}, \alpha_{1i}, \delta_{2i}, \alpha_{3i})\,\mathrm{d}\nu_{2i} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}, \alpha_{3i})\,\mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i}\,\mathrm{d}\alpha_{3i}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{\mathrm{II}}(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i},\, \nu_{2it}|\alpha_{1i}, \delta_{2i})$$
$$\cdot \int_{-\infty}^{-\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}} f_3^{\mathrm{II}}(\epsilon_{3it}|\epsilon_{1it} = y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i},\, \nu_{2it}, \alpha_{1i}, \delta_{2i}, \alpha_{3i})\,\mathrm{d}\epsilon_{3i}\,\mathrm{d}\nu_{2i}$$
$$\cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}, \alpha_{3i})\,\mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i}\,\mathrm{d}\alpha_{3i}$$

The function $g^{\mathrm{II}}(\cdot)$ is the distribution of the individual error components given by

$$\begin{pmatrix}\alpha_{1i}\\ \delta_{2i}\\ \alpha_{3i}\end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{11}^{\alpha} & \sigma_{12}^{\alpha,\delta} & \sigma_{13}^{\alpha}\\ \sigma_{12}^{\alpha,\delta} & \sigma_{22}^{\delta} & \sigma_{23}^{\alpha} + \gamma_2\sigma_{13}^{\alpha}\\ \sigma_{13}^{\alpha} & \sigma_{23}^{\alpha} + \gamma_2\sigma_{13}^{\alpha} & \sigma_{33}^{\alpha}\end{pmatrix}\right]$$

Case III: $y_{1it} = 0$ and $y_{2it} > 0$

This case can be developed in a similar way than the previous one. Since the first dependent variable is censored, we know that its latent counterpart is negative; on the other hand, the second variable is observed and this means that the dependent variable in the selection equation takes a positive value. It is straightforward to obtain the reduced form of the second equation since the simultaneous terms drops.

$$y_{2it} = \beta_{21}' x_{2it} + \beta_{22}' x_{it} + \alpha_{2i} + \epsilon_{2it} \Rightarrow \epsilon_{2it} = y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}$$

We write the probability for an observation from this case through the following development.

$$\begin{aligned}
&P(\nu_{1it} < -\Pi_1' X_{it} - \delta_{1i},\, \epsilon_{2it},\, y_{3it}^* > 0) = \\
&= P(\nu_{1it} < -\Pi_1' X_{it} - \delta_{1i},\, \epsilon_{2it},\, \epsilon_{3it} < \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i}) \\
&= \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} P(\nu_{1it},\, \epsilon_{2it},\, \epsilon_{3it} < \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i})\, \mathrm{d}\nu_{1i} \\
&= \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} P(\nu_{1it},\, \epsilon_{2it}) P(\epsilon_{3it} < \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i} | \nu_{1it},\, \epsilon_{2it})\, \mathrm{d}\nu_{1i}
\end{aligned}$$

The probability $P(\nu_{1it},\, \epsilon_{2it})$ is the same as $f^{\mathrm{III}}(\cdot)$. For the second part of the previous formula, we first require the joint distribution of the error terms.

$$\begin{pmatrix} \epsilon_{3it} \\ \nu_{1it} \\ \epsilon_{2it} \end{pmatrix} \overset{iid}{\sim} \mathbb{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{33}^{\epsilon} & \sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon} & \sigma_{23}^{\epsilon} \\ \sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon} & \sigma_{11}^{\nu} & \sigma_{12}^{\nu,\epsilon} \\ \sigma_{23}^{\epsilon} & \sigma_{12}^{\nu,\epsilon} & \sigma_{22}^{\epsilon} \end{pmatrix} \right]$$

With these information we can construct the required conditional distribution, which we will denote with $f_3^{\mathrm{III}}(\epsilon_{3it} | \nu_{1it}, \epsilon_{2it})$.

$$(\epsilon_{3it} | \nu_{1it}, \epsilon_{2it}) \overset{iid}{\sim} \mathbb{N} \left(\mu_3^{\mathrm{III}}, \Sigma_3^{\mathrm{III}} \right)$$

$$\mu_3^{\mathrm{III}} = \frac{[\sigma_{22}^{\epsilon}(\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon}) - \sigma_{12}^{\nu,\epsilon} \sigma_{23}^{\epsilon}] \nu_{1it} + [\sigma_{11}^{\nu} \sigma_{23}^{\epsilon} - \sigma_{12}^{\nu,\epsilon} (\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon})] \epsilon_{2it}}{\sigma_{11}^{\nu} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2}$$

$$\Sigma_3^{\mathrm{III}} = \sigma_{33}^{\epsilon} - \frac{\sigma_{22}^{\epsilon} (\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon})^2 + \sigma_{11}^{\nu} (\sigma_{23}^{\epsilon})^2 - 2\sigma_{12}^{\nu,\epsilon} (\sigma_{13}^{\epsilon} + \gamma_1 \sigma_{23}^{\epsilon}) \sigma_{23}^{\epsilon}}{\sigma_{11}^{\nu} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2}$$

Therefore, the probability of an observation from this case is obtained as follows.

$$\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it},\, \epsilon_{2it} = y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}) f_3^{\mathrm{III}}(\epsilon_{3it} < \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i} | \nu_{1it},\, \epsilon_{2it})\, \mathrm{d}\nu_{1i} =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it},\, \epsilon_{2it} = y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i} | \delta_{1i}, \alpha_{2i})$$

$$\cdot f_3^{\mathrm{III}}(\epsilon_{3it} < \beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i} | \nu_{1it}, \epsilon_{2it}, \delta_{1i}, \alpha_{2i}, \alpha_{3i})\, \mathrm{d}\nu_{1i} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}, \alpha_{3i})\, \mathrm{d}\delta_{1i}\, \mathrm{d}\alpha_{2i}\, \mathrm{d}\alpha_{3i}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it},\, \epsilon_{2it} = y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i} | \delta_{1i}, \alpha_{2i})$$

$$\cdot \int_{-\infty}^{\beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i}} f_3^{\mathrm{III}}(\epsilon_{3it} | \nu_{1it},\, \epsilon_{2it} = y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i},\, \delta_{1i}, \alpha_{2i}, \alpha_{3i})\, \mathrm{d}\epsilon_{3i}\, \mathrm{d}\nu_{1i}$$

$$\cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}, \alpha_{3i})\, \mathrm{d}\delta_{1i}\, \mathrm{d}\alpha_{2i}\, \mathrm{d}\alpha_{3i}$$

The distribution of the individual random effects, $g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}, \alpha_{3i})$, is given by

$$\begin{pmatrix} \delta_{1i} \\ \alpha_{2i} \\ \alpha_{3i} \end{pmatrix} \overset{iid}{\sim} \mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^{\delta} & \sigma_{12}^{\delta,\alpha} & \sigma_{13}^{\alpha} + \gamma_1 \sigma_{23}^{\alpha} \\ \sigma_{12}^{\delta,\alpha} & \sigma_{22}^{\alpha} & \sigma_{23}^{\alpha} \\ \sigma_{13}^{\alpha} + \gamma_1 \sigma_{23}^{\alpha} & \sigma_{23}^{\alpha} & \sigma_{33}^{\alpha} \end{pmatrix}\right]$$

Case IV: $y_{1it} = 0$ and $y_{2it} = 0$

This is the case with less information on the endogenous variables. With the observed outcome, we only know that the first and the third latent variables are smaller than zero and we have no information on the second. In order to write the probability of an observation from this case, we need to write the marginal probability of the first and third variables.

$$
\begin{aligned}
P(y^*_{1it} < 0,\, y^*_{3it} < 0) &= P(\epsilon_{1it} < -\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{3it} < -\beta'_{31}x_{3it} - \beta'_{32}x_{it} - \alpha_{3i}) \\
&= \int\limits_{-\infty}^{\infty} P(\epsilon_{1it} < -\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i},\, \epsilon_{2it},\, \epsilon_{3it} < -\beta'_{31}x_{3it} - \beta'_{32}x_{it} - \alpha_{3i})\, \mathrm{d}\epsilon 2i \\
&= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} P(\epsilon_{1it},\, \epsilon_{2it},\, \epsilon_{3it} < -\beta'_{31}x_{3it} - \beta'_{32}x_{it} - \alpha_{3i})\, \mathrm{d}\epsilon_{1i}\, \mathrm{d}\epsilon_{2i} \\
&= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} P(\epsilon_{1it},\, \epsilon_{2it}) P(\epsilon_{3it} < -\beta'_{31}x_{3it} - \beta'_{32}x_{it} - \alpha_{3i} | \epsilon_{1it},\, \epsilon_{2it})\, \mathrm{d}\epsilon_{1i}\, \mathrm{d}\epsilon_{2i}
\end{aligned}
$$

The first probability is given by the function $f^{\mathrm{IV}}(\cdot)$ first used in Chapter 3. The conditional probability of ϵ_{3it} given the two other error terms can be derived from the joint density of the error terms, (6.2).

$$
\begin{aligned}
(\epsilon_{3it}|\epsilon_{1it}, \epsilon_{2it}) &\overset{iid}{\sim} \mathbb{N}\left(\mu_3^{\mathrm{IV}}, \Sigma_3^{\mathrm{IV}}\right) \\
\mu_3^{\mathrm{IV}} &= \frac{(\sigma^\epsilon_{22}\sigma^\epsilon_{13} - \sigma^\epsilon_{12}\sigma^\epsilon_{23})\epsilon_{1it} + (\sigma^\epsilon_{11}\sigma^\epsilon_{23} - \sigma^\epsilon_{12}\sigma^\epsilon_{13})\epsilon_{2it}}{\sigma^\epsilon_{11}\sigma^\epsilon_{22} - (\sigma^\epsilon_{12})^2} \\
\Sigma_3^{\mathrm{IV}} &= \sigma^\epsilon_{33} - \frac{\sigma^\epsilon_{22}(\sigma^\epsilon_{13})^2 + \sigma^\epsilon_{11}(\sigma^\epsilon_{23})^2 - 2\sigma^\epsilon_{12}\sigma^\epsilon_{13}\sigma^\epsilon_{23}}{\sigma^\epsilon_{11}\sigma^\epsilon_{22} - (\sigma^\epsilon_{12})^2}
\end{aligned}
$$

The joint distribution of the three individual random effects, $g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}, \alpha_{3i})$ can also be obtained from (6.2) directly. With these elements, we can write the probability of one observation from this case.

$$\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it},\, \epsilon_{2it}) f_3^{\mathrm{IV}}(\epsilon_{3it} < -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i} | \epsilon_{1it},\, \epsilon_{2it}) \, \mathrm{d}\epsilon_{1i} \, \mathrm{d}\epsilon_{2i} =$$

$$= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it},\, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})$$

$$\cdot\, f_3^{\mathrm{IV}}(\epsilon_{3it} < -\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i} | \epsilon_{1it}, \epsilon_{2it}, \alpha_{1i}, \alpha_{2i}, \alpha_{3i}) \, \mathrm{d}\epsilon_{1i} \, \mathrm{d}\epsilon_{2i}$$

$$\cdot\, g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}, \alpha_{3i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \, \mathrm{d}\alpha_{3i}$$

$$= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it},\, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})$$

$$\cdot \int\limits_{-\infty}^{-\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}} f_3^{\mathrm{IV}}(\epsilon_{3it} | \epsilon_{1it}, \epsilon_{2it}, \alpha_{1i}, \alpha_{2i}, \alpha_{3i}) \, \mathrm{d}\epsilon_{3i} \, \mathrm{d}\epsilon_{1i} \, \mathrm{d}\epsilon_{2i}$$

$$\cdot\, g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}, \alpha_{3i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \, \mathrm{d}\alpha_{3i}$$

6.2.2 The likelihood function and its optimization

Once we have the probabilities for all observations from each of the four cases, we aggregate over individuals and time periods to construct the likelihood function. It is possible to do this at this stage since all functions are conditioned on the individual random effects. The complete log likelihood function of this model is given below. It is more complicated than the likelihood (3.15) we optimized in the preceding chapters: it also features four sections and multiple integrals but it is further complicated by a third layer introduced by the selection equation. An attempt to find the MLE directly would be very difficult, so one solution to work around this problem is the MCECM computational algorithm applied previously.

$$
\begin{aligned}
\log L = \sum_{i=1}^{N} & \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{I}}} \Bigg(f^{\mathrm{I}}(\Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i}, \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i} | \delta_{1i}, \delta_{2i}) \\
& \cdot \int_{-\infty}^{\beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i}} f_3^{\mathrm{I}}(\epsilon_{3it} | \nu_{1it} = \Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i},\, \nu_{2it} = \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i},\, \delta_{1i}, \delta_{2i}, \alpha_{3i}) \,\mathrm{d}\epsilon_{3i} \Bigg) \\
& \cdot g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}, \alpha_{3i}) \,\mathrm{d}\delta_{1i} \,\mathrm{d}\delta_{2i} \,\mathrm{d}\alpha_{3i} \\
& + \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{II}}} \Bigg(\int_{-\infty}^{\infty} f^{\mathrm{II}}(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i},\, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \\
& \cdot \int_{-\infty}^{-\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}} f_3^{\mathrm{II}}(\epsilon_{3it} | \epsilon_{1it} = y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i},\, \nu_{2it}, \alpha_{1i}, \delta_{2i}, \alpha_{3i}) \,\mathrm{d}\epsilon_{3i} \,\mathrm{d}\nu_{2i} \Bigg) \\
& \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}, \alpha_{3i}) \,\mathrm{d}\alpha_{1i} \,\mathrm{d}\delta_{2i} \,\mathrm{d}\alpha_{3i} \qquad (6.3) \\
& + \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{III}}} \Bigg(\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it},\, \epsilon_{2it} = y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i} | \delta_{1i}, \alpha_{2i}) \\
& \cdot \int_{-\infty}^{\beta_{31}' x_{3it} + \beta_{32}' x_{it} + \alpha_{3i}} f_3^{\mathrm{III}}(\epsilon_{3it} | \nu_{1it},\, \epsilon_{2it} = y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i},\, \delta_{1i}, \alpha_{2i}, \alpha_{3i}) \,\mathrm{d}\epsilon_{3i} \,\mathrm{d}\nu_{1i} \Bigg) \\
& \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}, \alpha_{3i}) \,\mathrm{d}\delta_{1i} \,\mathrm{d}\alpha_{2i} \,\mathrm{d}\alpha_{3i} \\
& + \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{IV}}} \Bigg(\int_{-\infty}^{\infty} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it},\, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \\
& \cdot \int_{-\infty}^{-\beta_{31}' x_{3it} - \beta_{32}' x_{it} - \alpha_{3i}} f_3^{\mathrm{IV}}(\epsilon_{3it} | \epsilon_{1it}, \epsilon_{2it}, \alpha_{1i}, \alpha_{2i}, \alpha_{3i}) \,\mathrm{d}\epsilon_{3i} \,\mathrm{d}\epsilon_{1i} \,\mathrm{d}\epsilon_{2i} \Bigg) \\
& \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}, \alpha_{3i}) \,\mathrm{d}\alpha_{1i} \,\mathrm{d}\alpha_{2i} \,\mathrm{d}\alpha_{3i}
\end{aligned}
$$

The Tobit type I-II model features a combination of different types of limited dependent variables. The first variable is censored at a fixed threshold: we only observe it when it is larger than that value and we observe the threshold in other cases. The second variable is truncated at a stochastic threshold: we only observe it when another random variable takes positive values. Finally, the third variable is observed only as a dichotomous variable: we never know the value of the latent variable, we only know if it is positive or negative because in the former case the second variable is also observed. This structure derives in the practically intractable likelihood function presented above. However, the MCECM algorithm applied in this document can also be employed here to deal with the complexity of the function that needs to be optimized. The fundamental idea is to use simulation to "complete" the data (the censored observations of the first dependent variable, the truncated values of the second and the full latent values taken by the third). Once this is done, the maximization task becomes significantly simpler. Algorithm 5 can be applied directly to this situation.

6.3 Dynamic models

In the models presented in Chapter 3, we made the assumption that the error terms of different time periods were uncorrelated. As a result, the only source of correlation over time was due to the presence of the individual effects. However, this may prove too restrictive in empirical applications where a shock produces an effect that is felt over longer time horizons. Such a situation would call for an error term that is serially correlated. Ignoring this phenomenon when it is present would lead to an estimation that, although still consistent, would be inefficient; on the other hand, the estimation of the variance of the error components would be biased.

To illustrate the consequences of including serially correlated errors in our model, take the random-effects model of the first section of Chapter 3 given by equations (3.4b) and (3.5b). Instead of making the assumption that the error terms are independent over time, we now add the following relationships to the model.

$$\epsilon_{1it} = \kappa_1 \epsilon_{1i,t-1} + \mu_{1it}$$
$$\epsilon_{2it} = \kappa_2 \epsilon_{2i,t-1} + \mu_{2it}$$

where μ_{1it} and μ_{2it} are iid normal random variables uncorrelated with the individual effects and the exogenous variables and κ_1 and κ_2 are assumed to be coefficients smaller than one in absolute value. In other words, we make the assumption that the errors follow a first-order (stationary) autoregressive process. If we substitute the equations of the error terms in the model, we obtain the following expressions.

$$\begin{aligned} y^*_{1it} &= \gamma_1 y_{2it} + \beta'_{11} x_{1it} + \beta'_{12} x_{it} + \alpha_{1i} + \kappa_1 (y^*_{1i,t-1} - \gamma_1 y_{2i,t-1} - \beta'_{11} x_{1i,t-1} - \beta'_{12} x_{i,t-1} - \alpha_{1i}) + \mu_{1it} \\ &= \kappa_1 y^*_{1i,t-1} + \gamma_1 (y_{2it} - \kappa_1 y_{2i,t-1}) + \beta'_{11} (x_{1it} - \kappa_1 x_{1i,t-1}) + \beta'_{12} (x_{it} - \kappa_1 x_{i,t-1}) + (1-\kappa_1)\alpha_{1i} + \mu_{1it} \end{aligned} \tag{6.4}$$

$$\begin{aligned} y^*_{2it} &= \gamma_2 y_{1it} + \beta'_{21} x_{2it} + \beta'_{22} x_{it} + \alpha_{2i} + \kappa_2 (y^*_{2i,t-1} - \gamma_2 y_{1i,t-1} - \beta'_{21} x_{2i,t-1} - \beta'_{22} x_{i,t-1} - \alpha_{2i}) + \mu_{2it} \\ &= \kappa_2 y^*_{2i,t-1} + \gamma_2 (y_{1it} - \kappa_2 y_{1i,t-1}) + \beta'_{21} (x_{2it} - \kappa_2 x_{2i,t-1}) + \beta'_{22} (x_{it} - \kappa_2 x_{i,t-1}) + (1-\kappa_2)\alpha_{2i} + \mu_{2it} \end{aligned} \tag{6.5}$$

In principle, it is possible to proceed as before: identify the possible cases that may be observed, construct the reduced form and the densities of the variables for each one of them, and use these elements to construct the likelihood function. However, two complications arise when considering serial correlation. First, since lagged values of the dependent variables are also included, instead of four possible cases for each individual i and period t as in Chapter 3, we would have sixteen.

I. $y_{1it} > 0$, $y_{1i,t-1} > 0$, $y_{2it} > 0$ and $y_{2i,t-1} > 0$

II. $y_{1it} > 0$, $y_{1i,t-1} > 0$, $y_{2it} > 0$ and $y_{2i,t-1} = 0$

III. $y_{1it} > 0$, $y_{1i,t-1} > 0$, $y_{2it} = 0$ and $y_{2i,t-1} > 0$

IV. $y_{1it} > 0$, $y_{1i,t-1} > 0$, $y_{2it} = 0$ and $y_{2i,t-1} = 0$

$\vdots$

XVI. $y_{1it} = 0$, $y_{1i,t-1} = 0$, $y_{2it} = 0$ and $y_{2i,t-1} = 0$

This would make the likelihood a lengthier expression. The second complication is more serious. Previously, we worked around the correlation across time periods by conditioning on the individual effects, the only source of time dependence in that model, and writing the likelihood using conditional distributions. However, in this case this is not enough because, as it can be verified in equations (6.4) and (6.5), there is an another source of correlation since both y^*_{1it} and $y^*_{1i,t-1}$ share some elements (and likewise for y^*_{2it} and $y^*_{2i,t-1}$). Therefore, due to the dynamic feature of the model, one has to write the joint likelihood for all time observations since only conditioning on the individual effects will not lead to time independence in this case.

However, it is possible to implement a MCECM-type procedure along the following lines. We can "complete" the vector of latent dependent variables by simulating the censored values according to the sixteen cases described above by using conditional distributions. In this way, the likelihood becomes somewhat simpler to handle and we could apply a MLE procedure for error components models with serially correlated errors, such as that developed by Anderson and Hsiao (1982). This is an interesting extension worth exploring.

Other than serial correlation in the error terms as described before, another source of time dynamics in the model could be the presence of lagged values of the dependent variables as regressors. This is an autoregressive process. We briefly present a first order autoregression to illustrate the situation (assuming there is no serial correlation in the errors).

$$y^*_{1it} = \kappa_1 y_{1i,t-1} + \gamma_1 y_{2it} + \beta'_{11} x_{1it} + \beta'_{12} x_{it} + \alpha_{1i} + \epsilon_{1it}$$
$$y^*_{2it} = \kappa_2 y_{2i,t-1} + \gamma_2 y_{1it} + \beta'_{21} x_{2it} + \beta'_{22} x_{it} + \alpha_{2i} + \epsilon_{2it}$$

Note that the lagged dependent variables in the right-hand side are the observed and not the latent values. This make sense if the censorship arises from an economic or logical reason (for example, expenditure being strictly nonnegative or wage always superior to the minimum wage). As in the previous example, the dependent variable y^*_{1it} and one of the regressors $y_{1i,t-1}$ are correlated through the individual effect, even if there is no serial correlation in the error terms (likewise for the second equation). Many alternatives have been developed for an uncensored model, but they rely on a transformation to wipe out the individual effect and then use instruments to correct for endogeneity. However, this is not possible in our situation because of the nonlinearity introduced by the limited character of the dependent variables.

An idea of how to proceed may start by sorting each observation according to one of the sixteen possible cases described above. We can obtain the reduced form for each one of these cases. In writing the likelihood, even after conditioning on individual effects, we would have to additionally condition on past values in order to decompose it into a product of conditional densities and then integrate out the individual effects. We then apply the MCECM methodology to "complete" each of the censored observations of the dependent variables. To do this, we generate a draw from the conditional distribution of a censored observation given the observed values of the other variables. We repeat this R times to obtain a Monte Carlo sample of the vector of latent dependent variables. We may then apply a (conditional) MLE procedure to obtain the estimator of the parameters. We can also apply solutions such as those proposed by Arellano and Bond (1991) or Blundell and Bond (1998) to the reduced form using the variables completed through simulation and taking into account the correlation between the error terms of both equations. This case needs a more thorough examination of all the complications involved and may prove to be a useful direction for further research in this area.

Chapter 7

Conclusions

The object of study of this thesis was a system of simultaneous equations in which some or all of the dependent variables have a limited character and where data follows a panel structure (both random components and fixed effects were considered). The derivation of the likelihood function for such a model is straightforward. However, this function presents multiple integrals, discontinuities and likely multiple modes; for this reason, standard optimization methods cannot be applied directly. Some other solutions, such as simulated MLE or metaheuristic algorithms, are prohibitive because they require many evaluations of such a complicated likelihood function. We proposed the use of an alternative methodology that combines the likelihood-maximizing approach of the EM algorithm with the feasibility of simulation methods and that is closely related to the method of simulated scores. This solution, called the MCECM algorithm, was applied following a special procedure to maintain the ascent-based property of the classical EM algorithm even if the Expectation-step is obtained through simulation. Two experiments were conducted, where data was simulated from panels with random and fixed effects.

The proposed estimator was tested under a variety of configurations, including different processes generating the exogenous variables, distributions of the error terms, functional forms, sample sizes, degree of censorship of the endogenous variables and other factors. Their results showed the good properties of this method in terms of estimation bias; in the majority of cases this was the estimator with the minimum bias among the alternatives calculated. It is an estimator based on simulation and this punishes it in terms of a higher estimation variance; however, this loss of efficiency was only moderate and frequently the proposed estimator was also the best in terms of MSE. It must be reminded that situations that naturally increase the variance of the estimation, such as a small sample size or a lower dispersion in the exogenous variables, can be magnified by the simulated nature of the estimators. However, in the experiment we showed that this was only a moderate threat and the estimator retained its acceptable performance even in those cases.

The development of the MCECM algorithm and the simulation exercise documented here concerned a bivariate system of equations with only individual components and censored (Tobit type-I) endogenous variables. This was chosen in order to exemplify the developments in a well-known, standard model in the body of limited dependent variable econometrics. However, we also showed how the same idea can be extended in a straightforward manner to other types of limited dependent variables, such as probit or Tobit type-II models. The idea remains the same: use simulation to render the Expectation-step of the EM algorithm feasible and then apply a conditional maximization procedure. Although the extension to a system with more than two equations is possible in theory, this would increase the dimensionality of the optimization problem and severely complicate the computational resources needed in the calculation of the MCECM estimator. The introduction of dynamics, be it through lagged dependent variables as regressors or serial correlation in the error term, complicates the elicitation of the likelihood function; however, the same simulation idea may be used to "complete" the vector of latent dependent variables and then apply standard estimation methods. This is an interesting idea that was only briefly studied here but that is worth further exploration.

The results from this thesis illustrate how powerful simulation methods have become for econometric applications. Faster processors and improved, more efficient versions of the simulation algorithms have contributed to this progress. These methods have been part of the standard toolkit of Bayesian econometrics for many years, but they have now extended to other areas where the standard solutions do not produce acceptable results or where they cannot be applied at all. The conclusions from this thesis show that a simulation-based alternative to a standard method (the EM algorithm) can perform remarkably well, yielding appropriate estimators in terms of bias with a limited cost in terms of loss of efficiency.

Appendix A

Results from the truncated bivariate normal distribution

Let (x, y) follow a bivariate normal distribution

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathbb{N} \left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \right]$$

with correlation coefficient

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

The density function can be written as

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$
$$\exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\}$$

Both variables are truncated from the left; i.e., x is only defined in the interval $[c, \infty)$ and y in the interval $[d, \infty)$. In this Appendix we will follow to results of Rosenbaum (1961) to obtain the expectations required for the calculation of the marginal effects.

We want to calculate the expected value of x given that we are in Case I of the four cases described in Chapter 3 (i.e., both variables are observed).

$$\mathrm{E}_x^I = \mathrm{E}(x|x>c, y>d) = \int_c^\infty x f(x|x>c, y>d)\,dx$$
$$= \frac{\int_d^\infty \int_c^\infty x f(x,y)\,dx\,dy}{\mathrm{P}(x>c, y>d)}$$

By naming $F^I = \mathrm{P}(x > c, y > d)$ we can write

$$F^I \mathrm{E}_x^I = \int_d^\infty \int_c^\infty x f(x,y)\,dx\,dy$$

If we apply the changes of variable

$$z = \frac{\frac{x-\mu_x}{\sigma_x} - \rho\frac{y-\mu_y}{\sigma_y}}{\sqrt{1-\rho^2}} \text{ and } c' = \frac{\frac{c-\mu_x}{\sigma_x} - \rho\frac{y-\mu_y}{\sigma_y}}{\sqrt{1-\rho^2}} \tag{A.1}$$

we obtain that

$$F^I \mathrm{E}_x^I = \frac{1}{2\pi\sigma_y}\int_d^\infty \int_{c'}^\infty \left[\mu_x + \sigma_x\left(z\sqrt{1-\rho^2} + \rho\left(\frac{y-\mu_y}{\sigma_y}\right)\right)\right] \exp\left\{-\frac{1}{2}\left[z^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dz\,dy$$

$$= \frac{1}{2\pi\sigma_y}\int_d^\infty \int_{c'}^\infty \sigma_x z\sqrt{1-\rho^2} \exp\left\{-\frac{1}{2}\left[z^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dz\,dy \tag{A.2a}$$

$$+ \frac{1}{2\pi\sigma_y}\int_d^\infty \int_{c'}^\infty \sigma_x \rho\left(\frac{y-\mu_y}{\sigma_y}\right) \exp\left\{-\frac{1}{2}\left[z^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dz\,dy \tag{A.2b}$$

$$+ \frac{1}{2\pi\sigma_y}\int_d^\infty \int_{c'}^\infty \mu_x \exp\left\{-\frac{1}{2}\left[z^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dz\,dy \tag{A.2c}$$

By integrating the first summand (A.2a) directly, we obtain

$$\frac{\sigma_x\sqrt{1-\rho^2}}{2\pi\sigma_y}\int_d^\infty \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{c-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{c-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dy \tag{A.3}$$

Now, integration by parts of (A.2b) gives one term equals to (A.3) times

${}^{\rho^2}/_{1-\rho^2}$ and another term equal to

$$\begin{aligned}
&\frac{\sigma_x\rho}{2\pi}\exp\left\{-\frac{1}{2}\left(\frac{d-\mu_y}{\sigma_y}\right)^2\right\}\int\limits_{c'|y=d}^{\infty}\exp\left\{-\frac{1}{2}z^2\right\}dz\\
&\quad=\sigma_x\rho\cdot\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{d-\mu_y}{\sigma_y}\right)^2\right\}\cdot\int\limits_{c'|y=d}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}z^2\right\}dz\\
&\quad=\sigma_x\rho\phi\left(\frac{d-\mu_y}{\sigma_y}\right)\left[1-\Phi\left(\frac{\frac{c-\mu_x}{\sigma_x}-\rho\frac{d-\mu_y}{\sigma_y}}{\sqrt{1-\rho^2}}\right)\right]
\end{aligned}$$

For the third summand (A.2c) we revert the change of variable of z applied before and we obtain

$$\mu_x \mathrm{P}(x>c,y>d)=\mu_x F^I \tag{A.4}$$

By putting together (A.3), the two terms of the integration by parts and (A.4), we obtain that

$$\begin{aligned}
F^I\mathrm{E}^I_x=\mu_x F^I+\sigma_x\Bigg\{&\phi\left(\frac{c-\mu_x}{\sigma_x}\right)\left[1-\Phi\left(\frac{\frac{d-\mu_y}{\sigma_y}-\rho\frac{c-\mu_x}{\sigma_x}}{\sqrt{1-\rho^2}}\right)\right]\\
&+\rho\phi\left(\frac{d-\mu_y}{\sigma_y}\right)\left[1-\Phi\left(\frac{\frac{c-\mu_x}{\sigma_x}-\rho\frac{d-\mu_y}{\sigma_y}}{\sqrt{1-\rho^2}}\right)\right]\Bigg\}
\end{aligned}$$

Finally, the conditional expectation will be given by

$$\begin{aligned}
\mathrm{E}^I_x=\mu_x+\frac{\sigma_x}{F^I}\Bigg\{&\phi\left(\frac{c-\mu_x}{\sigma_x}\right)\left[1-\Phi\left(\frac{\frac{d-\mu_y}{\sigma_y}-\rho\frac{c-\mu_x}{\sigma_x}}{\sqrt{1-\rho^2}}\right)\right]\\
&+\rho\phi\left(\frac{d-\mu_y}{\sigma_y}\right)\left[1-\Phi\left(\frac{\frac{c-\mu_x}{\sigma_x}-\rho\frac{d-\mu_y}{\sigma_y}}{\sqrt{1-\rho^2}}\right)\right]\Bigg\}
\end{aligned}\tag{A.5}$$

We will also need the expected value of x given that we are in Case II of the four cases described in Chapter 3 (in other words, x is observed but the other variable is smaller than its truncation level d).

$$\mathrm{E}_x^{II} = \mathrm{E}(x|x > c, y < d) = \int_c^{\infty} x f(x|x > c, y < d)\, dx$$
$$= \frac{\int_{-\infty}^{d} \int_c^{\infty} x f(x,y)\, dx\, dy}{\mathrm{P}(x > c, y < d)}$$

By naming $F^{II} = \mathrm{P}(x > c, y < d)$ we write

$$F^{II}\mathrm{E}_x^{II} = \int_{-\infty}^{d} \int_c^{\infty} x f(x,y)\, dx\, dy$$

Following a procedure similar to the one described in the preceding paragraphs, we obtain the following result.

$$\mathrm{E}_x^{II} = \mu_x + \frac{\sigma_x}{F^{II}} \left\{ \phi\left(\frac{c-\mu_x}{\sigma_x}\right) \Phi\left(\frac{\frac{d-\mu_y}{\sigma_y} - \rho\frac{c-\mu_x}{\sigma_x}}{\sqrt{1-\rho^2}}\right) - \rho\phi\left(\frac{d-\mu_y}{\sigma_y}\right)\left[1 - \Phi\left(\frac{\frac{c-\mu_x}{\sigma_x} - \rho\frac{d-\mu_y}{\sigma_y}}{\sqrt{1-\rho^2}}\right)\right]\right\} \quad \text{(A.6)}$$

Two other developments are needed in order to obtain the marginal effects. The first results is

$$\frac{\partial F_I}{\partial \mu_x} = \frac{\partial}{\partial \mu_x}\mathrm{P}(x > c, y > d)$$

It can be seen that, since the limits of the integrals do not depend on μ_x, we can reverse the order of the derivation and the integration.

$$\frac{\partial F_I}{\partial \mu_x} = \frac{\partial}{\partial \mu_x} \int_d^\infty \int_c^\infty \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dx\,dy$$
$$= \int_d^\infty \int_c^\infty \frac{\partial}{\partial \mu_x}\left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot\right.$$
$$\left.\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dx\,dy\right)$$
$$= \int_d^\infty \int_c^\infty \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} \cdot$$
$$\left[\frac{1}{\sigma_x(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x} - \rho\left(\frac{y-\mu_y}{\sigma_y}\right)\right)\right] dx\,dy$$

If we make the same change of variable described in (A.1), we can write the previous expression in a more compact form.

$$\frac{\partial F_I}{\partial \mu_x} = \int_d^\infty \int_{c'}^\infty \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left[z^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}\left(\frac{1}{\sigma_x\sqrt{1-\rho^2}} z\right)\sigma_x\sqrt{1-\rho^2}\,dz\,dy$$
$$= \int_d^\infty \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left[(c')^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dy$$

Both parts of the exponential depend on y so the integral cannot be calculated directly. We perform the following change of variable, parallel to that introduced before.

$$w = \frac{\frac{y-\mu_y}{\sigma_y} - \rho\frac{c-\mu_x}{\sigma_x}}{\sqrt{1-\rho^2}} \text{ and } d' = \frac{\frac{d-\mu_y}{\sigma_y} - \rho\frac{c-\mu_x}{\sigma_x}}{\sqrt{1-\rho^2}}$$

Through this change of variable, the preceding expression can be transformed as follows

$$\begin{aligned}\frac{\partial F_I}{\partial \mu_x} &= \int_{d'}^{\infty} \frac{1}{2\pi\sigma_x} \exp\left\{-\frac{1}{2}\left[w^2 + \left(\frac{c-\mu_x}{\sigma_x}\right)^2\right]\right\} dw \\ &= \frac{1}{\sigma_x} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{c-\mu_x}{\sigma_x}\right)^2\right\} \cdot \int_{d'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}w^2\right\} dw \\ &= \frac{1}{\sigma_x}\phi\left(\frac{c-\mu_x}{\sigma_x}\right)\left[1 - \Phi\left(\frac{\frac{d-\mu_y}{\sigma_y} - \rho\frac{c-\mu_x}{\sigma_x}}{\sqrt{1-\rho^2}}\right)\right] \qquad \text{(A.7)}\end{aligned}$$

In a parallel way, we obtain the following result for the derivative of $F_{II} = \mathrm{P}(x > c, y < d)$ with respect to μ_x.

$$\frac{\partial F_{II}}{\partial \mu_x} = \frac{1}{\sigma_x}\phi\left(\frac{c-\mu_x}{\sigma_x}\right)\Phi\left(\frac{\frac{d-\mu_y}{\sigma_y} - \rho\frac{c-\mu_x}{\sigma_x}}{\sqrt{1-\rho^2}}\right) \qquad \text{(A.8)}$$

Appendix B

Gradient of the log likelihood function

B.1 Random effects model

We have the likelihood function of the random effects model for the individual i given by equation (3.14). The objective is to calculate the derivative of this function with respect to each of the parameters. Aggregated over all individuals, these will be the components of the gradient vector (3.16). With this objective in mind, the logarithm of this function can be separated into four parts, each one corresponding to one of the cases described in Chapter 3.

$$\log L_i = \log L_i^{\mathrm{I}} + \log L_i^{\mathrm{II}} + \log L_i^{\mathrm{III}} + \log L_i^{\mathrm{IV}}$$

We will calculate the derivatives of each of these subcomponents of the likelihood function separately. In order to abbreviate the expressions, denote

$$\begin{aligned}
e^{\mathrm{I}}_{1it} &= \Gamma y_{1it} - \Pi_1' X_{it} - \delta_{1i} \\
e^{\mathrm{I}}_{2it} &= \Gamma y_{2it} - \Pi_2' X_{it} - \delta_{2i} \\
e^{\mathrm{II}}_{1it} &= y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i} \\
e^{\mathrm{III}}_{2it} &= y_{2it} - \beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}
\end{aligned}$$

We start with the first parameter, β_{11}. For the first section of the log likelihood,

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{I}}}{\partial \beta_{11}} &= \frac{1}{L_i^{\mathrm{I}}} \frac{\partial L_i^{\mathrm{I}}}{\partial \beta_{11}} \\
&= \frac{1}{L_i^{\mathrm{I}}} \frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \beta_{11}} \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right) \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i}
\end{aligned}$$

To calculate this derivative, we need to use the general version of the product rule:

$$\frac{\partial}{\partial x} \prod_{i=1}^{N} \Psi_i(x) = \left(\prod_{i=1}^{N} \Psi_i(x) \right) \left(\sum_{i=1}^{N} \frac{\frac{\partial}{\partial x} \Psi_i(x)}{\Psi_i(x)} \right)$$

By using this formula, the expression in parenthesis can be written in the following form. Note that in the last term (the expression with the summation), $g^{\mathrm{I}}(\delta_{1i}, \delta_{2i})$ does not depend on β_{11} and consequently becomes a factor outside the derivative in the numerator that can be canceled out with the same function in the denominator.

$$\frac{\partial}{\partial \beta_{11}} \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) =$$

$$= \left(\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right) \left(\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{\frac{\partial}{\partial \beta_{11}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i})}{f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i})} \right)$$

The derivative of the numerator of the last term is

$$\begin{aligned}
&\frac{\partial}{\partial \beta_{11}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) = \\
&\quad = \frac{\partial}{\partial \beta_{11}} \left(\frac{1}{\sqrt{2\pi \sigma_{11}^{\nu} \sigma_{22}^{\nu} (1 - \rho_{\nu}^2)}} \exp\left\{ -\frac{1}{2(1-\rho_{\nu}^2)} \left[\frac{(e_{1it}^{\mathrm{I}})^2}{\sigma_{11}^{\nu}} + \frac{(e_{2it}^{\mathrm{I}})^2}{\sigma_{22}^{\nu}} - \frac{2\rho_{\nu} e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}}}{\sqrt{\sigma_{11}^{\nu} \sigma_{22}^{\nu}}} \right] \right\} \right) \\
&\quad = \frac{1}{\sqrt{2\pi \sigma_{11}^{\nu} \sigma_{22}^{\nu} (1 - \rho_{\nu}^2)}} \exp\{\cdot\} \left(-\frac{1}{2(1-\rho_{\nu}^2)} \right) \left[\frac{2e_{1it}^{\mathrm{I}}(-x_{1it})}{\sigma_{11}^{\nu}} + \frac{2e_{2it}^{\mathrm{I}}(-\gamma_2 x_{1it})}{\sigma_{22}^{\nu}} \right. \\
&\qquad \left. - \frac{2\rho_{\nu}}{\sqrt{\sigma_{11}^{\nu} \sigma_{22}^{\nu}}} \Big((-x_{1it}) e_{2it}^{\mathrm{I}} + e_{iit}^{\mathrm{I}}(-\gamma_2 x_{1it}) \Big) \right] \\
&\quad = f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) \frac{x_{1it}}{1-\rho_{\nu}^2} \left[\frac{e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{\gamma_2 e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{\rho_{\nu} \left(e_{2it}^{\mathrm{I}} + \gamma_2 e_{1it}^{\mathrm{I}} \right)}{\sqrt{\sigma_{11}^{\nu} \sigma_{22}^{\nu}}} \right]
\end{aligned}$$

Putting these results together, we obtain the following result.

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{I}}}{\partial \beta_{11}} = \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} &\left(\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right) \\
&\left(\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{1it}}{1-\rho_{\nu}^2} \left[\frac{e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{\gamma_2 e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{\rho_{\nu} \left(e_{2it}^{\mathrm{I}} + \gamma_2 e_{1it}^{\mathrm{I}} \right)}{\sqrt{\sigma_{11}^{\nu} \sigma_{22}^{\nu}}} \right] \right) \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i}
\end{aligned}$$

For the second part of the likelihood,

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{II}}}{\partial \beta_{11}} &= \frac{1}{L_i^{\mathrm{II}}} \frac{\partial L_i^{\mathrm{II}}}{\partial \beta_{11}} \\
&= \frac{1}{L_i^{\mathrm{II}}} \frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \beta_{11}} \prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right) \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right) \\
&\qquad \left(\sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{\frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \, \mathrm{d}\nu_{2it}}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \, \mathrm{d}\nu_{2it}} \right) \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i}
\end{aligned}$$

The numerator of the sum in the previous formula features the derivative of an integral with an upper limit that depends on the variable of differentiation. This derivative is solved in the following way.

$$
\begin{aligned}
&\frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it} = \\
&= \frac{\partial}{\partial \beta_{11}} \left(-\Pi_2' X_{it} - \delta_{2i} \right) f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i} | \alpha_{1i}, \delta_{2i} \right) \\
&\quad + \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} \frac{\partial}{\partial \beta_{11}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it} \\
&= -\gamma_2 x_{1it} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i} | \alpha_{1i}, \delta_{2i} \right) \\
&\quad + \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} \frac{\partial}{\partial \beta_{11}} \left(\frac{1}{\sqrt{2\pi \sigma_{11}^{\epsilon} \sigma_{22}^{\nu} (1 - \rho_{\epsilon,\nu}^2)}} \exp \left\{ -\frac{1}{2(1 - \rho_{\epsilon,\nu}^2)} \left[\frac{(e_{1it}^{\mathrm{II}})^2}{\sigma_{11}^{\epsilon}} + \frac{\nu_{2it}^2}{\sigma_{22}^{\nu}} \right. \right. \right. \\
&\quad \left. \left. \left. - \frac{2\rho_{\epsilon,\nu} e_{1it}^{\mathrm{II}} \nu_{2it}}{\sqrt{\sigma_{11}^{\epsilon} \sigma_{22}^{\nu}}} \right] \right\} \right) \mathrm{d}\nu_{2it} \\
&= -\gamma_2 x_{1it} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i} | \alpha_{1i}, \delta_{2i} \right) \\
&\quad + \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \frac{x_{1it}}{1 - \rho_{\epsilon,\nu}^2} \left(\frac{e_{1it}^{\mathrm{II}}}{\sigma_{11}^{\epsilon}} - \frac{\rho_{\epsilon,\nu} \nu_{2it}}{\sqrt{\sigma_{11}^{\epsilon} \sigma_{22}^{\nu}}} \right) \mathrm{d}\nu_{2it}
\end{aligned}
$$

So that the derivative with respect to β_{11} of the second component of the likelihood function becomes

$$
\begin{aligned}
\frac{\partial \log L_i^{\mathrm{II}}}{\partial \beta_{11}} &= \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right) \\
&\left(\sum_{\mathbb{D}_i^{\mathrm{II}}} x_{1it} \frac{\frac{1}{1-\rho_{\epsilon,\nu}^2} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \cdot \right) \left(\frac{e_{1it}^{\mathrm{II}}}{\sigma_{11}^{\epsilon}} - \frac{\rho_{\epsilon,\nu} \nu_{2it}}{\sqrt{\sigma_{11}^{\epsilon} \sigma_{22}^{\nu}}} \right) \mathrm{d}\nu_{2it} - \gamma_2 f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i} | \cdot \right)}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it}} \right) \\
&\mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i}
\end{aligned}
$$

Continuing with the third part of the likelihood, corresponding to those observations for individual i that fall under Case III.

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{III}}}{\partial \beta_{11}} &= \frac{1}{L_i^{\mathrm{III}}} \frac{\partial L_i^{\mathrm{III}}}{\partial \beta_{11}} \\
&= \frac{1}{L_i^{\mathrm{III}}} \frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \beta_{11}} \prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right) \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right) \\
&\qquad \left(\sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{\frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it}}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it}} \right) \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}$$

The derivative in the numerator of the second factor can be calculated by following the next steps.

$$\frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} =$$

$$\begin{aligned}
&= \frac{\partial}{\partial \beta_{11}} \left(-\Pi_1' X_{it} - \delta_{1i} \right) f^{\mathrm{III}} \left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) + \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{\partial}{\partial \beta_{11}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} \\
&= -x_{1it} f^{\mathrm{III}} \left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}} | \alpha_{1i}, \delta_{2i} \right) + 0
\end{aligned}$$

The second summand from the last expression is equal to zero because the expression does not depend on β_{11}. The derivative of the third part of the likelihood with respect to this coefficients is then

$$\frac{\partial \log L_i^{\mathrm{III}}}{\partial \beta_{11}} = \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}\right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right)$$
$$\left(\sum_{\mathbb{D}_i^{\mathrm{III}}} x_{1it} \frac{-f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}\right)}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}\right) \mathrm{d}\nu_{1it}} \right) \mathrm{d}\delta_{1i}\, \mathrm{d}\alpha_{2i}$$

Finally, we have the last part of the likelihood,

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{IV}}}{\partial \beta_{11}} &= \frac{1}{L_i^{\mathrm{IV}}} \frac{\partial L_i^{\mathrm{IV}}}{\partial \beta_{11}} \\
&= \frac{1}{L_i^{\mathrm{IV}}} \frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \\
&\quad \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \beta_{11}} \prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \right. \\
&\quad \left. \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right) \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \cdot\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{\frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it}}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it}} \right) \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}
\end{aligned}$$

We calculate the derivative in the numerator of the preceding expression separately.

$$
\begin{aligned}
&\frac{\partial}{\partial \beta_{11}} \int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} = \\
&= \int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} \frac{\partial}{\partial \beta_{11}} \left(\int_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it} \right) \mathrm{d}\epsilon_{2it} \\
&= \int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} \Bigg[\frac{\partial}{\partial \beta_{11}} \left(-\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i}\right) f^{\mathrm{IV}}\left(-\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \\
&\qquad + \int_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} \frac{\partial}{\partial \beta_{11}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it} \Bigg] \mathrm{d}\epsilon_{2it} \\
&= \int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} (-x_{1it}) f^{\mathrm{IV}}\left(-\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{2it}
\end{aligned}
$$

If we insert this result in the expression preceding it, we obtain the derivative for the fourth part of the likelihood

$$
\begin{aligned}
&\frac{\partial \log L_i^{\mathrm{IV}}}{\partial \beta_{11}} = \\
&= \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \cdot\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{1it} \frac{- \int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} f^{\mathrm{IV}}\left(-\beta'_{11}x_{1it} - \beta'_{12}x_{it} - \alpha_{1i}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{2it}}{\int_{-\infty}^{-\beta'_{21}x_{2it}-\beta'_{22}x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta'_{11}x_{1it}-\beta'_{12}x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it}} \right) \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

By combining the results for the four parts of the likelihood, we obtain the full derivative of the likelihood with respect to β_{11}.

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \beta_{11}} = \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{1it}}{1-\rho_\nu^2} \left[\frac{e^{\mathrm{I}}_{1it}}{\sigma^\nu_{11}} + \frac{\gamma_2 e^{\mathrm{I}}_{2it}}{\sigma^\nu_{22}} - \frac{\rho_\nu \left(e^{\mathrm{I}}_{2it} + \gamma_2 e^{\mathrm{I}}_{1it} \right)}{\sqrt{\sigma^\nu_{11}\sigma^\nu_{22}}} \right] \right) \mathrm{d}\delta_{1i}\, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{II}}} x_{1it} \frac{ \frac{1}{1-\rho^2_{\epsilon,\nu}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e^{\mathrm{II}}_{1it}, \nu_{2it} | \cdot \right) \left(\frac{e^{\mathrm{II}}_{1it}}{\sigma^\epsilon_{11}} - \frac{\rho_{\epsilon,\nu} \nu_{2it}}{\sqrt{\sigma^\epsilon_{11}\sigma^\nu_{22}}} \right) \mathrm{d}\nu_{2it} - \gamma_2 f^{\mathrm{II}} \left(e^{\mathrm{II}}_{1it}, -\Pi_2' X_{it} - \delta_{2i} | \cdot \right) }{ \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it} } \right) \\
&\quad \mathrm{d}\alpha_{1i}\, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right) \qquad \text{(B.1)} \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{III}}} x_{1it} \frac{ -f^{\mathrm{III}} \left(-\Pi_1' X_{it} - \delta_{1i}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i} \right) }{ \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} } \right) \mathrm{d}\delta_{1i}\, \mathrm{d}\alpha_{2i} \\
&+ \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \cdot \right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{1it} \frac{ - \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} f^{\mathrm{IV}} \left(-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i} \right) \mathrm{d}\epsilon_{2it} }{ \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i} \right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it} } \right) \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

The derivatives with respect to the other slope parameters can be obtained with a parallel calculation. Here we will only report on the final result. The derivative of the log likelihood with respect to β_{12}, the coefficient of the common exogenous variable that appears in the first equation, is given by the following expression.

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \beta_{12}} = \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{it}}{1-\rho_\nu^2} \left[\frac{e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{\gamma_2 e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{\rho_\nu \left(e_{2it}^{\mathrm{I}} + \gamma_2 e_{1it}^{\mathrm{I}} \right)}{\sqrt{\sigma_{11}^{\nu} \sigma_{22}^{\nu}}} \right] \right) \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{II}}} x_{it} \frac{\frac{1}{1-\rho_{\epsilon,\nu}^2} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \cdot \right) \left(\frac{e_{1it}^{\mathrm{II}}}{\sigma_{11}^{\epsilon}} - \frac{\rho_{\epsilon,\nu} \nu_{2it}}{\sqrt{\sigma_{11}^{\epsilon} \sigma_{22}^{\nu}}} \right) \mathrm{d}\nu_{2it} - \gamma_2 f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i} | \cdot \right)}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it}} \right) \\
&\quad \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right) \qquad \text{(B.2)} \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{III}}} x_{it} \frac{-f^{\mathrm{III}} \left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right)}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}} \left(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i} \right) \mathrm{d}\nu_{1it}} \right) \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
&+ \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \cdot \right) \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{it} \frac{- \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} f^{\mathrm{IV}} \left(-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i} \right) \mathrm{d}\epsilon_{2it}}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i} \right) \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it}} \right) \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

Now, the derivative of the logarithm of the likelihood function with respect to the coefficient of the exogenous variable included only in the second equation, β_{21}, can be written as follows.

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \beta_{21}} = \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i},\delta_{2i}) g^{\mathrm{I}}(\delta_{1i},\delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{2it}}{1-\rho_\nu^2} \left[\frac{\gamma_1 e^{\mathrm{I}}_{1it}}{\sigma^\nu_{11}} + \frac{e^{\mathrm{I}}_{2it}}{\sigma^\nu_{22}} - \frac{\rho_\nu \left(e^{\mathrm{I}}_{1it} + \gamma_1 e^{\mathrm{I}}_{2it}\right)}{\sqrt{\sigma^\nu_{11}\sigma^\nu_{22}}} \right] \right) \mathrm{d}\delta_{1i}\,\mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e^{\mathrm{II}}_{1it}, \nu_{2it}|\alpha_{1i},\delta_{2i}\right) \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i},\delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{II}}} x_{2it} \frac{-f^{\mathrm{II}}\left(e^{\mathrm{II}}_{1it}, -\Pi_2' X_{it} - \delta_{2i}|\alpha_{1i},\delta_{2i}\right)}{\int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e^{\mathrm{II}}_{1it}, \nu_{2it}|\alpha_{1i},\delta_{2i}\right) \mathrm{d}\nu_{2it}} \right) \mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e^{\mathrm{III}}_{2it}|\delta_{1i},\alpha_{2i}\right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i},\alpha_{2i}) \right) \qquad \text{(B.3)} \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{III}}} x_{2it} \frac{\frac{1}{1-\rho^2_{\nu,\epsilon}} \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e^{\mathrm{III}}_{2it}|\cdot\right) \left(\frac{e^{\mathrm{III}}_{2it}}{\sigma^\epsilon_{22}} - \frac{\rho_{\nu,\epsilon}\nu_{1it}}{\sqrt{\sigma^\nu_{11}\sigma^\epsilon_{22}}} \right) \mathrm{d}\nu_{1it} - \gamma_1 f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e^{\mathrm{III}}_{2it}|\cdot\right)}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e^{\mathrm{III}}_{2it}|\delta_{1i},\alpha_{2i}\right) \mathrm{d}\nu_{1it}} \right) \\
&\quad \mathrm{d}\delta_{1i}\,\mathrm{d}\alpha_{2i} \\
&+ \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it},\epsilon_{2it}|\cdot\right) \mathrm{d}\epsilon_{1it}\,\mathrm{d}\epsilon_{2it} \cdot g^{\mathrm{IV}}(\alpha_{1i},\alpha_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{2it} \frac{- \int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}|\alpha_{1i},\alpha_{2i}\right) \mathrm{d}\epsilon_{1it}}{\int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it},\epsilon_{2it}|\alpha_{1i},\alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\,\mathrm{d}\epsilon_{2it}} \right) \mathrm{d}\alpha_{1i}\,\mathrm{d}\alpha_{2i}
\end{aligned}
$$

Finally, we obtain the derivative of the logarithm of the likelihood function with respect to β_{22}, the coefficient of the common exogenous variable that appears in the second equation.

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \beta_{22}} = \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i},\delta_{2i}) g^{\mathrm{I}}(\delta_{1i},\delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{it}}{1-\rho_\nu^2} \left[\frac{\gamma_1 e_{1it}^{\mathrm{I}}}{\sigma_{11}^\nu} + \frac{e_{2it}^{\mathrm{I}}}{\sigma_{22}^\nu} - \frac{\rho_\nu \left(e_{1it}^{\mathrm{I}} + \gamma_1 e_{2it}^{\mathrm{I}}\right)}{\sqrt{\sigma_{11}^\nu \sigma_{22}^\nu}} \right] \right) \mathrm{d}\delta_{1i}\,\mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}|\alpha_{1i},\delta_{2i}\right) \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i},\delta_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{II}}} x_{it} \frac{-f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}|\alpha_{1i},\delta_{2i}\right)}{\int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}|\alpha_{1i},\delta_{2i}\right) \mathrm{d}\nu_{2it}} \right) \mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i}\right) \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i},\alpha_{2i}) \right) \qquad \text{(B.4)} \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{III}}} x_{it} \frac{\frac{1}{1-\rho_{\nu,\epsilon}^2} \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}|\cdot\right) \left(\frac{e_{2it}^{\mathrm{III}}}{\sigma_{22}^\epsilon} - \frac{\rho_{\nu,\epsilon}\nu_{1it}}{\sqrt{\sigma_{11}^\nu \sigma_{22}^\epsilon}} \right) \mathrm{d}\nu_{1it} - \gamma_1 f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}|\cdot\right)}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i}\right) \mathrm{d}\nu_{1it}} \right) \\
&\quad \mathrm{d}\delta_{1i}\,\mathrm{d}\alpha_{2i} \\
&+ \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}|\cdot\right) \mathrm{d}\epsilon_{1it}\,\mathrm{d}\epsilon_{2it} \cdot g^{\mathrm{IV}}(\alpha_{1i},\alpha_{2i}) \right) \\
&\quad \left(\sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{it} \frac{- \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}|\alpha_{1i},\alpha_{2i}\right) \mathrm{d}\epsilon_{1it}}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}|\alpha_{1i},\alpha_{2i}\right) \mathrm{d}\epsilon_{1it}\,\mathrm{d}\epsilon_{2it}} \right) \mathrm{d}\alpha_{1i}\,\mathrm{d}\alpha_{2i}
\end{aligned}
$$

In order to derive the log likelihood with respect to the simultaneity parameters, γ_1 and γ_2, we first need to partially revert to the structural coefficients. This is because the individual effects of the reduced form are a function of the individual effects of the structural form and the simultaneity parameters.

$$\delta_1 = \alpha_1 + \gamma_1 \alpha_2$$

$$\delta_2 = \alpha_2 + \gamma_2 \alpha_1$$

We begin with the first of these parameters, γ_1. As before, we will proceed by obtaining the derivative for each of the four parts of the log likelihood. For the first part, we apply a change of variable to the density of the individual effects and we obtain that

$$h^{\mathrm{I}}(\alpha_1, \alpha_2) = \Gamma g^{\mathrm{I}}(\delta_1 = \alpha_1 + \gamma_1 \alpha_2, \delta_2 = \alpha_2 + \gamma_2 \alpha_1)$$

So that the derivative for this part of the log likelihood can be calculated according to the following expression.

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{I}}}{\partial \gamma_1} &= \frac{1}{L_i^{\mathrm{I}}} \frac{\partial L_i^{\mathrm{I}}}{\partial \gamma_1} \\
&= \frac{1}{L_i^{\mathrm{I}}} \frac{\partial}{\partial \gamma_1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \gamma_1} \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i}) \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i}) \right] \\
&\qquad \cdot \left[\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{\frac{\partial}{\partial \gamma_1} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i})}{f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i})} \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}$$

with $\delta_1 = \alpha_1 + \gamma_1 \alpha_2$ and $\delta_2 = \alpha_2 + \gamma_2 \alpha_1$. The derivative in the numerator of the last term can in turn be obtained by applying the product rule.

$$\frac{\partial}{\partial \gamma_1} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \delta_{2i} = \alpha_2 + \gamma_2\alpha_1) h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i}) =$$
$$= \left(\frac{\partial}{\partial \gamma_1} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \delta_{2i} = \alpha_2 + \gamma_2\alpha_1)\right) \Gamma g^{\mathrm{I}}(\delta_1 = \alpha_1 + \gamma_1\alpha_2, \delta_2 = \alpha_2 + \gamma_2\alpha_1)$$
$$+ f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \delta_{2i} = \alpha_2 + \gamma_2\alpha_1) \left(\frac{\partial}{\partial \gamma_1}\Gamma\right) g^{\mathrm{I}}(\delta_1 = \alpha_1 + \gamma_1\alpha_2, \delta_2 = \alpha_2 + \gamma_2\alpha_1)$$
$$+ f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \delta_{2i} = \alpha_2 + \gamma_2\alpha_1)\Gamma \left(\frac{\partial}{\partial \gamma_1} g^{\mathrm{I}}(\delta_1 = \alpha_1 + \gamma_1\alpha_2, \delta_2 = \alpha_2 + \gamma_2\alpha_1)\right)$$

We calculate each of the three derivatives in the previous equation.

- $\frac{\partial}{\partial\gamma_1} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \delta_{2i} = \alpha_2 + \gamma_2\alpha_1) =$

$$= \frac{\partial}{\partial\gamma_1}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma^{\nu}_{11}\sigma^{\nu}_{22}(1-\rho^2_{\nu})}}\exp\left\{-\frac{1}{2(1-\rho^2_{\nu})}\left[\frac{(e^{\mathrm{I}}_{1it})^2}{\sigma^{\nu}_{11}} + \frac{(e^{\mathrm{I}}_{2it})^2}{\sigma^{\nu}_{22}} - \frac{2\rho_{\nu}e^{\mathrm{I}}_{1it}e^{\mathrm{I}}_{2it}}{\sqrt{\sigma^{\nu}_{11}\sigma^{\nu}_{22}}}\right]\right\}\right)$$
$$= \frac{\partial}{\partial\gamma_1}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2}}\exp\left\{-\frac{\sigma^{\nu}_{22}(e^{\mathrm{I}}_{1it})^2 + \sigma^{\nu}_{11}(e^{\mathrm{I}}_{2it})^2 - 2\sigma^{\nu}_{12}e^{\mathrm{I}}_{1it}e^{\mathrm{I}}_{2it}}{2(\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2)}\right\}\right)$$
$$= -\frac{\sigma^{\nu}_{22}(\gamma_1\sigma^{\epsilon}_{22}+\sigma^{\epsilon}_{12}) - \sigma^{\nu}_{12}(\gamma_2\sigma^{\epsilon}_{12}+\sigma^{\epsilon}_{22})}{\sqrt{2\pi}(\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2)^{3/2}}\exp\{\cdot\} + \frac{1}{\sqrt{2\pi}\sqrt{\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2}}\exp\{\cdot\}$$
$$\cdot\left\{\frac{\sigma^{\nu}_{22}(\gamma_1\sigma^{\epsilon}_{22}+\sigma^{\epsilon}_{12}) - \sigma^{\nu}_{12}(\gamma_2\sigma^{\epsilon}_{12}+\sigma^{\epsilon}_{22})}{(\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2)^2}\left[\sigma^{\nu}_{22}(e^{\mathrm{I}}_{1it})^2 + \sigma^{\nu}_{11}(e^{\mathrm{I}}_{2it})^2 - 2\sigma^{\nu}_{12}e^{\mathrm{I}}_{1it}e^{\mathrm{I}}_{2it}\right]\right.$$
$$-\frac{1}{\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2}\left[\sigma^{\nu}_{22}e^{\mathrm{I}}_{1it}(-\gamma_2 y_{1it} - \beta_{21}x_{2it} - \beta_{22}x_{it} - \alpha_2) + (e^{\mathrm{I}}_{2it})^2(\gamma_1\sigma^{\epsilon}_{22}+\sigma^{\epsilon}_{12})\right.$$
$$+ 2\sigma^{\nu}_{11}e^{\mathrm{I}}_{2it}(-\gamma_2 y_{2it}) - e^{\mathrm{I}}_{1it}e^{\mathrm{I}}_{2it}(\gamma_2\sigma^{\epsilon}_{12}+\sigma^{\epsilon}_{22}) - \sigma^{\nu}_{12}e^{\mathrm{I}}_{2it}(-\gamma_2 y_{1it} - \beta_{12}x_{2it} - \beta_{22}x_{it} - \alpha_2)$$
$$\left.\left.-2\sigma^{\nu}_{12}e^{\mathrm{I}}_{1it}(-\gamma_2 y_{2it})\right]\right\}$$
$$= \frac{f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \delta_{2i} = \alpha_2 + \gamma_2\alpha_1)}{\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2}$$
$$\cdot\left\{\left(\frac{\sigma^{\nu}_{22}(e^{\mathrm{I}}_{1it})^2 + \sigma^{\nu}_{11}(e^{\mathrm{I}}_{2it})^2 - 2\sigma^{\nu}_{12}e^{\mathrm{I}}_{1it}e^{\mathrm{I}}_{2it}}{\sigma^{\nu}_{11}\sigma^{\nu}_{22}-(\sigma^{\nu}_{12})^2} - 1\right)[\sigma^{\nu}_{22}(\gamma_1\sigma^{\epsilon}_{22}+\sigma^{\epsilon}_{12}) - \sigma^{\nu}_{12}(\gamma_2\sigma^{\epsilon}_{12}+\sigma^{\epsilon}_{22})]\right.$$
$$-(\gamma_2 y_{1it} + \beta_{12}x_{2it} + \beta_{22}x_{it} + \alpha_2)(\sigma^{\nu}_{12}e^{\mathrm{I}}_{2it} - \sigma^{\nu}_{22}e^{\mathrm{I}}_{1it}) + e^{\mathrm{I}}_{2it}[e^{\mathrm{I}}_{1it}(\gamma_2\sigma^{\epsilon}_{12}+\sigma^{\epsilon}_{22}) - e^{\mathrm{I}}_{2it}(\sigma^{\epsilon}_{12}+\gamma_1\sigma^{\epsilon}_{22})]$$
$$\left.+\gamma_2 y_{2it}(\sigma^{\nu}_{11}e^{\mathrm{I}}_{2it} - \sigma^{\nu}_{12}e^{\mathrm{I}}_{1it})\right\}$$

- $\frac{\partial\Gamma}{\partial\gamma_1} = -\gamma_2$
- $\frac{\partial}{\partial\gamma_1} g^{\mathrm{I}}(\delta_1 = \alpha_1 + \gamma_1\alpha_2, \delta_2 = \alpha_2 + \gamma_2\alpha_1) =$

$$= \frac{\partial}{\partial\gamma_1}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2}}\exp\left\{-\frac{\sigma^{\delta}_{22}(\alpha_1+\gamma_1\alpha_2)^2 + \sigma^{\delta}_{11}(\alpha_2+\gamma_2\alpha_1)^2 - 2\sigma^{\delta}_{12}(\alpha_1+\gamma_1\alpha_2)(\alpha_2+\gamma_2\alpha_1)}{2(\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2)}\right\}\right)$$
$$= -\frac{\sigma^{\delta}_{22}(\gamma_1\sigma^{\alpha}_{22}+\sigma^{\alpha}_{12}) - \sigma^{\delta}_{12}(\gamma_2\sigma^{\alpha}_{12}+\sigma^{\alpha}_{22})}{\sqrt{2\pi}(\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2)^{3/2}}\exp\{\cdot\} + \frac{1}{\sqrt{2\pi}\sqrt{\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2}}\exp\{\cdot\}$$
$$\cdot\left\{\frac{\sigma^{\delta}_{22}(\gamma_1\sigma^{\alpha}_{22}+\sigma^{\alpha}_{12}) - \sigma^{\delta}_{12}(\gamma_2\sigma^{\alpha}_{12}+\sigma^{\alpha}_{22})}{(\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2)^2}\left[\sigma^{\delta}_{22}(\alpha_1+\gamma_1\alpha_2)^2 + \sigma^{\delta}_{11}(\alpha_2+\gamma_2\alpha_1)^2 - 2\sigma^{\delta}_{12}(\alpha_1+\gamma_1\alpha_2)(\alpha_2+\gamma_2\alpha_1)\right]\right.$$
$$-\frac{1}{\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2}\left[\sigma^{\delta}_{22}(\alpha_1+\gamma_1\alpha_2)\alpha_2 + (\alpha_2+\gamma_2\alpha_1)^2(\gamma_1\sigma^{\alpha}_{22}+\sigma^{\alpha}_{12})\right.$$
$$\left.\left.-(\alpha_1+\gamma_1\alpha_2)(\alpha_2+\gamma_2\alpha_1)(\gamma_2\sigma^{\alpha}_{12}+\sigma^{\alpha}_{22}) - \sigma^{\delta}_{12}(\alpha_2+\gamma_2\alpha_1)\alpha_1\right]\right\}$$
$$= \frac{g^{\mathrm{I}}(\alpha_1+\gamma_1\alpha_2, \alpha_2+\gamma_2\alpha_1)}{\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2}\left\{\left(\frac{\sigma^{\delta}_{22}(\alpha_1+\gamma_1\alpha_2)^2 + \sigma^{\delta}_{11}(\alpha_2+\gamma_2\alpha_1)^2 - 2\sigma^{\delta}_{12}(\alpha_1+\gamma_1\alpha_2)(\alpha_2+\gamma_2\alpha_1)}{\sigma^{\delta}_{11}\sigma^{\delta}_{22}-(\sigma^{\delta}_{12})^2} - 1\right)\right.$$
$$\cdot\left[\sigma^{\delta}_{22}(\gamma_1\sigma^{\alpha}_{22}+\sigma^{\alpha}_{12}) - \sigma^{\delta}_{12}(\gamma_2\sigma^{\alpha}_{12}+\sigma^{\alpha}_{22})\right] + \alpha_2[\sigma^{\delta}_{12}(\alpha_2+\gamma_2\alpha_1) - \sigma^{\delta}_{22}(\alpha_1+\gamma_1\alpha_2)]$$
$$\left.+(\alpha_2+\gamma_2\alpha_1)[(\alpha_1+\gamma_1\alpha_2)(\gamma_2\sigma^{\alpha}_{12}+\sigma^{\alpha}_{22}) - (\alpha_2+\gamma_2\alpha_1)(\sigma^{\alpha}_{12}+\gamma_1\sigma^{\alpha}_{22})]\right\}$$

With these elements we can write the derivative of the log likelihood for observations falling in Case I with respect to γ_1.

$$\frac{\partial \log L_i^{\mathrm{I}}}{\partial \gamma_1} = \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i} = \alpha_1 + \gamma_1 \alpha_2, \delta_{2i} = \alpha_2 + \gamma_2 \alpha_1) h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i}) \right]$$

$$\sum_{\mathbb{D}_i^{\mathrm{I}}} \left\{ \frac{1}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu} e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) [\sigma_{22}^{\nu}(\gamma_1 \sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{12}^{\nu}(\gamma_2 \sigma_{12}^{\epsilon} + \sigma_{22}^{\epsilon})] \right.\right.$$

$$- (\gamma_2 y_{1it} + \beta_{21} x_{2it} + \beta_{22} x_{it} + \alpha_2)(\sigma_{12}^{\nu} e_{2it}^{\mathrm{I}} - \sigma_{22}^{\nu} e_{1it}^{\mathrm{I}}) + e_{2it}^{\mathrm{I}}[e_{1it}^{\mathrm{I}}(\gamma_2 \sigma_{12}^{\epsilon} + \sigma_{22}^{\epsilon}) - e_{2it}^{\mathrm{I}}(\sigma_{12}^{\epsilon} + \gamma_1 \sigma_{22}^{\epsilon})]$$

$$\left. + \gamma_2 y_{2it}(\sigma_{11}^{\nu} e_{2it}^{\mathrm{I}} - \sigma_{12}^{\nu} e_{1it}^{\mathrm{I}}) \right] - \frac{\gamma_2}{\Gamma}$$

$$+ \frac{1}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} \left[\left(\frac{\sigma_{22}^{\delta}(\alpha_1 + \gamma_1\alpha_2)^2 + \sigma_{11}^{\delta}(\alpha_2 + \gamma_2\alpha_1)^2 - 2\sigma_{12}^{\delta}(\alpha_1 + \gamma_1\alpha_2)(\alpha_2 + \gamma_2\alpha_1)}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} - 1 \right) \right.$$

$$\left[\sigma_{22}^{\delta}(\gamma_1 \sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{12}^{\delta}(\gamma_2 \sigma_{12}^{\alpha} + \sigma_{22}^{\alpha})\right] + \alpha_2[\sigma_{12}^{\delta}(\alpha_2 + \gamma_2\alpha_1) - \sigma_{22}^{\delta}(\alpha_1 + \gamma_1\alpha_2)]$$

$$\left.\left. + (\alpha_2 + \gamma_2\alpha_1)[(\alpha_1 + \gamma_1\alpha_2)(\gamma_2\sigma_{12}^{\alpha} + \sigma_{22}^{\alpha}) - (\alpha_2 + \gamma_2\alpha_1)(\sigma_{12}^{\alpha} + \gamma_1\sigma_{22}^{\alpha})] \right] \right\} \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}$$

The second part of the logarithm of the likelihood can be written as follows.

$$\log L_i^{\mathrm{II}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it} | \alpha_{1i}, \delta_{2i} \right) \mathrm{d}\nu_{2it}$$
$$\cdot\, g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\delta_{2i}$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}} \left(y_{1it} - \beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \nu_{2it} | \alpha_{1i}, \delta_{2i} = \alpha_{2i} + \gamma_2 \alpha_{1i} \right) \mathrm{d}\nu_{2it}$$
$$\cdot\, h^{\mathrm{II}}(\alpha_{1i}, \alpha_{2i})\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}$$

where $h^{\mathrm{II}}(\alpha_{1i}, \alpha_{2i})$ is obtained by applying a change of variable to the density of the individual effects $g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i})$, so that

$$h^{\mathrm{II}}(\alpha_{1i}, \alpha_{2i}) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2}} \exp\left\{ - \frac{\sigma_{22}^{\delta}\alpha_1^2 + \sigma_{11}^{\alpha}(\alpha_2 + \gamma_2\alpha_1)^2 - 2\sigma_{12}^{\alpha,\delta}\alpha_1(\alpha_2 + \gamma_2\alpha_1)}{2(\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2)} \right\}$$

It can be seen that neither $f^{\mathrm{II}}(\cdot)$, $h^{\mathrm{II}}(\alpha_{1i}, \alpha_{2i})$ nor any other part of the log likelihood depend on γ_1, so the derivative of this second part of L_i equals zero. We move on to the third section of the log likelihood. We also perform a change of variable of the density function of the individual effects in order to explicit the presence of the variable of derivation, γ_1.

$$h^{\mathrm{III}}(\alpha_1, \alpha_2) = g^{\mathrm{III}}(\delta_1 = \alpha_1 + \gamma_1 \alpha_2, \alpha_2)$$

So that the derivative becomes

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{III}}}{\partial \gamma_1} &= \frac{1}{L_i^{\mathrm{III}}} \frac{\partial L_i^{\mathrm{III}}}{\partial \gamma_1} \\
&= \frac{1}{L_i^{\mathrm{III}}} \frac{\partial}{\partial \gamma_1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \gamma_1} \prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \\
&= \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \right] \\
&\qquad \cdot \left[\sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{\frac{\partial}{\partial \gamma_1} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i})}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i})} \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}$$

The derivative in the numerator of the last term is equal to

$$\begin{aligned}
&\frac{\partial}{\partial \gamma_1} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) = \\
&\quad = \frac{\partial}{\partial \gamma_1} \left(-\Pi_1' X_{it} - \delta_{1i} \right) f^{\mathrm{III}}(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \\
&\quad + \left(\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{\partial}{\partial \gamma_1} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \right) h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \\
&\quad + \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \left(\frac{\partial}{\partial \gamma_1} h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \right)
\end{aligned}$$

It can be seen that this last expression contains these three derivatives

- $\frac{\partial}{\partial\gamma_1}\left(-\Pi_1' X_{it} - \delta_{1i}\right) = -\beta_{21}x_{2it} - \beta_{22}x_{it} - \alpha_2$
- $\frac{\partial}{\partial\gamma_1} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i}, \delta_{2i}) =$

$$= \frac{\partial}{\partial\gamma_1}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2}}\exp\left\{-\frac{\sigma_{22}^{\epsilon}v_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}}{2(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2)}\right\}\right)$$

$$= -\frac{\sigma_{22}^{\epsilon}(\gamma_1\sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{12}^{\nu,\epsilon}\sigma_{22}^{\epsilon}}{\sqrt{2\pi}(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2)^{3/2}}\exp\{\cdot\} + \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2}}\exp\{\cdot\}$$
$$\cdot\left\{\frac{\sigma_{22}^{\epsilon}(\gamma_1\sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{12}^{\nu,\epsilon}\sigma_{22}^{\epsilon}}{(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2)^2}\left[\sigma_{22}^{\epsilon}\nu_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}\right]\right.$$
$$\left. - \frac{1}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2}\left[(e_{2it}^{\mathrm{III}})^2(\gamma_1\sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{22}^{\epsilon}v_{1it}e_{2it}^{\mathrm{III}}\right]\right\}$$
$$= \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \alpha_2)}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2}$$
$$\cdot\left\{\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} - 1\right)\left[\sigma_{22}^{\epsilon}(\gamma_1\sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{12}^{\nu,\epsilon}\sigma_{22}^{\epsilon}\right]\right.$$
$$\left. + e_{2it}^{\mathrm{III}}[\sigma_{22}^{\epsilon}\nu_{1it} - e_{2it}^{\mathrm{III}}(\sigma_{12}^{\epsilon} + \gamma_1\sigma_{22}^{\epsilon})]\right\}$$

- $\frac{\partial}{\partial\gamma_1} h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) =$

$$= \frac{\partial}{\partial\gamma_1}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2}}\exp\left\{-\frac{\sigma_{22}^{\alpha}(\alpha_1 + \gamma_1\alpha_2)^2 + \sigma_{11}^{\delta}\alpha_2^2 - 2\sigma_{12}^{\delta,\alpha}(\alpha_1 + \gamma_1\alpha_2)\alpha_2}{2(\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2)}\right\}\right)$$
$$= -\frac{\sigma_{22}^{\alpha}(\gamma_1\sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{12}^{\delta,\alpha}\sigma_{22}^{\alpha}}{\sqrt{2\pi}(\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2)^{3/2}}\exp\{\cdot\} + \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2}}\exp\{\cdot\}$$
$$\cdot\left\{\frac{\sigma_{22}^{\alpha}(\gamma_1\sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{12}^{\delta,\alpha}\sigma_{22}^{\alpha}}{(\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2)^2}\left[\sigma_{22}^{\alpha}(\alpha_1 + \gamma_1\alpha_2)^2 + \sigma_{11}^{\delta}\alpha_2^2 - 2\sigma_{12}^{\delta,\alpha}(\alpha_1 + \gamma_1\alpha_2)\alpha_2\right]\right.$$
$$\left. - \frac{\sigma_{22}^{\alpha}(\alpha_1 + \gamma_1\alpha_2)\alpha_2 + \alpha_2^2(\gamma_1\sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{22}^{\alpha}(\alpha_1 + \gamma_1\alpha_2)\alpha_2 - \sigma_{12}^{\delta,\alpha}\alpha_2^2}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2}\right\}$$
$$= \frac{h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i})}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2}\left\{\left(\frac{\sigma_{22}^{\alpha}(\alpha_1 + \gamma_1\alpha_2)^2 + \sigma_{11}^{\delta}\alpha_2^2 - 2\sigma_{12}^{\delta,\alpha}(\alpha_1 + \gamma_1\alpha_2)\alpha_2}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} - 1\right)\right.$$
$$\left.\cdot\left[\sigma_{22}^{\alpha}(\gamma_1\sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{12}^{\delta,\alpha}\sigma_{22}^{\alpha}\right] + \alpha_2^2[\sigma_{12}^{\delta,\alpha} - (\sigma_{12}^{\alpha} + \gamma_1\sigma_{22}^{\alpha})]\right\}$$

By putting these elements together, we obtain the derivative of the log likelihood with respect to the first simultaneity parameter for those observations under Case III.

$$\frac{\partial \log L_i^{\mathrm{III}}}{\partial \gamma_1} = \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \right]$$

$$\cdot \sum_{\mathbb{D}_i^{\mathrm{III}}} \left\{ \frac{1}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it}} \right.$$

$$\left\{ -(\beta_{21} x_{2it} + \beta_{22} x_{it} + \alpha_2) f^{\mathrm{III}}(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \right.$$

$$+ \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i} = \alpha_1 + \gamma_1 \alpha_2, \alpha_2)}{\sigma_{11}^{\nu} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2}$$

$$\cdot \left[\left(\frac{\sigma_{22}^{\epsilon} \nu_{1it}^2 + \sigma_{11}^{\nu} (e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon} \nu_{1it} e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} - 1 \right) \left[\sigma_{22}^{\epsilon} (\gamma_1 \sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{12}^{\nu,\epsilon} \sigma_{22}^{\epsilon} \right] \right.$$

$$\left. \left. + e_{2it}^{\mathrm{III}} [\sigma_{22}^{\epsilon} \nu_{1it} - e_{2it}^{\mathrm{III}} (\sigma_{12}^{\epsilon} + \gamma_1 \sigma_{22}^{\epsilon})] \right] \mathrm{d}\nu_{1it} \right\}$$

$$+ \frac{1}{\sigma_{11}^{\delta} \sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} \left[\left(\frac{\sigma_{22}^{\alpha} (\alpha_1 + \gamma_1 \alpha_2)^2 + \sigma_{11}^{\delta} \alpha_2^2 - 2\sigma_{12}^{\delta,\alpha} (\alpha_1 + \gamma_1 \alpha_2) \alpha_2}{\sigma_{11}^{\delta} \sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} - 1 \right) \right.$$

$$\left. \left. \cdot \left[\sigma_{22}^{\alpha} (\gamma_1 \sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{12}^{\delta,\alpha} \sigma_{22}^{\alpha} \right] + \alpha_2^2 [\sigma_{12}^{\delta,\alpha} - (\sigma_{12}^{\alpha} + \gamma_1 \sigma_{22}^{\alpha})] \right] \right\} \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}$$

Finally, the fourth and final section of the log likelihood is

$$\log L_i^{\mathrm{IV}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}} \left(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i} \right) \, \mathrm{d}\epsilon_{1it} \, \mathrm{d}\epsilon_{2it}$$

$$\cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}$$

It can be seen that this expression is constant with respect to γ_1 so its derivative with respect to this parameter is zero. The derivative of the second and fourth part of the derivative being equal to zero, the full derivative of the log likelihood with respect to γ_1 can be written in the following manner.

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \gamma_1} &= \frac{1}{L_i^{\mathrm{I}}} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i} = \alpha_1 + \gamma_1 \alpha_2, \delta_{2i} = \alpha_2 + \gamma_2 \alpha_1) \Gamma h^{\mathrm{I}}(\alpha_{1i}, \alpha_{2i}) \right] \\
& \sum_{\mathbb{D}_i^{\mathrm{I}}} \left\{ \frac{1}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu} e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) [\sigma_{22}^{\nu}(\gamma_1 \sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{12}^{\nu}(\gamma_2 \sigma_{12}^{\epsilon} + \sigma_{22}^{\epsilon})] \right.\right. \\
& \quad + (\gamma_2 y_{1it} + \beta_{21} x_{2it} + \beta_{22} x_{it} + \alpha_2)(\sigma_{22}^{\nu} e_{1it}^{\mathrm{I}} - \sigma_{12}^{\nu} e_{2it}^{\mathrm{I}}) + e_{2it}^{\mathrm{I}}[e_{1it}^{\mathrm{I}}(\gamma_2 \sigma_{12}^{\epsilon} + \sigma_{22}^{\epsilon}) - e_{2it}^{\mathrm{I}}(\sigma_{12}^{\epsilon} + \gamma_1 \sigma_{22}^{\epsilon})] \\
& \quad \left. + \gamma_2 y_{2it}(\sigma_{11}^{\nu} e_{2it}^{\mathrm{I}} - \sigma_{12}^{\nu} e_{1it}^{\mathrm{I}}) \right] - \frac{\gamma_2}{\Gamma} \\
& \quad + \frac{1}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} \left[\left(\frac{\sigma_{22}^{\delta}(\alpha_1 + \gamma_1\alpha_2)^2 + \sigma_{11}^{\delta}(\alpha_2 + \gamma_2\alpha_1)^2 - 2\sigma_{12}^{\delta}(\alpha_1 + \gamma_1\alpha_2)(\alpha_2 + \gamma_2\alpha_1)}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} - 1 \right) \right. \\
& \quad \left[\sigma_{22}^{\delta}(\gamma_1 \sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{12}^{\delta}(\gamma_2 \sigma_{12}^{\alpha} + \sigma_{22}^{\alpha}) \right] + \alpha_2[\sigma_{12}^{\delta}(\alpha_2 + \gamma_2\alpha_1) - \sigma_{22}^{\delta}(\alpha_1 + \gamma_1\alpha_2)] \\
& \quad \left.\left. + (\alpha_2 + \gamma_2\alpha_1)[(\alpha_1 + \gamma_1\alpha_2)(\gamma_2 \sigma_{12}^{\alpha} + \sigma_{22}^{\alpha}) - (\alpha_2 + \gamma_2\alpha_1)(\sigma_{12}^{\alpha} + \gamma_1 \sigma_{22}^{\alpha})] \right] \right\} \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i} \\
& + \frac{1}{L_i^{\mathrm{III}}} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int\limits_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it} \cdot h^{\mathrm{III}}(\alpha_{1i}, \alpha_{2i}) \right] \\
& \sum_{\mathbb{D}_i^{\mathrm{III}}} \left\{ \frac{1}{\int\limits_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{1it}} \right. \qquad \text{(B.5)} \\
& \quad \left\{ - (\beta_{21} x_{2it} + \beta_{22} x_{it} + \alpha_2) f^{\mathrm{III}}(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \delta_{2i}) \right. \\
& \quad + \int\limits_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i} = \alpha_1 + \gamma_1\alpha_2, \alpha_2)}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} \\
& \quad \cdot \left[\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon}\nu_{1it} e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} - 1 \right) [\sigma_{22}^{\epsilon}(\gamma_1 \sigma_{22}^{\epsilon} + \sigma_{12}^{\epsilon}) - \sigma_{12}^{\nu,\epsilon}\sigma_{22}^{\epsilon}] \right. \\
& \quad \left.\left. + e_{2it}^{\mathrm{III}}[\sigma_{22}^{\epsilon}\nu_{1it} - e_{2it}^{\mathrm{III}}(\sigma_{12}^{\epsilon} + \gamma_1 \sigma_{22}^{\epsilon})] \right] \mathrm{d}\nu_{1it} \right\} \\
& \quad + \frac{1}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} \left[\left(\frac{\sigma_{22}^{\alpha}(\alpha_1 + \gamma_1\alpha_2)^2 + \sigma_{11}^{\delta}\alpha_2^2 - 2\sigma_{12}^{\delta,\alpha}(\alpha_1 + \gamma_1\alpha_2)\alpha_2}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} - 1 \right) \right. \\
& \quad \left.\left. \cdot \left[\sigma_{22}^{\alpha}(\gamma_1 \sigma_{22}^{\alpha} + \sigma_{12}^{\alpha}) - \sigma_{12}^{\delta,\alpha}\sigma_{22}^{\alpha} \right] + \alpha_2^2[\sigma_{12}^{\delta,\alpha} - (\sigma_{12}^{\alpha} + \gamma_1 \sigma_{22}^{\alpha})] \right] \right\} \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

The derivative with respect to γ_2, the second simultaneity parameter, is obtained with a process parallel to that of the first simultaneity parameter shown above. In this case, the derivative of those observations under Case III and IV are equal to zero since their contribution to the log likelihood does not depend on γ_2.

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \gamma_2} = {} & \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Bigg[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it}|\delta_{1i}=\alpha_1+\gamma_1\alpha_2, \delta_{2i}=\alpha_2+\gamma_2\alpha_1)\Gamma h^{\mathrm{I}}(\alpha_{1i},\alpha_{2i})\Bigg] \\
& \sum_{\mathbb{D}_i^{\mathrm{I}}}\Bigg\{ \frac{1}{\sigma^\nu_{11}\sigma^\nu_{22}-(\sigma^\nu_{12})^2}\Bigg[\Bigg(\frac{\sigma^\nu_{22}(e^{\mathrm{I}}_{1it})^2+\sigma^\nu_{11}(e^{\mathrm{I}}_{2it})^2-2\sigma^\nu_{12}e^{\mathrm{I}}_{1it}e^{\mathrm{I}}_{2it}}{\sigma^\nu_{11}\sigma^\nu_{22}-(\sigma^\nu_{12})^2}-1\Bigg)[\sigma^\nu_{11}(\gamma_2\sigma^\epsilon_{11}+\sigma^\epsilon_{12})-\sigma^\nu_{12}(\gamma_1\sigma^\epsilon_{12}+\sigma^\epsilon_{11})] \\
& +(\gamma_1 y_{2it}+\beta_{11}x_{1it}+\beta_{12}x_{it}+\alpha_1)(\sigma^\nu_{11}e^{\mathrm{I}}_{2it}-\sigma^\nu_{12}e^{\mathrm{I}}_{1it})+e^{\mathrm{I}}_{1it}[e^{\mathrm{I}}_{2it}(\gamma_1\sigma^\epsilon_{12}+\sigma^\epsilon_{11})-e^{\mathrm{I}}_{1it}(\sigma^\epsilon_{12}+\gamma_2\sigma^\epsilon_{11})] \\
& +\gamma_1 y_{1it}(\sigma^\nu_{22}e^{\mathrm{I}}_{1it}-\sigma^\nu_{12}e^{\mathrm{I}}_{2it})\Bigg] - \frac{\gamma_1}{\Gamma} \\
& + \frac{1}{\sigma^\delta_{11}\sigma^\delta_{22}-(\sigma^\delta_{12})^2}\Bigg[\Bigg(\frac{\sigma^\delta_{22}(\alpha_1+\gamma_1\alpha_2)^2+\sigma^\delta_{11}(\alpha_2+\gamma_2\alpha_1)^2-2\sigma^\delta_{12}(\alpha_1+\gamma_1\alpha_2)(\alpha_2+\gamma_2\alpha_1)}{\sigma^\delta_{11}\sigma^\delta_{22}-(\sigma^\delta_{12})^2}-1\Bigg) \\
& \Big[\sigma^\delta_{11}(\gamma_2\sigma^\alpha_{11}+\sigma^\alpha_{12})-\sigma^\delta_{12}(\gamma_1\sigma^\alpha_{12}+\sigma^\alpha_{11})\Big]+\alpha_1[\sigma^\delta_{12}(\alpha_1+\gamma_1\alpha_2)-\sigma^\delta_{11}(\alpha_2+\gamma_2\alpha_1)] \\
& +(\alpha_1+\gamma_1\alpha_2)[(\alpha_1+\gamma_1\alpha_2)(\gamma_1\sigma^\alpha_{12}+\sigma^\alpha_{11})-(\alpha_2+\gamma_2\alpha_1)(\sigma^\alpha_{12}+\gamma_2\sigma^\alpha_{11})]\Bigg]\Bigg\}\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i} \\
& +\frac{1}{L_i^{\mathrm{II}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Bigg[\prod_{\mathbb{D}_i^{\mathrm{II}}}\int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} f^{\mathrm{II}}(e^{\mathrm{II}}_{1it},\nu_{2it}|\alpha_{1i},\delta_{2i})\,\mathrm{d}\nu_{2it}\cdot h^{\mathrm{II}}(\alpha_{1i},\alpha_{2i})\Bigg] \\
& \sum_{\mathbb{D}_i^{\mathrm{II}}}\Bigg\{\frac{1}{\int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} f^{\mathrm{II}}(e^{\mathrm{II}}_{1it},\nu_{2it}|\alpha_{1i},\delta_{2i})\,\mathrm{d}\nu_{2it}} \\
& \Bigg\{-(\beta_{11}x_{1it}+\beta_{12}x_{it}+\alpha_1)f^{\mathrm{II}}(e^{\mathrm{II}}_{1it},-\Pi_2'X_{it}-\delta_{2i}|\alpha_{1i},\delta_{2i}) \\
& +\int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} \frac{f^{\mathrm{II}}(e^{\mathrm{II}}_{1it},\nu_{2it}|\alpha_{1i},\delta_2=\alpha_2+\gamma_2\alpha_1)}{\sigma^\epsilon_{11}\sigma^\nu_{22}-(\sigma^{\epsilon,\nu}_{12})^2} \\
& \cdot\Bigg[\Bigg(\frac{\sigma^\nu_{22}(e^{\mathrm{II}}_{1it})^2+\sigma^\epsilon_{11}\nu^2_{2it}-2\sigma^{\epsilon,\nu}_{12}e^{\mathrm{II}}_{1it}\nu_{2it}}{\sigma^\epsilon_{11}\sigma^\nu_{22}-(\sigma^{\epsilon,\nu}_{12})^2}-1\Bigg)[\sigma^\epsilon_{11}(\gamma_2\sigma^\epsilon_{11}+\sigma^\epsilon_{12})-\sigma^{\epsilon,\nu}_{12}\sigma^\epsilon_{11}] \\
& +e^{\mathrm{II}}_{1it}[\sigma^\epsilon_{11}\nu_{2it}-e^{\mathrm{II}}_{1it}(\sigma^\epsilon_{12}+\gamma_2\sigma^\epsilon_{11})]\Bigg]\, \mathrm{d}\nu_{2it}\Bigg\} \\
& +\frac{1}{\sigma^\alpha_{11}\sigma^\delta_{22}-(\sigma^{\alpha,\delta}_{12})^2}\Bigg[\Bigg(\frac{\sigma^\delta_{22}\alpha_1^2+\sigma^\alpha_{11}(\alpha_2+\gamma_2\alpha_1)^2-2\sigma^{\alpha,\delta}_{12}\alpha_1(\alpha_2+\gamma_2\alpha_1)}{\sigma^\alpha_{11}\sigma^\delta_{22}-(\sigma^{\alpha,\delta}_{12})^2}-1\Bigg) \\
& \cdot\Big[\sigma^\alpha_{11}(\gamma_2\sigma^\alpha_{11}+\sigma^\alpha_{12})-\sigma^{\alpha,\delta}_{12}\sigma^\alpha_{11}\Big]+\alpha_1^2[\sigma^{\alpha,\delta}_{12}-(\sigma^\alpha_{12}+\gamma_2\sigma^\alpha_{11})]\Bigg]\Bigg\}\, \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}
\end{aligned} \tag{B.6}
$$

We now deal with the parameters found in the variance-covariance matrix of the error term. We begin with the variance of ϵ_{1it} found in the first equation. The derivative of the log likelihood with respect to this parameter can be written through the following expression.

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{I}}}{\partial \sigma_{11}^{\epsilon}} &= \frac{1}{L_i^{\mathrm{I}}} \frac{\partial L_i^{\mathrm{I}}}{\partial \sigma_{11}^{\epsilon}} \\
&= \frac{1}{L_i^{\mathrm{I}}} \frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \sigma_{11}^{\epsilon}} \prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \\
&\qquad \cdot \left[\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{\frac{\partial}{\partial \sigma_{11}^{\epsilon}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i})}{f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i})} \right] \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \left[\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{\frac{\partial}{\partial \sigma_{11}^{\epsilon}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i})}{f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i})} \right] \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i}
\end{aligned}$$

since $g^{\mathrm{I}}(\cdot)$ does not depend on σ_{11}^{ϵ}. We can obtain the derivative in the numerator of the last term according to the following development.

$$
\begin{aligned}
&\frac{\partial}{\partial \sigma_{11}^{\epsilon}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i}, \delta_{2i}) = \\
&= \frac{\partial}{\partial \sigma_{11}^{\epsilon}} \left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2}} \exp\left\{ -\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu}e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}}}{2(\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2)} \right\} \right) \\
&= -\frac{\gamma_2^2\sigma_{11}^{\nu} + \sigma_{22}^{\nu} - 2\gamma_2\sigma_{12}^{\nu}}{2\sqrt{2\pi}(\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2)^{3/2}} \exp\{\cdot\} + \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2}} \exp\{\cdot\} \\
&\quad \cdot \left\{ \frac{\gamma_2^2\sigma_{11}^{\nu} + \sigma_{22}^{\nu} - 2\gamma_2\sigma_{12}^{\nu}}{2(\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2)^2} \left[\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu}e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}} \right] \right. \\
&\qquad \left. - \frac{1}{2(\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2)} \left[\gamma_2^2(e_{1it}^{\mathrm{I}})^2 + (e_{2it}^{\mathrm{I}})^2 - 2\gamma_2 e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}} \right] \right\} \\
&= \frac{f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i}, \delta_{2i})}{2(\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2)} \left\{ \left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu}e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) [\gamma_2^2\sigma_{11}^{\nu} + \sigma_{22}^{\nu} - 2\gamma_2\sigma_{12}^{\nu}] \right. \\
&\quad \left. - \left[\gamma_2^2(e_{1it}^{\mathrm{I}})^2 + (e_{2it}^{\mathrm{I}})^2 - 2\gamma_2 e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}} \right] \right\}
\end{aligned}
$$

So that the derivative for this part of the log likelihood function becomes

$$
\begin{aligned}
\frac{\partial \log L_i^{\mathrm{I}}}{\partial \sigma_{11}^{\epsilon}} &= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \left[\sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{\frac{\partial}{\partial \sigma_{11}^{\epsilon}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i}, \delta_{2i})}{f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i}, \delta_{2i})} \right] \mathrm{d}\delta_{1i}\, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \\
&\quad \cdot \left\{ \frac{1}{2(\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2)} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu}e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) [\gamma_2^2\sigma_{11}^{\nu} + \sigma_{22}^{\nu} - 2\gamma_2\sigma_{12}^{\nu}] \right. \right. \\
&\qquad \left. \left. - \left[\gamma_2^2(e_{1it}^{\mathrm{I}})^2 + (e_{2it}^{\mathrm{I}})^2 - 2\gamma_2 e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}} \right] \right] \right\} \mathrm{d}\delta_{1i}\, \mathrm{d}\delta_{2i}
\end{aligned}
$$

We can calculate the derivative for the second part of the log likelihood function with a parallel procedure.

$$\begin{aligned}
\frac{\partial \log L_i^{\mathrm{II}}}{\partial \sigma_{11}^{\epsilon}} &= \frac{1}{L_i^{\mathrm{II}}} \frac{\partial L_i^{\mathrm{II}}}{\partial \sigma_{11}^{\epsilon}} \\
&= \frac{1}{L_i^{\mathrm{II}}} \frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \sigma_{11}^{\epsilon}} \prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right] \\
&\qquad \cdot \left[\sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{\frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i})}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i})} \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&= \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right] \\
&\qquad \cdot \left[\sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{\frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it}}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it}} \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i}
\end{aligned}$$

with

$$\frac{\partial}{\partial\sigma_{11}^{\epsilon}}\int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}},\nu_{2it}|\alpha_{1i},\delta_{2i})\,\mathrm{d}\nu_{2it} = \int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} \frac{\partial}{\partial\sigma_{11}^{\epsilon}}\Big(f^{\mathrm{II}}(e_{1it}^{\mathrm{II}},\nu_{2it}|\alpha_{1i},\delta_{2i})\Big)\,\mathrm{d}\nu_{2it}$$

$$= \int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} \frac{\partial}{\partial\sigma_{11}^{\epsilon}}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2}}\exp\left\{-\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2+\sigma_{11}^{\epsilon}\nu_{2it}^2-2\sigma_{12}^{\epsilon,\nu}e_{1it}^{\mathrm{II}}\nu_{2it}}{2(\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2)}\right\}\right)\mathrm{d}\nu_{2it}$$

$$= \int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}}\Bigg(-\frac{\gamma_2^2\sigma_{11}^{\epsilon}+\sigma_{22}^{\nu}-2\gamma_2\sigma_{12}^{\epsilon,\nu}}{2\sqrt{2\pi}(\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2)^{3/2}}\exp\{\cdot\}+\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2}}\exp\{\cdot\}$$

$$\cdot\left\{\frac{\gamma_2^2\sigma_{11}^{\epsilon}+\sigma_{22}^{\nu}-2\gamma_2\sigma_{12}^{\epsilon,\nu}}{2(\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2)^2}\left[\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2+\sigma_{11}^{\epsilon}\nu_{2it}^2-2\sigma_{12}^{\epsilon,\nu}e_{1it}^{\mathrm{II}}\nu_{2it}\right]\right.$$

$$\left.-\frac{1}{2(\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2)}\left[\gamma_2^2(e_{1it}^{\mathrm{II}})^2+\nu_{2it}^2-2\gamma_2e_{1it}^{\mathrm{II}}\nu_{2it}\right]\right\}\Bigg)\mathrm{d}\nu_{2it}$$

$$= \int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}}\frac{f^{\mathrm{II}}(e_{1it}^{\mathrm{II}},\nu_{2it}|\alpha_{1i},\delta_{2i})}{2(\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2)}\left\{\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2+\sigma_{11}^{\epsilon}\nu_{2it}^2-2\sigma_{12}^{\epsilon,\nu}e_{1it}^{\mathrm{II}}\nu_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2}-1\right)\left[\gamma_2^2\sigma_{11}^{\epsilon}+\sigma_{22}^{\nu}-2\gamma_2\sigma_{12}^{\epsilon,\nu}\right]\right.$$

$$\left.-\left[\gamma_2^2(e_{1it}^{\mathrm{II}})^2+\nu_{2it}^2-2\gamma_2e_{1it}^{\mathrm{II}}\nu_{2it}\right]\right\}\mathrm{d}\nu_{2it}$$

We arrive to the expression

$$\frac{\partial\log L_i^{\mathrm{II}}}{\partial\sigma_{11}^{\epsilon}} = \frac{1}{L_i^{\mathrm{II}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\prod_{\mathbb{D}_i^{\mathrm{II}}}\int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}},\nu_{2it}|\alpha_{1i},\delta_{2i})\,\mathrm{d}\nu_{2it}\cdot g^{\mathrm{II}}(\alpha_{1i},\delta_{2i})\right]$$

$$\left\{\sum_{\mathbb{D}_i^{\mathrm{II}}}\frac{1}{\int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}},\nu_{2it}|\alpha_{1i},\delta_{2i})\,\mathrm{d}\nu_{2it}}\right.$$

$$\cdot\int_{-\infty}^{-\Pi_2'X_{it}-\delta_{2i}}\frac{f^{\mathrm{II}}(e_{1it}^{\mathrm{II}},\nu_{2it}|\alpha_{1i},\delta_{2i})}{2(\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2)}\left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2+\sigma_{11}^{\epsilon}\nu_{2it}^2-2\sigma_{12}^{\epsilon,\nu}e_{1it}^{\mathrm{II}}\nu_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}-(\sigma_{12}^{\epsilon,\nu})^2}-1\right)\left[\gamma_2^2\sigma_{11}^{\epsilon}+\sigma_{22}^{\nu}-2\gamma_2\sigma_{12}^{\epsilon,\nu}\right]\right.$$

$$\left.\left.-\left[\gamma_2^2(e_{1it}^{\mathrm{II}})^2+\nu_{2it}^2-2\gamma_2e_{1it}^{\mathrm{II}}\nu_{2it}\right]\right]\mathrm{d}\nu_{2it}\right\}\mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i}$$

For the observations that fall under Case III, the derivative can be written as

$$\frac{\partial \log L_i^{\mathrm{III}}}{\partial \sigma_{11}^{\epsilon}} = \frac{1}{L_i^{\mathrm{III}}} \frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \, \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i}$$

$$= \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \sigma_{11}^{\epsilon}} \prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right] \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i}$$

$$= \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right]$$

$$\cdot \left[\sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{\frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i})}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i})} \right] \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i}$$

$$= \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right]$$

$$\cdot \left[\sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{\frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it}}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it}} \right] \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i}$$

The derivative of the last term of the previous expression is calculated as it was done in the previous section of the function.

$$\frac{\partial}{\partial\sigma_{11}^{\epsilon}}\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})\,\mathrm{d}\nu_{1it} = \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} \frac{\partial}{\partial\sigma_{11}^{\epsilon}}\left(f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})\right)\mathrm{d}\nu_{1it}$$

$$= \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} \frac{\partial}{\partial\sigma_{11}^{\epsilon}}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2}}\exp\left\{-\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2+\sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2-2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}}{2(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2)}\right\}\right)\mathrm{d}\nu_{1it}$$

$$= \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} \left(-\frac{\sigma_{22}^{\epsilon}}{2\sqrt{2\pi}(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2)^{3/2}}\exp\{\cdot\}+\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2}}\exp\{\cdot\}\right.$$

$$\cdot\left\{\frac{\sigma_{22}^{\epsilon}}{2(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2)^2}\left[\sigma_{22}^{\epsilon}\nu_{1it}^2+\sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2-2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}\right]\right.$$

$$\left.\left.-\frac{1}{2(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2)}\left(e_{2it}^{\mathrm{III}}\right)^2\right\}\right)\mathrm{d}\nu_{1it}$$

$$= \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})}{2(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2)}\left\{\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2+\sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2-2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2}-1\right)\sigma_{22}^{\epsilon}-\left(e_{2it}^{\mathrm{III}}\right)^2\right\}\mathrm{d}\nu_{1it}$$

$$\frac{\partial\log L_i^{\mathrm{III}}}{\partial\sigma_{11}^{\epsilon}} = \frac{1}{L_i^{\mathrm{III}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\prod_{\mathbb{D}_i^{\mathrm{III}}}\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})\,\mathrm{d}\nu_{1it}\cdot g^{\mathrm{III}}(\delta_{1i},\alpha_{2i})\right]$$

$$\left\{\sum_{\mathbb{D}_i^{\mathrm{III}}}\frac{1}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})\,\mathrm{d}\nu_{1it}}\right.$$

$$\cdot\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})}{2(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2)}\left[\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2+\sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2-2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}-(\sigma_{12}^{\nu,\epsilon})^2}-1\right)\sigma_{22}^{\epsilon}\right.$$

$$\left.\left.-\left(e_{2it}^{\mathrm{III}}\right)^2\right]\mathrm{d}\nu_{1it}\right\}\mathrm{d}\delta_{1i}\,\mathrm{d}\alpha_{2i}$$

Finally, we now calculate the derivative for the final section of the log likelihood.

$$\frac{\partial \log L_i^{\mathrm{IV}}}{\partial \sigma_{11}^{\epsilon}} = \frac{1}{L_i^{\mathrm{IV}}} \frac{\partial}{\partial \sigma_{11}^{\epsilon}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}$$

$$= \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial}{\partial \sigma_{11}^{\epsilon}} \prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}$$

$$= \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right] \left[\sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \frac{\partial}{\partial \sigma_{11}^{\epsilon}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it}}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it}} \right] \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}$$

To obtain with a final expression, we develop the derivative inside the double integral of the final term.

$$\frac{\partial}{\partial \sigma_{11}^{\epsilon}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it} =$$

$$= \frac{\partial}{\partial \sigma_{11}^{\epsilon}} \left(\frac{1}{\sqrt{2\pi} \sqrt{\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2}} \exp\left\{ - \frac{\sigma_{22}^{\epsilon} \epsilon_{1it}^2 + \sigma_{11}^{\epsilon} \epsilon_{2it}^2 - 2\sigma_{12}^{\epsilon} \epsilon_{1it} \epsilon_{2it}}{2(\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2)} \right\} \right)$$

$$= - \frac{\sigma_{22}^{\epsilon}}{2\sqrt{2\pi}(\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2)^{3/2}} \exp\{\cdot\} + \frac{1}{\sqrt{2\pi} \sqrt{\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2}} \exp\{\cdot\} \cdot \left\{ \frac{\sigma_{22}^{\epsilon}}{2(\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2)^2} \left[\sigma_{22}^{\epsilon} \epsilon_{1it}^2 + \sigma_{11}^{\epsilon} \epsilon_{2it}^2 - 2\sigma_{12}^{\epsilon} \epsilon_{1it} \epsilon_{2it} \right] - \frac{\epsilon_{2it}^2}{2(\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2)} \right\}$$

$$= \frac{f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})}{2(\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2)} \left\{ \left(\frac{\sigma_{22}^{\epsilon} \epsilon_{1it}^2 + \sigma_{11}^{\epsilon} \epsilon_{2it}^2 - 2\sigma_{12}^{\epsilon} \epsilon_{1it} \epsilon_{2it}}{\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2} - 1 \right) \sigma_{22}^{\epsilon} - \epsilon_{2it}^2 \right\}$$

so that

$$\frac{\partial \log L_i^{\mathrm{IV}}}{\partial \sigma_{11}^{\epsilon}} = \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right]$$

$$\left\{ \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it}} \right.$$

$$\cdot \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \frac{f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})}{2(\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2)}$$

$$\left. \left[\left(\frac{\sigma_{22}^{\epsilon} \epsilon_{1it}^2 + \sigma_{11}^{\epsilon} \epsilon_{2it}^2 - 2\sigma_{12}^{\epsilon} \epsilon_{1it} \epsilon_{2it}}{\sigma_{11}^{\epsilon} \sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2} - 1 \right) \sigma_{22}^{\epsilon} - \epsilon_{2it}^2 \right] \right\} \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}$$

Collecting the four sections calculated before, we obtain the derivative of the log likelihood with respect to σ_{11}^{ϵ}.

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \sigma^{\epsilon}_{11}} &= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \\
&\quad \cdot \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{2(\sigma^{\nu}_{11}\sigma^{\nu}_{22} - (\sigma^{\nu}_{12})^2)} \Bigg[\left(\frac{\sigma^{\nu}_{22}(e^{\mathrm{I}}_{1it})^2 + \sigma^{\nu}_{11}(e^{\mathrm{I}}_{2it})^2 - 2\sigma^{\nu}_{12} e^{\mathrm{I}}_{1it} e^{\mathrm{I}}_{2it}}{\sigma^{\nu}_{11}\sigma^{\nu}_{22} - (\sigma^{\nu}_{12})^2} - 1 \right) \left[\gamma_2^2 \sigma^{\nu}_{11} + \sigma^{\nu}_{22} - 2\gamma_2 \sigma^{\nu}_{12} \right] \\
&\quad - \left[\gamma_2^2 (e^{\mathrm{I}}_{1it})^2 + (e^{\mathrm{I}}_{2it})^2 - 2\gamma_2 e^{\mathrm{I}}_{1it} e^{\mathrm{I}}_{2it} \right] \Bigg] \Bigg\} \, \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right] \\
&\quad \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it}} \\
&\quad \cdot \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} \frac{f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i})}{2(\sigma^{\epsilon}_{11}\sigma^{\nu}_{22} - (\sigma^{\epsilon,\nu}_{12})^2)} \Bigg[\left(\frac{\sigma^{\nu}_{22}(e^{\mathrm{II}}_{1it})^2 + \sigma^{\epsilon}_{11}\nu^2_{2it} - 2\sigma^{\epsilon,\nu}_{12} e^{\mathrm{II}}_{1it} \nu_{2it}}{\sigma^{\epsilon}_{11}\sigma^{\nu}_{22} - (\sigma^{\epsilon,\nu}_{12})^2} - 1 \right) \left[\gamma_2^2 \sigma^{\epsilon}_{11} + \sigma^{\nu}_{22} - 2\gamma_2 \sigma^{\epsilon,\nu}_{12} \right] \\
&\quad - \left[\gamma_2^2 (e^{\mathrm{II}}_{1it})^2 + \nu^2_{2it} - 2\gamma_2 e^{\mathrm{II}}_{1it} \nu_{2it} \right] \Bigg] \, \mathrm{d}\nu_{2it} \Bigg\} \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right] \qquad \text{(B.7)} \\
&\quad \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it}} \\
&\quad \cdot \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i})}{2(\sigma^{\nu}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\nu,\epsilon}_{12})^2)} \Bigg[\left(\frac{\sigma^{\epsilon}_{22}\nu^2_{1it} + \sigma^{\nu}_{11}(e^{\mathrm{III}}_{2it})^2 - 2\sigma^{\nu,\epsilon}_{12} \nu_{1it} e^{\mathrm{III}}_{2it}}{\sigma^{\nu}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\nu,\epsilon}_{12})^2} - 1 \right) \sigma^{\epsilon}_{22} \\
&\quad - \left(e^{\mathrm{III}}_{2it} \right)^2 \Bigg] \, \mathrm{d}\nu_{1it} \Bigg\} \, \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
&+ \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right] \\
&\quad \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i}) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it}} \\
&\quad \cdot \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \frac{f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})}{2(\sigma^{\epsilon}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\epsilon}_{12})^2)} \\
&\quad \left[\left(\frac{\sigma^{\epsilon}_{22}\epsilon^2_{1it} + \sigma^{\epsilon}_{11}\epsilon^2_{2it} - 2\sigma^{\epsilon}_{12}\epsilon_{1it}\epsilon_{2it}}{\sigma^{\epsilon}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\epsilon}_{12})^2} - 1 \right) \sigma^{\epsilon}_{22} - \epsilon^2_{2it} \right] \mathrm{d}\epsilon_{1i} \, \mathrm{d}\epsilon_{2i} \Bigg\} \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

The derivatives with respect to σ^{ϵ}_{22} and σ^{ϵ}_{12} are calculated through a similar procedure and only the final result is included in this report.

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \sigma^{\epsilon}_{22}} &= \frac{1}{L^{\mathrm{I}}_i} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \left[\prod_{\mathbb{D}^{\mathrm{I}}_i} f^{\mathrm{I}}(e^{\mathrm{I}}_{1it}, e^{\mathrm{I}}_{2it} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \\
&\cdot \left\{ \sum_{\mathbb{D}^{\mathrm{II}}_i} \frac{1}{2(\sigma^{\nu}_{11}\sigma^{\nu}_{22} - (\sigma^{\nu}_{12})^2)} \left[\left(\frac{\sigma^{\nu}_{22}(e^{\mathrm{I}}_{1it})^2 + \sigma^{\nu}_{11}(e^{\mathrm{I}}_{2it})^2 - 2\sigma^{\nu}_{12} e^{\mathrm{I}}_{1it} e^{\mathrm{I}}_{2it}}{\sigma^{\nu}_{11}\sigma^{\nu}_{22} - (\sigma^{\nu}_{12})^2} - 1 \right) [\sigma^{\nu}_{11} + \gamma_1^2 \sigma^{\nu}_{22} - 2\gamma_1 \sigma^{\nu}_{12}] \right.\right. \\
&\left.\left. - \left[(e^{\mathrm{I}}_{1it})^2 + \gamma_1^2 (e^{\mathrm{I}}_{2it})^2 - 2\gamma_1 e^{\mathrm{I}}_{1it} e^{\mathrm{I}}_{2it} \right] \right] \right\} \mathrm{d}\delta_{1i}\, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L^{\mathrm{II}}_i} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \left[\prod_{\mathbb{D}^{\mathrm{II}}_i} \int\limits_{-\infty}^{-\Pi'_2 X_{it} - \delta_{2i}} f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i})\, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right] \\
&\left\{ \sum_{\mathbb{D}^{\mathrm{II}}_i} \frac{1}{\int_{-\infty}^{-\Pi'_2 X_{it} - \delta_{2i}} f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i})\, \mathrm{d}\nu_{2it}} \right. \\
&\cdot \int\limits_{-\infty}^{-\Pi'_2 X_{it} - \delta_{2i}} \frac{f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it} | \alpha_{1i}, \delta_{2i})}{2(\sigma^{\epsilon}_{11}\sigma^{\nu}_{22} - (\sigma^{\epsilon,\nu}_{12})^2)} \left[\left(\frac{\sigma^{\nu}_{22}(e^{\mathrm{II}}_{1it})^2 + \sigma^{\epsilon}_{11}\nu^2_{2it} - 2\sigma^{\epsilon,\nu}_{12} e^{\mathrm{II}}_{1it}\nu_{2it}}{\sigma^{\epsilon}_{11}\sigma^{\nu}_{22} - (\sigma^{\epsilon,\nu}_{12})^2} - 1 \right) \sigma^{\epsilon}_{11} \right. \\
&\left.\left. - \left(e^{\mathrm{II}}_{1it} \right)^2 \right] \mathrm{d}\nu_{2it} \right\} \mathrm{d}\alpha_{1i}\, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L^{\mathrm{III}}_i} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \left[\prod_{\mathbb{D}^{\mathrm{III}}_i} \int\limits_{-\infty}^{-\Pi'_1 X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i})\, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right] \qquad \text{(B.8)} \\
&\left\{ \sum_{\mathbb{D}^{\mathrm{III}}_i} \frac{1}{\int_{-\infty}^{-\Pi'_1 X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i})\, \mathrm{d}\nu_{1it}} \right. \\
&\cdot \int\limits_{-\infty}^{-\Pi'_1 X_{it} - \delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it} | \delta_{1i}, \alpha_{2i})}{2(\sigma^{\nu}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\nu,\epsilon}_{12})^2)} \left[\left(\frac{\sigma^{\epsilon}_{22}\nu^2_{1it} + \sigma^{\nu}_{11}(e^{\mathrm{III}}_{2it})^2 - 2\sigma^{\nu,\epsilon}_{12}\nu_{1it} e^{\mathrm{III}}_{2it}}{\sigma^{\nu}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\nu,\epsilon}_{12})^2} - 1 \right) [\sigma^{\nu}_{11} + \gamma_1^2 \sigma^{\epsilon}_{22} - 2\gamma_1 \sigma^{\nu,\epsilon}_{12}] \right. \\
&\left.\left. - \left[\nu^2_{1it} + \gamma_1^2 (e^{\mathrm{III}}_{2it})^2 - 2\gamma_1 \nu_{1it} e^{\mathrm{III}}_{2it} \right] \right] \mathrm{d}\nu_{1it} \right\} \mathrm{d}\delta_{1i}\, \mathrm{d}\alpha_{2i} \\
&+ \frac{1}{L^{\mathrm{IV}}_i} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \left[\prod_{\mathbb{D}^{\mathrm{IV}}_i} \int\limits_{-\infty}^{-\beta'_{21} x_{2it} - \beta'_{22} x_{it} - \alpha_{2i}} \int\limits_{-\infty}^{-\beta'_{11} x_{1it} - \beta'_{12} x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot)\, \mathrm{d}\epsilon_{2it}\, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right] \\
&\left\{ \sum_{\mathbb{D}^{\mathrm{IV}}_i} \frac{1}{\int_{-\infty}^{-\beta'_{21} x_{2it} - \beta'_{22} x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta'_{11} x_{1it} - \beta'_{12} x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})\, \mathrm{d}\epsilon_{2it}\, \mathrm{d}\epsilon_{1it}} \right. \\
&\cdot \int\limits_{-\infty}^{-\beta'_{21} x_{2it} - \beta'_{22} x_{it} - \alpha_{2i}} \int\limits_{-\infty}^{-\beta'_{11} x_{1it} - \beta'_{12} x_{it} - \alpha_{1i}} \frac{f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \alpha_{1i}, \alpha_{2i})}{2(\sigma^{\epsilon}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\epsilon}_{12})^2)} \\
&\left.\left[\left(\frac{\sigma^{\epsilon}_{22}\epsilon^2_{1it} + \sigma^{\epsilon}_{11}\epsilon^2_{2it} - 2\sigma^{\epsilon}_{12}\epsilon_{1it}\epsilon_{2it}}{\sigma^{\epsilon}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\epsilon}_{12})^2} - 1 \right) \sigma^{\epsilon}_{11} - \epsilon^2_{1it} \right] \mathrm{d}\epsilon_{1i}\, \mathrm{d}\epsilon_{2i} \right\} \mathrm{d}\alpha_{1i}\, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \sigma_{12}^{\epsilon}} = & \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}}|\delta_{1i},\delta_{2i}) g^{\mathrm{I}}(\delta_{1i},\delta_{2i})\right] \\
& \cdot \left\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu} e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) [\gamma_2\sigma_{11}^{\nu} + \gamma_1\sigma_{22}^{\nu} - (1+\gamma_1\gamma_2)\sigma_{12}^{\nu}] \right.\right. \\
& \left.\left. - \left[\gamma_2 (e_{1it}^{\mathrm{I}})^2 + \gamma_1 (e_{2it}^{\mathrm{I}})^2 - (1+\gamma_1\gamma_2) e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}}\right]\right]\right\} \mathrm{d}\delta_{1i}\,\mathrm{d}\delta_{2i} \\
& + \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it}|\alpha_{1i},\delta_{2i})\,\mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i},\delta_{2i})\right] \\
& \left\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it}|\alpha_{1i},\delta_{2i})\,\mathrm{d}\nu_{2it}} \right. \\
& \cdot \int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} \frac{f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it}|\alpha_{1i},\delta_{2i})}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2 + \sigma_{11}^{\epsilon}\nu_{2it}^2 - 2\sigma_{12}^{\epsilon,\nu} e_{1it}^{\mathrm{II}}\nu_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2} - 1 \right) (\gamma_2\sigma_{11}^{\epsilon} - \sigma_{12}^{\epsilon,\nu}) \right. \\
& \left.\left. - e_{1it}^{\mathrm{II}} \left(\gamma_2 e_{1it}^{\mathrm{II}} - \nu_{2it}\right)\right] \mathrm{d}\nu_{2it} \right\} \mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i} \\
& + \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})\,\mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i},\alpha_{2i})\right] \qquad \text{(B.9)} \\
& \left\{ \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})\,\mathrm{d}\nu_{1it}} \right. \\
& \cdot \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}}|\delta_{1i},\alpha_{2i})}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} \left[\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon}\nu_{1it} e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} - 1 \right) [\gamma_1\sigma_{22}^{\epsilon} - \sigma_{12}^{\nu,\epsilon}] \right. \\
& \left.\left. - e_{2it}^{\mathrm{III}} \left(\gamma_1 e_{2it}^{\mathrm{III}} - \nu_{1it}\right)\right] \mathrm{d}\nu_{1it} \right\} \mathrm{d}\delta_{1i}\,\mathrm{d}\alpha_{2i} \\
& + \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it},\epsilon_{2it}|\cdot)\,\mathrm{d}\epsilon_{2it}\,\mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i},\alpha_{2i})\right] \\
& \left\{ \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{\int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it},\epsilon_{2it}|\alpha_{1i},\alpha_{2i})\,\mathrm{d}\epsilon_{2it}\,\mathrm{d}\epsilon_{1it}} \right. \\
& \cdot \int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} \frac{f^{\mathrm{IV}}(\epsilon_{1it},\epsilon_{2it}|\alpha_{1i},\alpha_{2i})}{\sigma_{11}^{\epsilon}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2} \\
& \left.\left[\left(\frac{\sigma_{22}^{\epsilon}\epsilon_{1it}^2 + \sigma_{11}^{\epsilon}\epsilon_{2it}^2 - 2\sigma_{12}^{\epsilon}\epsilon_{1it}\epsilon_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2} - 1 \right) \sigma_{12}^{\epsilon} + \epsilon_{1it}\epsilon_{2it} \right] \mathrm{d}\epsilon_{1i}\,\mathrm{d}\epsilon_{2i} \right\} \mathrm{d}\alpha_{1i}\,\mathrm{d}\alpha_{2i}
\end{aligned}
$$

To complete the gradient vector of the log likelihood, we need the derivatives with respect to the variance and covariance parameters of the individual random effects. Since their calculation mimics those included before for the error term, we will not write the full developments here, and we will report only on the final expressions.

$$\begin{aligned}
\frac{\partial \log L_i}{\partial \sigma_{11}^{\alpha}} &= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i})\right] \\
&\quad \cdot \left\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{2(\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2)} \left[\left(\frac{\sigma_{22}^{\delta}\delta_{1i}^2 + \sigma_{11}^{\delta}\delta_{2i}^2 - 2\sigma_{12}^{\delta}\delta_{1i}\delta_{2i}}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} - 1 \right) \left[\gamma_2^2\sigma_{11}^{\delta} + \sigma_{22}^{\delta} - 2\gamma_2\sigma_{12}^{\delta}\right] \right.\right. \\
&\quad \left.\left. - \left[\gamma_2^2\delta_{1i}^2 + \delta_{2i}^2 - 2\gamma_2\delta_{1i}\delta_{2i}\right] \right] \right\} \mathrm{d}\delta_{1i}\,\mathrm{d}\delta_{2i} \\
&\quad + \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i})\,\mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i})\right] \\
&\quad \left\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{2(\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2)} \left[\left(\frac{\sigma_{22}^{\delta}\alpha_1^2 + \sigma_{11}^{\alpha}\delta_{2i}^2 - 2\sigma_{12}^{\alpha,\delta}\alpha_{1i}\delta_{2i}}{\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2} - 1 \right) \left[\gamma_2^2\sigma_{11}^{\alpha} + \sigma_{22}^{\delta} - 2\gamma_2\sigma_{12}^{\alpha,\delta}\right] \right.\right. \\
&\quad \left.\left. - \left[\gamma_2^2\alpha_{1i}^2 + \delta_{2i}^2 - 2\gamma_2\alpha_{1i}\delta_{2i}\right] \right] \right\} \mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i} \qquad \text{(B.10)} \\
&\quad + \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i})\,\mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i})\right] \\
&\quad \left\{ \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{2(\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2)} \left[\left(\frac{\sigma_{22}^{\alpha}\delta_{1i}^2 + \sigma_{11}^{\delta}\alpha_{2i}^2 - 2\sigma_{12}^{\delta,\alpha}\delta_{1i}\alpha_{2i}}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} - 1 \right) \sigma_{22}^{\alpha} - \alpha_{2i}^2 \right] \right\} \mathrm{d}\delta_{1i}\,\mathrm{d}\alpha_{2i} \\
&\quad + \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot)\,\mathrm{d}\epsilon_{2it}\,\mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i})\right] \\
&\quad \left\{ \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{2(\sigma_{11}^{\alpha}\sigma_{22}^{\alpha} - (\sigma_{12}^{\alpha})^2)} \left[\left(\frac{\sigma_{22}^{\alpha}\alpha_{1i}^2 + \sigma_{11}^{\alpha}\epsilon_{2i}^2 - 2\sigma_{12}^{\alpha}\alpha_{1i}\alpha_{2i}}{\sigma_{11}^{\alpha}\sigma_{22}^{\alpha} - (\sigma_{12}^{\alpha})^2} - 1 \right) \sigma_{22}^{\alpha} - \alpha_{2i}^2 \right] \right\} \mathrm{d}\alpha_{1i}\,\mathrm{d}\alpha_{2i}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \sigma_{22}^{\alpha}} &= \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \\
&\cdot \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{2(\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2)} \Bigg[\left(\frac{\sigma_{22}^{\delta}\delta_{1i}^2 + \sigma_{11}^{\delta}\delta_{2i}^2 - 2\sigma_{12}^{\delta}\delta_{1i}\delta_{2i}}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} - 1 \right) \left[\sigma_{11}^{\delta} + \gamma_1^2 \sigma_{22}^{\delta} - 2\gamma_1 \sigma_{12}^{\delta} \right] \\
&\quad - \left[\delta_{1i}^2 + \gamma_1^2 \delta_{2i}^2 - 2\gamma_1 \delta_{1i}\delta_{2i} \right] \Bigg] \Bigg\} \, \mathrm{d}\delta_{1i} \, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i}) \, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right] \\
&\quad \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{2(\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2)} \Bigg[\left(\frac{\sigma_{22}^{\delta}\alpha_{1i}^2 + \sigma_{11}^{\alpha}\delta_{2i}^2 - 2\sigma_{12}^{\alpha,\delta}\alpha_{1i}\delta_{2i}}{\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2} - 1 \right) \sigma_{11}^{\alpha} - \alpha_{1i}^2 \Bigg] \Bigg\} \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\delta_{2i} \\
&+ \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i}) \, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right] \qquad \text{(B.11)} \\
&\quad \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{2(\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2)} \Bigg[\left(\frac{\sigma_{22}^{\alpha}\delta_{1i}^2 + \sigma_{11}^{\delta}\alpha_{2i}^2 - 2\sigma_{12}^{\delta,\alpha}\delta_{1i}\alpha_{2i}}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} - 1 \right) \left[\sigma_{11}^{\delta} + \gamma_1^2 \sigma_{22}^{\alpha} - 2\gamma_1 \sigma_{12}^{\delta,\alpha} \right] \\
&\quad - \left[\delta_{1i}^2 + \gamma_1^2 \alpha_{2i}^2 - 2\gamma_1 \delta_{1i}\alpha_{2i} \right] \Bigg] \Bigg\} \, \mathrm{d}\delta_{1i} \, \mathrm{d}\alpha_{2i} \\
&+ \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right] \\
&\quad \Bigg\{ \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{2(\sigma_{11}^{\alpha}\sigma_{22}^{\alpha} - (\sigma_{12}^{\alpha})^2)} \Bigg[\left(\frac{\sigma_{22}^{\alpha}\alpha_{1i}^2 + \sigma_{11}^{\alpha}\alpha_{2i}^2 - 2\sigma_{12}^{\alpha}\alpha_{1i}\alpha_{2i}}{\sigma_{11}^{\alpha}\sigma_{22}^{\alpha} - (\sigma_{12}^{\alpha})^2} - 1 \right) \sigma_{11}^{\alpha} - \alpha_{1i}^2 \Bigg] \Bigg\} \, \mathrm{d}\alpha_{1i} \, \mathrm{d}\alpha_{2i}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \sigma_{12}^{\alpha}} = & \frac{1}{L_i^{\mathrm{I}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{I}}} f^{\mathrm{I}}(e_{1it}^{\mathrm{I}}, e_{2it}^{\mathrm{I}} | \delta_{1i}, \delta_{2i}) g^{\mathrm{I}}(\delta_{1i}, \delta_{2i}) \right] \\
& \cdot \left\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} \left[\left(\frac{\sigma_{22}^{\delta}\delta_{1i}^2 + \sigma_{11}^{\delta}\delta_{2i}^2 - 2\sigma_{12}^{\delta}\delta_{1i}\delta_{2i}}{\sigma_{11}^{\delta}\sigma_{22}^{\delta} - (\sigma_{12}^{\delta})^2} - 1 \right) \left[\gamma_2\sigma_{11}^{\delta} + \gamma_1\sigma_{22}^{\delta} - (1+\gamma_1\gamma_2)\sigma_{12}^{\delta} \right] \right. \right. \\
& \left. \left. - \left[\gamma_2\delta_{1i}^2 + \gamma_1\delta_{2i}^2 - (1+\gamma_1\gamma_2)\delta_{1i}\delta_{2i} \right] \right] \right\} \mathrm{d}\delta_{1i}\,\mathrm{d}\delta_{2i} \\
& + \frac{1}{L_i^{\mathrm{II}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{II}}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it} | \alpha_{1i}, \delta_{2i})\, \mathrm{d}\nu_{2it} \cdot g^{\mathrm{II}}(\alpha_{1i}, \delta_{2i}) \right] \\
& \left\{ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2} \left[\left(\frac{\sigma_{22}^{\delta}\alpha_{1i}^2 + \sigma_{11}^{\alpha}\delta_{2i}^2 - 2\sigma_{12}^{\alpha,\delta}\alpha_{1i}\delta_{2i}}{\sigma_{11}^{\alpha}\sigma_{22}^{\delta} - (\sigma_{12}^{\alpha,\delta})^2} - 1 \right) (\gamma_2\sigma_{11}^{\alpha} - \sigma_{12}^{\alpha,\delta}) \right. \right. \\
& \left. \left. -\alpha_{1i}\left(\gamma_2\alpha_{1i} - \delta_{2i}\right) \right] \right\} \mathrm{d}\alpha_{1i}\,\mathrm{d}\delta_{2i} \qquad \text{(B.12)} \\
& + \frac{1}{L_i^{\mathrm{III}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{III}}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}} | \delta_{1i}, \alpha_{2i})\, \mathrm{d}\nu_{1it} \cdot g^{\mathrm{III}}(\delta_{1i}, \alpha_{2i}) \right] \\
& \left\{ \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} \left[\left(\frac{\sigma_{22}^{\alpha}\delta_{1i}^2 + \sigma_{11}^{\delta}\alpha_{2i}^2 - 2\sigma_{12}^{\delta,\alpha}\delta_{1i}\alpha_{2i}}{\sigma_{11}^{\delta}\sigma_{22}^{\alpha} - (\sigma_{12}^{\delta,\alpha})^2} - 1 \right) \left[\gamma_1\sigma_{22}^{\alpha} - \sigma_{12}^{\delta,\alpha} \right] \right. \right. \\
& \left. \left. -\alpha_{2i}\left(\gamma_1\alpha_{2i} - \delta_{1i}\right) \right] \right\} \mathrm{d}\delta_{1i}\,\mathrm{d}\alpha_{2i} \\
& + \frac{1}{L_i^{\mathrm{IV}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\prod_{\mathbb{D}_i^{\mathrm{IV}}} \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it} | \cdot)\, \mathrm{d}\epsilon_{2it}\, \mathrm{d}\epsilon_{1it} \cdot g^{\mathrm{IV}}(\alpha_{1i}, \alpha_{2i}) \right] \\
& \left\{ \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{\sigma_{11}^{\alpha}\sigma_{22}^{\alpha} - (\sigma_{12}^{\alpha})^2} \left[\left(\frac{\sigma_{22}^{\alpha}\alpha_{1i}^2 + \sigma_{11}^{\alpha}\alpha_{2i}^2 - 2\sigma_{12}^{\alpha}\alpha_{1i}\alpha_{2i}}{\sigma_{11}^{\alpha}\sigma_{22}^{\alpha} - (\sigma_{12}^{\alpha})^2} - 1 \right) \sigma_{12}^{\alpha} + \alpha_{1i}\alpha_{2i} \right] \right\} \mathrm{d}\alpha_{1i}\,\mathrm{d}\alpha_{2i}
\end{aligned}
$$

B.2 Fixed effects model

The derivatives of the logarithm of the likelihood function of the fixed effects model (3.33) are relatively more straightforward to calculate. As a consequence, from the gradient vector we drop the derivatives with respect to the parameters of the density of the individual effects, but instead we directly add those with respect to α_{1i} and α_{2i}. As before, we partition the function into four segments, each corresponding to one of the cases described in Chapter 3, we obtain the derivatives separately, and we aggregate the results. Since the calculations are similar and simpler that those of the random effects model presented in the previous section, we will not describe the procedure in detail and we will present only the final expressions.

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \beta_{11}} &= \sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{1it}}{1-r_I^2}\left[\frac{e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{\gamma_2 e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{r_I\left(\gamma_2 e_{1it}^{\mathrm{I}} + e_{2it}^{\mathrm{I}}\right)}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu}}}\right] \\
&+ \sum_{\mathbb{D}_i^{\mathrm{II}}} x_{1it} \frac{\frac{1}{1-r_{II}^2}\int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right)\left(\frac{e_{1it}^{\mathrm{II}}}{\sigma_{11}^{\epsilon}} - \frac{r_{II}\nu_{2it}}{\sqrt{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}}}\right) \mathrm{d}\nu_{2it} - \gamma_2 f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}\right)}{\int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right) \mathrm{d}\nu_{2it}} \\
&- \sum_{\mathbb{D}_i^{\mathrm{III}}} x_{1it} \frac{f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}\right)}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right) \mathrm{d}\nu_{1it}} \\
&- \sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{1it} \frac{\int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} f^{\mathrm{IV}}\left(-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \epsilon_{2it}\right) \mathrm{d}\epsilon_{2it}}{\int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}\right) \mathrm{d}\epsilon_{1it}\, \mathrm{d}\epsilon_{2it}}
\end{aligned}
\tag{B.13}
$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \beta_{12}} &= \sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{it}}{1-r_I^2}\left[\frac{e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{\gamma_2 e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{r_I\left(\gamma_2 e_{1it}^{\mathrm{I}} + e_{2it}^{\mathrm{I}}\right)}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu}}}\right] \\
&+ \sum_{\mathbb{D}_i^{\mathrm{II}}} x_{it} \frac{\frac{1}{1-r_{II}^2}\int\limits_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right)\left(\frac{e_{1it}^{\mathrm{II}}}{\sigma_{11}^{\epsilon}} - \frac{r_{II}\nu_{2it}}{\sqrt{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}}}\right) \mathrm{d}\nu_{2it} - \gamma_2 f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}\right)}{\int\limits_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right) \mathrm{d}\nu_{2it}} \\
&- \sum_{\mathbb{D}_i^{\mathrm{III}}} x_{it} \frac{f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}\right)}{\int\limits_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right) \mathrm{d}\nu_{1it}} \\
&- \sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{it} \frac{\int\limits_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} f^{\mathrm{IV}}\left(-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \epsilon_{2it}\right) \mathrm{d}\epsilon_{2it}}{\int\limits_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} \int\limits_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}\right) \mathrm{d}\epsilon_{1it}\,\mathrm{d}\epsilon_{2it}}
\end{aligned} \tag{B.14}
$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \beta_{21}} &= \sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{2it}}{1-r_I^2}\left[\frac{\gamma_1 e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{r_I\left(e_{1it}^{\mathrm{I}} + \gamma_1 e_{2it}^{\mathrm{I}}\right)}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu}}}\right] \\
&- \sum_{\mathbb{D}_i^{\mathrm{II}}} x_{2it} \frac{f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}\right)}{\int\limits_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right) \mathrm{d}\nu_{2it}} \\
&+ \sum_{\mathbb{D}_i^{\mathrm{III}}} x_{2it} \frac{\frac{1}{1-r_{III}^2}\int\limits_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right)\left(\frac{e_{2it}^{\mathrm{III}}}{\sigma_{22}^{\epsilon}} - \frac{r_{III}\nu_{1it}}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}}}\right) \mathrm{d}\nu_{1it} - \gamma_1 f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}\right)}{\int\limits_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right) \mathrm{d}\nu_{1it}} \\
&- \sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{2it} \frac{\int\limits_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}\right) \mathrm{d}\epsilon_{1it}}{\int\limits_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} \int\limits_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}\right) \mathrm{d}\epsilon_{2it}\,\mathrm{d}\epsilon_{1it}}
\end{aligned} \tag{B.15}
$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \beta_{22}} &= \sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{x_{it}}{1-r_I^2}\left[\frac{\gamma_1 e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{r_I\left(e_{1it}^{\mathrm{I}} + \gamma_1 e_{2it}^{\mathrm{I}}\right)}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu}}}\right] \\
&- \sum_{\mathbb{D}_i^{\mathrm{II}}} x_{it} \frac{f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}\right)}{\int_{-\infty}^{-\Pi_2' X_{it}-\delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right)\,\mathrm{d}\nu_{2it}} \\
&+ \sum_{\mathbb{D}_i^{\mathrm{III}}} x_{it} \frac{\frac{1}{1-r_{III}^2}\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right)\left(\frac{e_{2it}^{\mathrm{III}}}{\sigma_{22}^{\epsilon}} - \frac{r_{III}\nu_{1it}}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}}}\right)\mathrm{d}\nu_{1it} - \gamma_1 f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}\right)}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right)\,\mathrm{d}\nu_{1it}} \\
&- \sum_{\mathbb{D}_i^{\mathrm{IV}}} x_{it} \frac{\int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}\right)\,\mathrm{d}\epsilon_{1it}}{\int_{-\infty}^{-\beta_{11}' x_{1it}-\beta_{12}' x_{it}-\alpha_{1i}} \int_{-\infty}^{-\beta_{21}' x_{2it}-\beta_{22}' x_{it}-\alpha_{2i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}\right)\,\mathrm{d}\epsilon_{2it}\,\mathrm{d}\epsilon_{1it}}
\end{aligned}
\tag{B.16}
$$

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \gamma_1} = \\
&\sum_{\mathbb{D}_i^{\mathrm{I}}}\Bigg\{\frac{1}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2}\Bigg[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu}e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1\right)[\sigma_{22}^{\nu}(\sigma_{12}^{\epsilon} + \gamma_1\sigma_{22}^{\epsilon}) - \sigma_{12}^{\nu}(\gamma_2\sigma_{12}^{\epsilon} + \sigma_{22}^{\epsilon})] \\
&\quad + (\gamma_2 y_{1it} + \beta_{21}x_{2it} + \beta_{22}x_{it} + \alpha_2)(\sigma_{22}^{\nu}e_{1it}^{\mathrm{I}} - \sigma_{12}^{\nu}e_{2it}^{\mathrm{I}}) + e_{2it}^{\mathrm{I}}[e_{1it}^{\mathrm{I}}(\gamma_2\sigma_{12}^{\epsilon} + \sigma_{22}^{\epsilon}) - e_{2it}^{\mathrm{I}}(\sigma_{12}^{\epsilon} + \gamma_1\sigma_{22}^{\epsilon})] \\
&\quad + \gamma_2 y_{2it}(\sigma_{11}^{\nu}e_{2it}^{\mathrm{I}} - \sigma_{12}^{\nu}e_{1it}^{\mathrm{I}})\Bigg]\Bigg\} \\
&+ \sum_{\mathbb{D}_i^{\mathrm{III}}}\Bigg\{\frac{1}{\int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}})\,\mathrm{d}\nu_{1it}} \\
&\quad \Bigg\{ - (\beta_{21}x_{2it} + \beta_{22}x_{it} + \alpha_2) f^{\mathrm{III}}(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}) + \int_{-\infty}^{-\Pi_1' X_{it}-\delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}})}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} \\
&\quad \cdot\Bigg[\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon}\nu_{1it}e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} - 1\right)[\sigma_{22}^{\epsilon}(\sigma_{12}^{\epsilon} + \gamma_1\sigma_{22}^{\epsilon}) - \sigma_{12}^{\nu,\epsilon}\sigma_{22}^{\epsilon}] \\
&\quad + e_{2it}^{\mathrm{III}}[\sigma_{22}^{\epsilon}\nu_{1it} - e_{2it}^{\mathrm{III}}(\sigma_{12}^{\epsilon} + \gamma_1\sigma_{22}^{\epsilon})]\Bigg]\,\mathrm{d}\nu_{1it}\Bigg\}\Bigg\}
\end{aligned}
\tag{B.17}
$$

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \gamma_2} = \\
&\sum_{\mathbb{D}_i^{\mathrm{I}}} \left\{ \frac{1}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu}e_{1it}^{\mathrm{I}}e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) [\sigma_{11}^{\nu}(\sigma_{12}^{\epsilon} + \gamma_2\sigma_{11}^{\epsilon}) - \sigma_{12}^{\nu}(\gamma_1\sigma_{12}^{\epsilon} + \sigma_{11}^{\epsilon})] \right.\right. \\
&\quad + (\gamma_1 y_{2it} + \beta_{11}x_{1it} + \beta_{12}x_{it} + \alpha_1)(\sigma_{11}^{\nu}e_{2it}^{\mathrm{I}} - \sigma_{12}^{\nu}e_{1it}^{\mathrm{I}}) + e_{1it}^{\mathrm{I}}[e_{2it}^{\mathrm{I}}(\gamma_1\sigma_{12}^{\epsilon} + \sigma_{11}^{\epsilon}) - e_{1it}^{\mathrm{I}}(\sigma_{12}^{\epsilon} + \gamma_2\sigma_{11}^{\epsilon})] \\
&\quad \left.\left. + \gamma_1 y_{1it}(\sigma_{22}^{\nu}e_{1it}^{\mathrm{I}} - \sigma_{12}^{\nu}e_{2it}^{\mathrm{I}}) \right] \right\} \\
&+ \sum_{\mathbb{D}_i^{\mathrm{II}}} \left\{ \frac{1}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it})\,\mathrm{d}\nu_{2it}} \right. \qquad \text{(B.18)} \\
&\quad \left\{ -(\beta_{11}x_{1it} + \beta_{12}x_{it} + \alpha_1) f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}) + \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} \frac{f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it})}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2} \right. \\
&\quad \cdot \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2 + \sigma_{11}^{\epsilon}\nu_{2it}^2 - 2\sigma_{12}^{\epsilon,\nu}e_{1it}^{\mathrm{II}}\nu_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2} - 1 \right) [\sigma_{11}^{\epsilon}(\sigma_{12}^{\epsilon} + \gamma_2\sigma_{11}^{\epsilon}) - \sigma_{12}^{\epsilon,\nu}\sigma_{11}^{\epsilon}] \right. \\
&\quad \left.\left.\left. + e_{1it}^{\mathrm{II}}[\sigma_{11}^{\epsilon}\nu_{2it} - e_{1it}^{\mathrm{II}}(\sigma_{12}^{\epsilon} + \gamma_2\sigma_{11}^{\epsilon})] \right] \mathrm{d}\nu_{2it} \right\} \right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \alpha_1} &= \sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{1}{1 - r_I^2} \left[\frac{e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{\gamma_2 e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{r_I\left(\gamma_2 e_{1it}^{\mathrm{I}} + e_{2it}^{\mathrm{I}}\right)}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu}}} \right] \\
&+ \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{\frac{1}{1 - r_{II}^2} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right) \left(\frac{e_{1it}^{\mathrm{II}}}{\sigma_{11}^{\epsilon}} - \frac{r_{II}\nu_{2it}}{\sqrt{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu}}} \right) \mathrm{d}\nu_{2it} - \gamma_2 f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}\right)}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right) \mathrm{d}\nu_{2it}} \\
&- \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}\right)}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right) \mathrm{d}\nu_{1it}} \qquad \text{(B.19)} \\
&- \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} f^{\mathrm{IV}}\left(-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}, \epsilon_{2it}\right) \mathrm{d}\epsilon_{2it}}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}\right) \mathrm{d}\epsilon_{1it}\,\mathrm{d}\epsilon_{2it}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \alpha_2} = & \sum_{\mathbb{D}_i^{\mathrm{I}}} \frac{1}{1-r_I^2}\left[\frac{\gamma_1 e_{1it}^{\mathrm{I}}}{\sigma_{11}^{\nu}} + \frac{e_{2it}^{\mathrm{I}}}{\sigma_{22}^{\nu}} - \frac{r_I\left(e_{1it}^{\mathrm{I}} + \gamma_1 e_{2it}^{\mathrm{I}}\right)}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\nu}}}\right] \\
& - \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, -\Pi_2' X_{it} - \delta_{2i}\right)}{\int\limits_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}\left(e_{1it}^{\mathrm{II}}, \nu_{2it}\right) \, \mathrm{d}\nu_{2it}} \\
& + \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{\frac{1}{1-r_{III}^2} \int\limits_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right)\left(\frac{e_{2it}^{\mathrm{III}}}{\sigma_{22}^{\epsilon}} - \frac{r_{III}\nu_{1it}}{\sqrt{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon}}}\right) \mathrm{d}\nu_{1it} - \gamma_1 f^{\mathrm{III}}\left(-\Pi_1' X_{it} - \delta_{1i}, e_{2it}^{\mathrm{III}}\right)}{\int\limits_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}\left(\nu_{1it}, e_{2it}^{\mathrm{III}}\right) \, \mathrm{d}\nu_{1it}} \\
& - \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{\int\limits_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, -\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}\right) \, \mathrm{d}\epsilon_{1it}}{\int\limits_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \int\limits_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} f^{\mathrm{IV}}\left(\epsilon_{1it}, \epsilon_{2it}\right) \, \mathrm{d}\epsilon_{2it} \, \mathrm{d}\epsilon_{1it}}
\end{aligned}
\tag{B.20}
$$

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \sigma^{\epsilon}_{11}} = \\
&\quad \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{2(\sigma^{\nu}_{11}\sigma^{\nu}_{22} - (\sigma^{\nu}_{12})^2)} \left[\left(\frac{\sigma^{\nu}_{22}(e^{\mathrm{I}}_{1it})^2 + \sigma^{\nu}_{11}(e^{\mathrm{I}}_{2it})^2 - 2\sigma^{\nu}_{12} e^{\mathrm{I}}_{1it} e^{\mathrm{I}}_{2it}}{\sigma^{\nu}_{11}\sigma^{\nu}_{22} - (\sigma^{\nu}_{12})^2} - 1 \right) \left[\gamma_2^2 \sigma^{\nu}_{11} + \sigma^{\nu}_{22} - 2\gamma_2 \sigma^{\nu}_{12}\right] \right. \\
&\qquad \left. - \left[\gamma_2^2 (e^{\mathrm{I}}_{1it})^2 + (e^{\mathrm{I}}_{2it})^2 - 2\gamma_2 e^{\mathrm{I}}_{1it} e^{\mathrm{I}}_{2it}\right] \right] \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it})\, \mathrm{d}\nu_{2it}} \int\limits_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} \frac{f^{\mathrm{II}}(e^{\mathrm{II}}_{1it}, \nu_{2it})}{2(\sigma^{\epsilon}_{11}\sigma^{\nu}_{22} - (\sigma^{\epsilon,\nu}_{12})^2)} \\
&\qquad \cdot \left[\left(\frac{\sigma^{\nu}_{22}(e^{\mathrm{II}}_{1it})^2 + \sigma^{\epsilon}_{11}\nu^2_{2it} - 2\sigma^{\epsilon,\nu}_{12} e^{\mathrm{II}}_{1it}\nu_{2it}}{\sigma^{\epsilon}_{11}\sigma^{\nu}_{22} - (\sigma^{\epsilon,\nu}_{12})^2} - 1 \right) \left[\gamma_2^2 \sigma^{\epsilon}_{11} + \sigma^{\nu}_{22} - 2\gamma_2 \sigma^{\epsilon,\nu}_{12}\right] \right. \\
&\qquad \left. - \left[\gamma_2^2 (e^{\mathrm{II}}_{1it})^2 + \nu^2_{2it} - 2\gamma_2 e^{\mathrm{II}}_{1it}\nu_{2it}\right] \right] \mathrm{d}\nu_{2it} \qquad \text{(B.21)} \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it})\, \mathrm{d}\nu_{1it}} \int\limits_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e^{\mathrm{III}}_{2it})}{2(\sigma^{\nu}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\nu,\epsilon}_{12})^2)} \\
&\qquad \cdot \left[\left(\frac{\sigma^{\epsilon}_{22}\nu^2_{1it} + \sigma^{\nu}_{11}(e^{\mathrm{III}}_{2it})^2 - 2\sigma^{\nu,\epsilon}_{12}\nu_{1it} e^{\mathrm{III}}_{2it}}{\sigma^{\nu}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\nu,\epsilon}_{12})^2} - 1 \right) \sigma^{\epsilon}_{22} - \left(e^{\mathrm{III}}_{2it}\right)^2 \right] \mathrm{d}\nu_{1it} \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})\, \mathrm{d}\epsilon_{2it}\, \mathrm{d}\epsilon_{1it}} \\
&\qquad \cdot \int\limits_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int\limits_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \frac{f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})}{2(\sigma^{\epsilon}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\epsilon}_{12})^2)} \\
&\qquad \left[\left(\frac{\sigma^{\epsilon}_{22}\epsilon^2_{1it} + \sigma^{\epsilon}_{11}\epsilon^2_{2it} - 2\sigma^{\epsilon}_{12}\epsilon_{1it}\epsilon_{2it}}{\sigma^{\epsilon}_{11}\sigma^{\epsilon}_{22} - (\sigma^{\epsilon}_{12})^2} - 1 \right) \sigma^{\epsilon}_{22} - \epsilon^2_{2it} \right] \mathrm{d}\epsilon_{1i}\, \mathrm{d}\epsilon_{2i}
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \sigma_{22}^{\epsilon}} = \\
&\quad \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{2(\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2)} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu} e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) \left[\sigma_{11}^{\nu} + \gamma_1^2 \sigma_{22}^{\nu} - 2\gamma_1 \sigma_{12}^{\nu}\right] \right. \\
&\qquad \left. - \left[(e_{1it}^{\mathrm{I}})^2 + \gamma_1^2 (e_{2it}^{\mathrm{I}})^2 - 2\gamma_1 e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}} \right] \right] \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it})\,\mathrm{d}\nu_{2it}} \int\limits_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} \frac{f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it})}{2(\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2)} \\
&\qquad \cdot \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2 + \sigma_{11}^{\epsilon}\nu_{2it}^2 - 2\sigma_{12}^{\epsilon,\nu} e_{1it}^{\mathrm{II}} \nu_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2} - 1 \right) \sigma_{11}^{\epsilon} - \left(e_{1it}^{\mathrm{II}}\right)^2 \right] \mathrm{d}\nu_{2it} \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}})\,\mathrm{d}\nu_{1it}} \int\limits_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}})}{2(\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2)} \qquad \text{(B.22)} \\
&\qquad \cdot \left[\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon} \nu_{1it} e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} - 1 \right) \left[\sigma_{11}^{\nu} + \gamma_1^2 \sigma_{22}^{\epsilon} - 2\gamma_1 \sigma_{12}^{\nu,\epsilon}\right] \right. \\
&\qquad \left. - \left[\nu_{1it}^2 + \gamma_1^2 (e_{2it}^{\mathrm{III}})^2 - 2\gamma_1 \nu_{1it} e_{2it}^{\mathrm{III}} \right] \right] \mathrm{d}\nu_{1it} \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})\,\mathrm{d}\epsilon_{2it}\,\mathrm{d}\epsilon_{1it}} \\
&\qquad \cdot \int\limits_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int\limits_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \frac{f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})}{2(\sigma_{11}^{\epsilon}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2)} \\
&\qquad \left[\left(\frac{\sigma_{22}^{\epsilon}\epsilon_{1it}^2 + \sigma_{11}^{\epsilon}\epsilon_{2it}^2 - 2\sigma_{12}^{\epsilon}\epsilon_{1it}\epsilon_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2} - 1 \right) \sigma_{11}^{\epsilon} - \epsilon_{1it}^2 \right] \mathrm{d}\epsilon_{1i}\,\mathrm{d}\epsilon_{2i}
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \sigma_{12}^{\epsilon}} = \\
&\quad \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{I}})^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{I}})^2 - 2\sigma_{12}^{\nu} e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}}}{\sigma_{11}^{\nu}\sigma_{22}^{\nu} - (\sigma_{12}^{\nu})^2} - 1 \right) [\gamma_2\sigma_{11}^{\nu} + \gamma_1\sigma_{22}^{\nu} - (1+\gamma_1\gamma_2)\sigma_{12}^{\nu}] \right. \\
&\qquad \left. - \left[\gamma_2 (e_{1it}^{\mathrm{I}})^2 + \gamma_1 (e_{2it}^{\mathrm{I}})^2 - (1+\gamma_1\gamma_2) e_{1it}^{\mathrm{I}} e_{2it}^{\mathrm{I}} \right] \right] \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{II}}} \frac{1}{\int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it})\, \mathrm{d}\nu_{2it}} \int_{-\infty}^{-\Pi_2' X_{it} - \delta_{2i}} \frac{f^{\mathrm{II}}(e_{1it}^{\mathrm{II}}, \nu_{2it})}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2} \\
&\qquad \cdot \left[\left(\frac{\sigma_{22}^{\nu}(e_{1it}^{\mathrm{II}})^2 + \sigma_{11}^{\epsilon}\nu_{2it}^2 - 2\sigma_{12}^{\epsilon,\nu} e_{1it}^{\mathrm{II}} \nu_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\nu} - (\sigma_{12}^{\epsilon,\nu})^2} - 1 \right) (\gamma_2\sigma_{11}^{\epsilon} - \sigma_{12}^{\epsilon,\nu}) - e_{1it}^{\mathrm{II}} \left(\gamma_2 e_{1it}^{\mathrm{II}} - \nu_{2it} \right) \right] \mathrm{d}\nu_{2it} \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{III}}} \frac{1}{\int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}})\, \mathrm{d}\nu_{1it}} \int_{-\infty}^{-\Pi_1' X_{it} - \delta_{1i}} \frac{f^{\mathrm{III}}(\nu_{1it}, e_{2it}^{\mathrm{III}})}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} \\
&\qquad \cdot \left[\left(\frac{\sigma_{22}^{\epsilon}\nu_{1it}^2 + \sigma_{11}^{\nu}(e_{2it}^{\mathrm{III}})^2 - 2\sigma_{12}^{\nu,\epsilon} \nu_{1it} e_{2it}^{\mathrm{III}}}{\sigma_{11}^{\nu}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\nu,\epsilon})^2} - 1 \right) (\gamma_1\sigma_{22}^{\epsilon} - \sigma_{12}^{\nu,\epsilon}) - e_{2it}^{\mathrm{III}} \left(\gamma_1 e_{2it}^{\mathrm{III}} - \nu_{1it} \right) \right] \mathrm{d}\nu_{1it} \\
&\quad + \sum_{\mathbb{D}_i^{\mathrm{IV}}} \frac{1}{\int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})\, \mathrm{d}\epsilon_{2it}\, \mathrm{d}\epsilon_{1it}} \\
&\qquad \cdot \int_{-\infty}^{-\beta_{21}' x_{2it} - \beta_{22}' x_{it} - \alpha_{2i}} \int_{-\infty}^{-\beta_{11}' x_{1it} - \beta_{12}' x_{it} - \alpha_{1i}} \frac{f^{\mathrm{IV}}(\epsilon_{1it}, \epsilon_{2it})}{\sigma_{11}^{\epsilon}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2} \\
&\qquad \left[\left(\frac{\sigma_{22}^{\epsilon}\epsilon_{1it}^2 + \sigma_{11}^{\epsilon}\epsilon_{2it}^2 - 2\sigma_{12}^{\epsilon}\epsilon_{1it}\epsilon_{2it}}{\sigma_{11}^{\epsilon}\sigma_{22}^{\epsilon} - (\sigma_{12}^{\epsilon})^2} - 1 \right) \sigma_{12}^{\epsilon} + \epsilon_{1it}\epsilon_{2it} \right] \mathrm{d}\epsilon_{1i}\, \mathrm{d}\epsilon_{2i}
\end{aligned} \tag{B.23}
$$

Appendix C

Computational considerations

The application of the MCECM algorithm (Algorithm 5 of Chapter 4) that was selected to estimate our model requires a few "enhancements" in order to ameliorate its functionality. First, an ascent-based version of the Monte Carlo experiment was implemented with the objective of improving and accelerating its convergence. Then, an optimal method to obtain random draws from a truncated normal distribution and the introduction of parallel computing were applied, resulting in a better efficiency in terms of computational cost. These solutions will be described in the following sections.

C.1 Ascent-based version of the MCECM algorithm

The standard EM algorithm has the ascent property in the sense that each iteration of the algorithm produces a likelihood equal or higher than the previous; in other words, each iteration produces a better estimator than (or, in the worst case, an estimator as good as) that obtained in the previous step. This property is derived from Jensen's inequality. In fact, the EM algorithm produces a sequence of estimator that, under regularity conditions, converges to the true value of the parameter. This is clearly an attractive property that we would like to preserve in other versions of the algorithm. However, when simulation is introduced, it comes along with a simulation error that may cause the ascent property to be lost. In order to recover the ascent property, a version of the Monte Carlo EM was proposed by Caffo et al. (2005). They argue that if the Monte Carlo sample size is left unchanged across iterations of the algorithm, then it will not converge due to a persistent simulation error. They devised an automatic strategy to increase the size of the simulation sample, so that it is used as efficiently as needed. Algorithm 6 introduces this idea into the pseudocode of the standard MCECM presented in Chapter 4.

At each step of this algorithm, a lower bound B is calculated; this bound takes into account the likelihood implied by a new estimator together with the variability that arises from the Monte Carlo error. If the gain in likelihood is large enough to be significant even in the presence of the variability introduced by the simulation, then the proposed estimator is accepted and the algorithm moves on. On the other hand, if this bound is negative, the

Algorithm 6 Ascent-based MCECM algorithm

1. Initialize the algorithm with a consistent initial estimator π_0. Choose an initial sample size for the simulation R_0. Set $i = 0$.

2. Monte Carlo E-step: given the current parameter estimate π_i and the current sample size R_i compute the statistic

$$\widehat{J}(\pi|\pi_i, Y) = \frac{1}{R_i}\sum_{r=1}^{R_i} \log f(\widetilde{Y}^{*(r)}|\pi)$$

   where $\widetilde{Y}^{*(r)}$ is the r-th simulation of the complete data.

3. Conditional M-step: compute the proposed estimator π_i' as

$$\pi_i' = \arg\max_{\pi} \widehat{J}(\pi|\pi_i, Y)$$

   and define the gain in likelihood by

$$\Delta\widehat{J}(\pi_i', \pi_i) = \widehat{J}(\pi_i'|\pi_i, Y) - \widehat{J}(\pi_i|\pi_i, Y)$$

4. Calculate the statistic

$$B = \Delta\widehat{J}(\pi_i', \pi_i) - \Phi(1-a)\sqrt{\frac{\hat{\sigma}^2}{R_i}}$$

   where $\Phi(\cdot)$ is the cumulative standard normal distribution and $\hat{\sigma}^2$ represents the sample variance over the R_i simulations of

$$\log f(\widetilde{Y}^{*(r)}|\pi_i') - \log f(\widetilde{Y}^{*(r)}|\pi_i)$$

5. If $B \geq 0$, then π_i' is accepted so that we take it as the new estimate: $\pi_{i+1} = \pi_i'$. Calculate the new sample size for the next iteration

$$R_{i+1} = \max\left[R_i, \hat{\sigma}^2\left(\frac{\Phi(1-a)+\Phi(1-b)}{\Delta\widehat{J}(\pi_{i+1}, \pi_i)}\right)^2\right]$$

   If $R_{i+1} > R_{\max}$ finish the algorithm. Otherwise, $i = i + 1$ and go to 2.

6. If $B < 0$, then the gain in likelihood from the proposed estimate is swamped with the Monte Carlo error. π_i' is rejected and the sample is expanded by appending R_i/s new observations to the existing sample,

$$R_i = R_i + \frac{R_i}{s}$$

   If $R_{i+1} > R_{\max}$ finish the algorithm. Otherwise, go to 2.

gain in likelihood is surpassed by the Monte Carlo error so the proposed estimator is rejected and the sample size of the simulation is increased in order to reduce the relative variability. Caffo et al. (2005) approximate the gain in likelihood through an asymptotic normal density as the simulation size increases. Based on this result, they define the bound B and propose a formula to calculate the sample size at every new iteration of the algorithm.

As it can be seen in Algorithm 6, the ascent-based MCECM algorithm has a number of parameters that defines its performance. Their importance varies according to the objective function and the data-generating process. Table C.1 summarizes these five parameters and the default values assigned to them for both the random-effects and the fixed-effects model. These values were determined according to the literature (see Caffo et al., 2005) and preliminary trials of the MCECM algorithm on the simulation studies described in Chapter 5.

Table C.1: Parameters of the ascent-based MCECM algorithm

Parameter	Definition	Value
R_0	Sample size of the initial simulation	50
$R_{\max}$	Maximum sample size	1000
s	Increase step of the sample size when the proposed estimator is rejected	3
a	Significance level for the gain in likelihood	0.1
b	Type II error to determine the sample size of the next iteration	0.1

The first parameter R_0 defines the initial sample size; in other words, the algorithm will generate R_0 random draws of the complete data for the first iteration and will use them to calculate the Monte Carlo E-step. A value too small will generate a highly variable estimate that will take longer to converge; on the other hand, since we are only at the beginning of the algorithm, a value too large will be inefficient since it will generate a precise estimate around an initial value of the coefficients which is likely to be still far from the optimal value. The idea is to start the algorithm with a sample small enough to let the algorithm move quickly in the initial iterations and gradually increase it as the algorithm converges to the maximum.

The second parameter $R_{\max}$ sets the maximum sample size of the simulation. It is this value that defines the stopping rule of the ascent-based version of the algorithm. Although a larger values of this parameter will produce more precise estimates, there is a constraint on the computational cost. For Chapter 5 it was fixed at 1000 for both simulation studies so that the exercises remained feasible. However, when a single application is estimated, it can be increased to the desired precision.

The third parameters s determines the step of the sample size increase when the proposed estimator is rejected. If this is the case, it is determined that the gain in likelihood is swamped with the simulation error, so the sample must be increased in order to augment the precision and confirm a gain in likelihood. The higher the value of s, the smaller the rate at which the sample size is increased. But choosing a very low value for this parameter is inefficient since the steps will be very large when such increases in sample size may not be yet needed. The default value was set to 3 which means that if an estimator is rejected, the sample size will be increased by one third.

The fourth parameter a is used in the calculation of the lower bound B in step 4 of Algorithm 6. In fact, what is done at this step is to obtain an interval for the gain in likelihood around its observed value (assuming a normal asymptotic distribution). If this interval contains negative values, we conclude that the simulation error is larger than the gain in likelihood and the proposed estimator is rejected. The parameter a defines the significance of this interval estimator. For smaller a, the interval becomes wider and a larger gain in likelihood is needed to take it as significantly larger than zero.

Finally, the fifth parameter b appears at step 5 of the algorithm and it determines the sample size for the next iteration once an estimator is accepted. It is in fact a type II error rate for the confidence interval used to calculate the bound. In the sake of computational efficiency, this parameter together with a should ensure that the sample size for each iteration is chosen so that the appending process (step 6 of the algorithm) only occurs infrequently[18].

C.2 Simulating from a truncated normal distribution

In step 2 of Algorithm 6, the complete data is obtained by replacing the censored data by the average of R_i random draws from its generating distribution. This is a truncated normal distribution restricted to values smaller than zero; i.e. $\mathrm{T}\mathbb{N}_{(-\infty,0)}$. This arises because for these values we only observe a zero but we know that the latent data is negative. Since R_i random draws are needed for each iteration of the algorithm, it is essential to obtain draws from the truncated normal distribution as efficiently as possible. There has been various solution to this problem (see for example Geweke,

[18]Notice also that at each iteration the sample size is at least as large as that of the previous iteration.

1991, and Rodriguez-Yam et al., 2004). Here we will employ the mixed rejection algorithm proposed by Geweke (1991) which, as we will see, translates in our case to a normal rejection sampler.

Geweke introduces a mixed rejection algorithm for a truncated normal distribution $\mathrm{T}\mathbb{N}_{(u_1,u_2)}(0,1)$ in which the type of sampling to employ depends on the bounds of truncation (u_1, u_2). The sampling procedure is the following.

1. If the bounds are (u_1, ∞) and $u_1 \geq 0$, use normal rejection sampling if $u_1 \leq 0.45$ and exponential rejection sampling otherwise.

2. If the bounds (u_1, u_2) are such that zero is included in the interval, use normal rejection sampling if $\phi(u_1) \leq 0.15$ or $\phi(u_2) \leq 0.15$ and uniform rejection sampling otherwise[19].

3. If the bounds (u_1, u_2) are such that they only include positive numbers, use uniform rejection sampling if ${}^{\phi(u_1)}/_{\phi(u_2)} \leq 2.18$; use half-normal rejection sampling if ${}^{\phi(u_1)}/_{\phi(u_2)} > 2.18$ and $u_1 < 0.725$; use exponential rejection sampling otherwise.

4. If the bounds are $(-\infty, u_2)$ and $u_2 \leq 0$, this case is symmetric to 1 and it is treated in the same manner.

5. If the bounds (u_1, u_2) are such that they only include negative numbers, this case is symmetric to 3 and it is treated in the same manner.

The principle behind this algorithm is to use the most efficient sampler depending on the configuration of the truncated normal density. For example, take case 1 of the list above. If the truncation occurs close to the mode

[19] $\phi(\cdot)$ indicates a normal standard distribution.

of the distribution ($u_1 \leq 0.45$), a normal rejection sampler would be enough. However, this solution would be very inefficient if the truncation occurs towards the tails of the distribution, since a lot of draws would be rejected until we obtain one for the desired interval. In such a situation, Geweke (1991) argues that the truncated normal distribution comes to resemble the exponential distribution as u_1 grows so we use exponential rejection sampling to generate the draws[20]. This idea is replicated in the other cases. The author performs an extensive experiment and proves that this sampler results in significant gains in computational efficiency relative to standard samplers.

For our model, we deal with normal distributions truncated at zero from the left; i.e. $\mathrm{T}\mathbb{N}_{(-\infty,0)}$. This configuration falls into number 4 for the mixed sampler described above. Since in this case $u_2 = 0 \geq -0.45$, a normal rejection sampling is used. In other words, according to Geweke's sampler, the most efficient way to sample from a $\mathrm{T}\mathbb{N}_{(-\infty,0)}$ is to obtain a random draw from a standard normal distribution and reject it until we obtain a value in the interval $(-\infty, 0)$. For other types of censorship in more general models, the procedure described above can be used to choose the most efficient method to obtain the required sample from the truncated normal

[20]Normal, exponential and other types of rejection sampling in the procedure refer to specific applications of the acceptance-rejection method in which the sampling distribution is normal, exponential or other, respectively. To apply them, we generate draws from the indicated distribution and we reject them until we obtain one in the desired interval $[u_1, u_2]$. For example, if we are in case 1 of the list above and $u_1 \leq 0.45$ it is recommended to use normal rejection sampling; we generate draws from a $\mathbb{N}(0,1)$ distribution and we reject them until we obtain a value in the interval $[u_1, \infty)$. The underlying distribution of the sampling mechanism are the following: for normal rejection sampling we generate values from a $\mathbb{N}(0,1)$; for half-normal rejection sampling we use a $\mathbb{N}(0,1)$ distribution and then take its absolute value; for uniform rejection sampling we draw from a $U(u_1, u_2)$; for exponential rejection sampling we generate values from an exponential distribution with parameter u_1 for the first case and ${}^{\phi(u_1)}/_{\phi(u_2)}$ for the third case.

distribution.

The sampler described above applies to a univariate truncated normal distribution. However, in our model we deal with a bivariate normal distribution for which one or both variables may be censored. Fortunately, the same procedure can be applied for a bivariate setting by using the properties of the conditional normal distribution. This will be demonstrated here for the fixed-effects model described in Section 3, although the same principle applies to the random-effects model. To study how to obtain a random sample from the censored observations, we have to consider the four possible combinations described in that section.

For case I, both variables are observed so there is no need to sample. In case II, the first variable is observed while the second is censored at 0[21]. According to Geweke's sampler, this configuration requires a normal rejection sampling mechanism. Since the first variable is already observed, we can generate random draws from y^*_{2it} given y_{1it} by drawing from

$$y^*_{2it}|y_{1it} \sim \mathbb{N}\left(\gamma_2 y_{1it} + \beta_2' x_{2it} + \alpha_{2i} + \frac{\sigma^\epsilon_{12}}{\sigma^\epsilon_{11}}(y_{1it} - \hat{y}_{1it}),\ \sigma^\epsilon_{22} - \frac{(\sigma^\epsilon_{12})^2}{\sigma^\epsilon_{11}}\right)$$

so that

$$y^*_{2it}|y_{1it} = \gamma_2 y_{1it} + \beta_2' x_{2it} + \alpha_{2i} + \frac{\sigma^\epsilon_{12}}{\sigma^\epsilon_{11}}(y_{1it} - \hat{y}_{1it}) + \sqrt{\sigma^\epsilon_{22} - \frac{(\sigma^\epsilon_{12})^2}{\sigma^\epsilon_{11}}} \cdot \theta \tag{C.1}$$

$$\hat{y}_{1it} = \gamma_1 y^*_{2it} + \beta_1' x_{1it} + \alpha_{1i} \tag{C.2}$$

where θ is a random draw from a standard normal distribution and all the coefficients are those at the current iteration of the algorithm. We can see that y^*_{2it} appears in the mean of its own conditional distribution; this arises from the simultaneity in the model. One solution is to work with the

[21]In other words, the latent variable lies in the interval $(-\infty, 0)$ and we want to simulate a truncated normal variable in this interval.

reduced form. However, we found it more convenient and easier to generate one random variate θ and then iterate (C.1) and (C.2) until they converged. This usually occurred very quickly, after only a few iterations. Once this process stabilized, we examine the random draw y^*_{2it} and we keep it if is smaller than zero; otherwise, we discard it and we repeat the process until we obtain a draw in the interval $(-\infty, 0)$. We then repeat this R_i times to obtain the desired number of simulations of the complete data.

Case III refers to a situation where the first variable is censored at 0 and the second variable is observed, so it is parallel to the case described above. We now want to "complete" the data by replacing the censored variable y_{1it} with simulated values. Since the other variable is observed, we can use the properties of the conditional normal distribution to deduce that

$$y^*_{1it}|y_{2it} \sim \mathbb{N}\left(\gamma_1 y_{2it} + \beta_1' x_{1it} + \alpha_{1i} + \frac{\sigma^\epsilon_{12}}{\sigma^\epsilon_{22}}(y_{2it} - \hat{y}_{2it}),\ \sigma^\epsilon_{11} - \frac{(\sigma^\epsilon_{12})^2}{\sigma^\epsilon_{22}}\right)$$

so that

$$y^*_{1it}|y_{2it} = \gamma_1 y_{2it} + \beta_1' x_{1it} + \alpha_{1i} + \frac{\sigma^\epsilon_{12}}{\sigma^\epsilon_{22}}(y_{2it} - \hat{y}_{2it}) + \sqrt{\sigma^\epsilon_{11} - \frac{(\sigma^\epsilon_{12})^2}{\sigma^\epsilon_{22}}} \cdot \theta \qquad \text{(C.3)}$$

$$\hat{y}_{2it} = \gamma_2 y^*_{1it} + \beta_2' x_{2it} + \alpha_{2i} \qquad \text{(C.4)}$$

where all the coefficients are those at the current iteration of the algorithm. Like before, we can iterate (C.3) and (C.4) until convergence to obtain a random draw from y^*_{1it} given the observed value of y_{2it}. If this value is negative, we keep it; otherwise we repeat the procedure until a random draw in the desired interval is generated. We then repeat the entire process R_i times.

Finally, both variables are censored in case IV. The solution applied in this situation was to generate a bivariate normal random variable for $(\epsilon_{1it}, \epsilon_{2it})$ according to the joint distribution of the errors described in Section 3. We than calculate

$$y^*_{1it} = \gamma_1 y^*_{2it} + \beta_1' x_{1it} + \alpha_{1i} + \epsilon_{1it} \tag{C.5}$$

$$y^*_{2it} = \gamma_2 y^*_{1it} + \beta_2' x_{2it} + \alpha_{2i} + \epsilon_{2it} \tag{C.6}$$

where the coefficients are those at the current iteration of the MCECM algorithm. We iterate (C.5) and (C.6) until convergence. Since both variables are censored, we want to obtain a bivariate draw such that both variables are negative. If this is not the case, we discard it and repeat the process until we obtain a draw where both variables are in the interval $(-\infty, 0)$. We repeat this process until we obtain the R_i desired simulations of the complete data.

To summarize, in order to obtain the simulated values of the complete data, we classify the observations according to the four cases described above. According to the specific case, we apply the appropriate procedure to generate the simulated draws needed to achieve the Monte Carlo E-step of the algorithm. Instead of working with the original model, we could instead pass to the reduced form to avoid the simultaneity problems that we encountered above. However, we found out that it was more practical and straightforward to found convergence through iteration for each simulation.

C.3 Parallel computing

Each step of the MCECM algorithm requires R_i simulation for the Monte Carlo E-step until we reach the maximum of $R_{\max}$ simulations. This entire procedure is repeated H times in the simulation study of Section 5 in order to deduce the properties of the estimator. It becomes evident that there is a significant simulation effort required and the computation cost of such an algorithm can be very large even for a fast computer and an efficient program.

However, this being a Monte Carlo experiment, the R repetitions of the algorithm are independent of one another. In a setting such as this one, there is considerable room for improvement from using parallel computing (see Creel, 2005). Modern processors with multiple cores can be used to run several repetitions of the experiment parallelly. Standard commands for parallel computing already incorporated in mathematical software can distribute the work among the cores and compile the results at the end of the experiment. The only thing that must be carefully programmed is the recording of the output so that it remains independent among repetitions. This solution, coded in Matlab, was used for this research.

Bibliography

E. H. Aarts, J. H. Korst, and P. J. V. Laarhoven. Simulated annealing. In E. Aarts and J. K. Lenstra, editors, *Local search in combinatorial optimization*, chapter 4, pages 91–120. John Wiley & Sons, Ltd., 1997.

C. Ai and C. Chen. Estimation of a fixed effects bivariate censored regression model. *Economics Letters*, 40(4):403–406, December 1992.

S. Alan, B. E. Honoré, and S. Leth-Petersen. Estimation of panel data models with two-sided censoring. CAM working papers, Centre for Applied Microeconometrics, University of Copenhagen, November 2008.

T. Amemiya. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42 (6):999–1012, November 1974.

T. Amemiya. The estimation of a simultaneous-equation Tobit model. *International Economic Review*, 20(1):169–181, February 1979.

T. Amemiya. Tobit models: a survey. *Journal of Econometrics*, 24(1-2): 3–61, 1984.

T. W. Anderson and C. Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82, 1982.

M. Arellano and S. Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2):277–97, April 1991.

M. Arellano and B. Honoré. Panel data models: Some recent developments. In J. J. Heckman and E. E. Leamer, editors, *Handbook of Econometrics*, volume 5, chapter 53, pages 3229–3296. Elsevier Science Publishers, 2001.

R. B. Avery. Error components and seemingly unrelated regressions. *Econometrica*, 45(1):199–209, January 1977.

A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, 61(3):579–602, 1999.

P. Balestra and J. Varadharajan-Krishnakumar. Full information estimations of a system of simultaneous equations with error component structure. *Econometric Theory*, 3(2):223–246, August 1987.

B. H. Baltagi. On seemingly unrelated regressions with error components. *Econometrica*, 48(6):1547–1551, September 1980.

B. H. Baltagi. Simultaneous equations with error components. *Journal of Econometrics*, 17(2):189–200, November 1981.

B. H. Baltagi. *Econometric Analysis of Panel Data.* John Wiley & Sons, Ltd., third edition, 2005.

R. Blundell and S. Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, August 1998.

R. W. Blundell and R. J. Smith. Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies*, 56(1):37–57, January 1989.

R. W. Blundell and R. J. Smith. Coherency and estimation in simultaneous models with censored or qualitative dependent variables. *Journal of Econometrics*, 64(1-2):355–373, 1994.

O. Bover and M. Arellano. Estimating dynamic limited dependent variable models from panel data. *Investigaciones Económicas*, 21(2):141–165, May 1997.

B. S. Caffo, W. Jank, and G. L. Jones. Ascent-based Monte Carlo expectation-maximization. *Journal of The Royal Statistical Society Series B*, 67(2):235–251, 2005.

J.-P. Chavas and K. Kim. A heteroskedastic multivariate Tobit analysis of price dynamics in the presence of price floors. *American Journal of Agricultural Economics*, 86(3):576–593, 8 2004.

S. Chen and X. Zhou. Semiparametric estimation of a bivariate Tobit model. *Journal of Econometrics*, 2011. (to appear).

S. Chib and E. Greenberg. Analysis of multivariate probit models. *Biometrika*, 85(2):357–361, 1998.

J. Cornick, T. L. Cox, and B. W. Gould. Fluid milk purchases: a multivariate Tobit analysis. *American Journal of Agricultural Economics*, 76(1): 74–82, Feburary 1994.

C. Cornwell, P. Schmidt, and D. Wyhowski. Simultaneous equations and panel data. *Journal of Econometrics*, 51(1-2):151–181, 1992.

M. Creel. User-friendly parallel computations with econometric examples. *Computational Economics*, 26(2):107–128, October 2005.

G. Criton. Estimation of simultaneous equation panel data models with censored endogenous variables. Master's thesis, University of Geneva, March 2007.

A. E. Gelfand. Gibbs sampling. *Journal of the American Statistical Association*, 95(452):1300–1304, December 2000.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–408, 1990.

A. E. Gelfand and A. F. M. Smith. Gibbs sampling for marginal posterior expectations. *Communications in Statistics - Theory and Methods*, 20(5): 1747–1766, 1991.

J. Geweke. Efficient simulation from the multivariate normal and Student t-distributions subject to linear constraints. *Computer Sciences and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 571–578, 1991.

M. Gilli and P. Winker. A review of heuristic optimization methods in econometrics. Swiss Finance Institute Research Paper Series 2008-12, Swiss Finance Institute, 2008.

C. Gouriéroux and A. Monfort. *Simulation-based econometric methods*. Oxford University Press, 1997.

W. Greene. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal*, 7(1):98–119, June 2004a.

W. Greene. Fixed effects and bias due to the incidental parameters problem in the Tobit model. *Econometric Reviews*, 23(2):125–147, 2004b.

W. H. Greene. *Econometric Analysis*. Pearson Education International, fifth edition, 2003.

M. Gregory and A. S. Kalwij. Overtime hours in Great Britain over the period 1975-1999: A panel data analysis. Economics Series Working Papers 27, University of Oxford, Department of Economics, 2000.

P. V. Grootendorst. Health care policy evaluation using longitudinal insurance claims data: An application of the panel Tobit estimator. *Health Economics*, 6(4):365–382, 1997.

V. A. Hajivassiliou. A simulation estimation analysis of the external debt crises of developing countries. *Journal of Applied Econometrics*, 9(2): 109–131, April-June 1994.

V. A. Hajivassiliou and P. A. Ruud. Classical estimation methods for ldv models using simulation. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 40, pages 2383–2441. Elsevier, 1994.

J. A. Hausman and D. A. Wise. Attrition bias in experimental and panel data: The Gary income maintenance experiment. *Econometrica*, 47(2): 455–473, March 1979.

J. J. Heckman and T. E. Macurdy. A life cycle model of female labour supply. *Review of Economic Studies*, 47(1):47–74, January 1980.

B. E. Honoré. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, 60(3): 533–565, May 1992.

B. E. Honoré. Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics*, 59(1-2):35–61, September 1993.

B. E. Honoré and L. Hu. Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics*, 122(2):293–316, October 2004.

B. E. Honoré, E. Kyriazidou, and J. L. Powell. Estimation of Tobit-type models with individual specific effects. *Econometric Reviews*, 19(3):341–366, 2000.

C. Hsiao. *Analysis of panel data*, volume 34 of *Econometric Society Monographs*. Cambridge University Press, second edition, 2003.

L. Hu. Estimation of a censored dynamic panel data model. *Econometrica*, 70(6):2499–2517, November 2002.

C. J. Huang, F. A. Sloan, and K. W. Adamache. Estimation of seemingly unrelated Tobit regressions via the EM algorithm. *Journal of Business & Economic Statistics*, 5(3):425–430, 1987.

H.-C. R. Huang. Estimation of the SUR Tobit model via the MCECM algorithm. *Economics Letters*, 64(1):25–30, July 1999.

J. Huguenin, F. Pelgrin, and A. Holly. Estimation of multivariate probit models by exact maximum likelihood. Working Papers 09-02, University of Lausanne, Institute of Health Economics and Management (IEMS), February 2009.

G. Jakubson. The sensitivity of labor-supply parameter estimates to unobserved individual effects: fixed- and random-effects estimates in a nonlinear model using panel data. *Journal of Labor Economics*, 6(3):302–329, July 1988.

A. S. Kalwij. A maximum likelihood estimator based on first differences for a panel data Tobit model with individual specific effects. *Economics Letters*, 81(2):165–172, 2003.

M. P. Keane. Simulation estimation for panel data models with limited dependent variables. In G. S. Maddala, C. R. Rao, and H. D. Vinod, editors, *Handbook of Statistics*, volume 11, chapter 20, pages 545–571. Elsevier Science Publishers, 1993.

M. P. Keane. A computationally practical simulation estimator for panel data. *Econometrica*, 62(1):95–116, January 1994.

B. S. Kim and G. S. Maddala. Estimation and specification analysis of models of dividend behavior based on censored panel data. *Empirical Economics*, 17(1):111–124, 1992.

T. Kinal and K. Lahiri. On the estimation of simultaneous-equations error-components models with an application to a model of developing country foreign trade. *Journal of Applied Econometrics*, 8(1):81–92, January-March 1993.

J. Krishnakumar. *Estimation of simultaneous equation models with error components structure*, volume 312 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1988.

F. Laisney and M. Lechner. Almost consistent estimation of panel probit models with "small" fixed effects. *Econometric Reviews*, 22(1):1–28, 2003.

T. Lancaster. *An introduction to modern Bayesian econometrics*. Wiley-Blackwell, 2004.

L.-F. Lee. Multivariate Tobit models in econometrics. In G. S. Maddala, C. R. Rao, and H. D. Vinod, editors, *Handbook of Statistics*, volume 11, chapter 6, pages 145–173. Elsevier Science Publishers, 1993.

M.-J. Lee. Semi-parametric estimation of simultaneous equations with limited dependent variables: a case study of female labour supply. *Journal of Applied Econometrics*, 10(2):187–200, April-June 1995.

T. A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistica Society, Series B*, 44(2):226–233, 1982.

G. S. Maddala. Limited dependent variable models using panel data. *Journal of Human Resources*, 22(3):307–338, 1987.

X.-L. Meng and D. B. Rubin. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, 80(2):267–278, June 1993.

F. Nelson and L. Olson. Specification and estimation of a simultaneous-equation model with limited dependent variables. *International Economic Review*, 19(3):695–709, October 1978.

W. K. Newey. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, 36(3): 231–250, November 1987.

I. R. Prucha. Maximum likelihood and instrumental variable estimation in simultaneous equation systems with error components. *International Economic Review*, 26(2):491–506, June 1985.

C. R. Reeves and J. E. Rowe. *Genetic algorithms: principles and perspectives*. Kluwer Academic Publishers, 2002.

G. Rodriguez-Yam, R. A. Davis, and L. L. Scharf. Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression. Unpublished, March 2004.

S. Rosenbaum. Moments of a truncated bivariate normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(2):405–408, 1961.

P. A. Ruud. Extensions of estimation methods using the em algorithm. *Journal of Econometrics*, 49(3):305–341, September 1991.

R. J. Smith and R. W. Blundell. An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. *Econometrica*, 54(3):679–685, May 1986.

M. A. Tanner and W. H. Wong. The calculation of posterior distribution by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, 1987.

D. A. Van Dyk and X.-L. Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 19(1):1–50, 2001.

F. Vella. A simple estimator for simultaneous models with censored endogenous regressors. *International Economic Review*, 34(2):441–457, May 1993.

F. Vella and M. Verbeek. Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics*, 90(2):239–263, June 1999.

G. H. Wan, W. E. Griffiths, and J. R. Anderson. Using panel data to estimate risk effects in seemingly unrelated production functions. *Empirical Economics*, 17(1):35–49, 1992.

G. C. G. Wei and M. A. Tanner. Posterior computations for censored regression data. *Journal of the American Statistical Association*, 85(411): 829–839, 1990a.

G. C. G. Wei and M. A. Tanner. A Monte Carlo implementaion of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association*, 85(411):699–704, 1990b.

G. C. G. Wei and M. A. Tanner. Applications of multiple imputation to the analysis of censored regression data. *Biometrics*, 47:1297–1309, 1991.

P. Winker. *Optimization heuristics in econometrics: Application of threshold accepting*. John Wiley & Sons, Ltd., 2001.

X. Zhou and X. Liu. The Monte Carlo EM method for estimating multinomial probit latent variable models. *Computational Statistics*, 23(2): 277–289, April 2008.