Simultaneous Equations Models with Limited Dependent Variables and Social Interactions

by

Sutianjie Zhou

B.S., Fudan University, 2015

M.A., Boston University, 2017

M.S., University of Colorado, 2023

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

2024

Committee Members:

Xiaodong Liu, Chair

Carlos Martins-Filho

Adam McCloskey

Scott Savage

Lijun Chen

Zhou, Sutianjie (Ph.D., Economics)

Simultaneous Equations Models with Limited Dependent Variables and Social Interactions

Thesis directed by Prof. Xiaodong Liu

My dissertation studies the behaviors of agents engaged in interconnected activities within a social network. This research comprehensively analyzes how agents' decisions across multiple activities are influenced by those of their peers. I characterize the Bayesian Nash Equilibrium of the underlying network game and provide a sufficient condition for the existence and uniqueness of the equilibrium. I also propose a computationally feasible estimation method to recover the structural parameters and investigate its finite sample performance through Monte Carlo simulations.

The first chapter, "Simultaneous Equations with Censored Outcomes and Social Interactions," considers a simultaneous equations model where agents participate in multiple activities with censored outcomes. An agent's decision in an activity depends on not only their latent incentives in other activities but also their rational expectation of other agents' decisions in all related activities. The second chapter, "Simultaneous Equations with Limited Dependent Variables and Social Interactions," further extends this model to allow the observed outcome in different activities to be continuous, censored, or binary. I provide a microfoundation of the econometric model and derive a sufficient condition for the existence and uniqueness of the equilibrium. I propose a two-stage estimation procedure for the model, where I estimate the reduced form parameter using the nested pseudo-likelihood method in the first stage and recover the structural parameters from the reduced form parameters in the second stage. I establish the asymptotic properties of the proposed estimator and conduct Monte Carlo simulations to study its finite sample performance.

The third chapter, "A Simultaneous Equation Tobit Model with Social Interactions," considers a similar model to the first chapter. The key difference is that we assume an agent's decision in an activity depends on their actual decisions instead of latent incentives in other activities. This introduces an additional complication to the analysis as it becomes impossible to derive the reduced form of the structural model. Hence, instead of using the two-stage estimation procedure developed in the first two chapters, I propose a new estimator that directly estimates the structural parameters. I show in Monte Carlo simulations that the new estimator works well in finite samples.

Dedication

To my beloved Father, Bin Zhou, and Mother, Fang Su.

Acknowledgements

I owe profound gratitude to my advisor, Professor Xiaodong Liu, for his invaluable guidance throughout my economic research and career development. His expertise was instrumental in shaping my ideas in econometric modeling and analysis and his influence has inspired me to become a diligent and proactive researcher.

I extend my heartfelt gratitude to Professor Carlos Martins-Filho for his expert guidance in advanced asymptotic probability and statistics. His encouragement to critically evaluate each step of my research and his essential nonparametric insights have been pivotal in clarifying my research direction.

My thanks also go to Professor Adam McCloskey, who has been instrumental in deepening my understanding of time series analysis and improving my skills in organizing econometric presentations. His career guidance has been a source of great encouragement to me.

I am grateful to Professor Scott Savage for his invaluable teaching in empirical industrial research and for assisting me in uncovering the microeconomic foundations behind econometric models, thus illuminating their potential real-world applications.

I also thank Professor Lijun Chen for introducing me to the theory and application of convex optimization. His teachings have not only benefited my research but have also provided me with a unique and insightful perspective on my work.

I extend my deepest gratitude to the entire staff in the Department of Economics for their diligent work and compassionate assistance. Their efficiency and warmth have significantly contributed to my experience.

I must also express my sincere thanks to my friends and classmates. As an international student, their support was crucial in making my graduate studies not only manageable but also enjoyable.

Lastly, but most importantly, I am profoundly thankful to my beloved mother and father for their unwavering support, deep concern, and continuous encouragement throughout this journey. Their strength and belief in me have been my guiding light.

Contents

Chapter 1

Simultaneous Equations with Censored Outcomes and Social Interactions

1.1 Introduction

In various real-world economic scenarios, the results of individual or agent activities are often censored. This means that the outcomes, such as a household's annual expenditure on travel, are restricted to either zero or a positive figure. Similarly, research into adolescent behavior focuses on quantifiable activities like cigarette smoking and alcohol consumption, with outcomes that are also non-negative. These outcomes can be influenced not only by an individual's other activities but also by the activities of others within their network. For instance, a household's spending on travel can affect its entertainment budget, and vice versa, with both potentially influenced by the spending habits of their peer group. Likewise, in adolescent behavior studies, the number of cigarettes smoked by a teenager can affect their alcohol consumption, and vice versa, with both behaviors also shaped by their social circle.

This interdependence of outcomes across different actions by the same agent, as well as the influence of others' actions, highlights a complex web of interactions. Our paper introduces a simultaneous equation model that accommodates censored outcomes and social interactions. This model aims to elucidate the interconnected nature of economic decision-making across different agents and activities, providing insights into how individual and collective behaviors shape economic outcomes.

Drawing from foundational works in the field, Amemiya (1974) introduced a multivariate Tobit model that extends the methodologies of Tobin (1958) and Amemiya (1973). This model simplifies consistent estimation by categorizing agents into two groups based on the outcomes of their activities. In a different vein, Nelson and Olson (1978) developed an alternative multivariate Tobit model that incorporates latent endogenous variables within a broader truncation framework, allowing for outcomes that are not exclusively zeros. This model is characterized by having one fully observed endogenous variable and another that is truncated, and introduces a two-stage estimation process that leverages both the least squares and likelihood methods. Building on these concepts, Amemiya (1979) further expanded the simultaneous-equation Tobit model to more comprehensively integrate both truncated and non-truncated endogenous outcomes, proposing a more efficient GLS-based estimator by optimizing the variance-covariance matrix. This approach was shown to be more effective than Nelson and Olson (1978) estimation technique.

Our econometric model aligns with the traditional framework of simultaneous equation models, but it distinctively relies on the unobserved reservation values rather than the observed outcomes for the endogenous variables. This approach underscores the importance of underlying factors not directly observable in influencing economic behaviors and outcomes.

Our research is predicated on the understanding that agents engage in multiple activities concurrently, with outcomes that are not visible to other agents prior to decision-making. However, the anticipation of the outcomes of others significantly impacts an agent's actions. This scenario necessitates that researchers unravel the complexities of the Bayesian Nash Equilibrium within the context of an incomplete information network game. Liu (2019) explores a binary simultaneous-equation model that incorporates social interactions under incomplete information, employing estimation techniques based on maximum likelihood and nested pseudo-likelihood as introduced by Aguirregabiria and Mira (2007). Yang et al. (2018) examines the dynamics of the single-equation Tobit model in scenarios characterized by incomplete information and social interactions, drawing comparisons with cases of complete information. This study builds upon Xu and Lee (2015), which developed a maximum likelihood estimator for the single-equation SAR Tobit model. Additional literature, including Lee et al. (2014), Lin and Xu (2017), and Yang and Lee (2017), delves into the existence of Bayesian Nash Equilibrium within social networks under incomplete information, where individuals harbor rational expectations regarding the outcomes of others.

The rest of this chapter is organized as follows. Section 1.2 clarifies the microeconomic foundation of this paper, a network multi-activity game under incomplete information case. Section 1.3 is the generation of the econometric model of this paper. Section 1.4 presents the estimation. Section 1.5 gives the Monte Carlo simulation. Section 1.6 concludes.

1.2 Incomplete Information Network Game

The microeconomic foundation of the econometric model in this paper is based on the network game with incomplete information. First of all, we need to clarify the network structure. Suppose there are n agents in a network. Each agent can interact with other agents. Then, without loss of generality, we will have a row-normalized (the sum of elements in each row equals one) network matrix $\mathbf{W} = [w_{ij}]$ for indices i and j from 1 to n, where the self-tied network weight is ruled out, i.e., $w_{ii} = 0$ for all i, and w_{ij} are known, predetermined, non-stochastic non-negative constants, representing agent j's influence on agent i. Therefore, w_{ij} is not necessarily equal to w_{ji}, and w_{ij} can be zero, meaning that agent j has no direct influence on agent i.

Suppose each agent i in the network participates in m activities. y^*_{ik} represents agent i's intention in the k-th activities. Other people can only observe agent i's outcome in activity k, i.e., $y_{ik} = y^*_{ik}\mathrm{I}(y^*_{ik} > 0)$ where $\mathrm{I}(\cdot)$ is an indicator function that will be 1 if the inside statement is true and 0 otherwise. For other agent $j \neq i$, he/she can only observe y_{ik} for $k = 1, ..., m$, and y^*_{ik} is known only by agent i him/herself. Then from Ballester et al. (2006), Calvó-Armengol et al. (2009) and Blume et al. (2015), we employ a linear-quadratic form utility function for agent i as

$$\mathcal{U}_i = \sum_{k=1}^{m}(\sum_{l=1}^{m} \varrho_{lk} \sum_{j=1}^{n} w_{ij} y_{jl} + \varpi_{ik} - \varepsilon_{ik}) y^*_{ik} - \frac{1}{2}\sum_{k=1}^{m}\sum_{l=1}^{m} \vartheta_{lk} y^*_{ik} y^*_{il} \tag{1.1}$$

where $\vartheta_{kl} = \vartheta_{lk}$, and $\vartheta_{kk} \neq 0$ for all k. $\{w_{ij}\}_{i=1,j=1}^{n,n}$ is the predetermined, non-stochastic, and exogenous network structure parameters. $\{\varpi_{ik}\}_{i=1,k=1}^{n,m}$ is the predetermined, non-stochastic, and exogenous attributes that directly influence agent i's utility from activity k. The utility function contains two items of activities' intention, the first item, $\sum_{k=1}^{m}(\sum_{l=1}^{m} \varrho_{lk} \sum_{j=1}^{n} w_{ij} y_{jl} + \varpi_{ik} - \varepsilon_{ik}) y^*_{ik}$, represents the payoff from $\{y_{ik}^*\}_{k=1}^m$ (the higher $\{y_{ik}^*\}_{k=1}^m$, the higher the utility $\mathcal{U}_i$). The second item, $\frac{1}{2}\sum_{k=1}^m \sum_{l=1}^m \vartheta_{lk} y_{ik}^* y_{il}^*$, represents the cost generated by $\{y_{ik}^*\}_{k=1}^m$ (the higher $\{y_{ik}^*\}_{k=1}^m$, the lower the utility $\mathcal{U}_i$ because there is a negative sign in front of the term in the utility function). We could figure out that the utility function is a payoff-cost-structure linear-quadratic function of $\{y_{ik}^*\}_{k=1}^m$. Given this form of the utility function, for each agent i in this network, other agents j's activities' outcome $\{y_{jk}\}_{k=1}^m$, where $w_{ij} \neq 0$, will influence the marginal benefit/payoff of agent i's activities' intention $\{y_{ik}^*\}_{k=1}^m$, the coefficient ϱ_{lk} in the first part of the utility function can be interpreted as the spillover effect for peers' activities outcomes on the marginal payoff of agent i's $\{y_{ik}^*\}_{k=1}^m$. Also, the marginal benefit/payoff is influenced by agent i's characteristics $\varpi_{ik} - \varepsilon_{ik}$, where ϖ_{ik} is public knowledge and known by all agents in the network, and ε_{ik} is the random error and only privately known by agent i. ε_{ik} are independent of $\{\varpi_{ik}\}_{i=1,k=1}^{n,m}$.

The utility (1.1) is proposed similarly to Liu (2019). The difference is the activities' outcomes in our model are censored instead of binary. The utility function we propose is also different from that in Cohen-Cole et al. (2018). First, the outcomes of agents' activities are censored instead of observable values (if y_{ik}^* is negative, other people can only observe $y_{ik} = 0$). Second, there is an unobserved random error term (shock) for each agent's activity, i.e., ε_{ik}.

Remark 1. *A linear quadratic form utility function is applied here because it incorporates a linear benefit component alongside a quadratic cost component. This structure ensures a consistent marginal benefit while leading to a rise in marginal cost, which fosters diminishing marginal utility, assuming other variables remain constant. A firm's research and development endeavors are advantageous within a real-world economic context. However, these benefits are also influenced by competitors' research outcomes, the firm's own attributes, and unique random shocks. Costs are categorized into intra-project and inter-project expenses, acknowledging that coordinated costs emerge across multiple projects as a company engages in various research and development works. Hsieh et al. (2022) utilizes a single-activity form of the linear quadratic utility function to examine a firm's research and development efforts within a social network.*

According to the network structure and the public knowledge that influences the activities' intention, each agent i ($i = 1, 2, ..., n$) chooses $\{y^*_{ik}\}_{k=1}^m$ simultaneously to maximize the conditional expected utility

$$E(\mathcal{U}_i|\{\varepsilon_{ik}\}_{k=1}^m) = \sum_{k=1}^m(\sum_{l=1}^m \varrho_{lk} \sum_{j=1}^n w_{ij}p_{jl} + \varpi_{ik} - \varepsilon_{ik})y^*_{ik} - \frac{1}{2}\sum_{k=1}^m\sum_{l=1}^m \vartheta_{lk}y^*_{ik}y^*_{il} \tag{1.2}$$

where $p_{jl} = \mathrm{E}(y_{jl})$. From the first-order condition in maximizing conditional expected utility, we have

$$\sum_{l=1}^m \frac{\vartheta_{lk}}{\vartheta_{kk}} y^*_{il} = \sum_{l=1}^m \frac{\varrho_{lk}}{\vartheta_{kk}} \sum_{j=1}^n w_{ij}p_{jl} + \frac{\varpi_{ik} - \varepsilon_{ik}}{\vartheta_{kk}}. \tag{1.3}$$

Let $\theta_{lk} = \frac{\vartheta_{lk}}{\vartheta_{kk}}$, $\lambda_{lk} = \frac{\varrho_{lk}}{\vartheta_{kk}}$, $\pi_{ik} = \frac{\varpi_{ik}}{\vartheta_{kk}}$, and $\epsilon_{ik} = \frac{\varepsilon_{ik}}{\vartheta_{kk}}$, then we can rewrite the first-order condition of maximizing conditional utility in the following form

$$\sum_{l=1}^m \theta_{lk}y^*_{il} = \sum_{l=1}^m \lambda_{lk} \sum_{j=1}^n w_{ij}p_{jl} + \pi_{ik} - \epsilon_{ik} \tag{1.4}$$

to write the vector form, we introduce the following notations,

$$\begin{aligned} \mathbf{y}^*_l =& \quad (y^*_{1l}, y^*_{2l}, ..., y^*_{nl})' \\ \mathbf{p}_l =& \quad (p_{1l}, p_{2l}, ..., p_{nl})' \\ \boldsymbol{\pi}_k =& \quad (\pi_{1k}, \pi_{2k}, ..., \pi_{nk})' \\ \boldsymbol{\epsilon}_k =& \quad (\epsilon_{1k}, \epsilon_{2k}, ..., \epsilon_{nk})' \end{aligned} \tag{1.5}$$

the vector form can be written as

$$\sum_{l=1}^m \theta_{lk}\mathbf{y}^*_l = \sum_{l=1}^m \lambda_{lk}\mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k - \boldsymbol{\epsilon}_k. \tag{1.6}$$

In the equation (1.6), the θ_{lk} represents the interdependence effect among activities conducted by the same agent. In other words, an agent's underlying intention in activity k will depend on his/her underlying intention in activity l. And obviously, we have $\theta_{kk} = 1$. λ_{kk} represents the same-activity's peer effect, which means an agent's intention of activity k may be impacted by the expected activities' outcomes of the peers in the same activity. λ_{kl} represents the cross-activity's peer effect, which means an agent's intention of activity k may be impacted by the expected activities' outcomes of the peers in activity l. Then, we introduce the following notations to write the matrix form

$$
\begin{aligned}
\mathbf{Y}^* = &\quad [\mathbf{y}_1^*, \mathbf{y}_2^*, ..., \mathbf{y}_m^*] \\
\mathbf{P} = &\quad [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m] \\
\mathbf{\Pi} = &\quad [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, ..., \boldsymbol{\pi}_m] \\
\mathbf{E} = &\quad [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, ..., \boldsymbol{\epsilon}_m]
\end{aligned}
\tag{1.7}
$$

then the matrix form can be written as

$$\mathbf{Y}^*\mathbf{\Theta} = \mathbf{WP\Lambda} + \mathbf{\Pi} - \mathbf{E} \tag{1.8}$$

Given $\mathbf{\Theta}$ is non-singular, we will have the reduced form of the matrix form of our model, which is

$$\mathbf{Y}^* = \mathbf{WP\Lambda}^* + \mathbf{\Pi}^* - \mathbf{E}^* \tag{1.9}$$

where $\mathbf{\Lambda}^* = \mathbf{\Lambda\Theta}^{-1}$, $\mathbf{\Pi}^* = \mathbf{\Pi\Theta}^{-1}$, and $\mathbf{E}^* = \mathbf{E\Theta}^{-1}$. According to the reduced matrix form of our model, we can derive the scalar reduced form as

$$y_{ik}^* = \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^* - \epsilon_{ik}^* \tag{1.10}$$

where $[\lambda_{lk}^*]_{l=1,k=1}^{m,m} = \mathbf{\Lambda}^*$, $[\pi_{ik}^*]_{i=1,k=1}^{n,m} = \mathbf{\Pi}^*$, and $[\epsilon_{ik}^*]_{i=1,k=1}^{n,m} = \mathbf{E}^*$. Then

$$p_{ik} = \mathrm{E}(y_{ik}|\{\varpi_{jl}\}_{j=1,l=1}^{n,m}) = \mathrm{E}(y_{ik}|\{\varpi_{jl}\}_{j=1,l=1}^{n,m}, y_{ik} > 0)\mathbf{Pr}(y_{ik} > 0|\{\varpi_{jl}\}_{j=1,l=1}^{n,m}). \tag{1.11}$$

In the current stage, we can not derive a similar form as the equation followed by (2.6) in Liu (2019). Therefore, we need to propose the following assumption.

Assumption 1.2.1. *The simultaneous effect (interdependence effect) matrix* Θ *is nonsingular.*

Remark 2. *Given the simultaneous effect matrix is nonsingular, we have the capability to extract the reduced form model. This foundation allows us to implement a two-step parameter estimation process: initially, we estimate the parameters of the reduced form, followed by the derivation of the structural form parameters. These procedures will be elaborately outlined in the sections dedicated to the econometric model and parameter estimation.*

Assumption 1.2.2. *For each agent i in our model, the random shock $(\epsilon_{i1}, \cdots, \epsilon_{im})'$ are jointly normally distributed with zero mean and variance-covariance matrix $\mathbf{\Sigma}$, and are independently among agents.*

Remark 3. *The derivation of our Bayesian Nash Equilibrium relies on the assumption that random shocks follow a joint-normal distribution. Yet, in cases where the outcomes are binary, similar to the model in Liu (2019), it is not necessary to presume that these random shocks are jointly normally distributed. Further elaboration on this point is provided in the subsequent parts of this section.*

With Assumption 1.2.1 and Assumption 1.2.2, we can derive that the reduced form random shock vector of our model, i.e., $(\epsilon^*_{i1}, \cdots, \epsilon^*_{im})'$, are jointly normally distributed with zeros mean and variance-covariance matrix $\mathbf{\Sigma}^* = \mathbf{\Theta}'^{-1}\mathbf{\Sigma}^*\mathbf{\Theta}^{-1} = [\rho^*_{kl}\sigma^*_k\sigma^*_l]^{m,m}_{k=1,l=1}$, all the diagonal elements of the matrix $\mathbf{\Sigma}^*$ are finite, $\rho^*_{kk} = 1$ for $k = 1, \cdots, m$ This can interpreted by the following proposition. Then equation 1.11 can be derived as

$$\begin{aligned} p_{ik} = &\Big(\sum_{l=1}^{m}\lambda^*_{lk}\sum_{j=1}^{n}w_{ij}p_{jl} + \pi^*_{ik}\Big)\mathbf{\Phi}\Big(\frac{\sum_{l=1}^{m}\lambda^*_{lk}\sum_{j=1}^{n}w_{ij}p_{jl} + \pi^*_{ik}}{\sigma^*_k}\Big) \\ &+ \sigma^*_k\phi\Big(\frac{\sum_{l=1}^{m}\lambda^*_{lk}\sum_{j=1}^{n}w_{ij}p_{jl} + \pi^*_{ik}}{\sigma^*_k}\Big) \end{aligned} \tag{1.12}$$

According to the expression of $\{p_{ik}\}^{n,m}_{i=1,k=1}$, we use the $vec()$ function where $\mathbf{p} = vec(\mathbf{P})$ and we have $\mathbf{p} = \vec{h}(\mathbf{p})$ in the Bayesian Nash Equilibrium (BNE). Then $\mathbf{p} = (\mathbf{p}'_1, \cdots, \mathbf{p}'_m)'$

$$\mathbf{p}_k = \Big(\sum_{l=1}^{m}\lambda^*_{lk}\mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k\Big) \odot \mathbf{\Phi}\Big(\frac{\sum_{l=1}^{m}\lambda^*_{lk}\mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma^*_k}\Big) + \sigma^*_k\phi\Big(\frac{\sum_{l=1}^{m}\lambda^*_{lk}\mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma^*_k}\Big) \tag{1.13}$$

then, we set the notation

$$\vec{h}(\mathbf{p}) = [\vec{h}_1(\mathbf{p})', \cdots, \vec{h}_m(\mathbf{p})']' \tag{1.14}$$

for each $k = 1, \cdots, m$, we have

$$\begin{aligned} \vec{h}_k(\mathbf{p}) = &(u_{1k}\mathbf{\Phi}_{1k} + \sigma^*_k\phi_{1k}, \cdots, u_{nk}\mathbf{\Phi}_{nk} + \sigma^*_k\phi_{nk})' \\ = &\Big(\sum_{l=1}^{m}\lambda^*_{lk}\mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k\Big) \odot \mathbf{\Phi}\Big(\frac{\sum_{l=1}^{m}\lambda^*_{lk}\mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma^*_k}\Big) + \sigma^*_k\phi\Big(\frac{\sum_{l=1}^{m}\lambda^*_{lk}\mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma^*_k}\Big) \end{aligned} \tag{1.15}$$

where

$$\begin{aligned} u_{ik} =& \sum_{l=1}^{m} \lambda_{lk}^{*} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^{*} \\ \mathbf{\Phi}_{ik} =& \mathbf{\Phi}\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^{*} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^{*}}{\sigma_k^{*}} \right) \\ \phi_{ik} =& \phi\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^{*} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^{*}}{\sigma_k^{*}} \right) \end{aligned} \tag{1.16}$$

We also have $p_{ik} = u_{ik}\mathbf{\Phi}_{ik} + \sigma_k^{*}\phi_{ik}$. To propose a sufficient condition for the existence of the uniqueness of the solution to $\mathbf{p} = \vec{h}(\mathbf{p})$, we need to add the following assumption to our previous assumptions.

Assumption 1.2.3. *The reduced form peer effect matrix $\mathbf{\Lambda}^{*}$, and the network structure matrix $\mathbf{W}$ should satisfy*

$$\min\{||\mathbf{\Lambda}^{*}||_1||\mathbf{W}||_\infty, ||\mathbf{\Lambda}^{*}||_\infty||\mathbf{W}||_1\} < 1$$

Where for any $(n \times m)$ matrix $\mathbf{A}$, $||\mathbf{A}||_\infty$ is the row-sum matrix norm and $||\mathbf{A}||_1$ is the column-sum matrix norm

$$\begin{aligned} ||\mathbf{A}||_\infty &= \max_{i=1,2,\ldots,n} \sum_{j=1}^{m} |a_{ij}| \\ ||\mathbf{A}||_1 &= \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} |a_{ij}| \end{aligned} \tag{1.17}$$

Remark 4. *Let us consider a two-activity case with agents in a row-normalize network, if $\lambda_{11} = \lambda_{22} = 0.9$ and $\lambda_{12} = \lambda_{21} = 0.6$ and $\theta_{21} = \theta_{12} = 0.5$. We can derive $\mathbf{\Lambda}^{*} = \mathbf{\Lambda}\mathbf{\Theta}^{-1}$ and have $\lambda_{11}^{*} = \lambda_{22}^{*} = 0.8$ and $\lambda_{12}^{*} = \lambda_{21}^{*} = 0.2$, then we have $\min\{||\mathbf{\Lambda}^{*}||_1||\mathbf{W}||_\infty, ||\mathbf{\Lambda}^{*}||_\infty||\mathbf{W}||_1\} = 1$, and the data generating process can not generate the Bayesian Nash Equilibrium (cannot stop at a fixed-point in the mapping 1.15). More geometric steps will be contained in the rest of this section. However, this parameter setting works well in the binary-outcome case, the simulation section of Liu (2019). I will also elaborate on the difference in the next chapter.*

Suppose we denote $\mathbf{p} = \text{vec}(\mathbf{P})$, $\mathbf{y}^{*} = \text{vec}(\mathbf{Y}^{*})$, and $\boldsymbol{\pi}^{*} = \text{vec}(\mathbf{\Pi}^{*})$, $\boldsymbol{\epsilon}^{*} = \text{vec}(\mathbf{E}^{*})$, then we have the following proposition

Proposition 1. *If Assumption 1.2.1, 1.2.2, and 1.2.3 holds, then the incomplete information network game has a unique pure strategy BNE, given the equilibrium strategy* $\mathbf{y}^\star$ *as*

$$\mathbf{y}^* = (\mathbf{\Lambda}^{*'} \otimes \mathbf{W})\mathbf{p}^\star + \boldsymbol{\pi}^* - \boldsymbol{\epsilon}^* \tag{1.18}$$

where the vector of equilibrium beliefs $p^\star$ *is the unique solution of*

$$\mathbf{p} = \vec{h}(\mathbf{p}) \tag{1.19}$$

where

$$\vec{h}(\mathbf{p}) = [\vec{h}_1(\mathbf{p})', \vec{h}_2(\mathbf{p})', \cdots, \vec{h}_m(\mathbf{p})']'$$

and

$$\vec{h}_k(p) = [\mathrm{F}_k(u_{1k}), \mathrm{F}_k(u_{2k}), \cdots, \mathrm{F}_k(u_{nk})]'$$

where

$$u_{ik} = \sum_{l=1}^{m} \lambda^*_{lk} \sum_{j \neq i}^{n} w_{ij} p_{jl} + \pi_{ik}$$

and

$$F_k(u) = u\Phi(\frac{u}{\sigma^*_k}) + \sigma^*_k \phi(\frac{u}{\sigma^*_k}) \tag{1.20}$$

where $\Phi(\cdot)$ *and* $\phi(\cdot)$ *is C.D.F. and P.D.F. of standard normal distribution.*

Proof: As $F_k(\cdot)$ is continuous for all $k \in \{1, 2, \cdots m\}$, therefore, $\vec{h}(\cdot)$ is continuous and according to contraction mapping theorem, there at least exist one solution to $\mathbf{p} = \vec{h}(\mathbf{p})$. According to the contraction mapping theorem, the solution to $\mathbf{p} = \vec{h}(\mathbf{p})$ is unique if there exists some kind of norm that the Hessian matrix norm less than one, i.e., $||\partial\vec{h}(\mathbf{p})/\partial\mathbf{p}'|| < 1$ for some $|| \cdot ||$ then we have

$$\frac{\partial\vec{h}(\mathbf{p})}{\partial\mathbf{p}'} = \begin{bmatrix} \frac{\partial\vec{h}_1(\mathbf{p})}{\partial\mathbf{p}'_1} & \cdots & \frac{\partial\vec{h}_1(\mathbf{p})}{\partial\mathbf{p}'_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial\vec{h}_m(\mathbf{p})}{\partial\mathbf{p}'_1} & \cdots & \frac{\partial\vec{h}_m(\mathbf{p})}{\partial\mathbf{p}'_m} \end{bmatrix}$$

then

$$\begin{aligned} \frac{F_k(u_{ik})}{\partial p_{jl}} &= \lambda^*_{lk}[w_{ij}\Phi(\frac{u_{ik}}{\sigma^*_k}) + \frac{w_{ij}}{\sigma^*_k} u_{ik}\phi(\frac{u_{ik}}{\sigma^*_k}) - \sigma^*_k w_{ij}\frac{u_{ik}}{\sigma^*_k}\frac{1}{\sigma^*_k}\phi(\frac{u_{ik}}{\sigma^*_k})] \\ &= \lambda^*_{lk} w_{ij}\Phi(\frac{u_{ik}}{\sigma^*_k}) \\ &\leq \lambda^*_{lk} w_{ij} \end{aligned}$$

It follows

$$\begin{aligned}||\frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'}||_\infty &\leq \max_{k=1,\cdots m}\sum_{l=1}^{m}|\lambda^*_{lk}|\max_{i=1,\cdots n}\sum_{j=1}^{n}|w_{ij}| = ||\mathbf{\Lambda}^*||_1||\mathbf{W}||_\infty \\ ||\frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'}||_1 &\leq \max_{l=1,\cdots m}\sum_{l=1}^{m}|\lambda^*_{lk}|\max_{j=1,\cdots n}\sum_{i=1}^{n}|w_{ij}| = ||\mathbf{\Lambda}^*||_\infty||\mathbf{W}||_1\end{aligned}$$

When the assumption 1.2.1, assumption 1.2.2, and assumption 1.2.3 hold, the contraction mapping property of $\mathbf{p} = \vec{h}(\mathbf{p})$ is ensured.

1.3 Econometric Model

Suppose $\pi_{ik} = \mathbf{x}_i'\boldsymbol{\beta}_k$, $\boldsymbol{\pi}_k = \mathbf{X}\boldsymbol{\beta}_k$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ is an $n \times q$ matrix, representing exogenous variables; and $\boldsymbol{\beta}_k$ is unknown q-dimension parameter in the model, then we can consider a simultaneous-equation model with a censored outcome variable $y_{ik} = y^*_{ik} \cdot \mathrm{I}(y^*_{ik} > 0)$, $p_{ik} = \mathrm{E}(y_{ik})$, where

$$y^*_{ik} = -\sum_{l=1,l\neq k}^{m}\theta_{lk}y^*_{il} + \sum_{l=1}^{m}\lambda_{lk}\sum_{j=1,j\neq i}^{n}w_{ij}p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}_k - \epsilon_{ik}, \tag{1.21}$$

for $i = 1, \cdots, n$ and $k = 1, \cdots, m$. Let $\mathbf{Y}^* = [\mathbf{y}^*_1, \mathbf{y}^*_2, \cdots, \mathbf{y}^*_m]$, $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_m]$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]'$, $\mathbf{E} = [\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_m]$, and $\mathbf{W} = [w_{ij}]$, where $\mathbf{y}^*_k = (y^*_{1k}, y^*_{2k}, \cdots, y^*_{nk})'$, $\mathbf{p}_k = (p_{1k}, \cdots, p_{nk})'$, and $\boldsymbol{\epsilon}_k = (\epsilon_{1k}, \epsilon_{2k}, \cdots, \epsilon_{nk})'$. Let $\boldsymbol{\Theta} = [\theta_{kl}]$, an $(m \times m)$-dimension matrix that reflects the inner effect cross activities of the same agent, where $\theta_{kk} = 1$ for all k, $\boldsymbol{\Lambda} = [\lambda_{kl}]$, an $(m \times m)$-dimension matrix that reflects the peer effects among agents and activities; and $\mathbf{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_m]$, which reflects the direct effects. In matrix form, Equation (1.21) can be written as

$$\mathbf{Y}^*\boldsymbol{\Theta} = \mathbf{WP\Lambda} + \mathbf{XB} - \mathbf{E}. \tag{1.22}$$

Based on our Assumption 1.2.2, we can extend our assumption that $\mathrm{vec}(\mathbf{E}) \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$. If $\boldsymbol{\Theta}$ is nonsingular, the reduced form of Equation (1.22) is

$$\mathbf{Y}^* = \mathbf{WP\Lambda}^* + \mathbf{XB}^* - \mathbf{E}^*, \tag{1.23}$$

where $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda\Theta}^{-1}$, $\mathbf{B}^* = \mathbf{B}\boldsymbol{\Theta}^{-1}$, and $\mathbf{E}^* = \mathbf{E}\boldsymbol{\Theta}^{-1}$. As $\mathrm{vec}(\mathbf{E}) \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, we have $\mathrm{vec}(\mathbf{E}^*) \sim N(\mathbf{0}, \boldsymbol{\Sigma}^* \otimes \mathbf{I}_n)$ where $\boldsymbol{\Sigma}^* = \boldsymbol{\Theta}'^{-1}\boldsymbol{\Sigma}\boldsymbol{\Theta}^{-1}$. Let the (k, l)-th element of $\boldsymbol{\Sigma}^*$ is $\rho^*_{kl}\sigma^*_k\sigma^*_l$, with $\rho^*_{kk} = 1$ for all $k = 1, \cdots, m$. From equation (1.23), we can derive the scalar reduced form of the econometric model as

$$y_{ik}^* = \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1, j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^* - \epsilon_{ik}^*. \tag{1.24}$$

Let $d_{ik} = 1$ if $y_{ik}^* > 0$. Then,

$$\Pr(d_{ik} = 1) = \Pr(y_{ik}^* > 0) = \Pr(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1, j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^* > \epsilon_{ik}^*) = \Phi_{ik},$$

where $\Phi_{ik} = \Phi[(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1, j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^*)/\sigma_k^*]$. And $\Phi(\cdot)$ is the cdf of standard normal distribution

$$\begin{aligned} p_{ik} &\equiv \mathrm{E}(y_{ik}) = \mathrm{E}[\mathrm{E}(y_{ik}|d_{ik})] \\ &= \mathrm{E}(y_{ik}|d_{ik} = 1) \Pr(d_{ik} = 1) + \mathrm{E}(y_{ik}|d_{ik} = 0) \Pr(d_{ik} = 0) \\ &= \mathrm{E}(y_{ik}^*|d_{ik} = 1) \Pr(d_{ik} = 1) \\ &= (\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1, j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^*) \Phi_{ik} + \sigma_k^* \phi_{ik} \end{aligned}$$

where $\phi_{ik} = \phi[(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1, j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^*)/\sigma_k^*]$. Let us consider the identification of the reduced form parameter first, i.e., $\boldsymbol{\Lambda}^* = [\lambda_{kl}^*]_{k=1,l=1}^{m,m}$, $\mathbf{B}^* = [\boldsymbol{\beta}_k^*]_{k=1}^{m}$, and $\{\sigma_k^*\}_{k=1}^{m}$. As the network agent connection matrix $\mathbf{W}$ and agents' characteristics matrix $\mathbf{X}$ are given, exogenous, and observable, we propose the case that the parameter group $\left(\boldsymbol{\Lambda}^*, \mathbf{B}^*, \{\sigma_k^*\}_{k=1}^{m}\right)$ and $\left(\widetilde{\boldsymbol{\Lambda}}^*, \widetilde{\mathbf{B}}^*, \{\widetilde{\sigma}_k^*\}_{k=1}^{m}\right)$ are observational equivalent if

$$(\sum_{l=1}^{m} \widetilde{\lambda}_{lk}^* \sum_{j=1, j\neq i}^{n} w_{ij} \widetilde{p}_{jl} + \mathbf{x}_i' \widetilde{\boldsymbol{\beta}}_k^*) \widetilde{\Phi}_{ik} + \widetilde{\sigma}_k^* \widetilde{\phi}_{ik} = (\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1, j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^*) \Phi_{ik} + \sigma_k^* \phi_{ik} \tag{1.25}$$

for all $i = 1, \cdots, n$ and $k = 1, \cdots, m$. Based on our Assumption 1.2.1, 1.2.2, and 1.2.3, p_{ik} and $\widetilde{p}_{ik}$ are fixed point solutions to the contraction mapping, which should be identical to each other due to the uniqueness. $p_{ik} = \widetilde{p}_{ik}$ for all $i = 1, \cdots, n$ and $k = 1, \cdots, m$. Therefore we have

$$\begin{aligned} &[\mathbf{WP}, \mathbf{X}][\boldsymbol{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]' \odot \boldsymbol{\Phi}([\mathbf{WP}, \mathbf{X}][\boldsymbol{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]' \mathbf{D}_m^{*-1}) + \phi([\mathbf{WP}, \mathbf{X}][\boldsymbol{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]' \mathbf{D}_m^{*-1}) \mathbf{D}_m^* \\ =&[\mathbf{WP}, \mathbf{X}][\widetilde{\boldsymbol{\Lambda}}^{*\prime}, \widetilde{\mathbf{B}}^{*\prime}]' \odot \boldsymbol{\Phi}([\mathbf{WP}, \mathbf{X}][\widetilde{\boldsymbol{\Lambda}}^{*\prime}, \widetilde{\mathbf{B}}^{*\prime}]' \widetilde{\mathbf{D}}_m^{*-1}) + \phi([\mathbf{WP}, \mathbf{X}][\widetilde{\boldsymbol{\Lambda}}^{*\prime}, \widetilde{\mathbf{B}}^{*\prime}]' \widetilde{\mathbf{D}}_m^{*-1}) \widetilde{\mathbf{D}}_m^* \end{aligned} \tag{1.26}$$

where $\odot$ is the Hadamard (Schur) product of matrices explained in section 7.5 of Horn and Johnson (2012). $\mathbf{D}_m^*$ and $\widetilde{\mathbf{D}}_m^*$ are diagonal matrices with diagonal elements $\{\sigma_k^{*2}\}_{k=1}^{m}$, $\{\widetilde{\sigma}_k^{*2}\}_{k=1}^{m}$, and other elements are zeros. Then we can derive the scalar form as

$$(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}\widetilde{p}_{jl}+\mathbf{x}_i\widetilde{\beta}_k^{*})+\widetilde{\sigma}_k^{*}\frac{\widetilde{\phi}_{ik}}{\widetilde{\Phi}_{ik}}=(\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}p_{jl}+\mathbf{x}_i\beta_k^{*})+\sigma_k^{*}\frac{\phi_{ik}}{\Phi_{ik}} \tag{1.27}$$

Suppose we denote $\mathbf{L}=[\phi_{ik}/\Phi_{ik}]_{i=1,k=1}^{n,m}$ (this is non-random because $\mathbf{X}$, $\mathbf{W}$, and other parameters are predetermined and non-random), then if the matrix $[\mathbf{WP},\mathbf{X},\mathbf{L}]$ is full column rank, the reduced form parameters $[\mathbf{\Lambda}^*,\mathbf{B}^*,\mathbf{D}_m^*]$ are identifiable. As

$$[\mathbf{WP},\mathbf{X},\mathbf{L}]\Big([\mathbf{\Lambda}^*,\mathbf{B}^*,\mathbf{D}_m^*]-[\widetilde{\mathbf{\Lambda}}^*,\widetilde{\mathbf{B}}^*,\widetilde{\mathbf{D}}_m^*]\Big)=\mathbf{0} \tag{1.28}$$

the scalar form results are

$$\begin{aligned}(\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}p_{jl}+\mathbf{x}_i\beta_k^{*})&=(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}\widetilde{p}_{jl}+\mathbf{x}_i\widetilde{\beta}_k^{*})\\ \sigma_k^{*2}&=\widetilde{\sigma}_k^{*2}\end{aligned} \tag{1.29}$$

And in matrix form, it will be

$$\begin{aligned}\mathbf{WP\Lambda}^*+\mathbf{XB}^*&=\mathbf{WP}\widetilde{\mathbf{\Lambda}}^*+\mathbf{X}\widetilde{\mathbf{B}}^*\\ \mathbf{D}_m^*&=\widetilde{\mathbf{D}}_m^*\end{aligned} \tag{1.30}$$

If $[\mathbf{WP},\mathbf{X},\mathbf{L}]$ has full column rank, the observational equivalence of $\left[\mathbf{\Lambda}^{*\prime},\mathbf{B}^{*\prime},\mathbf{D}_m^*\right]'$ and $\left[\widetilde{\mathbf{\Lambda}}^{*\prime},\widetilde{\mathbf{B}}^{*\prime},\widetilde{\mathbf{D}}_m^{*\prime}\right]$ implies that

$$\left[\mathbf{\Lambda}^{*\prime},\mathbf{B}^{*\prime},\mathbf{D}_m^*\right]'=\left[\widetilde{\mathbf{\Lambda}}^{*\prime},\widetilde{\mathbf{B}}^{*\prime},\widetilde{\mathbf{D}}_m^*\right]' \tag{1.31}$$

means that the reduced form parameters can be identified. Therefore, the following assumption is essential to our econometric model identification.

Assumption 1.3.1. $[\mathbf{WP},\mathbf{X},\mathbf{L}]$ *has full column rank.*

When the Assumption 1.3.1 holds, we will have the sufficient conditions for the identification of the reduced form parameters $\left[\mathbf{\Lambda}^{*\prime},\mathbf{B}^{*\prime},\mathbf{D}_m^*\right]'$, where $\mathbf{\Lambda}^*=\mathbf{\Lambda\Theta}^{-1}$, $\mathbf{B}=\mathbf{B\Theta}^{-1}$, and $\mathbf{D}_m^*$ is an $m\times m$ diagonal matrix whose diagonal elements are $\mathbf{\Theta}'^{-1}\mathbf{\Sigma\Theta}^{-1}$ and zeros in non-diagonal positions. To figure out the sufficient condition for the identification of structural form parameters, we need to propose some constraints to the structural form matrix $\mathbf{\Gamma} = [\mathbf{\Theta}', -\mathbf{\Lambda}', -\mathbf{B}']'$. Suppose $\boldsymbol{\gamma}_k$ is the k-th column of $\mathbf{\Gamma}$, and $\mathbf{R}_k$ is the matrix for the constraint that $\mathbf{R}_k\boldsymbol{\gamma}_k = 0$ and $\text{rank}(\mathbf{R}_k\mathbf{\Gamma}) = m-1$ for $k = 1, \cdots, m-1$, which is the sufficient rank condition to identify structural parameters from reduced form parameters by Schmidt (1976).

Assumption 1.3.2. *Let $\mathbf{\Gamma} = [\mathbf{\Theta}', -\mathbf{\Lambda}', -\mathbf{B}']'$, and $\boldsymbol{\gamma}_k$ is the k-th column of $\mathbf{\Gamma}$, and the $\mathbf{R}_k$ is the matrix for constraints $\mathbf{R}_k\boldsymbol{\gamma}_k = \mathbf{0}$, and rank$(\mathbf{R}_k\mathbf{\Gamma}) = m-1$ for $k = 1, \cdots, m$.*

Remark 5. *An easier way to understand this assumption is suppose a two activity case, and the dimension of $\mathbf{x}_i$ is two, i.e., $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{21})'$ and $\boldsymbol{\beta}_2 = (\beta_{12}, \beta_{22})'$. Even we can estimate all reduced form parameters, i.e., λ^*_{11}, λ^*_{12}, λ^*_{21}, λ^*_{22}, β^*_{11}, β^*_{12}, β^*_{21}, β^*_{22}. However, the number of structural parameter we need to estimate is ten – all λ, β, and θ_{12}, θ_{21}. This makes us use eight equations to solve ten unknowns. The derivation process will be unidentified. Therefore, the constraints to structural parameter space are necessary. In this, one constraint is $\beta_{12} = \beta_{21} = 0$, this is also applied in our Monte Carlo Simulation.*

Under all the assumptions above, we can derive a two-stage estimation process of the econometric model's structural parameters in the next section.

1.4 Estimation

After the discussion about the econometric model identification, we derive the estimation process. There are two steps of the estimation. The first step is to estimate the reduced form parameters $\mathbf{\Lambda}^*$, $\mathbf{B}^*$, and $(\sigma^*_1, \cdots, \sigma^*_m)$ by nested pseudo-likelihood (NPL) algorithm (we use "pseudo" here because, in each maximum likelihood estimation iteration, we cannot use the true value of $\mathbf{p}$ to estimate parameters. We calculate the $\mathbf{p}$ by parameters estimated from previous iteration, this is how our estimation differs from a typical maximum likelihood estimation), which is discussed in Aguirregabiria and Mira (2007). NPL is also applied in Lin and Xu (2017) in large network games and adopted in Liu (2019) for multi-activity network games with discrete outcomes. Suppose we denote $\mathbf{\Psi}^* = [\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}, (\sigma^*_1, \cdots, \sigma^*_m)']'$ and at $t = 0$ the NPL starts from an initial vector

$\mathbf{p}^{(0)} \in [0,1]^{mn}$ and conduct the following iterative steps:

Step 1 Given $\mathbf{p}^{(t-1)}$, obtain $\hat{\boldsymbol{\psi}}_k^{*(t)} = (\hat{\lambda}_{1k}^{*(t)}, \cdots, \hat{\lambda}_{mk}^{*(t)}, \boldsymbol{\beta}_k^{*(t)'}, \sigma_k^{*(t)})' = \arg\max \ln L(\boldsymbol{\psi}_k^*; \mathbf{p}^{(t-1)})$ where

$$\begin{aligned}
\ln L(\boldsymbol{\psi}_k^*; \mathrm{p}^{(t-1)}) &= \sum_{i=1}^{n} d_{ik} \ln\Big(\frac{1}{\sigma_k^*}\phi\Big(\frac{y_{ik} - u_{ik}^{(t-1)}}{\sigma_k^*}\Big)\Big) + (1-d_{ik})\ln\Big(1 - \Phi\Big(\frac{u_{ik}^{(t-1)}}{\sigma_k^*}\Big)\Big) \\
&= \sum_{i=1}^{n}\Big\{d_{ik}\Big[-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma_k^{*2} - \frac{1}{2\sigma_k^{*2}}\Big(y_{ik} - u_{ik}^{(t-1)}\Big)^2\Big] \\
&\quad +(1-d_{ik})\Big[1 - \Phi\Big(\frac{u_{ik}^{(t-1)}}{\sigma_k^*}\Big)\Big]\Big\}
\end{aligned}$$

where

$$u_{ik}^{(t-1)} = \sum_{l=1}^{m}\lambda_{lk}^*\sum_{j\neq i}^{n} w_{ij}p_{jl}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_k^*$$

for $k \in \{1, 2, \cdots, m\}$

Step 2 Given $\hat{\boldsymbol{\Psi}}^{*(t)} = [\hat{\boldsymbol{\psi}}_1^{*(t)}, \cdots, \hat{\boldsymbol{\psi}}_m^{*(t)}]$, obtain $\mathbf{p}^{(t)} = \vec{h}(\mathbf{p}^{(t-1)}; \hat{\boldsymbol{\Psi}}^{*(t)})$, where

$$\vec{h}(\mathbf{p}^{(t-1)}; \hat{\boldsymbol{\Psi}}^{*(t)}) = [\vec{h}_1(\mathbf{p}^{(t-1)}; \hat{\boldsymbol{\Psi}}^{*(t)})', \cdots, \vec{h}_m(\mathbf{p}^{(t-1)}; \hat{\boldsymbol{\Psi}}^{*(t)})']'$$

with

$$\begin{aligned}
&\overrightarrow{h}_k(\mathbf{p}^{(t-1)}; \hat{\Psi}^{*(t)}) \\
&= \begin{bmatrix}
\Big(\sum_{l=1}^m \hat{\lambda}_{1k}^{*(t)} \sum_{j=1}^n w_{1j}p_{jl}^{(t-1)} + x_1'\hat{\beta}_k^{*(t)'}\Big)\Phi\Big(\frac{\sum_{l=1}^m \hat{\lambda}_{1k}^{*(t)}\sum_{j=1}^n w_{1j}p_{jl}^{(t-1)} + x_1'\hat{\beta}_k^{*(t)'}}{\sigma_k^{*(t)}}\Big) \\
+\sigma_k^{*(t)}\phi\Big(\frac{\sum_{l=1}^m \hat{\lambda}_{1k}^{*(t)}\sum_{j=1}^n w_{1j}p_{jl}^{(t-1)} + x_1'\hat{\beta}_k^{*(t)'}}{\sigma_k^{*(t)}}\Big) \\
\vdots \\
\Big(\sum_{l=1}^m \hat{\lambda}_{1k}^{*(t)} \sum_{j=1}^n w_{nj}p_{jl}^{(t-1)} + x_n'\hat{\beta}_k^{*(t)'}\Big)\Phi\Big(\frac{\sum_{l=1}^m \hat{\lambda}_{1k}^{*(t)}\sum_{j=1}^n w_{nj}p_{jl}^{(t-1)} + x_n'\hat{\beta}_k^{*(t)'}}{\sigma_k^{*(t)}}\Big) \\
+\sigma_k^{*(t)}\phi(\frac{\sum_{l=1}^m \hat{\lambda}_{1k}^{*(t)}\sum_{j=1}^n w_{nj}p_{jl}^{(t-1)} + x_n'\hat{\beta}_k^{*(t)'}}{\sigma_k^{*(t)}})
\end{bmatrix}
\end{aligned}$$

for $k \in \{1, 2, \cdots, m\}$. Update $\mathbf{p}^{(t-1)}$ in **Step 1** to $\mathbf{p}^{(t)}$. Repeat these two steps until the estimation results of the reduced-form parameters converge.

The contraction mapping (1.19) with a fixed point is an important determinant of the convergence of the nested pseudo-likelihood algorithm (Kasahara and Shimotsu (2012)). The contraction property of the mapping $\mathbf{p} = \vec{h}(\mathbf{p})$ is ensured by assumption 1.2.1, assumption 1.2.2, and assumption 1.2.3. The NPL estimator is characterized by $\hat{\boldsymbol{\psi}}_k^* = \arg\max \ln L(\boldsymbol{\psi}_k^*, \widehat{\mathbf{p}})$ when NPL algorithm converges, where $\widehat{\mathbf{p}}$ is implicitly calculated through $\widehat{\mathbf{p}} = \vec{h}(\widehat{\mathbf{p}}; \widehat{\boldsymbol{\Psi}}^*)$. The details on the asymptotic distribution of the NPL algorithm estimator are in the Appendix.

Suppose we denote all the regressing variables as $\mathbf{Z} = [\mathbf{WP}, \mathbf{X}]$, then we can rewrite our reduced form model as following

$$\mathbf{Y}^* = \mathbf{Z}[\boldsymbol{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]' - \mathbf{E}^* = \mathbf{Z}\boldsymbol{\Psi} - \mathbf{E}^* \tag{1.32}$$

After the estimation of reduced form parameters $\boldsymbol{\Lambda}^*$, $\mathbf{B}^*$, and based on the rank condition of constraints of the structural form parameters, we can estimate the reduced form parameters $\boldsymbol{\Theta}$, $\boldsymbol{\Lambda}$, and $\mathbf{B}$ by the Amemiya Generalized Least Square (ALGS) estimation in Amemiya (1974) and Amemiya (1979) for simultaneous-equation Tobit model situation, and Amemiya (1978), Lee (1981), and Liu (2019) for simultaneous-equation Probit model situation. After the estimation of the structural form $\boldsymbol{\Theta}$, reduced form $\{\sigma_k^{*2}\}_{k=1}^m$ and $\{\rho_{kl}^*\}_{k=1,l=1,k\neq l}^{m,m}$, we can estimate the structural form parameters for random shock among agents and activities, i.e., $\{\sigma_k^2\}_{k=1}^m$ and $\{\rho_{kl}\}_{k=1,l=1,k\neq l}^{m,m}$. According to the constraints on the simultaneous effect matrix, we can derive the following model for $\mathbf{y}_1 = (y_{11}, \cdots, y_{n1})'$ as

$$\mathbf{y}_1^* = -\mathbf{Y}_1^*\boldsymbol{\theta}_1 + \mathbf{Z}_1\boldsymbol{\psi}_{\mathbf{1}} - \boldsymbol{\epsilon}_{\mathbf{1}} \tag{1.33}$$

where $\boldsymbol{\Psi} = [\boldsymbol{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]'$, and $\boldsymbol{\psi}_1$ is the first column of $\boldsymbol{\Psi}$. $\boldsymbol{\theta}_1$ is the first column of $\boldsymbol{\Theta}$. And $\mathbf{Y}_1^* = (\mathbf{y}_2^*, \cdots, \mathbf{y}_m^*)$ as we introduce selection matrix $\mathbf{Y}_1 = \mathbf{Y}\mathbf{J}_{Y_1}$, and $\mathbf{Z}_1 = \mathbf{Z}\mathbf{J}_{Z_1}$. Then, the reduced-form model can be written as

$$\mathbf{y}_1^* = -\mathbf{Y}^*\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 + \mathbf{Z}\mathbf{J}_{Z_1}\boldsymbol{\psi}_{\mathbf{1}} - \boldsymbol{\epsilon}_{\mathbf{1}} \tag{1.34}$$

when we combine this form and the original reduced-form result, we can get

$$\begin{aligned} \mathbf{y}_1^* =& -(\mathbf{Z}\boldsymbol{\Psi} - \mathbf{E}^*)\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 + \mathbf{Z}\mathbf{J}_{Z_1}\boldsymbol{\psi}_{\mathbf{1}} - \boldsymbol{\epsilon}_{\mathbf{1}} \\ =& -\mathbf{Z}(\boldsymbol{\Psi}\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 - \mathbf{J}_{Z_1}\boldsymbol{\psi}_{\mathbf{1}}) + \mathbf{E}^*\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 - \boldsymbol{\epsilon}_{\mathbf{1}} \end{aligned} \tag{1.35}$$

then we can derive the relation between the first column of the structural form parameter $\mathbf{\Psi}$, i.e., $\boldsymbol{\psi}_1$ and the first column of the reduced-form parameter $\mathbf{\Psi}^*$, i.e., $\boldsymbol{\psi}_1^*$, as following

$$\boldsymbol{\psi}_1^* = -\mathbf{\Psi}^* \mathbf{J}_{Y_1} \boldsymbol{\theta}_1 + \mathbf{J}_{Z_1} \boldsymbol{\psi}_{\mathbf{1}} \tag{1.36}$$

and the regression equation is

$$\widehat{\boldsymbol{\psi}}_1^* = \widehat{\mathbf{\Psi}}^* \mathbf{J}_{Y_1} \boldsymbol{\theta}_1 + \mathbf{J}_{Z_1} \boldsymbol{\psi}_{\mathbf{1}} + \mathbf{v}_1 \tag{1.37}$$

and the regression-error item is

$$\mathbf{v}_1 = (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) + (\widehat{\mathbf{\Psi}}^* - \mathbf{\Psi}^*) \mathbf{J}_{Y_1} \boldsymbol{\theta}_1 \tag{1.38}$$

If the asymptotic variance-covariance matrix of $\mathbf{v}_1$ is $\mathbf{\Omega}_{11}$, and $\widehat{\mathbf{\Omega}}_{11}$ is the estimator. Then the estimator of $(\boldsymbol{\theta}_1', \boldsymbol{\psi}_1^*)$ is

$$(\widehat{\boldsymbol{\theta}}_1', \widehat{\boldsymbol{\psi}}_1^*) = (\widehat{\mathbf{H}}_1' \widehat{\mathbf{\Omega}}_{11}^{-1} \widehat{\mathbf{H}}_1)^{-1} \widehat{\mathbf{H}}_1' \widehat{\mathbf{\Omega}}_{11}^{-1} \widehat{\boldsymbol{\psi}}_1^* \tag{1.39}$$

where $\widehat{\mathbf{H}}_1' = [-\widehat{\boldsymbol{\psi}}_1^* \mathbf{J}_{Y_1}, \mathbf{J}_{Z_1}]$. The estimation of the m-activity equation system can be derived by the same procedure. The detailed deriving process is in the Appendix.

1.5 Monte Carlo Simulation

1.5.1 Simulation Setup

We will simulate the performance of a finite sample based on the following two-equation model

$$\begin{aligned} \mathbf{y}_1^* &= -\theta_{21} \mathbf{y}_2^* + \lambda_{11} \mathbf{W} \mathbf{p}_1 + \lambda_{21} \mathbf{W} \mathbf{p}_2 + \mathbf{X}_1 \beta_1 - \epsilon_1 \\ \mathbf{y}_2^* &= -\theta_{12} \mathbf{y}_1^* + \lambda_{12} \mathbf{W} \mathbf{p}_1 + \lambda_{22} \mathbf{W} \mathbf{p}_2 + \mathbf{X}_2 \beta_2 - \epsilon_2 \end{aligned} \tag{1.40}$$

Then the matrix form of structural parameters are

$$\mathbf{\Theta} = \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix} \tag{1.41}$$

If we denote $\mathbf{Y}^* = [\mathbf{y}_1^*, \mathbf{y}_2^*]$, $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2]$, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, and $\mathbf{E} = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2]$, we can derive the matrix form of the two-equation model as

$$\mathbf{Y}^* \mathbf{\Theta} = \mathbf{W} \mathbf{P} \mathbf{\Lambda} + \mathbf{X} \mathbf{B} - \mathbf{E}. \tag{1.42}$$

then we can derive the reduced form

$$\mathbf{Y}^* = \mathbf{WP}\mathbf{\Lambda}^* + \mathbf{XB}^* - \mathbf{E}^*. \tag{1.43}$$

the matrix form of reduced-form parameters are

$$\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{\Theta}^{-1} = \begin{bmatrix} \lambda_{11}^* & \lambda_{12}^* \\ \lambda_{21}^* & \lambda_{22}^* \end{bmatrix} \qquad \mathbf{B}^* = \mathbf{B}\mathbf{\Theta}^{-1} = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix} \tag{1.44}$$

and the relation between reduced-form parameters and structural parameters can be expressed as

$$\begin{aligned} &\theta_{21} = -\frac{\beta_{12}^*}{\beta_{11}^*} \qquad \theta_{12} = -\frac{\beta_{21}^*}{\beta_{22}^*} \qquad \beta_1 = \beta_{11}^* - \frac{\beta_{12}^*\beta_{21}^*}{\beta_{22}^*} \qquad \beta_2 = \beta_{22}^* - \frac{\beta_{12}^*\beta_{21}^*}{\beta_{11}^*} \\ &\lambda_{11} = \lambda_{11}^* - \frac{\beta_{21}^*}{\beta_{22}^*}\lambda_{12}^* \qquad \lambda_{12} = \lambda_{12}^* - \frac{\beta_{12}^*}{\beta_{11}^*}\lambda_{11}^* \qquad \lambda_{21} = \lambda_{21}^* - \frac{\beta_{21}^*}{\beta_{22}^*}\lambda_{22}^* \qquad \lambda_{22} = \lambda_{22}^* - \frac{\beta_{12}^*}{\beta_{11}^*}\lambda_{21}^* \end{aligned} \tag{1.45}$$

the equation (1.45) gives an arithmetic method that reflects the relation between reduced-form parameters and structural-form parameters. That means if we know the true value of the reduced-form parameters, we can use this arithmetic method to calculate the true value of structural-form parameters. However, as we can only estimate the reduced-form parameters in our estimation and simulation, we need the following discussion of our algorithm estimator to show the relation between the reduced-form estimator and the structural-form estimator. All these details are included in the Appendix. As for the random shock $\text{vec}(\mathbf{E}^*) \sim N(\mathbf{0}, \mathbf{\Sigma}^* \otimes \mathbf{I_n})$ where

$$\mathbf{\Sigma}^* = \begin{bmatrix} \sigma_1^{*2} & \rho^*\sigma_1^*\sigma_2^* \\ \rho^*\sigma_1^*\sigma_2^* & \sigma_2^{*2} \end{bmatrix} \tag{1.46}$$

that means when we get the estimation results of reduced form parameters, then we can get all structural parameters. The vector form of the reduced-form model can be written as

$$\begin{aligned} \mathbf{y}_1^* &= \lambda_{11}^*\mathbf{W}\mathbf{p}_1 + \lambda_{21}^*\mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_{11}^* + \mathbf{X}_2\beta_{21}^* - \epsilon_1^* \\ \mathbf{y}_2^* &= \lambda_{12}^*\mathbf{W}\mathbf{p}_1 + \lambda_{22}^*\mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_{12}^* + \mathbf{X}_2\beta_{22}^* - \epsilon_2^* \end{aligned} \tag{1.47}$$

Note: As we will use the unconstrained toolbox during the optimization, therefore, both parameters and $(\rho^*, \sigma_1^*, \sigma_2^*)$ should be searched through the whole real line, however, we know that what we want to estimate $(\rho^*, \sigma_1^*, \sigma_2^*)$ are correlation and standard deviation, which means that

$\rho^* \in (-1, 1)$, $\sigma_1^* \in (0, +\infty)$, and $\sigma_2^* \in (0, +\infty)$, therefore, we introduce the following transformation during the optimization

$$\sigma_1^* = f(a_1) = e^{a_1} \qquad \sigma_2^* = f(a_2) = e^{a_2} \qquad \rho^* = g(a_3) = 1 - \frac{2}{1 + e^{a_3}} \tag{1.48}$$

after this transformation, we can search a_1, a_2, and a_3 all along the whole real line. And for any function, the first-order partial derivative with respect to a_1, a_2, and a_3 can be written as the first-order partial derivative with respect to σ_1^*, σ_2^*, ρ^*.

$$\begin{aligned} \frac{\partial \ln L}{\partial a_1} &= \frac{\partial \ln L}{\partial \sigma_1^*}\frac{\partial \sigma_1^*}{\partial a_1} = \sigma_1^* \frac{\partial \ln L}{\partial \sigma_1^*} \\ \frac{\partial \ln L}{\partial a_2} &= \frac{\partial \ln L}{\partial \sigma_2^*}\frac{\partial \sigma_2^*}{\partial a_2} = \sigma_2^* \frac{\partial \ln L}{\partial \sigma_2^*} \\ \frac{\partial \ln L}{\partial a_3} &= \frac{\partial \ln L}{\partial \rho^*}\frac{\partial \rho^*}{\partial a_3} = \rho^* \frac{\partial \ln L}{\partial \rho^*} \end{aligned} \tag{1.49}$$

The steps of the algorithm are the following.

• At step 0, suppose we initialize $\mathbf{p}_1^{(0)} = \mathbf{p}_2^{(0)} \in [0, 1]^n$, and initialize $\lambda_{pq}^{*(0)} = \beta_{pq}^{*(0)} = 0.1$,

• At step $t \geq 1$ and we define $d_{i1} = \mathrm{I}(y_{i1} > 0)$, $d_{i2} = \mathrm{I}(y_{i2} > 0)$, then we can write the log-likelihood as

$$\begin{aligned} \ln L_1 = \sum_{i=1}^{n} \Big[& d_{i1} \ln \Big[\frac{1}{\sigma_1^*} \phi\Big(\frac{\lambda_{11}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{21}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^* + x_{i2}\beta_{21}^* - y_{i1}}{\sigma_1^*}\Big)\Big] \\ & + (1 - d_{i1}) \ln \Big[1 - \Phi\Big(\frac{\lambda_{11}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{21}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^* + x_{i2}\beta_{21}^*}{\sigma_1^*}\Big)\Big]\Big] \\ \ln L_2 = \sum_{i=1}^{n} \Big[& d_{i2} \ln [\frac{1}{\sigma_2^*} \phi\Big(\frac{\lambda_{12}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{22}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^* + x_{i2}\beta_{22}^* - y_{i2}}{\sigma_2^*}\Big)] \\ & + (1 - d_{i2}) \ln \Big[1 - \Phi\Big(\frac{\lambda_{12}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{22}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^* + x_{i2}\beta_{22}^*}{\sigma_2^*}\Big)\Big]\Big] \end{aligned} \tag{1.50}$$

the results are

$$\begin{aligned} (\lambda_{11}^{*(t)}, \lambda_{21}^{*(t)}, \beta_{11}^{*(t)}, \beta_{21}^{*(t)}, \sigma_1^{*2(t)}) &= \arg\max \ln L_1(\lambda_{11}^*, \lambda_{21}^*, \beta_{11}^*, \beta_{21}^*, \sigma_1^*) \\ (\lambda_{12}^{*(t)}, \lambda_{22}^{*(t)}, \beta_{12}^{*(t)}, \beta_{22}^{*(t)}, \sigma_2^{*2(t)}) &= \arg\max \ln L_2(\lambda_{12}^*, \lambda_{22}^*, \beta_{12}^*, \beta_{22}^*, \sigma_2^*) \end{aligned} \tag{1.51}$$

then we have the $\mathbf{p}_1^{(t+1)} = (p_{11}^{(t+1)}, \cdots, p_{n1})'$ and $\mathbf{p}_2^{(t+1)} = (p_{12}^{(t+1)}, \cdots, p_{n2})'$ as

$$\begin{aligned}
p_{i1}^{(t+1)} =& (\lambda_{11}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{21}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)})\Phi_{i1}^{(t)} + \sigma_1^{*(t)} \phi_{i1}^{(t)} \\
p_{i2}^{(t+1)} =& (\lambda_{12}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{22}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^{*(t)} + x_{i2}\beta_{22}^{*(t)})\Phi_{i2}^{(t)} + \sigma_2^{*(t)} \phi_{i2}^{(t)}
\end{aligned} \tag{1.52}$$

for $i = 1, \cdots, n$ and we have

$$\begin{aligned}
\Phi_{i1}^{(t)} =& \Phi\Big(\frac{\lambda_{11}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{21}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)}}{\sigma_1^*}\Big) \\
\phi_{i1}^{(t)} =& \phi\Big(\frac{\lambda_{11}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{21}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)}}{\sigma_1^*}\Big) \\
\Phi_{i2}^{(t)} =& \Phi\Big(\frac{\lambda_{12}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{22}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^{*(t)} + x_{i2}\beta_{22}^{*(t)}}{\sigma_2^*}\Big) \\
\phi_{i2}^{(t)} =& \phi\Big(\frac{\lambda_{12}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{22}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^{*(t)} + x_{i2}\beta_{22}^{*(t)}}{\sigma_2^*}\Big)
\end{aligned} \tag{1.53}$$

the same as the previous notation, $\phi(\cdot)$ and $\Phi(\cdot)$ represent the PDF and CDF of the standard normal distribution. We repeat step t and $(t+1)$ until the parameters estimated converge. Then, use the estimated parameters, respectively, $\hat{\lambda}_{11}^*, \hat{\lambda}_{12}^*, \hat{\lambda}_{21}^*, \hat{\lambda}_{22}^*, \hat{\beta}_{11}^*, \hat{\beta}_{12}^*, \hat{\beta}_{21}^*, \hat{\beta}_{22}^*$ and $\hat{\sigma}_1^*, \hat{\sigma}_2^*$, calculated the equilibrium, $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$; then use all these to estimate ρ^*, the log-likelihood function is

$$\begin{aligned}
\ln L(\rho^*) \quad = \quad & \sum_{i=1}^{n} \Big[d_{i1} d_{i2} \ln \phi_2 \Big(\frac{\hat{\lambda}^*_{11} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{21} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{11} + x_{i2} \hat{\beta}^*_{21} - y_{i1}}{\hat{\sigma}^*_1}, \\
& \frac{\hat{\lambda}^*_{12} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{22} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{12} + x_{i2} \hat{\beta}^*_{22} - y_{i2}}{\hat{\sigma}^*_2}, \hat{\sigma}^*_1, \hat{\sigma}^*_2, \rho^* \Big) \\
& + d_{i1}(1 - d_{i2}) \ln \Big[\Phi \Big(\frac{\rho^* (\hat{\lambda}^*_{11} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{21} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{11} + x_{i2} \hat{\beta}^*_{21} - y_{i1})}{\hat{\sigma}^*_1 \sqrt{1 - \rho^{*2}}} \\
& - \frac{\hat{\lambda}^*_{12} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{22} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{12} + x_{i2} \hat{\beta}^*_{22}}{\hat{\sigma}^*_2 \sqrt{1 - \rho^{*2}}} \Big) \\
& \phi \Big(\frac{\hat{\lambda}^*_{11} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{21} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{11} + x_{i2} \hat{\beta}^*_{21} - y_{i1}}{\hat{\sigma}^*_1} \Big) \Big] \\
& + (1 - d_{i1}) d_{i2} \ln \Big[\Phi \Big(\frac{\rho^* (\hat{\lambda}^*_{12} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{22} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{12} + x_{i2} \hat{\beta}^*_{22} - y_{i2})}{\hat{\sigma}^*_2 \sqrt{1 - \rho^{*2}}} \\
& - \frac{\hat{\lambda}^*_{11} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{21} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{11} + x_{i2} \hat{\beta}^*_{21}}{\hat{\sigma}^*_1 \sqrt{1 - \rho^{*2}}} \Big) \\
& \phi \Big(\frac{\hat{\lambda}^*_{12} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{22} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{12} + x_{i2} \hat{\beta}^*_{22} - y_{i2}}{\hat{\sigma}^*_2} \Big) \Big] \\
& + (1 - d_{i1})(1 - d_{i2}) \ln \Phi_2 \Big(\frac{\hat{\lambda}^*_{11} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{21} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{11} + x_{i2} \hat{\beta}^*_{21}}{\hat{\sigma}^*_1}, \\
& \frac{\hat{\lambda}^*_{12} \sum_{j=1}^n w_{ij} \hat{p}_{i1} + \hat{\lambda}^*_{22} \sum_{j=1}^n w_{ij} \hat{p}_{i2} + x_{i1} \hat{\beta}^*_{12} + x_{i2} \hat{\beta}^*_{22}}{\hat{\sigma}^*_2}, \hat{\sigma}^*_1, \hat{\sigma}^*_2, \rho^* \Big) \Big]
\end{aligned}$$

where $\phi_2(\cdot, \cdot, \hat{\sigma}^*_1, \hat{\sigma}^*_2, \rho^*)$ and $\Phi_2(\cdot, \cdot, \hat{\sigma}^*_1, \hat{\sigma}^*_2, \rho^*)$ are PDF and CDF of bivariate normal distributed random variables with variance-covariance matrix

$$\begin{bmatrix} \hat{\sigma}^{*2}_1 & \rho^* \hat{\sigma}^*_1 \hat{\sigma}^*_2 \\ \rho^* \hat{\sigma}^*_1 \hat{\sigma}^*_2 & \hat{\sigma}^{*2}_2 \end{bmatrix} \tag{1.54}$$

and the estimation result of ρ^*, i.e., $\hat{\rho}^*$ can be estimated as

$$\hat{\rho}^* = \arg\max \ln L(\rho^*) \tag{1.55}$$

through a bivariate Probit Maximum Likelihood estimation in Greene (2017). Suppose we denote $\mathbf{Z} = [\mathbf{W}\mathbf{p_1}, \mathbf{W}\mathbf{p_2}, \mathbf{X}]$, and $\boldsymbol{\psi}^*_1 = (\lambda^*_{11}, \lambda^*_{21}, \beta^*_{11}, \beta^*_{21})'$ and $\boldsymbol{\psi}^*_2 = (\lambda^*_{12}, \lambda^*_{22}, \beta^*_{12}, \beta^*_{22})'$, then we have

$$\begin{aligned} \mathbf{y}_1 =& \mathbf{Z} \boldsymbol{\psi}^*_1 - \epsilon^*_1 \\ \mathbf{y}_2 =& \mathbf{Z} \boldsymbol{\psi}^*_2 - \epsilon^*_2 \end{aligned} \tag{1.56}$$

then given the NPL algorithm, for each step, suppose the current equilibrium $\widehat{\mathbf{p}} = (\widehat{\mathbf{p}}_1', \widehat{\mathbf{p}}_2')'$, and we use $\widehat{\mathbf{Z}} = [\mathbf{W}\widehat{\mathbf{p}}_1, \mathbf{W}\widehat{\mathbf{p}}_2, \mathbf{X}]$, then we have

$$\begin{aligned} (\hat{\boldsymbol{\psi}}_1^*, \hat{\sigma}_1^{*2}) &= \arg\max \ln L(\psi_1^*, \sigma_1^{*2}; \hat{\mathbf{p}}_1) \\ (\hat{\boldsymbol{\psi}}_2^*, \hat{\sigma}_2^{*2}) &= \arg\max \ln L(\psi_2^*, \sigma_2^{*2}; \hat{\mathbf{p}}_2) \end{aligned} \tag{1.57}$$

where

$$\begin{aligned} \ln L(\boldsymbol{\psi}_1^*, \sigma_1^*; \hat{\mathbf{p}}_1) &= \sum_{i=1}^{n} d_{i1} \ln \Big[\frac{1}{\sigma_1^*}\phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_1^* - y_{i1}}{\sigma_1^*}\Big)\Big] + (1-d_{i1}) \ln \Big[1 - \Phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Big] \\ \ln L(\boldsymbol{\psi}_2^*, \sigma_2^*; \hat{\mathbf{p}}_2) &= \sum_{i=1}^{n} d_{i2} \ln \Big[\frac{1}{\sigma_2^*}\phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_2^* - y_{i2}}{\sigma_2^*}\Big)\Big] + (1-d_{i2}) \ln \Big[1 - \Phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Big] \end{aligned} \tag{1.58}$$

the first-order conditions are

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\psi}_k^*} &= \sum_{i=1}^{n} -d_{ik}\frac{(\mathbf{z}_i'\boldsymbol{\psi}_k^* - y_{ik})\mathbf{z}_i}{\sigma_k^{*2}} - (1-d_{ik})\frac{\phi_{ik}\mathbf{z}_i}{(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{1}{2}}} = 0 \\ \frac{\partial L}{\partial(\sigma_k^{*2})} &= \sum_{i=1}^{n} d_{ik}\Big[-\frac{1}{2\sigma_k^{*2}} + \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{2(\sigma_k^{*2})^2}\Big] + (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} = 0 \end{aligned} \tag{1.59}$$

where $k = \{1, 2\}$, $\Phi_{ik} = \Phi(\mathbf{z}_i'\psi_k^*/\sigma_k^*)$ and $\phi_{ik} = \phi(\mathbf{z}_i'\psi_k^*/\sigma_k^*)$ then according to the results $\mathbf{E}(y_{ik} - p_{ik}) = 0$, i.e.,

$$\begin{aligned} \mathbf{E}\big[y_{ik} - (\mathbf{z}_i'\psi_k^*)\Phi_{ik} - \sigma_k^*\phi_{ik}\big] &= 0 \\ \mathbf{E}\big[d_{ik} - \Phi_{ik}\big] &= 0 \end{aligned} \tag{1.60}$$

this can be used to simplify the first-Taylor expansion of the first-order condition, which is the approach to derive the asymptotic variance-covariance matrix of reduced-form parameters' estimator discussed in Amemiya (1973), Amemiya (1985), and Maddala (1986). Detailed steps for deriving the variance-covariance matrix will be in the Appendix.

Remark 6. *(Binary Dependent Variable Case) Liu (2019) proposes the case in which all the decision outcomes are binary. There is no need to estimate the variance of reduced form error terms σ_1^* and σ_2^*. Therefore, the first-order conditions of the NPL estimator degenerate to*

$$\frac{\partial \ln L(\hat{\psi}_k^*; \hat{\mathbf{p}})}{\partial \psi_k^*} = \sum_{i=1}^{n} \frac{[d_{ik} - \Phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)]\phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)}{\Phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)[1 - \Phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)]}\hat{\mathbf{z}}_i = 0 \tag{1.61}$$

for $k = 1, 2$. The first-order Taylor expansion of the above equation around the ψ^* can draw the following equation

$$\begin{aligned}
& \sum_{i=1}^{n} \frac{(d_{ik}-\Phi_{ik})\phi_{ik}}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i - \frac{(d_{ik}-\Phi_{ik})\phi_{ik}}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i(\mathbf{z}_i'\psi_k^*)\left[\mathbf{z_i'} + \lambda_{\mathbf{1k}}^*\mathbf{w_i}\frac{\partial \mathbf{p_1}}{\partial \psi_{\mathbf{k}}^{*\prime}} + \lambda_{\mathbf{2k}}^*\mathbf{w_i}\frac{\partial \mathbf{p_2}}{\partial \psi_{\mathbf{k}}^{*\prime}}\right](\hat{\psi}_{\mathbf{k}}^* - \psi_{\mathbf{k}}^*) \\
& -\frac{\phi_{ik}^2}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i\left[\mathbf{z_i'} + \lambda_{\mathbf{1k}}^*\mathbf{w_i}\frac{\partial \mathbf{p_1}}{\partial \psi_{\mathbf{k}}^{*\prime}} + \lambda_{\mathbf{2k}}^*\mathbf{w_i}\frac{\partial \mathbf{p_2}}{\partial \psi_{\mathbf{k}}^{*\prime}}\right](\hat{\psi}_{\mathbf{k}}^* - \psi_{\mathbf{k}}^*) \\
& +\frac{(d_{ik}-\Phi_{ik})\phi_{ik}}{\Phi_{ik}(1-\Phi_{ik})}\left[\mathbf{0}', \mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}}, \mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right](\hat{\psi}_k^* - \psi_k^*) \\
& -\frac{(d_{ik}-\Phi_{ik})\phi_{ik}^2(1-2\Phi_{ik})}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i\left[\mathbf{z}_i' + \lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right](\hat{\psi}_k^* - \psi_k^*) \\
= & \sum_{i=1}^{n} \frac{(d_{ik}-\Phi_{ik})\phi_{ik}}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i - \frac{\phi_{ik}^2}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i\left[\mathbf{z}_i' + \lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right](\hat{\psi}_k^* - \psi_k^*) \\
= & O_p(1) \qquad (1.62)
\end{aligned}$$

where $\Phi_{ik} = \Phi(\mathbf{z}_i'\psi_k^*)$ and $\phi_{ik} = \phi(\mathbf{z}_i'\psi_k^*)$

1.5.2 Simulation Results

We conduct two types of normalized network structures in our simulation. The first type is the random network, which means agent i will be randomly affected by five other agents in the network. And each of the five agents' effects is identical. In this case, that is $w_{ij} = 1/5$ if agent j can affect agent i in the network. The other type is the circular network structure. In this case, the agent i will only connect with agent $i+1$ and $i-1$, and the peer effects are the same, i.e., $w_{i,i+1} = w_{i,i-1} = 1/2$. And for agent 1, we have $w_{12} = w_{1n} = 1/2$. And for agent n, we have $w_{n1} = w_{n,n-1} = 1/2$. The network graph is similar to a big circle in which each node only connects its two neighbors. That is where the network name 'circular' comes from. The parameter of the simulation are followings, $\theta_{12} = \theta_{21} = 0.5$, $\beta_1 = \beta_2 = 1$, $n = 2000$, and rep $= 1000$, $\sigma_1^* = \sigma_2^* = 1$, $\rho_{12}^* = 0.1$.

All the detailed simulation results are available in part B of the Appendix. In the following discussion, we will focus on three typical scenarios

- Case 1: Weak peer effect $\lambda_{11} = \lambda_{22} = 0.2$, $\lambda_{12} = \lambda_{21} = 0.1$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.500 (0.031)	0.502 (0.031)	0.226 (0.065)	0.121 (0.064)	0.126 (0.066)	0.230 (0.064)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1^*$	$\hat{\sigma}_2^*$	$\hat{\rho}_{12}^*$	
1.000 (0.046)	0.999 (0.047)	0.997 (0.029)	0.998 (0.028)	0.100 (0.033)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.500 (0.031)	0.500 (0.032)	0.225 (0.041)	0.122 (0.044)	0.123 (0.044)	0.227 (0.042)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
1.001 (0.047)	1.001 (0.047)	0.997 (0.028)	0.997 (0.028)	0.100 (0.032)	

- Case 2: Medium peer effect $\lambda_{11} = \lambda_{22} = 0.5$, $\lambda_{12} = \lambda_{21} = 0.3$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.500 (0.031)	0.502 (0.031)	0.542 (0.053)	0.329 (0.056)	0.332 (0.059)	0.546 (0.053)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
1.000 (0.047)	0.999 (0.047)	0.998 (0.029)	0.996 (0.029)	0.099 (0.029)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.515 (0.034)	0.513 (0.035)	0.540 (0.030)	0.333 (0.042)	0.338 (0.042)	0.544 (0.032)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.975 (0.048)	0.974 (0.046)	0.994 (0.030)	0.995 (0.030)	0.101 (0.029)	

• Case 3: Strong peer effect $\lambda_{11} = \lambda_{22} = 0.8$, $\lambda_{12} = \lambda_{21} = 0.5$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.502 (0.034)	0.503 (0.034)	0.813 (0.045)	0.488 (0.057)	0.484 (0.057)	0.810 (0.045)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.998 (0.051)	0.997 (0.052)	0.998 (0.032)	0.998 (0.031)	0.098 (0.024)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.533 (0.044)	0.529 (0.044)	0.811 (0.017)	0.503 (0.050)	0.507 (0.049)	0.812 (0.017)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.933 (0.049)	0.929 (0.051)	0.980 (0.039)	0.980 (0.037)	0.099 (0.025)	

From the results, we could find that the random network estimation of simultaneous effect matrix parameters (θ_{12} and θ_{21}) performs better than those in circular networks. In a random network, the estimator average accuracy of simultaneous effect matrix parameters (θ_{12} and θ_{21}) is always low, and the coefficient of variation is about 6% $\sim$ 7%. However, in a circular network, the estimator average accuracy of the simultaneous effect matrix parameter can be lower than 95%, and the coefficient of variation can be over 8%.

As the peer effect strengthens, the estimator's accuracy and consistency increase. In a strong peer effect case, the estimation average accuracy of the same-activity peer effect (θ_{11} and θ_{22}) is over 98%, and the coefficient of variation is about 2% in a circular network and 4% in a random network. As for the cross-activity peer effect, the accuracy is over 95%, and the coefficient of variation is about 10%. Both perspectives show under a strong peer effect case, the estimator works better.

1.6 Conclusion

This paper presents a simultaneous equation model that captures peer effects and rational expectations within an incomplete information network, alongside a discussion on its econometric model and microeconomic foundation. We explore the necessary conditions for the existence of a Bayesian Nash Equilibrium in games played over incomplete information networks. Additionally, we develop a method for nested pseudo-likelihood estimation and examine the asymptotic distribution characteristics of this estimator. Through Monte Carlo simulations, we demonstrate the estimator's consistency in scenarios of finite sample sizes and large network dimensions. In this work, networks are considered static and not influenced by agents' random shocks or characteristics. However, in practical economic scenarios, an agent's network connections often mirror both their observable and hidden traits. This suggests that incorporating an endogenous network structure could be a valuable direction for future research. Another possible extension identified in our study is the exploration of simultaneous effects stemming from an agent's various activities intentions, and how these may relate to the outcomes of other activities. This highlights the necessity of considering option models and addresses the imminent challenge of model selection in subsequent studies.

Chapter 2

Simultaneous Equations with Limited Dependent Variables and Social Interactions

2.1 Introduction

In this chapter, we consider an extensive case from the previous chapter. Given an agent make decisions on multiple activities and the outcomes of these activities are of three types: some are fully observable by analysts, such as an investor who faces decisions on adjusting investment levels, allocating funds for consultancy in investment, and the choice to venture into a new financial market. From an observational standpoint, we can track changes in her investment—whether it increases, remains unchanged, or decreases, showcasing fully observable outcomes; others are censored (partially observable) – for the inventor's consultancy expenses, we only discern between expenditure and non-expenditure, marking it as censored; and a few are binary – her decision to enter or not enter a new market falls into a binary category. The decision-making process of the investor is not isolated; it's influenced by the interconnected incentives across the different activities and further shaped by the dynamics with competitors and collaborators in her network.

Given the economic scenario, we consider a simultaneous equation model with continuous, censored, and binary outcomes. Previous chapter focus on the case that all outcomes are censored. Liu (2019) discusses the case that all outcomes are binary. Lee (1981) discusses the general form simultaneous equation system containing different types of outcomes but without peer effect. We combine previous work and propose the simultaneous equation model with continuous and limited dependent variables and social interaction.

The rest of this chapter will be organized in this way. Section 2.2 shows how we derive the Bayesian Nash Equilibrium and the difference from chapter 1. Section 2.3 derives the econometric model and the identification condition. Section 2.4 develops the estimation based on nested pseudo likelihood algorithm. Section 2.5 is the Monte Carlo Simulation. Section 2.6 is the chapter conclusion.

2.2 Bayesian Nash Equilibrium

Given the assumption on non-singularity of simultaneous effect matrix $\mathbf{\Theta}$, we have the reduced form model

$$y^*_{ik} = \sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik} - \epsilon^*_{ik} \tag{2.1}$$

where $p_{ik} = \mathrm{E}(y_{ik})$. Without loss of the generality, let $y_{ik} = y^*_{ik}$ for $k = 1, \cdots, m_1$, $y_{ik} = y^*_{ik}\mathrm{I}(y^*_{ik} > 0)$ for $k = m_1+1, \cdots, m_1+m_2$, and $y_{ik} = \mathrm{I}(y^*_{ik} > 0)$ for $k = m_1+m_2+1, \cdots, m$. $m_3 = m-m_2-m_1$. For all fully observable decision outcome activities, i.e., $k = 1, \cdots, m_1$, we have

$$p_{ik} = \mathrm{E}(y_{ik}) = \mathrm{E}(y^*_{ik}) = \mathrm{E}\Big(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik} - \epsilon^*_{ik}\Big) = \sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik} \tag{2.2}$$

We need an assumption related to continuous-outcome activities reduced-form random shock ϵ^*_{ik} for $k = 1, \cdots, m_1$ as following

Assumption 2.2.1. *The reduced-form random shock term ϵ^*_{ik} for all continuous-outcome activities, $k = 1, \cdots, m_1$ have zero mean and variance σ^{*2}_k*

Remark 7. *For continuous-outcome activities' reduced-form random shock terms, a finite variance for each activity $k = 1, \cdots, m$ is needed for deriving the rational expectation p_{ik}. For the maximum likelihood estimation step, we denote $F_k(\cdot)$ and $f_k(\cdot)$ as the CDF and PDF of ϵ^*_{ik}.*

For all binary outcomes, i.e., $k = m_1 + m_2 + 1, \cdots, m$, a unit-variance assumption is needed.

Assumption 2.2.2. *The variance of reduced form idiosyncratic shock term $\sigma^*_k = 1$, for all binary-outcome activities, i.e., $k = m_1 + m_2 + 1, \cdots, m$.*

Remark 8. *From Maddala (1986), the reduced-form idiosyncratic vector* $(\epsilon^*_{i1}, \cdots, \epsilon^*_{im})'$ *variance covariance matrix* $\mathbf{\Sigma}^*$ *has normalized diagonal elements. This helps the parametric identification and the maximum likelihood estimation steps in the following sections.*

Then, we have

$$\begin{aligned} p_{ik} &= \mathrm{E}(y_{ik}) = \mathrm{E}(y_{ik}|y^*_{ik} > 0) \\ &= \mathbf{Pr}(y^*_{ik} > 0) = \mathrm{F}_k(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}) \end{aligned} \tag{2.3}$$

where $\mathrm{F}_k(\cdot)$ is the cumulative distribution function, i.e., CDF, of the reduced form idiosyncratic shock ϵ^*_{ik}. We need the following assumption to derive the rational expectations of censored outcome activities, i.e., $k = m_1 + 1, \cdots, m_1 + m_2$.

Assumption 2.2.3. *For all agents, the reduced form idiosyncratic shock* ϵ^*_{ik} *satisfies the i.i.d normal distribution with zero means and variance-covariance matrix* $\mathbf{\Sigma}^*$*, for censored outcome activities, i.e.,* $k = m_1 + 1, \cdots, m_1 + m_2$.

Under the i.i.d. joint normal distribution assumption, we can derive the rational expectation on decision outcomes of activities $k = m_1 + 1, \cdots, m_1 + m_2$ in the same form as the chapter 1 as following (The diagonal elements of $\mathbf{\Sigma}^*$ is σ^{*2}_k, for all $k = m_1 + 1, \cdots, m_1 + m_2$.)

$$p_{ik} = \Big(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}\Big) \Phi\Big(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}}{\sigma^*_k}\Big) + \sigma^*_k \phi\Big(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}}{\sigma^*_k}\Big) \tag{2.4}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and $\phi(\cdot)$ is the probability density function of the standard normal distribution. Then, we can get the vector form of rational expectations on decision outcomes among all activities.

$$\begin{aligned} &\mathbf{p}_k = \sum_{l=1}^{m} \lambda^*_{lk} \mathbf{W} \mathbf{p}_l + \boldsymbol{\pi}_k \qquad k = 1, \cdots, m_1 \\ &\mathbf{p}_k = \Big(\sum_{l=1}^{m} \lambda^*_{lk} \mathbf{W} \mathbf{p}_l + \boldsymbol{\pi}_k\Big) \odot \mathbf{\Phi}\Big(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \mathbf{W} \mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma^*_k}\Big) + \sigma^*_k \phi\Big(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \mathbf{W} \mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma^*_k}\Big) \\ &\qquad\qquad k = m_1 + 1, \cdots, m_1 + m_2 \\ &\mathbf{p}_k = \mathrm{F}_k\Big(\sum_{l=1}^{m} \lambda^*_{lk} \mathbf{W} \mathbf{p}_l + \boldsymbol{\pi}_k\Big) \qquad k = m_1 + m_2 + 1, \cdots, m \end{aligned} \tag{2.5}$$

We apply the vec$(\cdot)$ function that transfer the rational expectation matrix $\mathbf{P}$ into $\mathbf{p} = \text{vec}(\mathbf{P}) = (\mathbf{p}_1', \cdots, \mathbf{p}_m')$. The process of developing Bayesian Nash Equilibrium (BNE) is $\mathbf{p} = \vec{h}(\mathbf{p}) = [\vec{h}_1(\mathbf{p})', \cdots, \vec{h}_m(\mathbf{p})']$, for each column $\vec{h}_k(\mathbf{p})$ can be written as

$$\begin{aligned}\vec{h}_k(\mathbf{p}) &= \sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k \qquad k = 1, \cdots, m_1 \\ &= \Big(\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k\Big) \odot \boldsymbol{\Phi}\Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma_k^*}\Big) + \sigma_k^* \phi\Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k}{\sigma_k^*}\Big) \\ &\qquad k = m_1 + 1, \cdots, m_1 + m_2 \\ &= \mathrm{F}_k\Big(\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \boldsymbol{\pi}_k\Big) \qquad k = m_1 + m_2 + 1, \cdots, m\end{aligned} \tag{2.6}$$

To guarantee the existence of the Bayesian Nash Equilibrium, we need the following assumption about the reduced-form model's parameter space.

Assumption 2.2.4. *The reduced-form simultaneous effect matrix* $\boldsymbol{\Lambda}^*$*, the network structure weight matrix* $\mathbf{W}$*, and the probability density function of reduced-form idiosyncratic shock for all binary-outcome activities, i.e.,* $\{f_k(\cdot)\}_{k=m_1+m_2+1}^{m}$*, need to satisfy*

$$\min\{||\boldsymbol{\Lambda}^*||_1||\mathbf{W}||_\infty, ||\boldsymbol{\Lambda}^*||_\infty||\mathbf{W}||_1\} < 1$$

$$\min\{||\boldsymbol{\Lambda}^*||_1||\mathbf{W}||_\infty \max_k \sup_u f_k(u), ||\boldsymbol{\Lambda}^*||_\infty||\mathbf{W}||_1 \max_k \sup_u f_k(u)\} < 1$$

Given the reduced form parameter space constraint, we will have sufficient conditions for the existence of Bayesian Nash Equilibrium according to the Contraction Mapping Theorem.

Proposition 2. *If assumptions 1.2.1, 2.2.2, 2.2.3, and 2.2.4 hold, then the incomplete information network game with the linear-quadratic form utility function will have a unique Bayesian Nash Equilibrium* $\mathbf{y}^\star$ *as*

$$\mathbf{y}^\star = (\boldsymbol{\Lambda}^{*\prime} \otimes \mathbf{W})\mathbf{p}^\star + \boldsymbol{\pi}^* - \boldsymbol{\epsilon}^* \tag{2.7}$$

where $\boldsymbol{\pi}^* = \text{vec}(\boldsymbol{\Pi})$*,* $\boldsymbol{\epsilon}^* = \text{vec}(\mathbf{E})$*, and the equilibrium from rational expectation of activities outcomes is the unique solution to* $\mathbf{p} = \vec{h}(\mathbf{p})$*.*

We use the first-order derivative of the mapping $\mathbf{p} = \vec{h}(\mathbf{p})$ to prove this proposition. For $k = 1, \cdots, m_1$, we have

$$|\frac{\partial p_{ik}}{\partial p_{jl}}| = |\lambda^*_{lk} w_{ij}| \tag{2.8}$$

For $k = m_1 + 1, \cdots, m_1 + m_2$, we have

$$|\frac{\partial p_{ik}}{\partial p_{jl}}| = |\lambda^*_{lk} w_{ij}| \Phi\Big(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}}{\sigma^*_k}\Big) < |\lambda^*_{lk} w_{ij}| \tag{2.9}$$

For $k = m_1 + m_2 + 1, \cdots, m$, we have

$$|\frac{\partial p_{ik}}{\partial p_{jl}}| = |\lambda^*_{lk} w_{ij}| f_k\Big(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}\Big) < |\lambda^*_{lk} w_{ij}| \sup_u f_k(u) \tag{2.10}$$

Given all these scalar-form first-order derivatives, we can derive the norm bound of the matrix form of the first-order derivative of $\mathbf{p} = \vec{h}(\mathbf{p})$. When $m_3 = 0$

$$\begin{aligned} ||\frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'}||_\infty \leq & \max_{k=1,\cdots m} \sum_{l=1}^{m} |\lambda^*_{lk}| \max_{i=1,\cdots n} \sum_{j=1}^{n} |w_{ij}| = ||\mathbf{\Lambda}^*||_1 ||\mathbf{W}||_\infty \\ ||\frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'}||_1 \leq & \max_{l=1,\cdots m} \sum_{l=1}^{m} |\lambda^*_{lk}| \max_{j=1,\cdots n} \sum_{i=1}^{n} |w_{ij}| = ||\mathbf{\Lambda}^*||_\infty ||\mathbf{W}||_1 \end{aligned} \tag{2.11}$$

and when $m_3 \neq 0$, we have

$$\begin{aligned} ||\frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'}||_\infty \leq \Big[\max_k \sum_{l=1}^{m} |\lambda^*_{lk}| \max_i \sum_{j=1}^{n} |w_{ij}|\Big] \max_k \sup_u f_k(u) = \Big[||\mathbf{\Lambda}^*||_1 ||\mathbf{W}||_\infty\Big] \max_k \sup_u f_k(u) \\ ||\frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'}||_1 \leq \Big[\max_l \sum_{l=1}^{m} |\lambda^*_{lk}| \max_j \sum_{i=1}^{n} |w_{ij}|\Big] \max_k \sup_u f_k(u) = \Big[||\mathbf{\Lambda}^*||_\infty ||\mathbf{W}||_1\Big] \max_k \sup_u f_k(u) \end{aligned} \tag{2.12}$$

Given the assumption 1.2.1, assumption 2.2.1, assumption 2.2.2, assumption 2.2.3, and assumption 2.2.4, the contraction property of the fixed-point mapping, $\mathbf{p} = \vec{h}(\mathbf{p})$ is hold.

2.3 Econometric Model

In this section, we will transfer the exogenous attributes that directly influence the intention of activities among all agents into a linear form, i.e., $\boldsymbol{\pi}_k = \mathbf{X}\boldsymbol{\beta}_k$, where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]'$ is an $n \times q$ matrix, represent exogenous variables that contribute to the direct effect of each agent $i = 1, \cdots, n$ intention of activities. $\boldsymbol{\beta}_k$ is the coefficient of direct effect on activity k. We have a simultaneous equation model as follows

$$y_{ik}^* = -\sum_{l=1,l\neq k}^{m} \theta_{lk} y_{il}^* + \sum_{l=1}^{m} \lambda_{lk} \sum_{j=1,j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k - \epsilon_{ik}, \tag{2.13}$$

for all agents $i = 1, \cdots, n$. And for $k = 1, \cdots, m_1$, $y_{ik} = y_{ik}^*$, for $k = m_1 + 1, \cdots, m_1 + m_2$, $y_{ik} = y_{ik}^* \mathrm{I}(y_{ik}^* > 0)$, for $k = m_1 + m_2 + 1, \cdots, m$, $y_{ik} = \mathrm{I}(y_{ik}^* > 0)$. We also use the following notation, the intention of all activities among all agents, $\mathbf{Y}^* = [\mathbf{y}_1^*, \mathbf{y}_2^*, \cdots, \mathbf{y}_m^*]$, the Bayesian Nash Equilibrium of rational expectations, $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_m]$, the exogenous attributes of all agents, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]'$, the idiosyncratic shocks in all activities among all agents, $\mathbf{E} = [\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_m]$, the network structure matrix $\mathbf{W} = [w_{ij}]$, $\mathbf{p}_k = (p_{1k}, \cdots, p_{nk})'$, and $\boldsymbol{\epsilon}_k = (\epsilon_{1k}, \epsilon_{2k}, \cdots, \epsilon_{nk})'$, the simultaneous effect matrix with unit diagonal elements $\boldsymbol{\Theta} = [\theta_{kl}]$, the direct effect coefficient matrix $\mathbf{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_m]$, then we can derive the matrix form of the model as

$$\mathbf{Y}^* \boldsymbol{\Theta} = \mathbf{W}\mathbf{P}\boldsymbol{\Lambda} + \mathbf{X}\mathbf{B} - \mathbf{E} \tag{2.14}$$

According to our assumption that $\boldsymbol{\Theta}$ is non-singular, we can derive the reduced form model

$$\mathbf{Y}^* = \mathbf{W}\mathbf{P}\boldsymbol{\Lambda}^* + \mathbf{X}\mathbf{B}^* - \mathbf{E}^* \tag{2.15}$$

where $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{\Theta}^{-1}$, $\mathbf{B}^* = \mathbf{B}\boldsymbol{\Theta}^{-1}$, and $\mathbf{E}^* = \mathbf{E}\boldsymbol{\Theta}^{-1}$. $\mathrm{vec}(\mathbf{E}^*) \sim N(0, \boldsymbol{\Sigma}^* \otimes \mathbf{I}_n)$. According to our previous assumptions, the k-th diagonal elements of $\boldsymbol{\Sigma}^*$ are bounded for $k = 1, \cdots, m_1 + m_2$, and are unit for $k = m_1 + m_2 + 1, \cdots, m$. The scalar form of the reduced-form econometric model is

$$y_{ik}^* = \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1,j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^* - \epsilon_{ik}^*. \tag{2.16}$$

If there are no censored-outcome activities, i.e., $m_2 = 0$, then the reduced form econometric model can be identified when $[\mathbf{WP}, \mathbf{X}]$ are full column rank. $\Phi(\cdot)$ is a monotonic increasing function. Let $\boldsymbol{\Psi}^* = [\boldsymbol{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]$, and an alternative parameters $\widetilde{\boldsymbol{\Psi}}^* = [\widetilde{\boldsymbol{\Lambda}}^{*\prime}, \widetilde{\mathbf{B}}^{*\prime}]$, then we will have the following equation for perfect-observed-outcome activities

$$\mathrm{E}(y_{ik}) = \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1,j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^* = \sum_{l=1}^{m} \widetilde{\lambda}_{lk}^* \sum_{j=1,j\neq i}^{n} w_{ij} \widetilde{p}_{jl} + \mathbf{x}_i' \widetilde{\boldsymbol{\beta}}_k^* \tag{2.17}$$

and also following equation for binary-outcome activities $\Pr(y_{ik}=1|\mathbf{W},\mathbf{X}) = \Pr(y_{ik}=1)$ as $\mathbf{W}$ and $\mathbf{X}$ are non-stochastic and predtermined.

$$\Pr(y_{ik}=1) = \Phi\Big(\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}_k^{*}\Big) = \Phi\Big(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}\widetilde{p}_{jl} + \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_k^{*}\Big) \tag{2.18}$$

then, according to our previous conclusion related to the uniqueness of the fixed point, we have, for perfect-observe-outcome activities,

$$\begin{aligned} p_{ik} &= \sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}_k^{*} \\ \widetilde{p}_{ik} &= \sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}\widetilde{p}_{jl} + \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_k^{*} \end{aligned} \tag{2.19}$$

and for the binary-outcome activities, we have

$$\begin{aligned} p_{ik} &= \Phi\Big(\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}_k^{*}\Big) \\ \widetilde{p}_{ik} &= \Phi\Big(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}\widetilde{p}_{jl} + \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_k^{*}\Big) \end{aligned} \tag{2.20}$$

as the fixed point are unique for the contraction mapping, then we have

$$\sum_{l=1}^{m}(\lambda_{lk}^{*} - \widetilde{\lambda}_{lk}^{*})\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'(\boldsymbol{\beta}_k^{*} - \widetilde{\boldsymbol{\beta}}_k^{*}) = 0 \tag{2.21}$$

which can be expressed in a matrix form as

$$[\mathbf{WP},\mathbf{X}](\boldsymbol{\Psi}^{*} - \widetilde{\boldsymbol{\Psi}}^{*}) = 0 \tag{2.22}$$

If $[\mathbf{WP},\mathbf{X}]$ is full column rank, then $[\mathbf{WP},\mathbf{X}](\boldsymbol{\Psi}^{*} - \widetilde{\boldsymbol{\Psi}}^{*}) = 0$ only when $\boldsymbol{\Psi}^{*} = \widetilde{\boldsymbol{\Psi}}^{*}$, which is the identification condition for the reduced form parameter $\boldsymbol{\Psi}^{*}$. When $m_2 \neq 0$, i.e., there are censored-outcome activities. Then, we need more conditions to make sure both parameters $\boldsymbol{\Psi}$ and σ_k^{*}, for $k = m_1+1,\cdots,m_1+m_2$, are identifiable. Before continuing our discussion, let $\Phi_{ik} = \Phi((\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}_k^{*})/\sigma_k^{*})$, $\phi_{ik} = \phi((\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}_k^{*})/\sigma_k^{*})$, and $\widetilde{\Phi}_{ik} = \Phi((\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_k^{*})/\widetilde{\sigma}_k^{*})$, $\widetilde{\phi}_{ik} = \phi((\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_k^{*})/\widetilde{\sigma}_k^{*})$. Then we have

$$\begin{aligned} \mathbf{Pr}(y_{ik}>0) = \Phi_{ik} &= \Phi((\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}_k^{*})/\sigma_k^{*}) \\ = \widetilde{\Phi}_{ik} &= \Phi((\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_k^{*})/\widetilde{\sigma}_k^{*}) \end{aligned} \tag{2.23}$$

Given $\Phi(\cdot)$ is a monotonic increasing function, we have

$$\sum_{l=1}^{m}\Big(\frac{\lambda_{lk}^*}{\sigma_k^*}-\frac{\widetilde{\lambda}_{lk}^*}{\widetilde{\sigma}_k^*}\Big)\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl}+\mathbf{x}_i'\Big(\frac{\boldsymbol{\beta}_k^*}{\sigma_k^*}-\frac{\widetilde{\boldsymbol{\beta}}_k^*}{\widetilde{\sigma}_k^*}\Big)=0. \tag{2.24}$$

If we have $[\mathbf{WP},\mathbf{X}]$ full column rank, then $\boldsymbol{\Psi}^*\mathbf{D}^{-1}=\widetilde{\boldsymbol{\Psi}}^*\widetilde{\mathbf{D}}^{-1}$, where $\mathbf{D}$ and $\widetilde{\mathbf{D}}$ are $m\times m$ diagonal matrices, for $k=1,\cdots,m_1,m_1+m_2+1,\cdots,m$, the k-th diagonal element of $\mathbf{D}$ and $\widetilde{\mathbf{D}}$ are one. For $k=m_1+1,\cdots,m_1+m_2$, the k-th element of $\mathbf{D}$ is σ_k^* and the k-th element of $\widetilde{\mathbf{D}}$ is $\widetilde{\sigma}_k^*$. To derive the identification condition for both $[\boldsymbol{\Lambda}^{*\prime},\mathbf{B}^{*\prime}]'$ and $\{\sigma_k^*\}_{k=m_1+1}^{m_1+m_2}$, we need the following assumption.

Assumption 2.3.1. *If $m_2=0$, i.e., no censored-outcome activities in the model, the sufficient condition for the reduced form econometric model identification is $[\mathbf{WP},\mathbf{X}]$ full column rank. Otherwise, $[\mathbf{WP},\mathbf{X},\mathbf{L}]$ should be full column rank. $\mathbf{L}$ is the $n\times m_2$ matrix, in which the (i,j)-th element is $\phi_{i(m_1+l)}/\Phi_{i(m_1+l)}$ for $l=1,\cdots,m_2$ and $i=1,\cdots,n$.*

To clarify the importance of this assumption, we start with the following. For $k=m_1+1,\cdots,m_1+m_2$, we have

$$\begin{aligned} p_{ik} =&\Big(\sum_{l=1}^{m}\lambda_{lk}^*\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl}+\mathbf{x}_i'\boldsymbol{\beta}_k^*\Big)\Phi_{ik}+\sigma_k^*\phi_{ik}\\ =&\Big(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^*\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl}+\mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_k^*\Big)\widetilde{\Phi}_{ik}+\widetilde{\sigma}_k^*\widetilde{\phi}_{ik}\end{aligned} \tag{2.25}$$

based on our previous discussion, we have $\Phi_{ik}=\widetilde{\Phi}_{ik}$ and $\phi_{ik}=\widetilde{\phi}_{ik}$, therefore we can derive

$$\sum_{l=1}^{m}(\lambda_{lk}^*-\widetilde{\lambda}_{lk}^*)\sum_{j=1,j\neq i}^{n} w_{ij}p_{jl}+\mathbf{x}_i'(\boldsymbol{\beta}_k^*-\widetilde{\boldsymbol{\beta}}_k^*)+(\sigma_k^*-\widetilde{\sigma}_k^*)\frac{\phi_{ik}}{\Phi_{ik}}=0 \tag{2.26}$$

and when $[\mathbf{WP},\mathbf{X},\mathbf{L}]$ is full column rank, the reduced form parameters $[\boldsymbol{\Lambda}^{*\prime},\mathbf{B}^{*\prime}]'$ and $\{\sigma_k^*\}_{k=1}^{m_1+m_2}$ are identifiable. With sufficient conditions for identifying the reduced form parameters, we need to derive the structural form parameters' identification condition. The structural form parameter matrix, $[\boldsymbol{\Lambda}^{*\prime},\mathbf{B}^{*\prime}]'$, is an $m\times(m+q)$ matrix. And the structural form parameter matrix, $[\boldsymbol{\Theta},-\boldsymbol{\Lambda}',-\mathbf{B}']'$, is an $m\times(2m+q)$ matrix. Even though we know all the diagonal elements are all one, we still need structural form parameter matrix rank condition to make it identifiable. According to Schmidt (1976), let $\boldsymbol{\gamma}_k$ is the k-th column of $\boldsymbol{\Gamma}=[\boldsymbol{\Theta},-\boldsymbol{\Lambda}',-\mathbf{B}']'$, and $\mathbf{R}_k$ is the constraint matrix, i.e., $\mathbf{R}_k\boldsymbol{\gamma}_k=\mathbf{0}$. And the rank condition as $\text{rank}(\mathbf{R}_k\boldsymbol{\Gamma})=m-1$ for $k=1,\cdots,m-1$.

Assumption 2.3.2. *Suppose* $\mathbf{\Gamma} = [\mathbf{\Theta}, -\mathbf{\Lambda}', -\mathbf{B}']'$, $\boldsymbol{\gamma}_k$ *is the* k*-th column of* $\mathbf{\Gamma}$, $\mathbf{R}_k$ *is the matrix for the constraint* $\mathbf{R}_k\boldsymbol{\gamma}_k = 0$, *and* $\text{rank}(\mathbf{R}_k\mathbf{\Gamma}) = m - 1$ *for* $k = 1, \cdots, m$.

Under all previous assumptions and discussions, we can derive sufficient conditions for the identification of both reduced form parameters and structural form parameters. In the next section, we will discuss the estimation method and steps.

2.4 Estimation

Given that we have figured out sufficient conditions for the identification of reduced-form and structural-form model parameters, we develop our estimation process in two steps. In the first step, we will use Nested Psedo Likelihood (NPL) estimation to estimate all reduced form parameters - $\{\lambda^*_{lk}\}^m_{l,k=1}$, $\{\boldsymbol{\beta}^*_k\}^m_{k=1}$, $\{\sigma^*_k\}^{m_1+m_2}_{k=1}$. NPL estimation method is discussed and applied in Aguirregabiria and Mira (2007), Lin and Xu (2017), and Liu (2019). During the estimation, the true value of the rational expectations, i.e., the Bayesian Nash Equilibrium, is unknown. We must initialize an $nm \times 1$ vector $\mathbf{p}$, i.e., $\text{vec}(\mathbf{P})$, the first $n(m_1 + m_2)$ elements should locate in $(0, +\infty)$, and the rest nm_3 elements should locate in $(0, 1)$. The iterative steps conducted in NPL estimation after initializing $\mathbf{p}$ are the following.

Step 1 Given $\mathbf{p}^{(t-1)}$, obtain $\hat{\boldsymbol{\psi}}^{*(t)}_k = (\hat{\lambda}^{*(t)}_{1k}, \cdots, \hat{\lambda}^{*(t)}_{mk}, \hat{\boldsymbol{\beta}}^{*(t)'}_k, \sigma^{*(t)}_k)' = \arg\max \ln L(\boldsymbol{\psi}^*_k; \mathbf{p}^{(t-1)})$, for $k = 1, \cdots, m_1$

$$\begin{aligned}\ln L(\boldsymbol{\psi}^*_k; \mathbf{p}^{(t-1)}) &= \sum_{i=1}^{n} \left[\ln\left(\frac{1}{\sigma^*_k} \phi\left(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p^{(t-1)}_{jl} + \mathbf{x}'_i \boldsymbol{\beta}^*_k - y_{ik}}{\sigma^*_k} \right)\right)\right] \\ &= \sum_{i=1}^{n} \Big[-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^{*2}_k - \frac{1}{2\sigma^{*2}_k}\Big(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p^{(t-1)}_{jl} + \mathbf{x}'_i \boldsymbol{\beta}^*_k - y_{ik}\Big)^2 \Big]\end{aligned} \tag{2.27}$$

for $k = m_1 + 1, \cdots, m_1 + m_2$

$$\begin{aligned}
\ln L(\boldsymbol{\psi}_k^*; \mathbf{p}^{(t-1)}) =& \sum_{i=1}^{n} \Bigg\{ d_{ik} \ln \Big(\frac{1}{\sigma_k^*} \phi \Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^* - y_{ik}}{\sigma_k^*} \Big) \Big) \\
&+ (1 - d_{ik}) \ln \Bigg(1 - \Phi \Bigg(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^*}{\sigma_k^*} \Bigg) \Bigg) \Bigg\} \\
=& \sum_{i=1}^{n} \Bigg\{ d_{ik} \Big[-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma_k^{*2} - \frac{1}{2\sigma_k^{*2}} \Big(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^* - y_{ik} \Big)^2 \Big] \\
&+ (1 - d_{ik}) \Big[1 - \Phi \Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^*}{\sigma_k^*} \Big) \Big] \Bigg\}
\end{aligned} \tag{2.28}$$

for $k = m_1 + m_2 + 1, \cdots, m$

$$\begin{aligned}
\ln L(\boldsymbol{\psi}_k^*; \mathbf{p}^{(t-1)}) = \sum_{i=1}^{n} \Bigg[& d_{ik} \ln \Phi \Big(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^* \Big) \\
&+ (1 - d_{ik}) \ln \Big(1 - \Phi \Big(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^* \Big) \Big) \Bigg]
\end{aligned} \tag{2.29}$$

Step 2 Given $\widehat{\boldsymbol{\Psi}}^{*(t)} = [\widehat{\boldsymbol{\psi}}_1^{*(t)}, \cdots, \widehat{\boldsymbol{\psi}}_m^{*(t)}]$, we can derive $\mathbf{p}^{(t)} = \vec{h}(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\Psi}}^{*(t)})$ where $\vec{h}(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\Psi}}^{*(t)}) = [\vec{h}_1(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\Psi}}^{*(t)})', \cdots, \vec{h}_m(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\Psi}}^{*(t)})']$. For $k = 1, \cdots, m_1$

$$\vec{h}_k(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\Psi}}^{*(t)}) = \begin{bmatrix} \sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{1j} p_{jl}^{(t-1)} + \mathbf{x}_1' \boldsymbol{\beta}_k^{*(t)} \\ \vdots \\ \sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{nj} p_{jl}^{(t-1)} + \mathbf{x}_n' \boldsymbol{\beta}_k^{*(t)} \end{bmatrix} \tag{2.30}$$

For $k = m_1 + 1, \cdots, m_1 + m_2$

$$\vec{h}_k(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\Psi}}^{*(t)}) = \begin{bmatrix} (\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{1j} p_{jl}^{(t-1)} + \mathbf{x}_1' \boldsymbol{\beta}_k^{*(t)}) \Phi \Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{1j} p_{jl}^{(t-1)} + \mathbf{x}_1' \boldsymbol{\beta}_k^{*(t)}}{\sigma_k^{*(t)}} \Big) \\ + \sigma_k^{*(t)} \phi \Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{1j} p_{jl}^{(t-1)} + \mathbf{x}_1' \boldsymbol{\beta}_k^{*(t)}}{\sigma_k^{*(t)}} \Big) \\ \vdots \\ (\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{nj} p_{jl}^{(t-1)} + \mathbf{x}_n' \boldsymbol{\beta}_k^{*(t)}) \Phi \Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{nj} p_{jl}^{(t-1)} + \mathbf{x}_n' \boldsymbol{\beta}_k^{*(t)}}{\sigma_k^{*(t)}} \Big) \\ + \sigma_k^{*(t)} \phi \Big(\frac{\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{nj} p_{jl}^{(t-1)} + \mathbf{x}_n' \boldsymbol{\beta}_k^{*(t)}}{\sigma_k^{*(t)}} \Big) \end{bmatrix} \tag{2.31}$$

For $k = m_1 + m_2 + 1, \cdots, m$

$$\vec{h}_k(\mathbf{p}^{(t-1)}; \widehat{\mathbf{\Psi}}^{*(t)}) = \begin{bmatrix} \Phi(\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{1j} p_{jl}^{(t-1)} + \mathbf{x}_1' \boldsymbol{\beta}_k^{*(t)}) \\ \vdots \\ \Phi(\sum_{l=1}^{m} \lambda_{lk}^{*(t)} \sum_{j=1}^{n} w_{nj} p_{jl}^{(t-1)} + \mathbf{x}_n' \boldsymbol{\beta}_k^{*(t)}) \end{bmatrix} \tag{2.32}$$

Then, update $\mathbf{p}^{(t-1)}$ in **Step 1** to $\mathbf{p}^{(t)}$, repeat the iterations until estimation results convergence.

The contraction mapping 2.6 with a fixed point is an essential determinant of the convergence of the nested pseudo likelihood algorithm by Kasahara and Shimotsu (2012). The contraction property of the mapping $\mathbf{p} = \vec{h}(\mathbf{p})$ is ensured by assumption 1.2.1, assumption 2.2.1, assumption 2.2.2, assumption 2.2.3, and assumption 2.2.4. The NPL estimator is characterized by $\hat{\boldsymbol{\psi}}_k^* = \arg\max \ln L(\boldsymbol{\psi}_k^*, \widehat{\mathbf{p}})$ when NPL algorithm converges, where $\widehat{\mathbf{p}}$ is implicitly calculated through $\widehat{\mathbf{p}} = \vec{h}(\widehat{\mathbf{p}}; \widehat{\mathbf{\Psi}}^*)$. The details on the asymptotic distribution of the NPL algorithm estimator are in the appendix. A Monte Carlo simulation is conducted in the next section to offer a straightforward perspective of our estimation.

2.5 Monte Carlo Simulation

We will develop our simulation based on the following two-equation Tobit-Probit model.

$$\begin{aligned} y_{i1}^* &= -\theta_{21} y_{i2}^* + \lambda_{11} \sum_{j=1}^{n} w_{ij} p_{j1} + \lambda_{21} \sum_{j=1}^{n} w_{ij} p_{j2} + x_{i1}\beta_1 - \epsilon_{i1} \\ y_{i2}^* &= -\theta_{12} y_{i1}^* + \lambda_{12} \sum_{j=1}^{n} w_{ij} p_{j1} + \lambda_{22} \sum_{j=1}^{n} w_{ij} p_{j2} + x_{i2}\beta_2 - \epsilon_{i2} \end{aligned} \tag{2.33}$$

the observed decision outcomes are $y_{i1} = y_{i1}^* \mathrm{I}(y_{i1}^* > 0)$ and $y_{i2} = \mathrm{I}(y_{i2}^* > 0)$. The vector form is

$$\begin{aligned} \mathbf{y}_1^* &= -\,\theta_{21}\mathbf{y}_2^* + \lambda_{11}\mathbf{W}\mathbf{p}_1 + \lambda_{21}\mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_1 - \boldsymbol{\epsilon}_1 \\ \mathbf{y}_2^* &= -\,\theta_{12}\mathbf{y}_1^* + \lambda_{12}\mathbf{W}\mathbf{p}_1 + \lambda_{22}\mathbf{W}\mathbf{p}_2 + \mathbf{X}_2\beta_2 - \boldsymbol{\epsilon}_2 \end{aligned} \tag{2.34}$$

and the matrix form is

$$\mathbf{Y}^*\mathbf{\Theta} = \mathbf{W}\mathbf{P}\mathbf{\Lambda} + \mathbf{X}\mathbf{B} - \mathbf{E}. \tag{2.35}$$

where

$$\mathbf{\Theta} = \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix} \tag{2.36}$$

and $\mathbf{Y}^* = [\mathbf{y}_1^*, \mathbf{y}_2^*]$, $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2]$, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, and $\mathbf{E} = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2]$. Then, we can derive the reduced form of the model in the matrix style

$$\mathbf{Y}^* = \mathbf{WP}\boldsymbol{\Lambda}^* + \mathbf{XB}^* - \mathbf{E}^*. \tag{2.37}$$

where

$$\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{\Theta}^{-1} = \begin{bmatrix} \lambda_{11}^* & \lambda_{12}^* \\ \lambda_{21}^* & \lambda_{22}^* \end{bmatrix} \qquad \mathbf{B}^* = \mathbf{B}\boldsymbol{\Theta}^{-1} = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix} \tag{2.38}$$

are reduced form parameters. The scalar and vector expression of the reduced form model is

$$\begin{aligned} \mathbf{y}_1^* &= \lambda_{11}^* \mathbf{W}\mathbf{p}_1 + \lambda_{21}^* \mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_{11}^* + \mathbf{X}_2\beta_{21}^* - \boldsymbol{\epsilon}_1^* \\ \mathbf{y}_2^* &= \lambda_{12}^* \mathbf{W}\mathbf{p}_1 + \lambda_{22}^* \mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_{12}^* + \mathbf{X}_2\beta_{22}^* - \boldsymbol{\epsilon}_2^* \end{aligned} \tag{2.39}$$

and

$$\begin{aligned} y_{i1}^* &= \lambda_{11}^* \sum_{j=1}^n w_{ij} p_{j1} + \lambda_{21}^* \sum_{j=1}^n w_{ij} p_{j2} + x_{i1}\beta_{11}^* + x_{i2}\beta_{21}^* - \epsilon_{i1}^* \\ y_{i2}^* &= \lambda_{12}^* \sum_{j=1}^n w_{ij} p_{j1} + \lambda_{22}^* \sum_{j=1}^n w_{ij} p_{j2} + x_{i1}\beta_{12}^* + x_{i2}\beta_{22}^* - \epsilon_{i2}^* \end{aligned} \tag{2.40}$$

the reduced form error term $(\epsilon_{i1}, \epsilon_{i2})'$ satisfies the joint normal distribution with zero means and variance-covariance matrix $\boldsymbol{\Sigma}^*$, where

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \sigma_1^{*2} & \rho^*\sigma_1^* \\ \rho^*\sigma_1^* & 1 \end{bmatrix} \tag{2.41}$$

as our previous assumption, $\sigma_2^{*2} = 1$, without loss of generality, we set $\sigma_1^{*2} = 1$, and consider cases in which $\rho^* = 0.2, 0.5,$ and 0.8. We follow the steps of the NPL algorithm.

- At step 0, initialize $\mathbf{p}_1 \in (0, c)^n$ where $c > 0$, and $\mathbf{p}_2 \in (0, 1)^n$. We also initialize all parameters (λ_{11}^*, λ_{12}^*, λ_{21}^*, λ_{22}^*, σ_1^*, β_{11}^*, β_{12}^*, β_{21}^*, β_{22}^*) as 0.1 before iteration.

• For steps $t > 0$, we can derive the log-likelihood function as

$$
\begin{aligned}
\ln L_1 = & \sum_{i=1}^{n}\Big[\mathbf{I}(y_{i1} > 0)\ln\Big[\frac{1}{\sigma_1^*}\phi\Big(\frac{\lambda_{11}^*\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{21}^*\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{11}^* + x_{i2}\beta_{21}^* - y_{i1}}{\sigma_1^*}\Big)\Big] \\
& + \mathbf{I}(y_{i1} = 0)\ln\Big[1 - \Phi\Big(\frac{\lambda_{11}^*\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{21}^*\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{11}^* + x_{i2}\beta_{21}^*}{\sigma_1^*}\Big)\Big]\Big] \\
\ln L_2 = & \sum_{i=1}^{n}\Big\{y_{i2}\ln\Big[\Phi\Big(\lambda_{12}^*\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{22}^*\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{12}^* + x_{i2}\beta_{22}^*\Big)\Big] \\
& + (1 - y_{i2})\ln\Big[1 - \Phi\Big(\lambda_{12}^*\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{22}^*\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{12}^* + x_{i2}\beta_{22}^*\Big)\Big]\Big\}
\end{aligned}
\tag{2.42}
$$

and the results as

$$
\begin{aligned}
(\lambda_{11}^{*(t)}, \lambda_{21}^{*(t)}, \beta_{11}^{*(t)}, \beta_{21}^{*(t)}, \sigma_1^{*(t)}) &= \arg\max \ln L_1(\lambda_{11}^*, \lambda_{21}^*, \beta_{11}^*, \beta_{21}^*, \sigma_1^*) \\
(\lambda_{12}^{*(t)}, \lambda_{22}^{*(t)}, \beta_{12}^{*(t)}, \beta_{22}^{*(t)}) &= \arg\max \ln L_2(\lambda_{12}^*, \lambda_{22}^*, \beta_{12}^*, \beta_{22}^*)
\end{aligned}
\tag{2.43}
$$

use the result from Maximum Likelihood Estimator, we derive the Bayesian Nash Equilibrium $\mathbf{p}_1 = (p_{11}, \cdots, p_{n1})'$ and $\mathbf{p}_2 = (p_{12}, \cdots, p_{n2})'$ where

$$
\begin{aligned}
p_{i1}^{(t+1)} =& (\lambda_{11}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{21}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)})\Phi_{i1}^{(t)} + \sigma_1^{*(t)}\phi_{i1}^{(t)} \\
p_{i2}^{(t+1)} =& \Phi(\lambda_{12}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{22}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{12}^{*(t)} + x_{i2}\beta_{22}^{*(t)})
\end{aligned}
\tag{2.44}
$$

and

$$
\begin{aligned}
\Phi_{i1}^{(t)} =& \Phi\Big(\frac{\lambda_{11}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{21}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)}}{\sigma_1^*}\Big) \\
\phi_{i1}^{(t)} =& \phi\Big(\frac{\lambda_{11}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i1}^{(t)} + \lambda_{21}^{*(t)}\sum_{j=1}^{n} w_{ij}p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)}}{\sigma_1^*}\Big)
\end{aligned}
\tag{2.45}
$$

the step t and step $t+1$ will be repeatedly executed until the estimated parameters converge. The estimation results after all the iterations, i.e., $\hat{\lambda}_{11}^*, \hat{\lambda}_{12}^*, \hat{\lambda}_{21}^*, \hat{\lambda}_{22}^*, \hat{\beta}_{11}^*, \hat{\beta}_{12}^*, \hat{\beta}_{21}^*, \hat{\beta}_{22}^*$ and $\hat{\sigma}_1^*$, can be used to calculate the Bayesian Nash Equilibrium, i.e., $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$. Then we can use the Maximum

Likelihood estimation to estimate ρ^*. The log-likelihood is the following

$$
\begin{aligned}
\ln L(\rho^*) \quad = \quad & \sum_{i=1}^{n}\Bigg[\mathbf{I}(y_{i1}>0)y_{i2} \\
& \ln\Big[\Phi\Big(\frac{\hat{\lambda}_{12}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{22}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{12}^*+x_{i2}\hat{\beta}_{22}^*}{\sqrt{1-\rho^{*2}}} \\
& -\frac{\rho^*(\hat{\lambda}_{11}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{21}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{11}^*+x_{i2}\hat{\beta}_{21}^*-y_{i1})}{\hat{\sigma}_1^*\sqrt{1-\rho^{*2}}}\Big) \\
& \frac{1}{\hat{\sigma}_1^*}\phi\Big(\frac{\hat{\lambda}_{11}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{21}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{11}^*+x_{i2}\hat{\beta}_{21}^*-y_{i1}}{\hat{\sigma}_1^*}\Big)\Big] \\
& +\mathbf{I}(y_{i1}>0)(1-y_{i2}) \\
& \ln\Big[\Phi\Big(\frac{\rho^*(\hat{\lambda}_{11}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{21}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{11}^*+x_{i2}\hat{\beta}_{21}^*-y_{i1})}{\hat{\sigma}_1^*\sqrt{1-\rho^{*2}}} \\
& -\frac{\hat{\lambda}_{12}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{22}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{12}^*+x_{i2}\hat{\beta}_{22}^*}{\sqrt{1-\rho^{*2}}}\Big) \\
& \frac{1}{\hat{\sigma}_1^*}\phi\Big(\frac{\hat{\lambda}_{11}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{21}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{11}^*+x_{i2}\hat{\beta}_{21}^*-y_{i1}}{\hat{\sigma}_1^*}\Big)\Big] \\
& +\mathbf{I}(y_{i1}=0)y_{i1} \\
& \ln\Phi_2\Big(-\frac{\hat{\lambda}_{11}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{21}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{11}^*+x_{i2}\hat{\beta}_{21}^*}{\hat{\sigma}_1^*}, \\
& \hat{\lambda}_{12}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{22}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{12}^*+x_{i2}\hat{\beta}_{22}^*,-\rho^*\Big) \\
& +\mathbf{I}(y_{i1}=0)(1-y_{i2}) \\
& \ln\Phi_2\Big(-\frac{\hat{\lambda}_{11}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{21}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{11}^*+x_{i2}\hat{\beta}_{21}^*}{\hat{\sigma}_1^*}, \\
& -(\hat{\lambda}_{12}^*\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}_{22}^*\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}_{12}^*+x_{i2}\hat{\beta}_{22}^*),\rho^*\Big)\Bigg]
\end{aligned}
$$

where $\Phi_2(\cdot,\cdot,\rho)$ is the cumulative distribution function (CDF) of the standard joint normal distribution with unit variance and correlation ρ. The MLE estimation result of ρ^* is $\hat{\rho}^*=\arg\max\ln L(\rho^*)$ from the Maximum Likelihood estimation of a bivariate normal distribution case in Greene (2017).

We can use the similar approach discussed in Amemiya (1973), Amemiya (1985), and Maddala (1986) to figure out the asymptotic property of our NPL algorithm to estimate the reduced form parameters. Let $\mathbf{Z}=[\mathbf{W}\mathbf{p}_1,\mathbf{W}\mathbf{p}_2,\mathbf{X}]$, and $\boldsymbol{\psi}_1^*=(\lambda_{11}^*,\lambda_{21}^*,\beta_{11}^*,\beta_{21}^*)'$ and $\boldsymbol{\psi}_2^*=(\lambda_{12}^*,\lambda_{22}^*,\beta_{12}^*,\beta_{22}^*)'$, the reduced form model of our simulation can be written in the following vector form

$$\begin{aligned} \mathbf{y}_1^* =& \mathbf{Z}\boldsymbol{\psi}_1^* - \boldsymbol{\epsilon}_1^* \\ \mathbf{y}_2^* =& \mathbf{Z}\boldsymbol{\psi}_2^* - \boldsymbol{\epsilon}_2^* \end{aligned} \tag{2.46}$$

with the NPL estimation algorithm, we calculate the $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_1$ for each step, and the regressors $\mathbf{Z} = [\mathbf{W}\hat{\mathbf{p}}_1, \mathbf{W}\hat{\mathbf{p}}_1, \mathbf{X}]$, and results of Maximum Likelihood Estimation as following

$$\begin{aligned} (\hat{\boldsymbol{\psi}}_1^*, \hat{\sigma}_1^{*2}) &= \arg\max \ln L_1(\boldsymbol{\psi}_1^*, \sigma_1^{*2}; \hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) \\ \hat{\boldsymbol{\psi}}_2^* &= \arg\max \ln L_2(\boldsymbol{\psi}_2^*; \hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) \end{aligned} \tag{2.47}$$

where

$$\begin{aligned} \ln L_1(\boldsymbol{\psi}_1^*, \sigma_1^*; \hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) &= \sum_{i=1}^{n} \mathbf{I}(y_{i1} > 0) \ln \Big[\frac{1}{\sigma_1^*}\phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_1^* - y_{i1}}{\sigma_1^*}\Big)\Big] + \mathbf{I}(y_{i1} = 0) \ln \Big[1 - \Phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Big] \\ \ln L_2(\boldsymbol{\psi}_2^*, \sigma_2^*; \hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) &= \sum_{i=1}^{n} y_{i2} \ln \Big[\Phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Big] + (1 - y_{i2}) \ln \Big[1 - \Phi\Big(\frac{\hat{\mathbf{z}}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Big] \end{aligned} \tag{2.48}$$

and we can write the first-order derivative to $\boldsymbol{\psi}_1^*$, σ_1^*, and $\boldsymbol{\psi}_2^*$ as

$$\begin{aligned} \frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*} &= \sum_{i=1}^{n} -\mathbf{I}(y_{i1} > 0)\frac{(\hat{\mathbf{z}}_i'\hat{\boldsymbol{\psi}}_1^* - y_{i1})\hat{\mathbf{z}}_i}{\hat{\sigma}_1^{*2}} - \mathbf{I}(y_{i1} = 0)\frac{\hat{\phi}_{i1}\hat{\mathbf{z}}_i}{(1 - \hat{\Phi}_{i1})\hat{\sigma}_1^*} = 0 \\ \frac{\partial \ln L_1}{\partial \sigma_1^*} &= \sum_{i=1}^{n} \mathbf{I}(y_{i1} > 0)\Big[- \frac{1}{\hat{\sigma}_1^*} - \frac{(\hat{\mathbf{z}}_i'\hat{\boldsymbol{\psi}}_1^* - y_{i1})^2}{\hat{\sigma}_1^{*3}}\Big] + \mathbf{I}(y_{i1} = 0)\frac{(\hat{\mathbf{z}}_i'\hat{\boldsymbol{\psi}}_1^*)\hat{\phi}_{i1}}{2(1 - \hat{\Phi}_{i1})\hat{\sigma}_1^{*2}} = 0 \\ \frac{\partial \ln L_2}{\partial \boldsymbol{\psi}_2^*} &= \sum_{i=1}^{n} \frac{[y_{i2} - \Phi(\hat{\mathbf{z}}_i'\hat{\boldsymbol{\psi}}_2^*)]\phi(\hat{\mathbf{z}}_i'\hat{\boldsymbol{\psi}}_2^*)}{\Phi(\hat{\mathbf{z}}_i'\hat{\boldsymbol{\psi}}_2^*)[1 - \Phi(\hat{\mathbf{z}}_i'\hat{\boldsymbol{\psi}}_2^*)]}\hat{\mathbf{z}}_i = 0 \end{aligned} \tag{2.49}$$

where $\hat{\Phi}_{i1} = \Phi(\hat{\mathbf{z}}_i'\hat{\psi}_1^*/\hat{\sigma}_1^*)$ and $\hat{\phi}_{i1} = \phi(\hat{\mathbf{z}}_i'\hat{\psi}_1^*/\hat{\sigma}_1^*)$. We implement two types of normalized network structures in our simulation. The first type is the random network, where each agent i is randomly influenced by five other agents in the network, with each of these five agents exerting an identical effect. Thus, in this scenario, $w_{ij} = 1/5$ if agent j can influence agent i in the network.

The second type is the circular network structure. In this setup, each agent i is only connected to agents $i+1$ and $i-1$, and the peer effects are consistent, specifically $w_{i,i+1} = w_{i,i-1} = 1/2$. For agent 1, we have $w_{12} = w_{1n} = 1/2$, and for agent n, we have $w_{n1} = w_{n,n-1} = 1/2$. This network graph resembles a large circle where each node only connects to its two neighboring nodes, hence the name 'circular' network.

The simulation parameters are as follows: $\theta_{12} = \theta_{21} = 0.5$, $\beta_1 = \beta_2 = 1$, $n = 1000$, rep $= 1000$, $\sigma_1^* = \sigma_2^* = 1$, and $\rho_{12}^* = 0.2, 0.5,$ and 0.8.

• Case 1: Weak peer effect $\lambda_{11} = \lambda_{22} = 0.2$, $\lambda_{12} = \lambda_{21} = 0.1$. Weak reduced-form error term correlation $\rho^* = 0.2$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.498	0.506	0.203	0.102	0.101	0.198	0.997	1.007
(0.051)	(0.050)	(0.069)	(0.069)	(0.122)	(0.121)	(0.068)	(0.080)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.497	0.505	0.198	0.100	0.103	0.199	1.003	1.009
(0.051)	(0.054)	(0.100)	(0.106)	(0.165)	(0.168)	(0.073)	(0.082)

• Case 2: Medium peer effect $\lambda_{11} = \lambda_{22} = 0.5$, $\lambda_{12} = \lambda_{21} = 0.3$. Weak reduced-form error term correlation $\rho^* = 0.2$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.497	0.506	0.498	0.295	0.305	0.504	1.000	1.009
(0.051)	(0.053)	(0.058)	(0.055)	(0.115)	(0.116)	(0.068)	(0.081)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.497	0.505	0.499	0.297	0.301	0.500	1.002	1.008
(0.050)	(0.052)	(0.096)	(0.095)	(0.179)	(0.174)	(0.069)	(0.079)

• Case 3: Strong peer effect $\lambda_{11} = \lambda_{22} = 0.8$, $\lambda_{12} = \lambda_{21} = 0.5$. Weak reduced-form error term correlation $\rho^* = 0.2$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.498	0.507	0.796	0.496	0.511	0.806	1.000	1.009
(0.058)	(0.061)	(0.053)	(0.051)	(0.099)	(0.103)	(0.076)	(0.081)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.496	0.506	0.798	0.499	0.512	0.806	1.001	1.012
(0.055)	(0.053)	(0.087)	(0.085)	(0.190)	(0.174)	(0.071)	(0.086)

• Case 4: Weak peer effect $\lambda_{11} = \lambda_{22} = 0.2$, $\lambda_{12} = \lambda_{21} = 0.1$. Medium reduced-form error term correlation $\rho^* = 0.5$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.498	0.504	0.203	0.102	0.101	0.199	0.998	1.008
(0.052)	(0.051)	(0.072)	(0.071)	(0.127)	(0.123)	(0.071)	(0.082)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.498	0.505	0.197	0.098	0.102	0.197	1.002	1.008
(0.053)	(0.055)	(0.105)	(0.110)	(0.172)	(0.175)	(0.075)	(0.084)

• Case 5: Medium peer effect $\lambda_{11} = \lambda_{22} = 0.5$, $\lambda_{12} = \lambda_{21} = 0.3$. Medium reduced-form error term correlation $\rho^* = 0.5$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.499	0.506	0.498	0.296	0.306	0.506	0.999	1.006
(0.052)	(0.055)	(0.059)	(0.057)	(0.120)	(0.119)	(0.071)	(0.081)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.498	0.504	0.500	0.300	0.300	0.498	1.002	1.008
(0.053)	(0.054)	(0.100)	(0.100)	(0.189)	(0.187)	(0.071)	(0.083)

• Case 6: Strong peer effect $\lambda_{11} = \lambda_{22} = 0.8$, $\lambda_{12} = \lambda_{21} = 0.5$. Medium reduced-form error term correlation $\rho^* = 0.5$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.499	0.505	0.795	0.496	0.509	0.804	1.002	1.007
(0.060)	(0.064)	(0.054)	(0.053)	(0.102)	(0.106)	(0.083)	(0.084)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.502	0.503	0.795	0.501	0.512	0.804	1.000	1.003
(0.057)	(0.055)	(0.092)	(0.089)	(0.196)	(0.194)	(0.074)	(0.090)

• Case 7: Weak peer effect $\lambda_{11} = \lambda_{22} = 0.2$, $\lambda_{12} = \lambda_{21} = 0.1$. Strong reduced-form error term correlation $\rho^* = 0.8$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.498	0.507	0.203	0.102	0.104	0.203	0.996	1.005
(0.054)	(0.054)	(0.074)	(0.072)	(0.132)	(0.128)	(0.073)	(0.083)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.498	0.505	0.196	0.097	0.104	0.201	1.003	1.008
(0.054)	(0.056)	(0.109)	(0.111)	(0.180)	(0.181)	(0.077)	(0.084)

• Case 8: Medium peer effect $\lambda_{11} = \lambda_{22} = 0.5$, $\lambda_{12} = \lambda_{21} = 0.3$. Strong reduced-form error term correlation $\rho^* = 0.8$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.500	0.504	0.497	0.295	0.305	0.506	1.000	1.004
(0.056)	(0.057)	(0.062)	(0.058)	(0.122)	(0.123)	(0.074)	(0.086)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.500	0.505	0.500	0.300	0.301	0.499	1.000	1.003
(0.056)	(0.055)	(0.105)	(0.104)	(0.195)	(0.190)	(0.074)	(0.086)

• Case 9: Strong peer effect $\lambda_{11} = \lambda_{22} = 0.8$, $\lambda_{12} = \lambda_{21} = 0.5$. Strong reduced-form error term correlation $\rho^* = 0.8$

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.499	0.506	0.795	0.497	0.510	0.806	1.002	1.007
(0.061)	(0.066)	(0.054)	(0.053)	(0.109)	(0.112)	(0.085)	(0.085)

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.501	0.505	0.794	0.498	0.515	0.807	0.999	1.004
(0.058)	(0.058)	(0.095)	(0.092)	(0.206)	(0.200)	(0.077)	(0.088)

From the results, we could find the estimation of simultaneous effect is very stable and accurate, with an error rate $\pm 1\%$ and a coefficient of variation 10%. The strength of peer effect and error-term correlation doesn't influence the estimation of simultaneous effect parameters. This shows the accuracy and stability of our two-stage approach to the estimation of structural form parameters. We could find that the estimation of peer effect caused by censored-outcome activity is better than the estimation of peer effect caused by binary-outcome activity. Especially the standard deviation of $\hat{\lambda}_{21}$ and $\hat{\lambda}_{22}$ are significant higher than the standard deviation of estimation of $\hat{\lambda}_{11}$ and $\hat{\lambda}_{12}$. The overall peer effect estimation is more consistent in the circular network than in a random network. Each agent will be treated equally and predictable in the circular network. However, in a random network, some agents may be overweight and may cause the overall peer effect estimation to be less consistent. All p_{i1} are between zero and positive infinite, but all p_{i2} are between zero and one. Therefore, $\mathbf{Wp}_1$ will be more variant than $\mathbf{Wp}_2$ given the same network weighted structure. And this will lead to a more consistent estimation in λ_{11} and λ_{12} than λ_{21} and λ_{22}. The standard deviation of the peer effect estimation doesn't change much as we increase the peer effect, which means the estimation results will be more reliable under a stronger peer effect case. As for the estimation of direct effect, $\hat{\beta}_1$ is more accurate $\hat{\beta}_2$ ($\pm 0.3\%$ and $\pm 1\%$ in error rate), as the censored outcomes contain more numerical information than binary outcomes. All the results show our estimation approach's accuracy, consistency, and reliability numerically.

2.6 Conclusion

This paper proposes a simultaneous equation model with limited dependent variables and social interactions. We develop the game theoretical foundation as an incomplete network game and figure out the parameter space condition for the existence of the Bayesian Nash Equilibrium. A Nested Pseudo Likelihood estimation is proposed to estimate the reduced form parameters without knowing the true values of agents' rational expectations of others' decision outcomes. We develop different forms of Bayesian Nash Equilibrium for different activity types and separately identify their parameter space constraints. Monte Carlo simulation shows the accuracy, consistency, and reliability of the NPL algorithm and Amemiya General Least square estimation under a sufficient parameter space condition. Future work is necessary for potential model selection problems or a different mechanism in simultaneous effect.

Chapter 3

A Simultaneous Equation Tobit Model with Social Interactions

3.1 Introduction

In this chapter, we consider a simultaneous equation model, but the interdependent effect is caused by other activity outcomes instead of incentives.

Amemiya (1974) develops a computation-relief approach to handle the two-equation Tobit model with censored outcomes and proves the consistency of the estimation. Agents in this scenario are categorized into two groups - positive outcome values in both activities and the rest of them. The consistency of such estimation is proved. However, as the number of activities increases, the information loss problem will be more severe as we group all the agents into two instead of 2^m (where m is the number of activities). When agents have two activities, they can be categorized into four groups if both activities' outcomes are censored at zero. (positive-positive, positive-zero, zero-positive, zero-zero). Amemiya (1974) estimation categorizes agents into two groups (positive-positive and non-positive-positive) and provides consistency proof. However, when the number of activities becomes larger, i.e., m, there will be 2^m groups for agents based on their activity outcomes. Suppose we still categorize them into an all-positive group and a non-all-positive group. In that case, the number of agents in the first group will take a much lower percentage of all agents if the data is randomly distributed. Even the estimation process will be simplified if we categorize all agents into two groups, and the estimation results will be influenced. Another concern is when other agents influence an agent's decision, and such peer effect is through rational expectations on other agents' decision outcomes, then Amemiya (1974) simplified-categorization estimation cannot handle. Amemiya (1974) is important to the study of simultaneous equation systems with truncated outcomes and elaborated in book chapters by Amemiya (1985) and Maddala (1986). Our model differs from Amemiya (1974) as we contain peer effect among agents through rational expectations. Another difference is we don't use the simplified version of maximum likelihood estimation raised by Amemiya (1974) as we need to use nested-pseudo likelihood estimation and calculate the Bayesian Nash Equilibrium for every iteration.

The single-equation Tobit SAR model also develops fast, and several papers enlighten our research. Qu and Lee (2012) propose two types of SAR Tobit model, the difference is peer effect is through actual decision outcomes (positive or zero) or reservation value (positive, zero, or negative) of the decision outcome. They also provide the hypothesis test for LM statistics for testing models. Qu and Lee (2013) propose a more locally powerful test for Spatial Tobit models. Xu and Lee (2015) develop the maximum likelihood estimation for the Spatial Tobit model in which peer effect is through reservation value of decision outcome. They discuss the asymptotic properties of the estimation and prove the consistency. Xu and Lee (2018) extend the maximum likelihood estimation to sieve maximum likelihood estimation to handle distribution-free estimation cases. All of this recent progress contributes to my current work and my next-step research - as we can make both peer effect and simultaneous effect through other agents' decision outcomes of different activities.

The rest of the paper is organized as follows. Section 3.2 is the econometric model, identification, and parameter space discussion. Section 3.3 shows the estimation procedure. Section 3.4 presents the Simulation and the discussion of simulation results. Section 3.5 is the conclusion.

3.2 Econometric Model

In the econometric model, agents cannot see other agents' decision outcomes when they make their own decisions (incomplete information; this is the same as in previous chapters), and they need to make predictions (rational expectations) on other agents' decision outcomes. Such rational expectations of other agents' outcomes will influence the certain agent's reservation value of each activity (peer effect). Peer effect can exist in the same activity (inner-activity peer effect) and among different activities (cross-/inter-activity peer effects). Moreover, for each agent, her reservation value of each activity is also influenced by her decision outcomes (this is different from the previous two chapters. In the structural form, the right side of the equation is y_{ik} instead of y_{ik}^*) of other activities (simultaneous effect) and her exogenous attributes (direct effect).

Suppose there are n agents belonging to a network, and each agent needs to make m decisions. The econometric model is

$$y_{ik}^* = -\sum_{l=1, l\neq k}^{m} \theta_{lk} y_{il} + \sum_{l=1}^{m} \lambda_{lk} \sum_{j=1, j\neq i}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k - \epsilon_{ik}, \tag{3.1}$$

where y_{ik}^* is agent i's reservation value toward activity k, it can only be observed as positive or zero, i.e., $y_{ik} = y_{ik}^* \mathrm{I}(y_{ik}^* > 0)$. In other words, y_{ik} is agent i's decision outcomes toward activity k. $p_{jl} = E(y_{jl})$. This is observable to all other agents and researchers. θ_{lk} represent the simultaneous effect of agent i's decision outcome on activity l on agent i's reservation value of activity k. w_{ij} reflect the network structure, and the strength of agent j influence agent i. λ_{lk} represents the peer effect of the weighted average of agent i's rational expectations on other agents' decision outcomes in activity l on agent i's reservation value of activity k. $\boldsymbol{\beta}_k$ represents the direct effect on activity k. $\mathbf{x}_i$ contains all the exogenous characteristics of agent i that directly influence the reservation values of different activities' decision-making processes. ϵ_{ik} represents unobserved idiosyncratic term that influence agent i's reservation value on activity k. We need the following assumption with respect to $(\epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{im})'$ for each agent i

Assumption 3.2.1. *$(\epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{im})'$ are i.i.d for $i = 1, 2, \cdots, n$ and satisfies a joint normal distribution with zero means and variance-covariance matrix $\boldsymbol{\Sigma}$.*

Remark 9. *The joint normal distribution will assist in deriving the Bayesian Nash Equilibrium (BNE) and the solution to the fixed point mapping problem. However, the form of the BNE will differ from Chapter 1 and will be discussed in the rest of this section.*

Assumption 3.2.2. *All principal minors of $\boldsymbol{\Theta}$ are positive.*

Remark 10. *The assumption of $\boldsymbol{\Theta}$ is stronger in this chapter compared to the previous two chapters. We use the condition for $\boldsymbol{\Theta}$ in Schmidt (1981). Cases discussed Amemiya (1974) can clarify* *the necessity of this assumption. Suppose in a two-activity simultaneous equation system, in which* $\theta_{12} = \theta_{21} = 2$, *and* $\sum_{l=1}^{2} \lambda_{lk} \sum_{j=1,j\neq i}^{n} w_{ij} p_{j1} + \mathbf{x}_i' \boldsymbol{\beta}_1 - \epsilon_{i1} = \sum_{l=1}^{2} \lambda_{lk} \sum_{j=1,j\neq i}^{n} w_{ij} p_{j2} + \mathbf{x}_i' \boldsymbol{\beta}_2 - \epsilon_{i2} = 1$, *then* (y_{i1}, y_{i2}) *can be* $(\frac{1}{3}, \frac{1}{3})$, $(1, 0)$, *or* $(0, 1)$, *instead of a unique result. Suppose* $\theta_{12} = \theta_{21} = -2$, $\sum_{l=1}^{2} \lambda_{lk} \sum_{j=1,j\neq i}^{n} w_{ij} p_{j1} + \mathbf{x}_i' \boldsymbol{\beta}_1 - \epsilon_{i1} > 0$, *and* $\sum_{l=1}^{2} \lambda_{lk} \sum_{j=1,j\neq i}^{n} w_{ij} p_{j2} + \mathbf{x}_i' \boldsymbol{\beta}_2 - \epsilon_{i2} > 0$, *then there will be no* (y_{i1}, y_{i2}) *satisfies the model 3.1.*

The assumption (3.2.1) clarifies our model environment and the precondition for unobserved error terms. They should be independent among different agents. The unobserved error vector satisfies a joint normal distribution for each agent, and the variance is finite. The distribution of the unobserved error vector for each agent is identical and known by all the agents (public knowledge). This assumption is necessary because the Bayesian Nash Equilibrium is derived in our following steps. This is different from Liu (2019) because our outcomes are censored instead of binary, we need to restrict the function form of idiosyncratic terms to derive the rational expectations on all agents decision outcomes, i.e., Bayesian Nash Equilibrium. Then, the matrix form is

$$\mathbf{Y}^* = -\mathbf{Y}\overline{\boldsymbol{\Theta}} + \mathbf{W}\mathbf{P}\boldsymbol{\Lambda} + \mathbf{X}\mathbf{B} - \mathbf{E} \tag{3.2}$$

where $\overline{\boldsymbol{\Theta}} = [\bar{\theta}_{kl}]$, $\bar{\theta}_{kk} = 0$ for $k = 1, \cdots, m$, represents the simultaneous effect matrix. $\boldsymbol{\Lambda} = [\lambda_{lk}]$ represents the peer effect matrix. $\mathbf{W} = [w_{ij}]_{i,j=1}^{n}$ is the network structure matrix. $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m]$, represents the decision outcomes matrix of all m activities among all n agents. $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_m]$, represents rational expectations on decision outcomes of all m activities, among all n agents. $\mathbf{B} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m]$, represents the direct effect matrix. $\mathbf{X}$ contains all the exogenous variables influencing the reservation value of all m activities of all n agents. And $\mathbf{E} = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \cdots, \boldsymbol{\epsilon}_m]$ represent the matrix containing all unobserved error term related to all m activities' decision-making process of all n agents.

Then we can define $\boldsymbol{\Theta} = \mathbf{I}_m + \overline{\boldsymbol{\Theta}}$, and $\boldsymbol{\Theta} = [\theta_{kl}]$, where $\theta_{kk} = 1$ for all $k = 1, \cdots, m$. According to whether each agent's decision outcome on each activity is positive or zero, we can categorize all the n agents into at most 2^m groups. Then, without loss of generality, there are n_g agents in the group g. And we have

$$\mathbf{Y}_g^* \boldsymbol{\Theta}_g \quad = \quad \mathbf{W}_g \mathbf{P} \boldsymbol{\Lambda} + \mathbf{X}_g \mathbf{B} - \mathbf{E}_g. \tag{3.3}$$

We denote the 1st group containing all agents with positive outcomes in each activity. and the 2^m-th group containing all agents with 0 outcome values in every activity. The diagonal elements in the k-row of $\boldsymbol{\Theta}_g$ are all ones; other elements are the same as $\boldsymbol{\Theta}$ or zeros. We give a brief discussion of what each $\boldsymbol{\Theta}_g$ looks like when $m = 2$ and $m = 3$. $\mathbf{Y}_g^*$ is $n_g \times m$, i.e., only containing agents belonging to group g, and $\mathbf{W}_g$, a revised $n_g \times n$ network matrix, only contains the rows of original $n \times n$ network matrix related to agents in group g. In the following part of this section, we will consider two basic scenarios of how to develop $\boldsymbol{\Theta}_g$

- $k \in \{1, 2\}$ - then we can categorize all the n agents into 4 groups

$$\begin{aligned}
&\text{Group 1:} \quad g = 1 \quad y_{i1} > 0, y_{i2} > 0 \\
&\text{Group 2:} \quad g = 2 \quad y_{i1} > 0, y_{i2} = 0 \\
&\text{Group 3:} \quad g = 3 \quad y_{i1} = 0, y_{i2} > 0 \\
&\text{Group 4:} \quad g = 4 \quad y_{i1} = 0, y_{i2} = 0
\end{aligned}$$

and the $\boldsymbol{\Theta}_g$ matrix for each group of agents can be expressed as

$$\boldsymbol{\Theta}_1 = \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix} \quad \boldsymbol{\Theta}_2 = \begin{bmatrix} 1 & \theta_{12} \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{\Theta}_3 = \begin{bmatrix} 1 & 0 \\ \theta_{21} & 1 \end{bmatrix} \quad \boldsymbol{\Theta}_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For agents in group 1, there exist simultaneous effect of activity 2 decision outcome on the reservation value of activity 1, and simultaneous effect of activity 1 decision outcome on the reservation value of activity 2. So $\boldsymbol{\Theta}_1$ is the same as $\boldsymbol{\Theta}$. For agents in group 2, as $y_{i2} = 0$, there only exists the simultaneous effect of activity 1 outcomes on the reservation value of activity 2, but doesn't exist the simultaneous effect of activity 2 outcomes on the reservation value of activity 1. Therefore, the simultaneous effect matrix $\boldsymbol{\Theta}_2$ is upper diagonal. For agents in group 3, as $y_{i1} = 0$, there only exists the simultaneous effect of activity 2 outcomes on the reservation value of activity 1, but doesn't exist the simultaneous effect of activity 1 outcomes on the reservation value of activity 2. Therefore, the simultaneous effect matrix $\boldsymbol{\Theta}_3$ is lower diagonal. There is no simultaneous effect for agents in group 4 as both activities' outcomes are zeros and $\mathbf{\Theta}_4$ is an identity matrix. For each activity, reservation value is only determined by peer effect, direct effect, and unobserved idiosyncratic terms.

- $k \in \{1, 2, 3\}$ - then we can categorize all the n agents into 8 groups

$$\begin{aligned}
&\text{Group 1:} \quad g = 1 \quad y_{i1} > 0, \quad y_{i2} > 0, \quad y_{i3} > 0 \\
&\text{Group 2:} \quad g = 2 \quad y_{i1} > 0, \quad y_{i2} > 0, \quad y_{i3} = 0 \\
&\text{Group 3:} \quad g = 3 \quad y_{i1} > 0, \quad y_{i2} = 0, \quad y_{i3} > 0 \\
&\text{Group 4:} \quad g = 4 \quad y_{i1} = 0, \quad y_{i2} > 0, \quad y_{i3} > 0 \\
&\text{Group 5:} \quad g = 5 \quad y_{i1} > 0, \quad y_{i2} = 0, \quad y_{i3} = 0 \\
&\text{Group 6:} \quad g = 6 \quad y_{i1} = 0, \quad y_{i2} > 0, \quad y_{i3} = 0 \\
&\text{Group 7:} \quad g = 7 \quad y_{i1} = 0, \quad y_{i2} = 0, \quad y_{i3} > 0 \\
&\text{Group 8:} \quad g = 8 \quad y_{i1} = 0, \quad y_{i2} = 0, \quad y_{i3} = 0
\end{aligned}$$

We can derive the simultaneous effect matrix for each group of agents through a similar approach to $k \in \{1, 2\}$ and have the following results. For agents in group 1, there exist all types of directions' simultaneous effect, and $\mathbf{\Theta}_1$ is the same as $\mathbf{\Theta}$. For agents in group 8, there will be no simultaneous effect between different activities as all activities' outcomes are zero. Therefore, the simultaneous effect matrix $\mathbf{\Theta}_8$ is an identity matrix. For agents in group 2, simultaneous effects only exist from activity 1 and 2 toward other activities as the outcome of activity 3 is zero. Therefore, the non-diagonal non-zero elements in $\mathbf{\Theta}_2$ are θ_{12}, θ_{13}, θ_{21}, and θ_{23}. For agents in group 3, simultaneous effects only exist from activity 1 and 3 toward other activities as the outcome of activity 2 is zero. Therefore, the non-diagonal non-zero elements in $\mathbf{\Theta}_3$ are θ_{12}, θ_{13}, θ_{31}, and θ_{32}. For agents in group 4, simultaneous effects only exist from activity 2 and 3 toward other activities as the outcome of activity 1 is zero. Therefore, the non-diagonal non-zero elements in $\mathbf{\Theta}_4$ are θ_{21}, θ_{23}, θ_{31}, and θ_{32}. For agents in group 5, simultaneous effects only exist from activity 1 toward other activities, as the outcome of activity 2 and 3 are zero. Therefore, the non-diagonal non-zero elements in $\mathbf{\Theta}_5$ are θ_{12} and θ_{13}. For agents in group 6, simultaneous effects only exist from activity 2 toward other activities, as the outcome of activity 1 and 3 are zero. Therefore, the non-diagonal non-zero elements in $\mathbf{\Theta}_6$ are θ_{21} and θ_{23}. For agents in group 7, simultaneous effects only exist from activity 3 toward other activities, as the outcome of activity 1 and 2 are zero. Therefore, the non-diagonal non-zero elements in $\mathbf{\Theta}_7$ are θ_{31} and θ_{32}.

$$\mathbf{\Theta}_1 = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} \\ \theta_{21} & 1 & \theta_{23} \\ \theta_{31} & \theta_{32} & 1 \end{bmatrix} \quad \mathbf{\Theta}_2 = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} \\ \theta_{21} & 1 & \theta_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{\Theta}_3 = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} \\ 0 & 1 & 0 \\ \theta_{31} & \theta_{32} & 1 \end{bmatrix} \mathbf{\Theta}_4 = \begin{bmatrix} 1 & 0 & 0 \\ \theta_{21} & 1 & \theta_{23} \\ \theta_{31} & \theta_{32} & 1 \end{bmatrix}$$

$$\mathbf{\Theta}_5 = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{\Theta}_6 = \begin{bmatrix} 1 & 0 & 0 \\ \theta_{21} & 1 & \theta_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{\Theta}_7 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \theta_{31} & \theta_{32} & 1 \end{bmatrix} \mathbf{\Theta}_8 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For a general number of activities m for each agent, all the agents can be divided into 2^m different groups. We have

$$\begin{aligned} \mathbf{Y}_g^* &= \mathbf{W}_g \mathbf{P} \mathbf{\Lambda} \mathbf{\Theta}_g^{-1} + \mathbf{X}_g \mathbf{B} \mathbf{\Theta}_g^{-1} - \mathbf{E}_g \mathbf{\Theta}_g^{-1} \\ &= \mathbf{W}_g \mathbf{P} \mathbf{\Lambda}_g^* + \mathbf{X}_g \mathbf{B}_g^* - \mathbf{E}_g^* \end{aligned} \tag{3.4}$$

where $\mathbf{\Lambda}_g^* = \mathbf{\Lambda}\mathbf{\Theta}_g^{-1}$, $\mathbf{B}_g^* = \mathbf{B}_g\mathbf{\Theta}_g^{-1}$, and $\mathbf{E}_g^* = \mathbf{E}_g\mathbf{\Theta}_g^{-1}$ and we have

$$[\mathbf{Y}_g^*]_{g=1}^{2^m} \quad = \quad [\mathbf{W}_g]_{g=1}^{2^m} \mathbf{P} [\mathbf{\Lambda}_g^*]_{g=1}^{2^m} + [\mathbf{X}_g \mathbf{B}_g^*]_{g=1}^{2^m} - [\mathbf{E}_g^*]_{g=1}^{2^m} \tag{3.5}$$

so for each p_{ik}. That means we will have a group-specific reduced form model similar to Liu (2019) for each group of agents. We can derive the rational expectation of activity outcomes, i.e., p_{ik} as following.

$$\begin{aligned} p_{ik} &= \mathrm{E}(y_{ik}) = \mathrm{E}(y_{ik} | y_{ik} > 0) \mathbf{Pr}(y_{ik} > 0) \\ &= \sum_{g=1}^{2^m} \mathrm{E}(y_{ik} | \text{ agent } i \text{ in group } g) \mathbf{Pr}(\text{agent } i \text{ in group } g). \end{aligned} \tag{3.6}$$

It is obvious that some $\mathrm{E}(y_{ik}|$ agent i in group $g)$ are zeros, but to make the formula represent a general case, we still use this form. According to our previous discussion about the group-specific reduced form model, we need the following assumption.

Assumption 3.2.3. *All group-specific reduced form simultaneous effect matrix $\mathbf{\Theta}_g$, for $g = 1, \cdots, 2^m$, should be non-singular.*

Given $k \in \{1,2\}$, we have the following simultaneous equation Tobit model with social interactions.

$$\begin{aligned} y_{i1}^* &= -\theta_{21} y_{i2} + \lambda_{11} \sum_{j=1}^{n} w_{ij} p_{j1} + \lambda_{21} \sum_{j=1}^{n} w_{ij} p_{j2} + \mathbf{x}_i' \boldsymbol{\beta}_1 - \epsilon_{i1} \\ y_{i2}^* &= -\theta_{12} y_{i1} + \lambda_{12} \sum_{j=1}^{n} w_{ij} p_{j1} + \lambda_{22} \sum_{j=1}^{n} w_{ij} p_{j2} + \mathbf{x}_i' \boldsymbol{\beta}_2 - \epsilon_{i2} \end{aligned} \tag{3.7}$$

for each agent i, neither y_{i1}^* or y_{i2}^* can be observed. However, the outcome y_{i1} and y_{i2} can be observed as zero or positive.

$$\begin{aligned} y_{i1} &= y_{i1}^* \mathrm{I}(y_{i1}^* > 0) \\ y_{i2} &= y_{i2}^* \mathrm{I}(y_{i2}^* > 0) \end{aligned} \tag{3.8}$$

then, we derive the rational expectation from our previous group discussion and the property of conditional expectation values.

$$\begin{aligned} p_{i1} &= \mathrm{E}(y_{i1}) = \mathrm{E}(y_{i1}|y_{i1} > 0)\mathbf{Pr}(y_{i1} > 0) + \mathrm{E}(y_{i1}|y_{i1} = 0)\mathbf{Pr}(y_{i1} = 0) \\ &= \mathrm{E}(y_{i1}^*|y_{i1}^* > 0)\mathbf{Pr}(y_{i1}^* > 0) \\ &= \mathrm{E}(y_{i1}^*|y_{i1}^* > 0, y_{i2}^* > 0)\mathbf{Pr}(y_{i1}^* > 0, y_{i2}^* > 0) + \mathrm{E}(y_{i1}^*|y_{i1}^* > 0, y_{i2}^* \le 0)\mathbf{Pr}(y_{i1}^* > 0, y_{i2}^* \le 0) \\ p_{i2} &= \mathrm{E}(y_{i2}) = \mathrm{E}(y_{i2}|y_{i2} > 0)\mathbf{Pr}(y_{i2} > 0) + \mathrm{E}(y_{i2}|y_{i2} = 0)\mathbf{Pr}(y_{i2} = 0) \\ &= \mathrm{E}(y_{i2}^*|y_{i2}^* > 0)\mathbf{Pr}(y_{i2}^* > 0) \\ &= \mathrm{E}(y_{i2}^*|y_{i1}^* > 0, y_{i2}^* > 0)\mathbf{Pr}(y_{i1}^* > 0, y_{i2}^* > 0) + \mathrm{E}(y_{i2}^*|y_{i1}^* \le 0, y_{i2}^* > 0)\mathbf{Pr}(y_{i1}^* \le 0, y_{i2}^* > 0). \end{aligned} \tag{3.9}$$

For the possible case that agent i belong to group 1, i.e., $y_{i1}^* > 0$ and $y_{i2}^* > 0$, we have

$$\begin{aligned} y_{i1}^* &= -\theta_{21} y_{i2}^* + \mathbf{z}_i' \boldsymbol{\psi}_1 - \epsilon_{i1} > 0 \\ y_{i2}^* &= -\theta_{12} y_{i1}^* + \mathbf{z}_i' \boldsymbol{\psi}_2 - \epsilon_{i2} > 0 \end{aligned} \tag{3.10}$$

following is our notation of previous equation system

$$\begin{aligned} \mathbf{z}_i &= \Big(\sum_{j=1}^{n} w_{ij} p_{j1}, \sum_{j=1}^{n} w_{ij} p_{j2}, \mathbf{x}_i' \Big)' \\ \boldsymbol{\psi}_1 &= (\lambda_{11}, \lambda_{21}, \beta_1')' \\ \boldsymbol{\psi}_2 &= (\lambda_{12}, \lambda_{22}, \beta_2')' \end{aligned} \tag{3.11}$$

and the inequality, when $y^*_{i1} > 0$ and $y^*_{i2} > 0$, can be rewritten as the vector form

$$(y^*_{i1}, y^*_{i2}) \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix} = \mathbf{z}'_i(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) - (\epsilon_{i1}, \epsilon_{i2}) > (0, 0) \tag{3.12}$$

the simultaneous effect matrix under this case is the same as in the structural model. Then, according to the assumption made on the parameter space, we have $1 - \theta_{12}\theta_{21} \neq 0$, we can rewrite the inequality of (y^*_{i1}, y^*_{i2}) as

$$\begin{aligned} (y^*_{i1}, y^*_{i2}) &= \mathbf{z}'_i(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix}^{-1} - (\epsilon_{i1}, \epsilon_{i2}) \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix}^{-1} \\ &= \mathbf{z}'_i(\boldsymbol{\psi}^*_1, \boldsymbol{\psi}^*_2) - (\epsilon^*_{i1}, \epsilon^*_{i2}) > (0, 0) \end{aligned} \tag{3.13}$$

$(\boldsymbol{\psi}^*_1, \boldsymbol{\psi}^*_2)$ are group 1 specific reduced form parameters, and $(\epsilon^*_{i1}, \epsilon^*_{i2})$ are group 1 specific reduced form idiosyncratic error terms, which is similar to Lee (1981), Liu (2019) and Chapters 1 and 2. The formula of $(\boldsymbol{\psi}^*_1, \boldsymbol{\psi}^*_2)$ and $(\epsilon^*_{i1}, \epsilon^*_{i2})$ can be written as

$$\begin{aligned} (\boldsymbol{\psi}^*_1, \boldsymbol{\psi}^*_2) &= (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix}^{-1} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \frac{1}{1 - \theta_{12}\theta_{21}} \begin{bmatrix} 1 & -\theta_{12} \\ -\theta_{21} & 1 \end{bmatrix} \\ &= \Big(\frac{\boldsymbol{\psi}_1 - \theta_{21}\boldsymbol{\psi}_2}{1 - \theta_{12}\theta_{21}}, \frac{\boldsymbol{\psi}_2 - \theta_{12}\boldsymbol{\psi}_1}{1 - \theta_{12}\theta_{21}} \Big) \\ (\epsilon^*_{i1}, \epsilon^*_{i2}) &= (\epsilon_{i1}, \epsilon_{i2}) \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix}^{-1} = (\epsilon_1, \epsilon_2) \frac{1}{1 - \theta_{12}\theta_{21}} \begin{bmatrix} 1 & -\theta_{12} \\ -\theta_{21} & 1 \end{bmatrix} \\ &= \Big(\frac{\epsilon_1 - \theta_{21}\epsilon_2}{1 - \theta_{12}\theta_{21}}, \frac{\epsilon_2 - \theta_{12}\epsilon_1}{1 - \theta_{12}\theta_{21}} \Big). \end{aligned} \tag{3.14}$$

Suppose for each agent $i = 1, 2, \cdots, n$, the structural form model idiosyncratic vector $(\epsilon_{i1}, \epsilon_{i2})$ has zero mean and variance-covariance matrix as

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{3.15}$$

and according to the property of joint normal distributed random vector, we can derive the variance-covariance matrix of the group 1 specific reduced form idiosyncratic vector, i.e., $(\epsilon_{i1}^*, \epsilon_{i2}^*)$, as

$$\begin{bmatrix} \sigma_1^{*2} & \rho^*\sigma_1^*\sigma_2^* \\ \rho^*\sigma_1^*\sigma_2^* & \sigma_2^{*2} \end{bmatrix} = \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix}^{\prime -1} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} \sigma_1^2 - 2\theta_{21}\rho\sigma_1\sigma_2 + \theta_{21}^2\sigma_2^2 & (1+\theta_{12}\theta_{21})\rho\sigma_1\sigma_2 - \theta_{12}\sigma_1^2 - \theta_{21}\sigma_2^2 \\ (1+\theta_{12}\theta_{21})\rho\sigma_1\sigma_2 - \theta_{12}\sigma_1^2 - \theta_{21}\sigma_2^2 & \sigma_2^2 - 2\theta_{12}\rho\sigma_1\sigma_2 + \theta_{12}^2\sigma_1^2 \end{bmatrix} \tag{3.16}$$

σ_1^*, σ_2^*, and ρ^* are the reduced form parameters related to the joint normal distribution of group 1 specific reduced form idiosyncratic vector, their formula are

$$\begin{aligned} \sigma_1^* &= \sqrt{\sigma_1^2 - 2\theta_{21}\rho\sigma_1\sigma_2 + \theta_{21}^2\sigma_2^2} \\ \sigma_2^* &= \sqrt{\sigma_2^2 - 2\theta_{12}\rho\sigma_1\sigma_2 + \theta_{12}^2\sigma_1^2} \\ \rho^* &= \frac{(1+\theta_{12}\theta_{21})\rho\sigma_1\sigma_2 - \theta_{12}\sigma_1^2 - \theta_{21}\sigma_2^2}{\sqrt{(\sigma_1^2 - 2\theta_{21}\rho\sigma_1\sigma_2 + \theta_{21}^2\sigma_2^2)(\sigma_2^2 - 2\theta_{12}\rho\sigma_1\sigma_2 + \theta_{12}^2\sigma_1^2)}}. \end{aligned} \tag{3.17}$$

Then, we can use the group 1 specific reduced form model's parameters to derive the probability that agent i belongs to group 1, as follows

$$\mathbf{Pr}(y_{i1} > 0, y_{i2} > 0) = \mathbf{Pr}(y_{i1}^* > 0, y_{i2}^* > 0) = \mathbf{Pr}(\mathbf{z}_i'\boldsymbol{\psi}_1^* > \epsilon_{i1}^*, \mathbf{z}_i'\boldsymbol{\psi}_2^* > \epsilon_{i2}^*) = \Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}, \rho^*\Big) = \Phi_2^* \tag{3.18}$$

Where $\Phi_2(\cdot,\cdot,\rho)$ is the CDF of the standard bivariate normal distributed vector with correlation coefficient ρ.

Then we consider the possible case that agent i belongs to group 2, that means the activities' outcomes are $y_{i1} > 0$ and $y_{i2} = 0$, i.e., the activities' reservation values are $y_{i1}^* > 0$ and $y_{i2}^* \leq 0$

$$\begin{aligned} y_{i1}^* &= -\theta_{21}y_{i2} + \mathbf{z}_i'\boldsymbol{\psi}_1 - \epsilon_{i1}0 \\ &= \mathbf{z}_i'\boldsymbol{\psi}_1 - \epsilon_{i1} > 0 \\ y_{i2}^* &= -\theta_{12}y_{i1} + \mathbf{z}_i'\boldsymbol{\psi}_2 - \epsilon_{i2} \\ &= -\theta_{12}y_{i1}^* + \mathbf{z}_i'\boldsymbol{\psi}_2 - \epsilon_{i2} \leq 0 \end{aligned} \tag{3.19}$$

when we write the two-equation system in a vector-matrix form, we can also derive the group 2 specific simultaneous effect matrix

$$(y_{i1}^*, y_{i2}^*)\begin{bmatrix}1 & \theta_{12}\\ 0 & 1\end{bmatrix} = \mathbf{z}_i'(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) - (\epsilon_{i1}, \epsilon_{i2}) \qquad (>0, \leq 0) \tag{3.20}$$

according to our previous assumption, if all of the group-specific simultaneous effect matrices are non-singular, then the vector-matrix form of the two-equation system can be rewritten as

$$\begin{aligned}(y_{i1}^*, y_{i2}^*) &= \mathbf{z}_i'(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\begin{bmatrix}1 & \theta_{12}\\ 0 & 1\end{bmatrix}^{-1} - (\epsilon_{i1}, \epsilon_{i2})\begin{bmatrix}1 & \theta_{12}\\ 0 & 1\end{bmatrix}^{-1} \qquad (>0, \leq 0)\\ &= \mathbf{z}_i'(\boldsymbol{\psi}_1^{**}, \boldsymbol{\psi}_2^{**}) - (\epsilon_{i1}^{**}, \epsilon_{i2}^{**}) \qquad (>0, \leq 0)\end{aligned} \tag{3.21}$$

the group 2 specific reduced form parameters $(\boldsymbol{\psi}_1^{**}, \boldsymbol{\psi}_2^{**})$ and idiosyncratic vector $(\epsilon_{i1}^{**}, \epsilon_{i2}^{**})$ can be written as

$$(\boldsymbol{\psi}_1^{**}, \boldsymbol{\psi}_2^{**}) = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\begin{bmatrix}1 & \theta_{12}\\ 0 & 1\end{bmatrix}^{-1} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2 - \theta_{12}\boldsymbol{\psi}_1) \tag{3.22}$$

$$(\epsilon_{i1}^{**}, \epsilon_{i2}^{**}) = (\epsilon_{i1}, \epsilon_{i2})\begin{bmatrix}1 & \theta_{12}\\ 0 & 1\end{bmatrix}^{-1} = (\epsilon_1, \epsilon_2 - \theta_{12}\epsilon_1) \tag{3.23}$$

according to the property of bivariate normal distribution random vector, the variance-covariance matrix of $(\epsilon_{i1}^{**}, \epsilon_{i2}^{**})$ is

$$\begin{aligned}\begin{bmatrix}\sigma_1^{**2} & \rho^{**}\sigma_1^{**}\sigma_2^{**}\\ \rho^{**}\sigma_1^{**}\sigma_2^{**} & \sigma_2^{**2}\end{bmatrix} &= \begin{bmatrix}1 & \theta_{12}\\ 0 & 1\end{bmatrix}^{\prime -1}\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\begin{bmatrix}1 & \theta_{12}\\ 0 & 1\end{bmatrix}^{-1}\\ &= \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 - \theta_{12}\sigma_1^2\\ \rho\sigma_1\sigma_2 - \theta_{12}\sigma_1^2 & \sigma_2^2 - 2\theta_{12}\rho\sigma_1\sigma_2 + \theta_{12}^2\sigma_1^2\end{bmatrix}\end{aligned} \tag{3.24}$$

σ_1^{**}, σ_2^{**}, and ρ_1^{**} are parameters related to the distribution of $(\epsilon_{i1}^{**}, \epsilon_{i2}^{**})$ and can be written as following

$$\begin{aligned}\sigma_1^{**} &= \sigma_1\\ \sigma_2^{**} &= \sqrt{\sigma_2^2 - 2\theta_{12}\rho\sigma_1\sigma_2 + \theta_{12}^2\sigma_1^2}\\ \rho^{**} &= \frac{\rho\sigma_1\sigma_2 - \theta_{12}\sigma_1^2}{\sigma_1\sqrt{\sigma_2^2 - 2\theta_{12}\rho\sigma_1\sigma_2 + \theta_{12}^2\sigma_1^2}} = \frac{\rho\sigma_2 - \theta_{12}\sigma_1}{\sqrt{\sigma_2^2 - 2\theta_{12}\rho\sigma_1\sigma_2 + \theta_{12}^2\sigma_1^2}}\end{aligned} \tag{3.25}$$

Then, we can use the group 2 specific reduced form model's parameters to derive the probability that agent i belongs to group 2, as follows

$$\begin{aligned}\mathbf{Pr}(y_{i1} > 0, y_{i2} = 0) &= \mathbf{Pr}(y^*_{i1} > 0, y^*_{i2} \leq 0)\\ &= \mathbf{Pr}(\mathbf{z}_i'\boldsymbol{\psi}_1^{**} > \epsilon_{i1}^{**}, \mathbf{z}_i'\boldsymbol{\psi}_2^{**} \leq \epsilon_{i2}^{**}) = \Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}, -\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}, -\rho^{**}\Big) = \Phi_2^{**}.\end{aligned} \tag{3.26}$$

Then we consider the possible case that agent i belongs to group 3, which means the activities' outcomes are $y_{i1} = 0$ and $y_{i2} > 0$, i.e., the activities' reservation values are $y^*_{i1} \leq 0$ and $y^*_{i2} > 0$

$$\begin{aligned} y^*_{i1} &= -\theta_{21}y_{i2} + \mathbf{z}_i'\boldsymbol{\psi}_1 - \epsilon_{i1} = -\theta_{21}y^*_{i2} + \mathbf{z}_i'\boldsymbol{\psi}_1 - \epsilon_{i1} \leq 0\\ y^*_{i2} &= -\theta_{12}y_{i1} + \mathbf{z}_i'\boldsymbol{\psi}_2 - \epsilon_{i2} = \mathbf{z}_i'\boldsymbol{\psi}_2 - \epsilon_{i2} > 0\end{aligned} \tag{3.27}$$

when we write the two-equation system in a vector-matrix form, we can also derive the group 3 specific simultaneous effect matrix in the vector-matrix form is

$$(y^*_{i1}, y^*_{i2})\begin{bmatrix} 1 & 0 \\ \theta_{21} & 1\end{bmatrix} = \mathbf{z}_i'(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) - (\epsilon_{i1}, \epsilon_{i2}) \qquad (\leq 0, > 0) \tag{3.28}$$

according to our previous assumption, if all of the group-specific simultaneous effect matrices are non-singular, then the vector-matrix form of the two-equation system can be written as

$$(y^*_{i1}, y^*_{i2}) = \mathbf{z}_i'(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\begin{bmatrix} 1 & 0 \\ \theta_{21} & 1\end{bmatrix}^{-1} - (\epsilon_{i1}, \epsilon_{i2})\begin{bmatrix} 1 & 0 \\ \theta_{21} & 1\end{bmatrix}^{-1} \qquad (\leq 0, > 0) \tag{3.29}$$

we can derive the group 3 specific parameters and idiosyncratic vector as follows

$$\begin{aligned}(\boldsymbol{\psi}_1^{***}, \boldsymbol{\psi}_2^{***}) &= (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\begin{bmatrix} 1 & 0 \\ \theta_{21} & 1\end{bmatrix}^{-1} = (\boldsymbol{\psi}_1 - \theta_{21}\boldsymbol{\psi}_2, \boldsymbol{\psi}_2)\\ (\epsilon_{i1}^{***}, \epsilon_{i2}^{***}) &= (\epsilon_{i1}, \epsilon_{i2})\begin{bmatrix} 1 & 0 \\ \theta_{21} & 1\end{bmatrix}^{-1} = (\epsilon_1 - \theta_{21}\epsilon_2, \epsilon_2)\end{aligned} \tag{3.30}$$

according to the property of bivariate normal distribution, we can derive the variance-covariance matrix of $(\epsilon_{i1}^{***}, \epsilon_{i2}^{***})$ as follows

$$\begin{bmatrix} \sigma_1^{***2} & \rho^{***}\sigma_1^{***}\sigma_2^{***} \\ \rho^{***}\sigma_1^{***}\sigma_2^{***} & \sigma_2^{***2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \theta_{21} & 1 \end{bmatrix}^{\prime -1} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \theta_{21} & 1 \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} \sigma_1^2 - 2\theta_{21}\rho\sigma_1\sigma_2 + \theta_{21}^2\sigma_2^2 & \rho\sigma_1\sigma_2 - \theta_{21}\sigma_2^2 \\ \rho\sigma_1\sigma_2 - \theta_{21}\sigma_2^2 & \sigma_2^2 \end{bmatrix} \tag{3.31}$$

σ_1^{***}, σ_2^{***}, and ρ^{***} are parameters related to the distribution of $(\epsilon_{i1}^{***}, \epsilon_{i2}^{***})$ and can be written as following

$$\begin{aligned} \sigma_1^{***} &= \sqrt{\sigma_1^2 - 2\theta_{21}\rho\sigma_1\sigma_2 + \theta_{21}^2\sigma_2^2} \\ \sigma_2^{***} &= \sigma_2 \\ \rho^{***} &= \frac{\rho\sigma_1\sigma_2 - \theta_{21}\sigma_2^2}{\sigma_2\sqrt{\sigma_1^2 - 2\theta_{21}\rho\sigma_1\sigma_2 + \theta_{21}^2\sigma_2^2}} = \frac{\rho\sigma_1 - \theta_{21}\sigma_2}{\sqrt{\sigma_1^2 - 2\theta_{21}\rho\sigma_1\sigma_2 + \theta_{21}^2\sigma_2^2}} \end{aligned} \tag{3.32}$$

Then, we can use the group 3 specific reduced form model's parameters to derive the probability that agent i belongs to group 3, as follows

$$\begin{aligned} &\mathbf{Pr}(y_{i1} = 0, y_{i2} > 0) = \mathbf{Pr}(y_{i1}^* \leq 0, y_{i2}^* > 0) \\ =&\mathbf{Pr}(\mathbf{z}_i'\boldsymbol{\psi}_1^{***} \leq \epsilon_{i1}^{***}, \mathbf{z}_i'\boldsymbol{\psi}_2^{***} > \epsilon_{i2}^{***}) = \Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}, -\rho^{***}\Big) = \Phi_2^{***}. \end{aligned} \tag{3.33}$$

The last possible case to discuss is when agent i belongs to group 4, means the activity outcomes $y_{i1} = 0$ and $y_{i2} = 0$, i.e., $y_{i1}^* \leq 0$ and $y_{i2}^* \leq 0$, we have

$$\begin{aligned} y_{i1}^* &= -\theta_{21}y_{i2} + \mathbf{z}_i'\boldsymbol{\psi}_1 - \epsilon_{i1} = \mathbf{z}_i'\boldsymbol{\psi}_1 - \epsilon_{i1} \leq 0 \\ y_{i2}^* &= -\theta_{12}y_{i1} + \mathbf{z}_i'\boldsymbol{\psi}_2 - \epsilon_{i2} = \mathbf{z}_i'\boldsymbol{\psi}_2 - \epsilon_{i2} \leq 0 \end{aligned} \tag{3.34}$$

we can find no simultaneous effect as both y_{i1} and y_{i2} are zeros. Then, we can derive the probability of agent i belongs to group 4 as follows

$$\begin{aligned} \mathbf{Pr}(y_{i1} = 0, y_{i2} = 0) &= \mathbf{Pr}(y_{i1}^* \leq 0, y_{i2}^* \leq 0) \\ &= \mathbf{Pr}(\mathbf{z}_i'\boldsymbol{\psi}_1 \leq \epsilon_{i1}, \mathbf{z}_i'\boldsymbol{\psi}_2 \leq \epsilon_{i2}) = \Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1}{\sigma_1}, -\frac{\mathbf{z}_i'\boldsymbol{\psi}_2}{\sigma_2}, \rho\Big). \end{aligned} \tag{3.35}$$

According to the results of probability for different cases and the method of conditional expectation, we can derive the rational expectation values as

$$
\begin{aligned}
p_{i1} =&(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\Phi_2^* + (\mathbf{z}_i'\boldsymbol{\psi}_1^{**})\Phi_2^{**} \\
&+ \sigma_1^*\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) + \rho^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\Big] \\
&+ \sigma_1^{**}\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}\Big)\Phi\Big(\frac{\rho^{**}\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big) \\
&- \rho^{**}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\Phi\Big(\frac{\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\Big] \\
p_{i2} =&(\mathbf{z}_i'\boldsymbol{\psi}_2^*)\Phi_2^* + (\mathbf{z}_i'\boldsymbol{\psi}_2^{***})\Phi_2^{***} \\
&+ \sigma_2^*\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) + \rho^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\Big] \\
&+ \sigma_2^{***}\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}\Big)\Phi\Big(\frac{\rho^{***}\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&- \rho^{***}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big)\Phi\Big(\frac{\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big)\Big]
\end{aligned}
\tag{3.36}
$$

Suppose we denote $\mathbf{p} = \text{vec}(\mathbf{P}) = (\mathbf{p}_1', \mathbf{p}_2')'$, and $\mathbf{p} = \vec{h}(\mathbf{p})$, then we have

$$
\vec{h}(\mathbf{p}) = [\vec{h}_1(\mathbf{p}), \vec{h}_2(\mathbf{p})] \tag{3.37}
$$

where

$$
\begin{aligned}
&\vec{h}_1(\mathbf{p}) = \left(\sum_{l=1}^{2}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*}\right) \\
&\odot\boldsymbol{\Phi}\left(\frac{\sum_{l=1}^{2}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*}}{\sigma_1^{*}}, \frac{\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*}}{\sigma_2^{*}}, \rho^{*}\right) \\
&+\left(\sum_{l=1}^{2}\lambda_{l1}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{**}\right) \\
&\odot\boldsymbol{\Phi}\left(\frac{\sum_{l=1}^{2}\lambda_{l1}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{**}}{\sigma_1^{**}}, -\frac{\sum_{l=1}^{2}\lambda_{l2}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{**}}{\sigma_2^{**}}, -\rho^{**}\right) \\
&+\sigma_1^{*}\Bigg[\phi\left(\frac{\sum_{l=1}^{2}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*}}{\sigma_1^{*}}\right) \\
&\odot\Phi\left(\frac{\sigma_1^{*}(\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*}) - \rho^{*}\sigma_2^{*}(\sum_{l=1}^{2}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*})}{\sigma_1^{*}\sigma_1^{*}\sqrt{1-\rho^{*2}}}\right) \\
&+\rho^{*}\phi\left(\frac{\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*}}{\sigma_2^{*}}\right) \\
&\odot\Phi\left(\frac{\sigma_2^{*}(\sum_{l=1}^{m}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*}) - \rho^{*}\sigma_1^{*}(\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*})}{\sigma_1^{*}\sigma_1^{*}\sqrt{1-\rho^{*2}}}\right)\Bigg] \\
&+\sigma_1^{**}\Bigg[\phi\left(\frac{\sum_{l=1}^{2}\lambda_{l1}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{**}}{\sigma_1^{**}}\right) \\
&\odot\Phi\left(\frac{\rho^{**}\sigma_2^{**}(\sum_{l=1}^{2}\lambda_{l1}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{**}) - \sigma_1^{**}(\sum_{l=1}^{2}\lambda_{l2}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{**})}{\sigma_1^{**}\sigma_1^{**}\sqrt{1-\rho^{**2}}}\right) \\
&-\rho^{**}\phi\left(\frac{\sum_{l=1}^{2}\lambda_{l2}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{**}}{\sigma_2^{**}}\right) \\
&\odot\Phi\left(\frac{\sigma_2^{**}(\sum_{l=1}^{m}\lambda_{l1}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{**}) - \rho^{**}\sigma_1^{**}(\sum_{l=1}^{2}\lambda_{l2}^{**}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{**})}{\sigma_1^{**}\sigma_1^{**}\sqrt{1-\rho^{**2}}}\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\vec{h}_2(\mathbf{p}) = &\bigg(\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*}\bigg)\\
&\odot\boldsymbol{\Phi}\bigg(\frac{\sum_{l=1}^{2}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*}}{\sigma_1^{*}}, \frac{\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*}}{\sigma_2^{*}}, \rho^{*}\bigg)\\
&+\bigg(\sum_{l=1}^{2}\lambda_{l2}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{***}\bigg)\\
&\odot\boldsymbol{\Phi}\bigg(-\frac{\sum_{l=1}^{2}\lambda_{l1}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{***}}{\sigma_1^{***}}, \frac{\sum_{l=1}^{2}\lambda_{l2}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{***}}{\sigma_2^{***}}, -\rho^{***}\bigg)\\
&+\sigma_1^{*}\Bigg[\phi\bigg(\frac{\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*}}{\sigma_2^{*}}\bigg)\\
&\odot\Phi\bigg(\frac{\sigma_2^{*}(\sum_{l=1}^{m}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*}) - \rho^{*}\sigma_1^{*}(\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*})}{\sigma_1^{*}\sigma_1^{*}\sqrt{1-\rho^{*2}}}\bigg)\\
&+\rho^{*}\phi\bigg(\frac{\sum_{l=1}^{2}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*}}{\sigma_1^{*}}\bigg)\\
&\odot\Phi\bigg(\frac{\sigma_1^{*}(\sum_{l=1}^{2}\lambda_{l2}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{*}) - \rho^{*}\sigma_2^{*}(\sum_{l=1}^{2}\lambda_{l1}^{*}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{*})}{\sigma_1^{*}\sigma_1^{*}\sqrt{1-\rho^{*2}}}\bigg)\Bigg]\\
&+\sigma_2^{***}\Bigg[\phi\bigg(\frac{\sum_{l=1}^{2}\lambda_{l2}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{***}}{\sigma_2^{***}}\bigg)\\
&\odot\Phi\bigg(\frac{\rho^{***}\sigma_1^{***}(\sum_{l=1}^{2}\lambda_{l2}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{***}) - \sigma_2^{***}(\sum_{l=1}^{m}\lambda_{l1}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{***})}{\sigma_1^{***}\sigma_1^{***}\sqrt{1-\rho^{***2}}}\bigg)\\
&-\rho^{***}\phi\bigg(\frac{\sum_{l=1}^{2}\lambda_{l1}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{***}}{\sigma_1^{***}}\bigg)\\
&\odot\Phi\bigg(\frac{\sigma_1^{***}(\sum_{l=1}^{2}\lambda_{l2}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_2^{***}) - \rho^{***}\sigma_2^{***}(\sum_{l=1}^{2}\lambda_{l1}^{***}\mathbf{W}\mathbf{p}_l + \mathbf{X}\boldsymbol{\beta}_1^{***})}{\sigma_1^{***}\sigma_1^{***}\sqrt{1-\rho^{***2}}}\bigg)\Bigg]
\end{aligned}
$$

and the first-order derivative is

$$
\frac{\partial\vec{h}(\mathbf{p})}{\partial\mathbf{p}'} = \begin{pmatrix} \frac{\partial h_1(\mathbf{p})}{\partial\mathbf{p}_1'} & \frac{\partial h_1(\mathbf{p})}{\partial\mathbf{p}_2'} \\ \frac{\partial h_2(\mathbf{p})}{\partial\mathbf{p}_1'} & \frac{\partial h_2(\mathbf{p})}{\partial\mathbf{p}_2'} \end{pmatrix}. \tag{3.38}
$$

We can derive the formula of the first-order derivative by figuring out the expression of $\partial p_{i1}/\partial\mathbf{p}'$

and $\partial p_{i2}/\partial \mathbf{p}'$. Then we can get results similar in forms to Liu (2019) but with some extra terms.

$$\begin{aligned}\frac{\partial p_{i1}}{\partial \mathbf{p}'} =&[\lambda_{11}^{*}\mathbf{w}_i, \lambda_{21}^{*}\mathbf{w}_i]\Phi_2^{*} \\ &+ \left(\mathbf{z}_i'\boldsymbol{\psi}_1^{*} - \frac{\rho^{*}\sigma_1^{*}}{\sigma_2^{*}}\mathbf{z}_i'\boldsymbol{\psi}_2^{*}\right)\frac{[\lambda_{12}^{*}\mathbf{w}_i, \lambda_{22}^{*}\mathbf{w}_i]}{\sigma_2^{*}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_2^{*}}\Big)\Phi\Big(\frac{\sigma_2^{*}\mathbf{z}_i'\boldsymbol{\psi}_1^{*} - \rho^{*}\sigma_1^{*}\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_1^{*}\sigma_2^{*}\sqrt{1-\rho^{*2}}}\Big) \\ &+ (1-\rho^{*2})\frac{\sigma_1^{*}[\lambda_{12}^{*}\mathbf{w}_i, \lambda_{22}^{*}\mathbf{w}_i]}{\sigma_2^{*}}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_1^{*}}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_2^{*}}, \rho^{*}\Big) + [\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]\Phi_2^{**} \\ &- \left(\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \frac{\rho^{**}\sigma_1^{**}}{\sigma_2^{**}}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}\right)\frac{[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\Phi\Big(\frac{\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{***}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big) \\ &- (1-\rho^{**2})\frac{\sigma_1^{**}[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}, \rho^{**}\Big) \\ =&[\lambda_{11}^{*}\mathbf{w}_i, \lambda_{21}^{*}\mathbf{w}_i]\Phi_2^{*} + [\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]\Phi_2^{**} + \mathbf{K}_{i1}\end{aligned} \tag{3.39}$$

$$\begin{aligned}\frac{\partial p_{i2}}{\partial \mathbf{p}'} =&[\lambda_{12}^{*}\mathbf{w}_i, \lambda_{22}^{*}\mathbf{w}_i]\Phi_2^{*} \\ &+ \left(\mathbf{z}_i'\boldsymbol{\psi}_2^{*} - \frac{\rho^{*}\sigma_2^{*}}{\sigma_1^{*}}\mathbf{z}_i'\boldsymbol{\psi}_1^{*}\right)\frac{[\lambda_{11}^{*}\mathbf{w}_i, \lambda_{21}^{*}\mathbf{w}_i]}{\sigma_1^{*}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_1^{*}}\Big)\Phi\Big(\frac{\sigma_1^{*}\mathbf{z}_i'\boldsymbol{\psi}_2^{*} - \rho^{*}\sigma_2^{*}\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_1^{*}\sigma_2^{*}\sqrt{1-\rho^{*2}}}\Big) \\ &+ (1-\rho^{*2})\frac{\sigma_2^{*}[\lambda_{11}^{*}\mathbf{w}_i, \lambda_{21}^{*}\mathbf{w}_i]}{\sigma_1^{*}}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_1^{*}}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_2^{*}}, \rho^{*}\Big) + [\lambda_{12}^{***}\mathbf{w}_i, \lambda_{22}^{***}\mathbf{w}_i]\Phi_2^{***} \\ &- \left(\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \frac{\rho^{***}\sigma_2^{***}}{\sigma_1^{***}}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}\right)\frac{[\lambda_{11}^{***}\mathbf{w}_i, \lambda_{21}^{***}\mathbf{w}_i]}{\sigma_1^{***}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big) \\ &\Phi\Big(\frac{\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\ &- (1-\rho^{***2})\frac{\sigma_2^{***}[\lambda_{11}^{***}\mathbf{w}_i, \lambda_{21}^{***}\mathbf{w}_i]}{\sigma_1^{***}}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}, \rho^{***}\Big) \\ =&[\lambda_{12}^{*}\mathbf{w}_i, \lambda_{22}^{*}\mathbf{w}_i]\Phi_2^{*} + [\lambda_{12}^{***}\mathbf{w}_i, \lambda_{22}^{***}\mathbf{w}_i]\Phi_2^{***} + \mathbf{K}_{i2}\end{aligned} \tag{3.40}$$

To prove both $\partial p_{i1}/\partial \mathbf{p}'$ and $\partial p_{i2}/\partial \mathbf{p}'$ are bounded, the key part is to prove the extra terms are bounded. Then, we can conclude that under our parameter space restriction condition, at least one type of norm of the first-order derivative matrix is bounded.

$$\begin{aligned}
\mathbf{K}_{i1} =& \Big(\mathbf{z}_i'\boldsymbol{\psi}_1^* - \frac{\rho^*\sigma_1^*}{\sigma_2^*}\mathbf{z}_i'\boldsymbol{\psi}_2^*\Big)\frac{[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]}{\sigma_2^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) \\
&- \Big(\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \frac{\rho^{**}\sigma_1^{**}}{\sigma_2^{**}}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}\Big)\frac{[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\Phi\Big(\frac{\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big) \\
\mathbf{K}_{i2} =& \Big(\mathbf{z}_i'\boldsymbol{\psi}_2^* - \frac{\rho^*\sigma_2^*}{\sigma_1^*}\mathbf{z}_i'\boldsymbol{\psi}_1^*\Big)\frac{[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]}{\sigma_1^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) \\
&- \Big(\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \frac{\rho^{***}\sigma_2^{***}}{\sigma_1^{***}}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}\Big)\frac{[\lambda_{11}^{***}\mathbf{w}_i, \lambda_{21}^{***}\mathbf{w}_i]}{\sigma_1^{***}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big) \\
&\Phi\Big(\frac{\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big)
\end{aligned} \tag{3.41}$$

To fulfill the condition of the existence of fixed-point in rational expectations of outcomes among different activities. We need the following assumption related to the parameter space of the model. The parameter space assumption for our model is different from what is presented in Liu (2019) as our restriction condition on the model parameter space is related to the agent's exogenous variables.

Assumption 3.2.4. *The group-specific reduced form peer effect matrix* $||\mathbf{\Lambda}^*||$, $||\mathbf{\Lambda}^{**}||$, $||\mathbf{\Lambda}^{***}||$, *the network weight matrix* $\mathbf{W}$, *and the extra term determined by both the parameters and the exogenous variables, should satisfy the any either of the following conditions,*

$$\begin{aligned}
\min\Big\{&\max\{||\mathbf{\Lambda}^*||_1, ||\mathbf{\Lambda}^{**}||_1, ||\mathbf{\Lambda}^{***}||_1\}||\mathbf{W}||_\infty[1+f(\mathbf{X},\boldsymbol{\psi})], \\
&\max\{||\mathbf{\Lambda}^*||_\infty, ||\mathbf{\Lambda}^{**}||_\infty, ||\mathbf{\Lambda}^{***}||_\infty\}||\mathbf{W}||_1[1+f(\mathbf{X},\boldsymbol{\psi})]\Big\} < 1
\end{aligned} \tag{3.42}$$

Proposition 3. *If assumption 3.2.1, assumption 3.2.3, and assumption 3.2.4 hold, then the incomplete information network game will hold the Bayesian Nash Equilibrium, and the fixed point of* $\mathbf{p} = h(\mathbf{p})$ *will exist.*

To achieve this, we can start with $[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]\mathbf{\Phi}_2^* + [\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]\mathbf{\Phi}_2^{**}$ and $[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]\mathbf{\Phi}_2^* + [\lambda_{12}^{***}\mathbf{w}_i, \lambda_{22}^{***}\mathbf{w}_i]\mathbf{\Phi}_2^{***}$ to derive the following two inequalities.

$$\begin{aligned}
&[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]\mathbf{\Phi}_2^* + [\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]\mathbf{\Phi}_2^{**} < [\max\{\lambda_{11}^*, \lambda_{11}^{**}\}\mathbf{w}_i, \max\{\lambda_{21}^*, \lambda_{21}^{**}\}\mathbf{w}_i] \\
&[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]\mathbf{\Phi}_2^* + [\lambda_{12}^{***}\mathbf{w}_i, \lambda_{22}^{***}\mathbf{w}_i]\mathbf{\Phi}_2^{***} < [\max\{\lambda_{12}^*, \lambda_{12}^{***}\}\mathbf{w}_i, \max\{\lambda_{22}^*, \lambda_{22}^{***}\}\mathbf{w}_i]
\end{aligned} \tag{3.43}$$

then we can derive the inequalities for $\mathbf{K}_{i1}$ and $\mathbf{K}_{i2}$ as following

$$\begin{aligned}\mathbf{K}_{i1} <&\sqrt{\frac{2}{\pi}}[\max\{\lambda_{11}^{*}, \lambda_{11}^{**}\}\mathbf{w}_i, \max\{\lambda_{21}^{*}, \lambda_{21}^{**}\}\mathbf{w}_i] \\ &\max\left\{\left(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_2^{*}} - \frac{\rho^{*}\sigma_1^{*}\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_2^{*2}}\right), \left(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_2^{**}} - \frac{\rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**2}}\right)\right\} \\ \mathbf{K}_{i2} <&\sqrt{\frac{2}{\pi}}[\max\{\lambda_{12}^{*}, \lambda_{12}^{***}\}\mathbf{w}_i, \max\{\lambda_{22}^{*}, \lambda_{22}^{***}\}\mathbf{w}_i] \\ &\max\left\{\left(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_1^{*}} - \frac{\rho^{*}\sigma_2^{*}\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_1^{*}}\right), \left(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_1^{***}} - \frac{\rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***2}}\right)\right\}\end{aligned} \tag{3.44}$$

That means if any of the two following conditions are satisfied, we can conclude the existence of a fixed point according to contraction mapping theorem.

$$\begin{aligned}&\max\{||\boldsymbol{\Lambda}^{*}||_1, ||\boldsymbol{\Lambda}^{**}||_1, ||\boldsymbol{\Lambda}^{***}||_1\}||\mathbf{W}||_\infty[1 + f(\mathbf{X}, \boldsymbol{\psi})] < 1 \\ &\max\{||\boldsymbol{\Lambda}^{*}||_\infty, ||\boldsymbol{\Lambda}^{**}||_\infty, ||\boldsymbol{\Lambda}^{***}||_\infty\}||\mathbf{W}||_1[1 + f(\mathbf{X}, \boldsymbol{\psi})] < 1\end{aligned} \tag{3.45}$$

Where $\mathbf{X}$ are all exogenous independent variables of all agents related to all activities. $\boldsymbol{\psi}$ contains all the parameters of the model, i.e., $\boldsymbol{\psi} = (\theta_{12}, \theta_{21}, vec(\boldsymbol{\Lambda})', \boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \sigma_1, \sigma_2, \rho)$ and

$$\begin{aligned}f(\mathbf{X}, \boldsymbol{\psi}) =&\sqrt{\frac{2}{\pi}}\max_i\left\{\max\left\{\left|\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_2^{*}} - \frac{\rho^{*}\sigma_1^{*}\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_2^{*2}}\right|, \left|\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_2^{**}} - \frac{\rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**2}}\right|,\right.\right. \\ &\left.\left.\left|\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{*}}{\sigma_1^{*}} - \frac{\rho^{*}\sigma_2^{*}\mathbf{z}_i'\boldsymbol{\psi}_1^{*}}{\sigma_1^{*}}\right|, \left|\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_1^{***}} - \frac{\rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***2}}\right|\right\}\right\}\end{aligned} \tag{3.46}$$

When the assumption 3.2.1, assumption 3.2.3, assumption 3.2.4, the contraction mapping property of $\mathbf{p} = \vec{h}(\mathbf{p})$ is ensured. The assumption 3.2.2 is the precondition for identifying the two-equation Tobit Model by Schmidt (1981). The discussion of the econometric model identification in this chapter differs from that in Chapters 1 and 2 because we don't need to consider estimating the reduced form parameters. All structural form parameters are estimated directly.

3.3 Estimation

After figuring out the parameter space condition, we derive the estimation process by Nested Pseudo Likelihood (NPL) Estimation developed by Aguirregabiria and Mira (2007) and has been applied in incomplete information network game estimation in Lin and Xu (2017) and Liu (2019). In our estimation, the true value of rational expectation value of agents' activities' outcomes cannot be observed directly. So we need to initialize $\mathbf{p}_1$ and $\mathbf{p}_2$ at the beginning. Then use the initialized $\mathbf{p} = (\mathbf{p}_1', \mathbf{p}_2')' \in [0,1]^{2n}$ to estimate model structural parameters $\boldsymbol{\psi} = (\theta_{12}, \theta_{21}, vec(\boldsymbol{\Lambda})', \boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \sigma_1, \sigma_2, \rho)$, then use the estimated model structural parameters to update the rational expectation value of agents' outcomes $\mathbf{p} = (\mathbf{p}_1', \mathbf{p}_2')'$. Then use the updated $\mathbf{p}$ to estimate $\boldsymbol{\psi}$ by maximum likelihood estimation (MLE). We repeat the previous two steps until our estimated model structural parameters converge. Here are the steps:

Step 1 Given $\mathbf{p}^{(t-1)}$, we estimate the model structural parameters $\boldsymbol{\psi}$, notation $\widehat{\boldsymbol{\psi}}^{(t)}$, by MLE, i.e., $\widehat{\boldsymbol{\psi}}^{(t)} = \arg\max \ln L(\boldsymbol{\psi}; \mathbf{p}^{(t-1)})$ where

$$
\begin{aligned}
\ln L(\boldsymbol{\psi}; \mathbf{p}^{(t-1)}) = & \sum_{i \in \text{Group 1}} \log \frac{1-\theta_{12}\theta_{21}}{\sigma_1 \sigma_2} \\
& \phi_2\Big(\frac{\lambda_{11}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{21}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_1 - \theta_{21} y_{i2} - y_{i1}}{\sigma_1}, \\
& \frac{\lambda_{12}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{22}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_2 - \theta_{12} y_{i1} - y_{i2}}{\sigma_2}, \rho\Big) \\
& + \sum_{i \in \text{Group 2}} \log \frac{1}{\sigma_1} \phi\Big(\frac{\lambda_{11}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{21}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_1 - y_{i1}}{\sigma_1}\Big) \\
& \Phi\Big(\frac{\rho(\lambda_{11}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{21}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_1 - y_{i1})}{\sigma_1\sqrt{1-\rho^2}} \\
& - \frac{\lambda_{12}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{22}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_2 - \theta_{12} y_{i1}}{\sigma_2\sqrt{1-\rho^2}}\Big) \\
& + \sum_{i \in \text{Group 3}} \log \frac{1}{\sigma_2} \phi\Big(\frac{\lambda_{12}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{22}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_2 - y_{i2}}{\sigma_2}\Big) \\
& \Phi\Big(\frac{\rho(\lambda_{12}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{22}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_2 - y_{i2})}{\sigma_2\sqrt{1-\rho^2}} - \\
& \frac{\lambda_{11}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{21}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_1 - \theta_{21} y_{i2}}{\sigma_1\sqrt{1-\rho^2}}\Big) \\
& + \sum_{i \in \text{Group 4}} \log \Phi_2\Big(-\frac{\lambda_{11}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{21}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_1}{\sigma_1}, \\
& - \frac{\lambda_{12}\sum_{j=1}^n w_{ij} p_{j1}^{(t-1)} + \lambda_{22}\sum_{j=1}^n w_{ij} p_{j2}^{(t-1)} + \mathbf{x}_i'\boldsymbol{\beta}_2}{\sigma_2}, \rho\Big)
\end{aligned}
\tag{3.47}
$$

Step 2 Given $\widehat{\boldsymbol{\psi}}^{(t)} = (\hat{\theta}_{12}^{(t)}, \hat{\theta}_{21}^{(t)}, vec(\widehat{\boldsymbol{\Lambda}}^{(t)})', \widehat{\boldsymbol{\beta}}_1^{(t)\prime}, \widehat{\boldsymbol{\beta}}_2^{(t)\prime}, \hat{\sigma}_1^{(t)}, \hat{\sigma}_2^{(t)}, \hat{\rho}^{(t)})$ and calculate $\mathbf{p}^{(t)}$ by

$$\mathbf{p}^{(t)} = \vec{h}(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\psi}}^{(t)}) = [\vec{h}_1(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\psi}}^{(t)})', \vec{h}_2(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\psi}}^{(t)})']' \tag{3.48}$$

both $\vec{h}_1(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\psi}}^{(t)})$ and $\vec{h}_2(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\psi}}^{(t)})$ are $n \times 1$ vectors. For $i = 1, \cdots, n$, the i-th element of $\vec{h}_1(\mathbf{p}^{(t-1)}; \widehat{\boldsymbol{\psi}}^{(t)})$ is

$$\begin{aligned}
p_{i1}^{(t)} =& (\sum_{l=1}^{2} \widehat{\lambda}_{l1}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{*(t)}) \Phi_2 \Big(\frac{\sum_{l=1}^2 \widehat{\lambda}_{l1}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{*(t)}}{\sigma_1^{*(t)}}, \\
& \frac{\sum_{l=1}^2 \widehat{\lambda}_{l2}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{*(t)}}{\sigma_2^{*(t)}}, \rho^{*(t)} \Big) \\
& + (\sum_{l=1}^{2} \widehat{\lambda}_{l1}^{**(t)} \sum_{j=1}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{**(t)}) \Phi_2 \Big(\frac{\sum_{l=1}^2 \widehat{\lambda}_{l1}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{**(t)}}{\sigma_1^{**(t)}}, \\
& - \frac{\mathbf{z}_i' \boldsymbol{\psi}_2^{**} = \sum_{l=1}^2 \widehat{\lambda}_{l2}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{**(t)}}{\sigma_2^{**(t)}}, -\rho^{**(t)} \Big) \\
& + \sigma_1^{*(t)} \Big[\phi \Big(\frac{\sum_{l=1}^2 \widehat{\lambda}_{l1}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{*(t)}}{\sigma_1^{*(t)}} \Big) \\
& \Phi \Big(\frac{\sigma_1^{*(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l2}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{*(t)}) - \rho^{*(t)} \sigma_2^{*(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l1}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{*(t)})}{\sigma_1^{*(t)} \sigma_2^{*(t)} \sqrt{1 - \rho^{*(t)2}}} \Big) \\
& + \rho^{*(t)} \phi \Big(\frac{\sum_{l=1}^2 \widehat{\lambda}_{l2}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{*(t)}}{\sigma_2^{*(t)}} \Big) \\
& \Phi \Big(\frac{\sigma_2^{*(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l1}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{*(t)}) - \rho^{*(t)} \sigma_1^{*(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l2}^{*(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{*(t)})}{\sigma_1^{*(t)} \sigma_2^{*(t)} \sqrt{1 - \rho^{*(t)2}}} \Big) \Big] \\
& + \sigma_1^{**(t)} \Big[\phi \Big(\frac{\sum_{l=1}^2 \widehat{\lambda}_{l1}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{**(t)}}{\sigma_1^{**(t)}} \Big) \\
& \Phi \Big(\frac{\rho^{**(t)} \sigma_2^{**(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l1}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{**(t)}) - \sigma_1^{**(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l2}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{**(t)})}{\sigma_1^{**(t)} \sigma_2^{**(t)} \sqrt{1 - \rho^{**(t)2}}} \Big) \\
& - \rho^{**(t)} \phi \Big(\frac{(\sum_{l=1}^2 \widehat{\lambda}_{l2}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{**(t)})}{\sigma_2^{**(t)}} \Big) \\
& \Phi \Big(\frac{\sigma_2^{**(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l1}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_1^{**(t)}) - \rho^{**(t)} \sigma_1^{**(t)} (\sum_{l=1}^2 \widehat{\lambda}_{l2}^{**(t)} \sum_{j=1}^n w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_2^{**(t)})}{\sigma_1^{**(t)} \sigma_2^{**(t)} \sqrt{1 - \rho^{**(t)2}}} \Big) \Big]
\end{aligned} \tag{3.49}$$

the i-th element of $\vec{h}_2(\mathbf{p}^{(t-1)};\widehat{\boldsymbol{\psi}}^{(t)})$ is

$$
\begin{aligned}
p_{i2} =&(\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{*(t)})\Phi_2\Big(\frac{\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{*(t)}}{\sigma_1^{*(t)}},\\
&\frac{\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{*(t)}}{\sigma_2^{*(t)}},\rho^{*(t)}\Big)\\
&+(\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{***(t)})\Phi_2\Big(-\frac{\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{***(t)}}{\sigma_1^{***(t)}},\\
&\frac{\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{***(t)}}{\sigma_2^{***(t)}},-\rho^{***(t)}\Big)\\
&+\sigma_2^{*(t)}\Big[\phi\Big(\frac{\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{*(t)}}{\sigma_2^{*(t)}}\Big)\\
&\Phi\Big(\frac{\sigma_2^{*(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{*(t)})-\rho^{*(t)}\sigma_1^{*(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{*(t)})}{\sigma_1^{*(t)}\sigma_2^{*(t)}\sqrt{1-\rho^{*(t)2}}}\Big)\\
&+\rho^{*(t)}\phi\Big(\frac{\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{*(t)}}{\sigma_1^{*(t)}}\Big)\\
&\Phi\Big(\frac{\sigma_1^{*(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{*(t)})-\rho^{*(t)}\sigma_2^{*(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{*(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{*(t)})}{\sigma_1^{*(t)}\sigma_2^{*(t)}\sqrt{1-\rho^{*(t)2}}}\Big)\Big]\\
&+\sigma_2^{***(t)}\Big[\phi\Big(\frac{\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{***(t)}}{\sigma_2^{***(t)}}\Big)\\
&\Phi\Big(\frac{\rho^{***(t)}\sigma_1^{***(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{***(t)})-\sigma_2^{***(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{***(t)})}{\sigma_1^{***(t)}\sigma_2^{***(t)}\sqrt{1-\rho^{***(t)2}}}\Big)\\
&-\rho^{***(t)}\phi\Big(\frac{(\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{***(t)})}{\sigma_1^{***(t)}}\Big)\\
&\Phi\Big(\frac{\sigma_1^{***(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l2}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_2^{***(t)})-\rho^{***(t)}\sigma_2^{***(t)}(\sum_{l=1}^{2}\widehat{\lambda}_{l1}^{***(t)}\sum_{j=1}^{n}w_{ij}p_{jl}^{(t-1)}+\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1^{***(t)})}{\sigma_1^{***(t)}\sigma_2^{***(t)}\sqrt{1-\rho^{***(t)2}}}\Big)\Big]
\end{aligned}
\tag{3.50}
$$

then we updated $\mathbf{p}^{(t-1)}$ in **Step 1** to $\mathbf{p}^{(t)}$. We repeat the two steps until the estimated model structural parameters $\widehat{\boldsymbol{\psi}}^{(t)}$ converge. The contraction mapping 3.37 with a fixed point is an important determinant of the nested pseudo likelihood algorithm convergence according to Kasahara and Shimotsu (2012). Given assumption 3.2.1, assumption 3.2.3, assumption 3.2.4, the contraction property of $\mathbf{p} = \vec{h}(\mathbf{p})$ is ensured. The NPL estimator is characterized by $\hat{\boldsymbol{\psi}}_k^* = \arg\max \ln L(\boldsymbol{\psi}_k^*, \widehat{\mathbf{p}})$ when NPL algorithm converges, where $\widehat{\mathbf{p}}$ is implicitly calculated through $\widehat{\mathbf{p}} = \vec{h}(\widehat{\mathbf{p}}; \widehat{\boldsymbol{\Psi}}^*)$. (The structural parameters are directly estimated in our NPL. There is no need to derive the process from estimated reduced-form parameters to structural-form parameters. This is the difference in the estimation process from Liu (2019))A Monte Carlo simulation is conducted in the next section to offer a straightforward perspective of our estimation.

3.4 Monte Carlo Simulation

3.4.1 Simulation Setup

In our simulation, we build a n-agent network, and each agent makes decisions in two activities simultaneously. The model can be expressed for each agent i in the following two equations.

$$\begin{aligned} y_{i1}^* &= -\theta_{21} y_{i2} + \lambda_{11} \sum_{j=1} w_{ij} p_{j1} + \lambda_{21} \sum_{j=1} w_{ij} p_{j2} + x_{i1}\beta_1 - \varepsilon_{i1} \\ y_{i2}^* &= -\theta_{12} y_{i1} + \lambda_{12} \sum_{j=1} w_{ij} p_{j1} + \lambda_{22} \sum_{j=1} w_{ij} p_{j2} + x_{i2}\beta_2 - \varepsilon_{i2} \end{aligned} \tag{3.51}$$

where y_{i1}^* and y_{i2}^* are reservation values of activity outcomes and can only be observed as positive results or zeros, i.e., $y_{i1} = y_{i1}^* \mathrm{I}(y_{i1}^* > 0)$ and $y_{i2} = y_{i2}^* \mathrm{I}(y_{i2}^* > 0)$. x_{i1} and x_{i2} are independent variables related to different activities. ε_{i1} and ε_{i2} are unobserved idiosyncratic terms among agents in the two activities. $(\varepsilon_{i1}, \varepsilon_{i2})'$ is identical and independently distributed, satisfying joint normal distribution with zeros means and variance-covariance matrix as

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

the vector form is

$$\begin{aligned} \mathbf{y}_1^* &= -\theta_{21}\mathbf{y}_2 + \lambda_{11}\mathbf{W}\mathbf{p}_1 + \lambda_{21}\mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_1 - \boldsymbol{\varepsilon}_1 \\ \mathbf{y}_2^* &= -\theta_{12}\mathbf{y}_1 + \lambda_{12}\mathbf{W}\mathbf{p}_1 + \lambda_{22}\mathbf{W}\mathbf{p}_2 + \mathbf{X}_2\beta_2 - \boldsymbol{\varepsilon}_2 \end{aligned} \tag{3.52}$$

where $\mathbf{y}_1^* = (y_{11}^*, \cdots, y_{n1}^*)'$, $\mathbf{y}_2^* = (y_{12}^*, \cdots, y_{n2}^*)'$, $\mathbf{y}_1 = (y_{11}, \cdots, y_{n1})'$, $\mathbf{y}_2 = (y_{12}, \cdots, y_{n2})'$, $\mathbf{p}_1 = (p_{11}, \cdots, p_{n1})'$, $\mathbf{p}_2 = (p_{12}, \cdots, p_{n2})'$, $\mathbf{X}_1 = (x_{11}, \cdots, x_{n1})'$, $\mathbf{X}_2 = (x_{12}, \cdots, x_{n2})'$, $\boldsymbol{\varepsilon}_1 = (\varepsilon_{11}, \cdots, \varepsilon_{n1})'$, $\boldsymbol{\varepsilon}_2 = (\varepsilon_{12}, \cdots, \varepsilon_{n2})'$. In the data-generating process, we need to use group-specific reduce form parameter value to iterate calculate $\mathbf{p}_1$ and $\mathbf{p}_2$, suppose we denote $\boldsymbol{\psi}_1 = (\lambda_{11}, \lambda_{21}, \beta_1, 0)'$ and $\boldsymbol{\psi}_2 = (\lambda_{12}, \lambda_{22}, 0, \beta_2)'$, and $\mathbf{z}_i = (\sum_{j=1}^n w_{ij} p_{j1}, \sum_{j=1}^n w_{ij} p_{j2}, x_{i1}, x_{i2})'$ then we will have the following reduced form parameters. The reduced form parameters for Group 1 are

$$\begin{aligned}
\boldsymbol{\psi}_1^* =& \frac{\boldsymbol{\psi}_1 - \theta_{21}\boldsymbol{\psi}_2}{1-\theta_{12}\theta_{21}} \\
\boldsymbol{\psi}_2^* =& \frac{\boldsymbol{\psi}_2 - \theta_{12}\boldsymbol{\psi}_1}{1-\theta_{12}\theta_{21}} \\
\sigma_1^* =& \sqrt{1-2\theta_{21}\rho+\theta_{21}^2} \\
\sigma_2^* =& \sqrt{1-2\theta_{12}\rho+\theta_{12}^2} \\
\rho^* =& \frac{(1+\theta_{12}\theta_{21}\rho-\theta_{12}-\theta_{21})}{\sqrt{(1-2\theta_{21}\rho+\theta_{21}^2)(1-2\theta_{12}\rho+\theta_{12}^2)}}
\end{aligned} \tag{3.53}$$

The reduced form parameters for Group 2 are

$$\begin{aligned}
\boldsymbol{\psi}_1^{**} =& \boldsymbol{\psi}_1 \\
\boldsymbol{\psi}_2^{**} =& \boldsymbol{\psi}_2 - \theta_{12}\boldsymbol{\psi}_1 \\
\sigma_1^{**} =& 1 \\
\sigma_2^{**} =& \sqrt{1-2\theta_{12}\rho+\theta_{12}^2} \\
\rho^{**} =& \frac{\rho-\theta_{12}}{\sqrt{1-2\theta_{12}\rho+\theta_{12}^2}}
\end{aligned} \tag{3.54}$$

The reduced form parameters for Group 3 are

$$\begin{aligned}
\boldsymbol{\psi}_1^{***} =& \boldsymbol{\psi}_1 - \theta_{21}\boldsymbol{\psi}_2 \\
\boldsymbol{\psi}_2^{***} =& \boldsymbol{\psi}_2 \\
\sigma_1^{***} =& \sqrt{1-2\theta_{21}\rho+\theta_{21}^2} \\
\sigma_2^{***} =& 1 \\
\rho^{***} =& \frac{\rho-\theta_{21}}{\sqrt{1-2\theta_{21}\rho+\theta_{21}^2}}
\end{aligned} \tag{3.55}$$

then, we can calculate the rational expectations by equation (3.36), until results converge. In the estimation process, when we obtain the rational expectation values, we write the log-likelihood function of structural form parameters, and we will get the following result by Amemiya (1974) and Maddala (1986)

$$
\begin{aligned}
&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\mathbf{Pr}(y_1=y_{i1},y_2=y_{i2})+d_{i1}(1-d_{i2})\ln\mathbf{Pr}(y_1=y_{i1},y_2=0)\\
&+(1-d_{i1})d_{i2}\ln\mathbf{Pr}(y_1=0,y_2=y_{i2})+(1-d_{i1})(1-d_{i2})\ln\mathbf{Pr}(y_1=0,y_2=0)\Big]\\
=&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\mathbf{Pr}(\epsilon_{i1}=\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}y_{i2}-y_{i1},\epsilon_{i2}=\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}y_{i1}-y_{i2})\\
&+d_{i1}(1-d_{i2})\ln\mathbf{Pr}(\epsilon_{i1}=\mathbf{z}_i'\boldsymbol{\psi}_1-y_{i1},\epsilon_{i2}\geq\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}y_{i1})\\
&+(1-d_{i1})d_{i2}\ln\mathbf{Pr}(\epsilon_{i1}\geq\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}y_{i2},\epsilon_{i2}=\mathbf{z}_i'\boldsymbol{\psi}_2-y_{i2})\\
&+(1-d_{i1})(1-d_{i2})\ln\mathbf{Pr}(\epsilon_{i1}\geq\mathbf{z}_i'\boldsymbol{\psi}_1,\epsilon_{i2}\geq\mathbf{z}_i'\boldsymbol{\psi}_2)\Big]\\
=&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\frac{1-\theta_{12}\theta_{21}}{\sigma_1\sigma_2}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}y_{i2}-y_{i1}}{\sigma_1},\frac{\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}y_{i1}-y_{i2}}{\sigma_2},\rho\Big)\\
&+d_{i1}(1-d_{i2})\ln\frac{1}{\sigma_1}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1-y_{i1}}{\sigma_1}\Big)\Phi\Big(\frac{\rho(\mathbf{z}_i'\boldsymbol{\psi}_1-y_{i1})}{\sigma_1\sqrt{1-\rho^2}}-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}y_{i1}}{\sigma_2\sqrt{1-\rho^2}}\Big)\\
&+(1-d_{i1})d_{i2}\ln\frac{1}{\sigma_2}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2-y_{i2}}{\sigma_2}\Big)\Phi\Big(\frac{\rho(\mathbf{z}_i'\boldsymbol{\psi}_2-y_{i2})}{\sigma_2\sqrt{1-\rho^2}}-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}y_{i2}}{\sigma_1\sqrt{1-\rho^2}}\Big)\\
&+(1-d_{i1})(1-d_{i2})\ln\Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1}{\sigma_1},-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2}{\sigma_2},\rho\Big)\Big]\\
=&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln L_1+d_{i1}(1-d_{i2})\ln L_2+(1-d_{i1})d_{i2}\ln L_3+(1-d_{i1})(1-d_{i2})\ln L_4\Big]
\end{aligned}
\tag{3.56}
$$

The Maximum Likelihood Estimation (MLE) step in our simulation is a standard version of MLE in handling the two-equation simultaneous Tobit Model in Amemiya (1974), Amemiya (1985) and Maddala (1986). We estimate the structural parameters directly in each iteration of NPL estimation, unlike the practice in Liu (2019), in which the reduced-form parameters are estimated. However, we use the group-specific reduced form in our rational expectation obtaining step for each iteration to accelerate the computation and reduce the programming complexity.

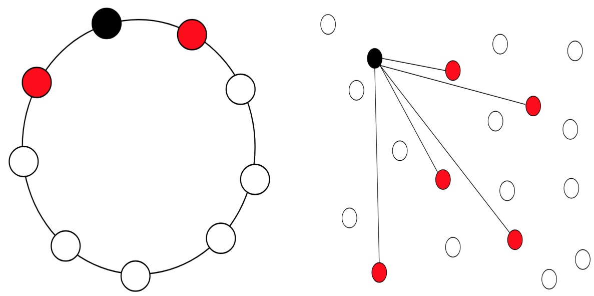
Note From a group-specific reduced form parameter perspective, the likelihood function can be written as

$$
\begin{aligned}
&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\mathbf{Pr}(y_1=y_{i1},y_2=y_{i2})+d_{i1}(1-d_{i2})\ln\mathbf{Pr}(y_1=y_{i1},y_2=0)\\
&\qquad+(1-d_{i1})d_{i2}\ln\mathbf{Pr}(y_1=0,y_2=y_{i2})+(1-d_{i1})(1-d_{i2})\ln\mathbf{Pr}(y_1=0,y_2=0)\Big]\\
=&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\mathbf{Pr}(\epsilon_{i1}^{*}=\mathbf{z}_i'\boldsymbol{\psi}_1^{*}-y_{i1},\epsilon_{i2}^{*}=\mathbf{z}_i'\boldsymbol{\psi}_2^{*}-y_{i2})\\
&\qquad+d_{i1}(1-d_{i2})\ln\mathbf{Pr}(\epsilon_{i1}^{**}=\mathbf{z}_i'\boldsymbol{\psi}_1^{**}-y_{i1},\epsilon_{i2}^{**}\geq\mathbf{z}_i'\boldsymbol{\psi}_2^{**})\\
&\qquad+(1-d_{i1})d_{i2}\ln\mathbf{Pr}(\epsilon_{i1}^{***}\geq\mathbf{z}_i'\boldsymbol{\psi}_1^{***},\epsilon_{i2}^{***}=\mathbf{z}_i'\boldsymbol{\psi}_2^{***}-y_{i2})\\
&\qquad+(1-d_{i1})(1-d_{i2})\ln\mathbf{Pr}(\epsilon_{i1}\geq\mathbf{z}_i'\boldsymbol{\psi}_1,\epsilon_{i2}\geq\mathbf{z}_i'\boldsymbol{\psi}_2)\Big]\\
=&\sum_{i=1}^{n}\Bigg\{d_{i1}d_{i2}\ln\frac{1}{\sigma_1^{*}\sigma_2^{*}}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{*}-y_{i1}}{\sigma_1^{*}},\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{*}-y_{i2}}{\sigma_2^{*}},\rho^{*}\Big)\\
&\qquad+d_{i1}(1-d_{i2})\ln\left[\frac{1}{\sigma_1^{**}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}-y_{i1}}{\sigma_1^{**}}\Big)\Phi\Big(\frac{\rho^{**}(\mathbf{z}_i'\boldsymbol{\psi}_1^{**}-y_{i1})}{\sigma_1^{**}\sqrt{1-\rho^{**2}}}-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\right]\\
&\qquad+(1-d_{i1})d_{i2}\ln\left[\frac{1}{\sigma_2^{***}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}-y_{i2}}{\sigma_2^{***}}\Big)\Phi\Big(\frac{\rho^{***}(\mathbf{z}_i'\boldsymbol{\psi}_2^{***}-y_{i2})}{\sigma_2^{***}\sqrt{1-\rho^{***2}}}-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sqrt{1-\rho^{***2}}}\Big)\right]\\
&\qquad+(1-d_{i1})(1-d_{i2})\ln\Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1}{\sigma_1},-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2}{\sigma_2},\rho\Big)\Bigg\}
\end{aligned}
\tag{3.57}
$$

The group-specific reduced-form parameters approach will be useful in rational expectation value calculation in each iteration of our NPL estimation step, following the equation (3.36). The group-specific reduced form's likelihood function will finally be used to estimate structural-form parameters only. If the number of activities increases, we can use reduced-form parameters to assist the programming process.

3.4.2 Simulation Results

The network in our simulation contains n agents, each with two exogenous independent variables. They are normally distributed with zero means and unit variance, and no correlation exists among agents' independent variables. As for the network structure, we consider both circular and random networks. In a circular network, an agent will only be friends with her neighbors. In a random network, an agent will be friends with five other agents randomly. The following are the graphs of a circular network example and a random network example.



The number of replications is 1000. The unit direct effect and the unit variance of idiosyncratic terms are applied, i.e., $\beta_1 = \beta_2 = 1$ and $\sigma_1 = \sigma_2 = 1$. Following are our results with different values in peer effect and idiosyncratic shock correlation.

• Case 1: Weak peer effect, weak error-term correlation. $\lambda_{11} = \lambda_{22} = 0.2$, $\lambda_{12} = \lambda_{21} = 0.1$, $\rho = 0.2$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.108 (0.072)	0.108 (0.071)	0.190 (0.111)	0.122 (0.110)	0.121 (0.114)	0.188 (0.126)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.997 (0.058)	0.998 (0.058)	0.996 (0.045)	0.995 (0.045)	0.211 (0.084)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.110 (0.075)	0.107 (0.076)	0.203 (0.090)	0.118 (0.092)	0.107 (0.090)	0.194 (0.093)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.997 (0.061)	0.999 (0.062)	0.996 (0.046)	0.997 (0.043)	0.209 (0.092)	

• Case 2: Medium peer effect, weak error-term correlation. $\lambda_{11} = \lambda_{22} = 0.3$, $\lambda_{12} = \lambda_{21} = 0.2$, $\rho = 0.2$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.105 (0.069)	0.108 (0.071)	0.298 (0.120)	0.203 (0.125)	0.209 (0.137)	0.300 (0.135)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.998 (0.060)	1.000 (0.060)	0.996 (0.043)	0.997 (0.043)	0.208 (0.089)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.104 (0.074)	0.102 (0.069)	0.299 (0.083)	0.205 (0.094)	0.200 (0.089)	0.298 (0.085)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.997 (0.062)	1.001 (0.058)	0.996 (0.042)	0.999 (0.043)	0.206 (0.090)	

• Case 3: Strong peer effect, weak error-term correlation. $\lambda_{11} = \lambda_{22} = 0.5$, $\lambda_{12} = \lambda_{21} = 0.3$, $\rho = 0.2$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.103 (0.061)	0.105 (0.062)	0.491 (0.106)	0.302 (0.118)	0.313 (0.116)	0.503 (0.106)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.999 (0.055)	1.000 (0.056)	0.996 (0.040)	0.996 (0.041)	0.207 (0.080)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.103 (0.063)	0.103 (0.064)	0.503 (0.052)	0.301 (0.074)	0.300 (0.073)	0.502 (0.050)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.998 (0.059)	1.000 (0.058)	0.999 (0.041)	0.997 (0.040)	0.201 (0.084)	

• Case 4: Weak peer effect, medium error-term correlation. $\lambda_{11} = \lambda_{22} = 0.2$, $\lambda_{12} = \lambda_{21} = 0.1$, $\rho = 0.5$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.108	0.107	0.193	0.117	0.118	0.194
(0.066)	(0.068)	(0.108)	(0.104)	(0.111)	(0.122)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.995	0.998	0.997	0.996	0.509	
(0.058)	(0.058)	(0.047)	(0.047)	(0.060)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.108	0.106	0.204	0.116	0.106	0.197
(0.069)	(0.072)	(0.087)	(0.090)	(0.090)	(0.090)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.996	0.998	0.996	0.996	0.506	
(0.060)	(0.062)	(0.048)	(0.047)	(0.066)	

• Case 5: Medium peer effect, medium error-term correlation. $\lambda_{11} = \lambda_{22} = 0.3$, $\lambda_{12} = \lambda_{21} = 0.2$, $\rho = 0.5$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.103	0.106	0.295	0.200	0.209	0.303
(0.063)	(0.068)	(0.112)	(0.117)	(0.131)	(0.127)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
1.001	0.998	0.999	0.996	0.507	
(0.057)	(0.060)	(0.044)	(0.045)	(0.062)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.105 (0.070)	0.105 (0.067)	0.300 (0.083)	0.204 (0.094)	0.204 (0.086)	0.299 (0.077)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.999 (0.063)	0.998 (0.060)	0.997 (0.044)	0.995 (0.045)	0.502 (0.069)	

• Case 6: Strong peer effect, medium error-term correlation. $\lambda_{11} = \lambda_{22} = 0.5$, $\lambda_{12} = \lambda_{21} = 0.3$, $\rho = 0.5$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.100 (0.057)	0.101 (0.060)	0.500 (0.089)	0.299 (0.101)	0.299 (0.105)	0.500 (0.095)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
1.001 (0.053)	1.001 (0.056)	0.995 (0.041)	0.997 (0.040)	0.500 (0.060)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.101 (0.065)	0.106 (0.069)	0.499 (0.057)	0.302 (0.084)	0.306 (0.086)	0.497 (0.062)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
1.001 (0.060)	0.999 (0.058)	0.999 (0.050)	0.997 (0.045)	0.501 (0.069)	

• Case 7: Weak peer effect, strong error-term correlation. $\lambda_{11} = \lambda_{22} = 0.2$, $\lambda_{12} = \lambda_{21} = 0.1$, $\rho = 0.8$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.105 (0.061)	0.108 (0.065)	0.195 (0.098)	0.106 (0.098)	0.114 (0.108)	0.202 (0.111)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.999 (0.057)	0.996 (0.060)	1.001 (0.052)	0.997 (0.051)	0.804 (0.029)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.106 (0.066)	0.109 (0.064)	0.203 (0.087)	0.108 (0.087)	0.108 (0.081)	0.201 (0.080)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.997 (0.061)	0.996 (0.060)	0.998 (0.052)	0.995 (0.050)	0.802 (0.031)	

• Case 8: Medium peer effect, strong error-term correlation. $\lambda_{11} = \lambda_{22} = 0.3$, $\lambda_{12} = \lambda_{21} = 0.2$, $\rho = 0.8$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.101 (0.058)	0.103 (0.062)	0.299 (0.103)	0.199 (0.111)	0.200 (0.115)	0.300 (0.112)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
1.001 (0.055)	1.000 (0.059)	0.995 (0.049)	0.996 (0.049)	0.800 (0.029)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.100 (0.063)	0.108 (0.065)	0.299 (0.075)	0.200 (0.090)	0.206 (0.093)	0.298 (0.080)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
1.002 (0.059)	0.997 (0.059)	0.999 (0.052)	0.995 (0.052)	0.801 (0.030)	

- Case 9: Strong peer effect, strong error-term correlation. $\lambda_{11} = \lambda_{22} = 0.5$, $\lambda_{12} = \lambda_{21} = 0.3$, $\rho = 0.8$

⋆ Random Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.104 (0.056)	0.103 (0.057)	0.504 (0.103)	0.305 (0.110)	0.300 (0.106)	0.499 (0.096)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.998 (0.054)	0.999 (0.054)	0.997 (0.048)	0.996 (0.047)	0.802 (0.028)	

⋆ Circular Network

$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$
0.108 (0.071)	0.106 (0.069)	0.502 (0.078)	0.304 (0.094)	0.302 (0.090)	0.502 (0.078)
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$	
0.997 (0.062)	0.999 (0.065)	1.002 (0.064)	1.005 (0.068)	0.803 (0.032)	

From the results, we could find that the estimation of peer effect in the circular network have a smaller standard deviation than that from the random network. The structure of a circular network matrix is more predictable than a random network. The rational expectation values of each agent is weighted identically in a circular network, making circular network work structure more robust to an outlier. When the peer effect is strong, the estimation of the simultaneous effect parameter performs better. Estimating the direct effect parameter is always more accurate and has less standard deviation than the simultaneous effect and peer effect parameters' estimation. The network structure's influence on estimating direct effect and error term parameters is not obvious in our simulation, neither in the bias or the standard deviation of the estimation. The increase in error-term correlation will increase the accuracy and decrease the standard deviation of the estimation result. In conclusion, the simulation results shows that the NPL estimation of model structural parameters performs well in a finite sample.

3.5 Conclusion

This paper proposes a simultaneous equation Tobit model to handle the incomplete information network game, in which the individual decision-making process is influenced by the same agent's outcomes in other activities (simultaneous effect), the rational expectation outcomes of the agents' friends (peer effect), and the agents own attributes (direct effect). The constraints of the model parameters for the existence of the fixed point in the contraction mapping and the identification of the econometric model are discussed. A nested pseudo-likelihood (NPL) estimation of the model structural parameters is developed. The Monte Carlo simulation shows the consistency. Future studies are needed to clarify the form of parameter space constraints as an expression related to the exogenous variables of agents in the network.

Bibliography

Aguirregabiria, V. and P. Mira (2007). Sequential estimation of dynamic discrete games. Econometrica 75(1), 1–53.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. Econometrica: Journal of the Econometric Society, 997–1016.

Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. Econometrica: Journal of the Econometric Society, 999–1012.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. Econometrica: Journal of the Econometric Society, 1193–1205.

Amemiya, T. (1979). The estimation of a simultaneous-equation tobit model. International economic review, 169–181.

Amemiya, T. (1985). Advanced Econometrics. Harvard university press.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's who in networks. wanted: The key player. Econometrica 74(5), 1403–1417.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear social interactions models. Journal of Political Economy 123(2), 444–496.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. The review of economic studies 76(4), 1239–1267.

Cohen-Cole, E., X. Liu, and Y. Zenou (2018). Multivariate choices and identification of social interactions. Journal of Applied Econometrics 33(2), 165–178.

Greene, W. H. (2017). Econometric analysis 8th edition. Pearson, 807–819.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. Econometrica: Journal of the Econometric Society, 931–959.

Horn, R. A. and C. R. Johnson (2012). Matrix analysis. Cambridge university press.

Hsieh, C.-S., M. D. König, and X. Liu (2022). A structural model for the coevolution of networks and behavior. Review of Economics and Statistics 104(2), 355–367.

Kasahara, H. and K. Shimotsu (2012). Sequential estimation of structural models with a fixed point constraint. Econometrica 80(5), 2303–2319.

Lee, L. (1993). Multivariate tobit models in econometrics. Handbook of statistics 11, 145–173.

Lee, L.-F. (1978). Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. International economic review, 415–433.

Lee, L.-F. (1979). Identification and estimation in binary choice models with limited (censored) dependent variables. Econometrica: Journal of the Econometric Society, 977–996.

Lee, L.-F. (1981). Simultaneous equations models with discrete and censored dependent variables. Structural analysis of discrete data with econometric applications 346, 364.

Lee, L.-F. (1982). Health and wage: a simultaneous equation model with multiple discrete indicators. International Economic Review, 199–221.

Lee, L.-f., J. Li, and X. Lin (2014). Binary choice models with social network under heterogeneous rational expectations. Review of Economics and Statistics 96(3), 402–417.

Lin, Z. and H. Xu (2017). Estimation of social-influence-dependent peer pressure in a large network game. The Econometrics Journal 20(3), S86–S102.

Liu, X. (2019). Simultaneous equations with binary outcomes and social interactions. Econometric Reviews 38(8), 921–937.

Maddala, G. and L.-F. Lee (1976). Recursive models with qualitative endogenous variables. In Annals of Economic and Social Measurement, Volume 5, number 4, pp. 525–545. NBER.

Maddala, G. S. (1986). Limited-dependent and qualitative variables in econometrics. Number 3. Cambridge university press.

Nelson, F. and L. Olson (1978). Specification and estimation of a simultaneous-equation model with limited dependent variables. International Economic Review, 695–709.

Qu, X. and L.-f. Lee (2012). Lm tests for spatial correlation in spatial models with limited dependent variables. Regional Science and Urban Economics 42(3), 430–445.

Qu, X. and L.-f. Lee (2013). Locally most powerful tests for spatial interactions in the simultaneous sar tobit model. Regional Science and Urban Economics 43(2), 307–321.

Schmidt, P. (1976). Econometrics. Marcel Dekker, New York.

Schmidt, P. (1981). Constraints on the parameters in simultaneous tobit and probit models. Structural analysis of discrete data with econometric applications 1, 423–434.

Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. Journal of the Royal Statistical Society: Series B (Methodological) 23(1), 223–229.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. The Review of Economic Studies 70(1), 147–165.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica: journal of the Econometric Society, 24–36.

Xu, X. and L.-f. Lee (2015). Maximum likelihood estimation of a spatial autoregressive tobit model. Journal of Econometrics 188(1), 264–280.

Xu, X. and L.-f. Lee (2018). Sieve maximum likelihood estimation of the spatial autoregressive tobit model. Journal of Econometrics 203(1), 96–112.

Yang, C. and L.-f. Lee (2017). Social interactions under incomplete information with heterogeneous expectations. Journal of Econometrics 198(1), 65–83.

Yang, C., L.-f. Lee, and X. Qu (2018). Tobit models with social interactions: Complete vs incomplete information. Regional Science and Urban Economics 73, 30–50.

Appendix A

Appendix to Chapter 1

Let $m = 2$, according to Lee (1978, 1979, 1981, 1982), Maddala and Lee (1976), Nelson and Olson (1978), and Liu (2019), we have the structural model as

$$\begin{aligned} \mathbf{y}_1^* &= -\mathbf{y}_2^*\theta_{21} + \mathbf{Z}_1\psi_1 - \epsilon_1 \\ \mathbf{y}_2^* &= -\mathbf{y}_1^*\theta_{12} + \mathbf{Z}_2\psi_2 - \epsilon_2 \end{aligned} \tag{A.1}$$

where $\mathbf{Z}_1 = [\mathbf{W}\mathbf{p}_1, \mathbf{W}\mathbf{p}_2, \mathbf{X}_1]$, $\mathbf{Z}_2 = [\mathbf{W}\mathbf{p}_1, \mathbf{W}\mathbf{p}_2, \mathbf{X}_2]$, and $\mathbf{Z} = [\mathbf{W}\mathbf{p}_1, \mathbf{W}\mathbf{p}_2, \mathbf{X}]$. Then, we can write the reduced form as

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{Z}\psi_1^* - \epsilon_1^* \\ \mathbf{y}_2 &= \mathbf{Z}\psi_2^* - \epsilon_2^* \end{aligned} \tag{A.2}$$

For agent i, $\mathbf{z}_i' = (\mathbf{w}_i\mathbf{p}_1, \mathbf{w}_i\mathbf{p}_2, \mathbf{x}_i')$ is the i-th row of $\mathbf{Z}$. And $\mathbf{w}_i = (w_{i1}, \cdots, w_{in})'$ is the i-th row of the network structure matrix $\mathbf{W}$. As for the random error term of agent i, $(\epsilon_{i1}, \epsilon_{i2})$, it satisfies the normal distribution $N(\mathbf{0}, \mathbf{\Sigma}^*)$, where

$$\mathbf{\Sigma}^* = \begin{bmatrix} \sigma_1^{*2} & \rho^*\sigma_1^*\sigma_2^* \\ \rho^*\sigma_1^*\sigma_2^* & \sigma_2^{*2} \end{bmatrix}$$

And the random shock vector $(\epsilon_1^{*\prime}, \epsilon_2^{*\prime})' \sim N(\mathbf{0}, \mathbf{\Sigma}^* \otimes \mathbf{I}_n)$ according to the identical independently distribution (i.i.d.) assumption among all the agents in the network. The reduced form parameters of the model of equation system (B.2), i.e., $\psi^* = (\psi_1^{*\prime}, \psi_2^{*\prime})'$ can be estimated by the NPL estimator. The estimation result is denoted as $\widehat{\psi}^* = (\widehat{\psi}_1^{*\prime}, \widehat{\psi}_2^{*\prime})'$. Suppose we denote $\widehat{\mathbf{p}} = (\widehat{\mathbf{p}}_1', \widehat{\mathbf{p}}_2')$ as the fixed point result from NPL estimator for $k = 1, 2$, where $\widehat{\mathbf{Z}} = [\mathbf{W}\widehat{\mathbf{p}}_1, \mathbf{W}\widehat{\mathbf{p}}_2, \mathbf{X}]$ and

$$\mathbf{p}_k = (\widehat{\mathbf{Z}}\widehat{\psi}_k^*) \odot \mathbf{\Phi}(\widehat{\mathbf{Z}}\widehat{\psi}_k^*/\widehat{\sigma}_k^*) + \widehat{\sigma}_k^*\phi(\widehat{\mathbf{Z}}\widehat{\psi}_k^*/\widehat{\sigma}_k^*) \tag{A.3}$$

We can get the NPL estimation result by $(\widehat{\psi}_k^{*\prime}, \widehat{\sigma}_k^*)' = \arg\max \ln L(\psi_k^*, \sigma_k^*; \widehat{\mathbf{p}})$ and

$$\ln L(\psi_k^*, \sigma_k^*; \widehat{\mathbf{p}}) = \sum_{i=1}^{n} \left\{ d_{ik} \ln \left[\phi((\widehat{\mathbf{z}}_i' \psi_k^* - y_{ik})/\sigma_k^*)/\sigma_k^* \right] + (1 - d_{ik}) \ln \left[1 - \mathbf{\Phi}(\widehat{\mathbf{z}}_i' \psi_k^* / \sigma_k^*) \right] \right\} \tag{A.4}$$

Now, suppose we introduce the vectors of ones and zeros to form the selection matrices $\mathbf{J}_1$ and $\mathbf{J}_2$, subject to $\mathbf{Z}_1 = \mathbf{Z}\mathbf{J}_1$ and $\mathbf{Z}_2 = \mathbf{Z}\mathbf{J}_2$, we can rewrite the structural model as

$$\begin{aligned} \mathbf{y}_1 &= -\mathbf{y}_2\theta_{21} + \mathbf{Z}\mathbf{J}_1\psi_1 - \epsilon_1 \\ \mathbf{y}_2 &= -\mathbf{y}_1\theta_{12} + \mathbf{Z}\mathbf{J}_2\psi_2 - \epsilon_2 \end{aligned} \tag{A.5}$$

then we put the reduced-form equations in the model (B.2) into the reorganized structural model (A.5), we can get

$$\begin{aligned} \mathbf{y}_1 &= -(\mathbf{Z}\psi_1^* - \epsilon_1^*)\theta_{21} + \mathbf{Z}\mathbf{J}_1\psi_1 - \epsilon_1 = \mathbf{Z}(-\psi_1^*\theta_{21} + \mathbf{J}_1\psi_1) + \epsilon_1^*\theta_{21} - \epsilon_1 \\ \mathbf{y}_2 &= -(\mathbf{Z}\psi_2^* - \epsilon_2^*)\theta_{12} + \mathbf{Z}\mathbf{J}_2\psi_2 - \epsilon_2 = \mathbf{Z}(-\psi_2^*\theta_{12} + \mathbf{J}_2\psi_2) + \epsilon_2^*\theta_{12} - \epsilon_2 \end{aligned} \tag{A.6}$$

then, we can derive the relation between reduced-form parameters and structural form parameters as following

$$\begin{aligned} \psi_1^* &= -\psi_2^*\theta_{21} + \mathbf{J}_1\psi_1 \\ \psi_2^* &= -\psi_1^*\theta_{12} + \mathbf{J}_2\psi_2 \end{aligned} \tag{A.7}$$

then, we can derive the relation between the estimation of reduced-form parameters and the true value of structural parameters as following

$$\begin{aligned} \widehat{\psi}_1^* &= -\widehat{\psi}_2^*\theta_{21} + \mathbf{J}_1\psi_1 + \mathbf{v}_1 \\ \widehat{\psi}_2^* &= -\widehat{\psi}_1^*\theta_{12} + \mathbf{J}_2\psi_2 + \mathbf{v}_2 \end{aligned} \tag{A.8}$$

where

$$\begin{aligned} \mathbf{v}_1 &= (\widehat{\psi}_1^* - \psi_1^*) + (\widehat{\psi}_2^* - \psi_2^*)\theta_{21} \\ \mathbf{v}_2 &= (\widehat{\psi}_2^* - \psi_2^*) + (\widehat{\psi}_1^* - \psi_1^*)\theta_{12} \end{aligned} \tag{A.9}$$

Suppose we apply Ω as the notation of the asymptotic covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$. Suppose we have $\delta_1 = (\theta_{21}, \psi_1')'$ and $\delta_2 = (\theta_{12}, \psi_2')'$. Then the estimator $\delta = (\delta_1', \delta_2')'$ is

$$\widehat{\delta} \;\;=\;\; (\widehat{\mathbf{H}}'\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\widehat{\mathbf{\Omega}}^{-1}\widehat{\psi}^* \tag{A.10}$$

where

$$\widehat{\mathbf{H}} = \begin{bmatrix} \widehat{\mathbf{H}}_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{H}}_2 \end{bmatrix} \tag{A.11}$$

in which $\widehat{\mathbf{H}}_1 = [-\widehat{\psi}_2^*, \mathbf{J}_1]$ and $\widehat{\mathbf{H}}_2 = [-\widehat{\psi}_1^*, \mathbf{J}_2]$. And $\widehat{\mathbf{\Omega}}$ is a consistent estimator of $\mathbf{\Omega}$. To derive the detailed form of $\mathbf{\Omega}$, we need to derive the asymptotic variance-covariance matrix of $(\psi_k^{*\prime}, \sigma_k^{*2})'$. We start with the first-order condition of our NPL estimator

$$\frac{\partial \ln L(\widehat{\psi}_k^*, \widehat{\sigma}_k^*; \widehat{\mathbf{p}})}{\partial \psi_k^*} = \sum_{i=1}^n \left\{ -d_{ik} \frac{(\widehat{\mathbf{z}}_i' \widehat{\psi}_k^* - y_{ik})}{\widehat{\sigma}_k^{*2}} - (1-d_{ik}) \frac{\phi(\widehat{\mathbf{z}}_i^{*\prime} \widehat{\psi}_k^* / \widehat{\sigma}_k^*)}{(1-\Phi(\widehat{\mathbf{z}}_i^{*\prime} \widehat{\psi}_k^* / \widehat{\sigma}_k^*))(\widehat{\sigma}_k^{*2})^{\frac{1}{2}}} \right\} \widehat{\mathbf{z}}_i = 0$$

$$\frac{\partial \ln L(\widehat{\psi}_k^*, \widehat{\sigma}_k^*; \widehat{\mathbf{p}})}{\partial (\sigma_k^{*2})} = \sum_{i=1}^n \left\{ d_{ik} \left[-\frac{1}{2\widehat{\sigma}_k^{*2}} + \frac{(\widehat{\mathbf{z}}_i' \widehat{\psi}_k^* - y_{ik})^2}{2(\widehat{\sigma}_k^{*2})^2} \right] + (1-d_{ik}) \frac{(\widehat{\mathbf{z}}_i' \widehat{\psi}_k^*) \phi(\widehat{\mathbf{z}}_i^{*\prime} \widehat{\psi}_k^* / \widehat{\sigma}_k^*))}{2(1-\Phi(\widehat{\mathbf{z}}_i^{*\prime} \widehat{\psi}_k^* / \widehat{\sigma}_k^*))(\widehat{\sigma}_k^{*2})^{\frac{3}{2}}} \right\} = 0$$

The first-order Taylor expansion is

$$\begin{aligned} \frac{\partial \ln L}{\partial \psi_k^*} + \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) (\widehat{\psi}_k^* - \psi_k^*) + \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) (\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) = O_p(1) \\ \frac{\partial \ln L}{\partial \sigma_k^{*2}} + \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) (\widehat{\psi}_k^* - \psi_k^*) + \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) (\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) = O_p(1) \end{aligned} \tag{A.12}$$

the matrix form is

$$\begin{pmatrix} \frac{\partial \ln L}{\partial \psi_k^*} \\ \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} + \begin{bmatrix} \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) & \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) \\ \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) & \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) \end{bmatrix} \begin{bmatrix} (\widehat{\psi}_k^* - \psi_k^*) \\ (\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) \end{bmatrix} = \begin{bmatrix} O_p(1) \\ O_p(1) \end{bmatrix} \tag{A.13}$$

then the asymptotic results can be written as

$$\begin{bmatrix} \sqrt{n}(\widehat{\psi}_k^* - \psi_k^*) \\ \sqrt{n}(\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) \end{bmatrix} \overset{A}{=} - \begin{bmatrix} \frac{1}{n} \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) & \frac{1}{n} \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) \\ \frac{1}{n} \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) & \frac{1}{n} \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) \end{bmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} \frac{\partial \ln L}{\partial \psi_k^*} \\ \frac{1}{\sqrt{n}} \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} \tag{A.14}$$

The second-order full-term derivatives are

$$
\begin{aligned}
\frac{\partial^2 L_k}{\partial \psi_k^* \partial \psi_k^{*\prime}} &= \sum_{i=1}^{n} -d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})}{\sigma_k^{*2}} \begin{pmatrix} \mathbf{0}' \\ \mathbf{w}_i \frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} \\ \mathbf{w}_i \frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}} \end{pmatrix} - (1-d_{ik})\frac{\phi_{ik}}{(1-\Phi_{ik})\sigma_k^*} \begin{pmatrix} \mathbf{0}' \\ \mathbf{w}_i \frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} \\ \mathbf{w}_i \frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}} \end{pmatrix} \\
&\quad - \frac{d_{ik}}{\sigma_k^{*2}}\mathbf{z}_i\left(\mathbf{z}_i' + \lambda_{1k}\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right) \\
&\quad - \frac{(1-d_{ik})\phi_{ik}^2}{(1-\Phi_{ik})^2\sigma_k^{*2}}\mathbf{z}_i\left(\mathbf{z}_i' + \lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right) \\
&\quad + \frac{(1-d_{ik})\phi_{ik}(\mathbf{z}_i'\psi_k^*)}{(1-\Phi_{ik})\sigma_k^{*3}}\mathbf{z}_i\left(\mathbf{z}_i' + \lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right) \\
&= \sum_{i=1}^{n} -\frac{d_{ik}}{\sigma_k^{*2}}\mathbf{z}_i\mathbf{z}_i' - \frac{(1-d_{ik})\phi_{ik}^2}{(1-\Phi_{ik})^2\sigma_k^{*2}}\mathbf{z}_i\mathbf{z}_i' + \frac{(1-d_{ik})\phi_{ik}}{(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}}\mathbf{z}_i\mathbf{z}_i' \\
&\quad - \frac{d_{ik}}{\sigma_k^{*2}}\mathbf{z}_i\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right) \\
&\quad - \frac{(1-d_{ik})\phi_{ik}^2}{(1-\Phi_{ik})^2\sigma_k^{*2}}\mathbf{z}_i\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right) \\
&\quad + \frac{(1-d_{ik})\phi_{ik}(\mathbf{z}_i'\psi_k^*)}{(1-\Phi_{ik})\sigma_k^{*3}}\mathbf{z}_i\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right) \\
&\quad - d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})}{\sigma_k^{*2}} \begin{pmatrix} \mathbf{0}' \\ \mathbf{w}_i \frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} \\ \mathbf{w}_i \frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}} \end{pmatrix} - (1-d_{ik})\frac{\phi_{ik}}{(1-\Phi_{ik})\sigma_k^*} \begin{pmatrix} \mathbf{0}' \\ \mathbf{w}_i \frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} \\ \mathbf{w}_i \frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}} \end{pmatrix} \\
&= \sum_{i=1}^{n} -\left[\frac{\Phi_{ik}}{\sigma_k^{*2}} + \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}}\mathbf{z}_i\mathbf{z}_i' - \frac{\phi_{ik}}{(\sigma_k^{*2})^{\frac{3}{2}}}\right]\mathbf{z}_i\mathbf{z}_i' \\
&\quad - \left[\frac{\Phi_{ik}}{\sigma_k^{*2}} + \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}} - \frac{\phi_{ik}(\mathbf{z}_i'\psi_k^*)}{\sigma_k^{*3}}\right]\mathbf{z}_i\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\right) \\
&\quad - \left[\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{\sigma_k^{*2}} + \frac{\phi_{ik}}{\sigma_k^*}\right] \begin{pmatrix} \mathbf{0}' \\ \mathbf{w}_i \frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} \\ \mathbf{w}_i \frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}} \end{pmatrix}
\end{aligned}
\tag{A.15}
$$

$$
\begin{aligned}
\frac{\partial^2 L}{\partial(\sigma_k^{*2})^2} =& \sum_{i=1}^{n} d_{ik}\Bigg[\frac{1}{2(\sigma_k^{*2})^2} - \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{(\sigma_k^{*2})^3} \\
&+ \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2}\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial(\sigma_k^{*2})} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial(\sigma_k^{*2})}\right)\Bigg] \\
&+ (1-d_{ik})\Bigg\{\left[\frac{(\mathbf{z}_i'\psi_k^*)^3\phi_{ik}}{4(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{7}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{4(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{4(1-\Phi_{ik})^2(\sigma_k^{*2})^3}\right] \\
&+ \left[\frac{\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})^2(\sigma_k^{*2})^2}\right]\Bigg(\lambda_{1k}^*\mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial(\sigma_k^{*2})} \\
&+ \lambda_{2k}^*\mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial(\sigma_k^{*2})}\Bigg)\Bigg\} \\
=& \sum_{i=1}^{n} -\frac{\Phi_{ik}}{2(\sigma_k^{*2})^2} + \frac{(\mathbf{z}_i'\psi_k^*)^3\phi_{ik}}{4(\sigma_k^{*2})^{\frac{7}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{4(\sigma_k^{*2})^{\frac{5}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{4(1-\Phi_{ik})(\sigma_k^{*2})^3} \\
&+ \left[\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2}\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\right] \\
&\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial(\sigma_k^{*2})} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial(\sigma_k^{*2})}\right)
\end{aligned} \tag{A.16}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \sigma_k^{*2}}\Big(\frac{\partial L}{\partial \psi_k^*}\Big) &= \sum_{i=1}^{n} d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\mathbf{z}_i}{(\sigma_k^{*2})^2} + (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} \\
&\quad - (1-d_{ik})\frac{\mathbf{z}_i}{(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{1}{2}}}\frac{\partial \phi_{ik}}{\partial(\sigma_k^{*2})} - (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}}{(1-\Phi_{ik})^2(\sigma_k^{*2})^{\frac{1}{2}}}\frac{\partial \Phi_{ik}}{\partial(\sigma_k^{*2})} \\
&= \sum_{i=1}^{n} d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\mathbf{z}_i}{(\sigma_k^{*2})^2} + (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} \\
&\quad - (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}}{(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{1}{2}}}\Bigg[\frac{(\mathbf{z}_i'\psi_k^*)^2}{2(\sigma_k^{*2})^2} - \frac{(\mathbf{z}_i'\psi_k^*)}{\sigma_k^{*2}}\Big(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \sigma_k^{*2}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \sigma_k^{*2}}\Big)\Bigg] \\
&\quad - (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}^2}{(1-\Phi_{ik})^2(\sigma_k^{*2})^{\frac{1}{2}}}\Bigg[-\frac{(\mathbf{z}_i'\psi_k^*)}{2(\sigma_k^{*2})^{\frac{3}{2}}} + \frac{1}{(\sigma_k^{*2})^{\frac{1}{2}}}\Big(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \sigma_k^{*2}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \sigma_k^{*2}}\Big)\Bigg] \\
&= \sum_{i=1}^{n} d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\mathbf{z}_i}{(\sigma_k^{*2})^2} + (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} \\
&\quad - (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}\mathbf{z}_i}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} + (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2\mathbf{z}_i}{2(1-\Phi_{ik})^2(\sigma_k^{*2})^2} \\
&\quad + (1-d_{ik})\Bigg[\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}\mathbf{z}_i}{(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{\phi_{ik}^2\mathbf{z}_i}{(1-\Phi_{ik})^2\sigma_k^{*2}}\Bigg]\Big(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \sigma_k^{*2}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \sigma_k^{*2}}\Big) \\
&= \sum_{i=1}^{n} -\frac{\mathbf{z}_i\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}\mathbf{z}_i}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2\mathbf{z}_i}{2(1-\Phi_{ik})(\sigma_k^{*2})^2} \\
&\quad + \Bigg[-\frac{\Phi_{ik}\mathbf{z}_i}{\sigma_k^{*2}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}\mathbf{z}_i}{(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{\phi_{ik}^2\mathbf{z}_i}{(1-\Phi_{ik})\sigma_k^{*2}}\Bigg]\Big(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \sigma_k^{*2}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \sigma_k^{*2}}\Big) \\
&\quad - \Bigg[\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\Phi_{ik}}{\sigma_k^{*2}} + \frac{\phi_{ik}}{(\sigma_k^{*2})^{\frac{1}{2}}}\Bigg]\Big(0, \mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \sigma_k^{*2}}, \mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \sigma_k^{*2}}\Big)'
\end{aligned} \tag{A.17}
$$

$$\begin{aligned}
\frac{\partial}{\partial \psi_k^*}\left(\frac{\partial L}{\partial \sigma_k^{*2}}\right) =& \sum_{i=1}^{n}\Bigg[d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2} + (1-d_{ik})\frac{\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} \\
& - (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} + (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2}{2(1-\Phi_{ik})^2(\sigma_k^{*2})^2}\Bigg] \\
& \left(\mathbf{z}_i + \lambda_{1k}^*\frac{\partial \mathbf{p}_1'}{\partial \psi_k^*}\mathbf{w}_i' + \lambda_{2k}^*\frac{\partial \mathbf{p}_2'}{\partial \psi_k^*}\mathbf{w}_i'\right) \\
=& \sum_{i=1}^{n} d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\mathbf{z}_i}{(\sigma_k^{*2})^2} + (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} \\
& - (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}\mathbf{z}_i}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} + (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2\mathbf{z}_i}{2(1-\Phi_{ik})^2(\sigma_k^{*2})^2} \\
& + \Bigg[d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2} + (1-d_{ik})\frac{\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} \\
& - (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} + (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2}{2(1-\Phi_{ik})^2(\sigma_k^{*2})^2}\Bigg] \\
& \left(\lambda_{1k}^*\frac{\partial \mathbf{p}_1'}{\partial \psi_k^*}\mathbf{w}_i' + \lambda_{2k}^*\frac{\partial \mathbf{p}_2'}{\partial \psi_k^*}\mathbf{w}_i'\right) \\
=& \sum_{i=1}^{n} -\frac{\mathbf{z}_i\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}\mathbf{z}_i}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2\mathbf{z}_i}{2(1-\Phi_{ik})(\sigma_k^{*2})^2} \\
& + \left[\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2} + \frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\right] \\
& \left(\lambda_{1k}^*\frac{\partial \mathbf{p}_1'}{\partial \psi_k^*}\mathbf{w}_i' + \lambda_{2k}^*\frac{\partial \mathbf{p}_2'}{\partial \psi_k^*}\mathbf{w}_i'\right)
\end{aligned} \tag{A.18}$$

According to $p_{ik} = (\mathbf{x}_i'\psi_k^*)\Phi_{ik} + \sigma_k^*\phi_{ik}$, we can derive $\partial \mathbf{p}_k/\partial \psi_l^*$ and $\partial \mathbf{p}_k/\partial \sigma_l^{*2}$. ($k = 1, 2$ and $l = 1, 2$)

$$\begin{aligned}
\frac{\partial \mathbf{p}_1}{\partial \psi_1^{*\prime}} &= \mathbf{K}_1^{-1}\mathbf{A}_1\mathbf{Z} \\
\frac{\partial \mathbf{p}_2}{\partial \psi_2^{*\prime}} &= \mathbf{K}_2^{-1}\mathbf{A}_2\mathbf{Z} \\
\frac{\partial \mathbf{p}_1}{\partial \psi_2^{*\prime}} &= (\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})\lambda_{21}^*\mathbf{A}_1\mathbf{W}\mathbf{K}_2^{-1}\mathbf{A}_2\mathbf{Z} \\
\frac{\partial \mathbf{p}_2}{\partial \psi_1^{*\prime}} &= (\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})\lambda_{12}^*\mathbf{A}_2\mathbf{W}\mathbf{K}_1^{-1}\mathbf{A}_1\mathbf{Z} \\
\frac{\partial \mathbf{p}_1}{\partial \sigma_1^{*2}} &= \mathbf{K}_1^{-1}\frac{\phi_1}{2(\sigma_1^{*2})^{1/2}} \\
\frac{\partial \mathbf{p}_2}{\partial \sigma_2^{*2}} &= \mathbf{K}_2^{-1}\frac{\phi_2}{2(\sigma_2^{*2})^{1/2}} \\
\frac{\partial \mathbf{p}_1}{\partial \sigma_2^{*2}} &= (\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})\lambda_{21}^*\mathbf{A}_1\mathbf{W}\mathbf{K}_2^{-1}\frac{\phi_2}{2(\sigma_2^{*2})^{1/2}} \\
\frac{\partial \mathbf{p}_2}{\partial \sigma_1^{*2}} &= (\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})\lambda_{12}^*\mathbf{A}_2\mathbf{W}\mathbf{K}_1^{-1}\frac{\phi_1}{2(\sigma_1^{*2})^{1/2}}
\end{aligned} \tag{A.19}$$

where

$$\mathbf{K}_1 = \mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{A}_1\mathbf{W}(\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})^{-1}\mathbf{A}_2\mathbf{W}$$
$$\mathbf{K}_2 = \mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{A}_2\mathbf{W}(\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})^{-1}\mathbf{A}_1\mathbf{W}$$
$$\mathbf{A}_1 = \mathrm{diag}(\Phi_{11}, \Phi_{21}, \cdots \Phi_{n1})$$
$$\mathbf{A}_2 = \mathrm{diag}(\Phi_{12}, \Phi_{22}, \cdots \Phi_{n2}) \tag{A.20}$$

Then we can derive $\partial(\partial \ln L/\partial\psi_k^*)/\partial\psi_k^*$, $\partial(\partial \ln L/\partial\psi_k^*)/\partial\sigma_k^{*2}$, $\partial(\partial \ln L/\partial\sigma_k^{*2})/\partial\psi_k^*$, and $\partial(\partial \ln L/\partial\sigma_k^{*2})/\partial\sigma_k^{*2}$ by previous results. According to the algebra results, we can derive the asymptotic variance of $(\widehat{\psi}_1^{*\prime}, \widehat{\sigma}_1^{*2}, \widehat{\psi}_2^{*\prime}, \widehat{\sigma}_2^{*2})'$ is

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{bmatrix} \tag{A.21}$$

as for the diagonal element of $\mathbf{V}$, i.e., $\mathbf{V}_{kk}$, can be derived as following

$$\mathbf{V}_{kk} = \begin{bmatrix} \mathbf{A}_{kk} & \mathbf{b}_{kk} \\ \mathbf{c}_{kk}' & g_{kk} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}_{kk} & \mathbf{p}_{kk} \\ \mathbf{q}_{kk}' & s_{kk} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{kk} & \mathbf{b}_{kk} \\ \mathbf{c}_{kk}' & g_{kk} \end{bmatrix}^{\prime -1} \tag{A.22}$$

all the elements to derive $\mathbf{V}_{kk}$, i.e., $\mathbf{A}_{kk}$, $\mathbf{b}_{kk}$, $\mathbf{c}_{kk}$, g_{kk}, $\mathbf{H}_{kk}$, $\mathbf{p}_{kk}$, $\mathbf{q}_{kk}$, and s_{kk} can be derived as following

$$\mathbf{A}_{kk} = \mathbf{Z}'\left[\mathrm{diag}_{i=1}^n\left(\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{\Phi_{ik}}{\sigma_k^{*2}} - \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}}\right)\right]\left(\mathbf{Z} + \lambda_{11}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\psi_k^{*\prime}} + \lambda_{21}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\psi_k^{*\prime}}\right) \tag{A.23}$$

$$\begin{aligned}\mathbf{b}_{kk} =& \mathbf{Z}'\left[\mathrm{diag}_{i=1}^n\left(-\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\right)\right]\iota_n \\ &+ \mathbf{Z}'\left[\mathrm{diag}_{i=1}^n\left(-\frac{\Phi_{ik}}{\sigma_k^{*2}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}}\right)\right]\left(\lambda_{1k}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\sigma_k^{*2}} + \lambda_{2k}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\sigma_k^{*2}}\right) \\ &- \left(\mathbf{0}, \mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\sigma_k^{*2}}, \mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\sigma_k^{*2}}\right)'\left[\mathrm{diag}_{i=1}^n\left(\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\Phi_{ik}}{\sigma_k^{*2}} + \frac{\phi_{ik}}{(\sigma_k^{*2})^{\frac{1}{2}}}\right)\right]\iota_n\end{aligned} \tag{A.24}$$

$$\begin{aligned}\mathbf{c}_{kk} =& \mathbf{Z}'\left[\mathrm{diag}_{i=1}^n\left(-\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\right)\right]\iota_n \\ &+ \left(\lambda_{1k}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\psi_k^{*\prime}}\right)' \\ &\left[\mathrm{diag}_{i=1}^n\left(\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2} + \frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\right)\right]\iota_n\end{aligned} \tag{A.25}$$

$$
\begin{aligned}
g_{kk} =& \iota_n'\Big[\text{diag}_{i=1}^{1}\Big(-\frac{\Phi_{ik}}{2(\sigma_k^{*2})^2} + \frac{(\mathbf{z}_i'\psi_k^*)^3\phi_{ik}}{4(\sigma_k^{*2})^{\frac{7}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{4(\sigma_k^{*2})^{\frac{5}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{4(1-\Phi_{ik})(\sigma_k^{*2})^3}\Big)\Big]\iota_n \\
&+ \iota_n'\Big[\text{diag}_{i=1}^{n}\Big(\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2}\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\Big)\Big] \\
&\Big(\lambda_{1k}^*\mathbf{W}\frac{\partial \mathbf{p}_1}{\partial(\sigma_k^{*2})} + \lambda_{2k}^*\mathbf{W}\frac{\partial \mathbf{p}_2}{\partial(\sigma_k^{*2})}\Big)
\end{aligned} \tag{A.26}
$$

$$
\mathbf{H}_{kk} = \mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2\Phi_{ik}}{(\sigma_k^{*2})^2} + \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}}\Big)\Big]\mathbf{Z} \tag{A.27}
$$

$$
\mathbf{p}_{kk} = \mathbf{q}_{kk} = \mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\Phi_{ik}}{2(\sigma_k^{*2})^2} - \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^3\Phi_{ik}}{2(\sigma_k^{*2})^3} - \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\Big)\Big]\iota_n \tag{A.28}
$$

$$
s_{kk} = \iota_n'\Big[\text{diag}_{i=1}^{n}\Big(\frac{\Phi_{ik}}{4(\sigma_k^{*2})^4} - \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2\Phi_{ik}}{2(\sigma_k^{*2})^3} + \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^4\Phi_{ik}}{4(\sigma_k^{*2})^4} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^3}\Big)\Big]\iota_n \tag{A.29}
$$

and the non-diagonal elements, i.e., $\mathbf{V}_{kl}$ can be derived as following

$$
\mathbf{V}_{kl} = \begin{bmatrix}\mathbf{A}_{kk} & \mathbf{b}_{kk} \\ \mathbf{c}_{kk}' & g_{kk}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{H}_{kl} & \mathbf{p}_{kl} \\ \mathbf{q}_{kl}' & s_{kl}\end{bmatrix}\begin{bmatrix}\mathbf{A}_{ll} & \mathbf{b}_{ll} \\ \mathbf{c}_{ll}' & g_{ll}\end{bmatrix}^{\prime -1} \tag{A.30}
$$

and the elements $\mathbf{H}_{kl}$, $\mathbf{p}_{kl}$, $\mathbf{q}_{kl}$, and s_{kl} can be derived as following

$$
\begin{aligned}
\mathbf{H}_{kl} =& \mathbf{Z}'\Big\{\text{diag}_{i=1}^{n}\Big[\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})(\mathbf{z}_i'\psi_l^* - y_{il})}{\sigma_k^{*2}\sigma_l^{*2}}\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big) \\
&+ \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\phi_{il}}{(1-\Phi_{il})\sigma_k^{*2}\sigma_l^*}\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\
&+ \frac{(\mathbf{z}_i'\psi_l^* - y_{il})\phi_{ik}}{(1-\Phi_{ik})\sigma_k^*\sigma_l^{*2}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\
&+ \frac{\phi_{ik}\phi_{il}}{(1-\Phi_{ik})(1-\Phi_{il})\sigma_k^*\sigma_l^*}\Big]\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big)\Big\}\mathbf{Z}
\end{aligned} \tag{A.31}
$$

$$
\begin{aligned}
\mathbf{p}_{kl} =& -\mathbf{Z}'\text{diag}_{i=1}^{n}\Big\{\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})}{\sigma_k^{*2}}\Big[\frac{(\mathbf{z}_i'\psi_l^* - y_{il})^2}{2(\sigma_l^{*2})^2} - \frac{1}{2\sigma_l^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big) \\
&+ \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})(\mathbf{z}_i'\psi_l^*)\phi_{il}}{2(1-\Phi_{il})\sigma_k^{*2}(\sigma_l^{*2})^{\frac{3}{2}}}\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\
&+ \frac{\phi_{ik}}{(1-\Phi_{ik})\sigma_k^*}\Big[\frac{(\mathbf{z}_i'\psi_l^* - y_{il})^2}{2(\sigma_l^{*2})^2} - \frac{1}{2\sigma_l^{*2}}\Big]\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\
&+ \frac{(\mathbf{z}_i'\psi_l^*)\phi_{ik}\phi_{il}}{2(1-\Phi_{ik})(1-\Phi_{il})\sigma_k^*(\sigma_l^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big)\Big\}\iota_n
\end{aligned} \tag{A.32}
$$

$$\begin{aligned}\mathbf{q}_{kl} = & - \mathbf{Z}'\text{diag}_{i=1}^{n}\Big\{ \frac{(\mathbf{z}_i'\psi_l^* - y_{il})}{\sigma_l^{*2}}\Big[\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{2(\sigma_k^{*2})^2} - \frac{1}{2\sigma_k^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big) \\ & + \frac{\phi_{il}}{(1-\Phi_{il})\sigma_l^*}\Big[\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{2(\sigma_k^{*2})^2} - \frac{1}{2\sigma_k^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\ & + \frac{(\mathbf{z}_i'\psi_l^* - y_{il})(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})\sigma_l^{*2}(\sigma_k^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\ & + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}\phi_{il}}{2(1-\Phi_{ik})(1-\Phi_{il})\sigma_l^*(\sigma_k^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big)\Big\}\iota_n\end{aligned} \tag{A.33}$$

$$\begin{aligned}s_{kl} = & \iota_n'\text{diag}_{i=1}^{n}\Big\{\Big[\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{2(\sigma_k^{*2})^2} - \frac{1}{2\sigma_k^{*2}}\Big]\Big[\frac{(\mathbf{z}_i'\psi_l^* - y_{il})^2}{2(\sigma_l^{*2})^2} - \frac{1}{2\sigma_l^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big) \\ & + \Big[\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{2(\sigma_k^{*2})^2} - \frac{1}{2\sigma_k^{*2}}\Big]\Big[\frac{(\mathbf{z}_i'\psi_l^*)\phi_{il}}{2(1-\Phi_{il})(\sigma_l^{*2})^{\frac{3}{2}}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\ & + \Big[\frac{(\mathbf{z}_i'\psi_l^* - y_{il})^2}{2(\sigma_l^{*2})^2} - \frac{1}{2\sigma_l^{*2}}\Big]\Big[\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}}\Big]\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, \frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, -\rho^*\Big) \\ & \frac{(\mathbf{z}_i'\psi_k^*)(\mathbf{z}_i'\psi_l^*)\phi_{ik}\phi_{il}}{4(1-\Phi_{ik})(1-\Phi_{il})(\sigma_k^{*2})^{\frac{3}{2}}(\sigma_l^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*}, -\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*}, \rho^*\Big)\Big\}\iota_n\end{aligned} \tag{A.34}$$

where $\Phi_2(\cdot,\cdot,\rho)$ is the standard bivariate normal distribution C.D.F. with a coefficient ρ. Then we need to figure out the asymptotic variance-covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$. For notation convenience, we redefine the asymptotic covariance matrix of $\widehat{\psi}^* = (\widehat{\psi}_1^{*\prime}, \widehat{\psi}_2^{*\prime})'$ is

$$\widetilde{\mathbf{V}} = \begin{bmatrix} \widetilde{\mathbf{V}}_{11} & \widetilde{\mathbf{V}}_{12} \\ \widetilde{\mathbf{V}}_{12}' & \widetilde{\mathbf{V}}_{22} \end{bmatrix} \tag{A.35}$$

where $\widetilde{\mathbf{V}}_{kk}$ is the upper-left corner sub-matrix of $\mathbf{V}_{kk}$, i.e., only remove the last column and the last row from $\mathbf{V}_{kk}$. And $\widetilde{\mathbf{V}}_{kl}$ is the upper-left corner sub-matrix of $\mathbf{V}_{kl}$, i.e., only remove the last column and the last row from $\mathbf{V}_{kl}$. Therefore, from Liu (2019), the asymptotic covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$ is

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{bmatrix} \tag{A.36}$$

where

$$\Omega_{11} = \widetilde{\mathbf{V}}_{11} + \theta_{21}^2\widetilde{\mathbf{V}}_{22} + \theta_{21}(\widetilde{\mathbf{V}}_{12} + \widetilde{\mathbf{V}}_{12}') \tag{A.37}$$

$$\Omega_{22} = \widetilde{\mathbf{V}}_{22} + \theta_{12}^2\widetilde{\mathbf{V}}_{11} + \theta_{12}(\widetilde{\mathbf{V}}_{12} + \widetilde{\mathbf{V}}_{12}') \tag{A.38}$$

$$\Omega_{12} = \theta_{12}\widetilde{\mathbf{V}}_{11} + \theta_{21}\widetilde{\mathbf{V}}_{22} + \widetilde{\mathbf{V}}_{12} + \theta_{12}\theta_{21}\widetilde{\mathbf{V}}_{12}' \tag{A.39}$$

Appendix B

Appendix to Chapter 2

To derive the asymptotic variance covariance matrix of the AGLS estimator, we start with $m = 2$ case, in which $y_{i1} = y^*_{i1}\mathbf{I}(y^*_{i1} > 0)$ and $y_{i2} = \mathbf{I}(y^*_{i2} > 0)$ for $i = 1, \cdots, n$. Let $\mathbf{y}^*_1 = (y^*_{11}, \cdots, y^*_{n1})'$ and $\mathbf{y}^*_2 = (y^*_{12}, \cdots, y^*_{n2})'$, according to Lee (1978, 1979, 1981, 1982), Maddala and Lee (1976), Nelson and Olson (1978) , and Liu (2019), we have the following vector form structural model

$$\begin{aligned} \mathbf{y}^*_1 &= -\theta_{21}\mathbf{y}^*_2 + \mathbf{Z}_1\boldsymbol{\psi}_1 - \boldsymbol{\epsilon}_1 \\ \mathbf{y}^*_2 &= -\theta_{12}\mathbf{y}^*_1 + \mathbf{Z}_2\boldsymbol{\psi}_2 - \boldsymbol{\epsilon}_2 \end{aligned} \tag{B.1}$$

where $\mathbf{Z}_1 = [\mathbf{W}\mathbf{p}_1, \mathbf{W}\mathbf{p}_2, \mathbf{X}_1]$, $\mathbf{Z}_2 = [\mathbf{W}\mathbf{p}_1, \mathbf{W}\mathbf{p}_2, \mathbf{X}_2]$, and $\mathbf{Z} = [\mathbf{W}\mathbf{p}_1, \mathbf{W}\mathbf{p}_2, \mathbf{X}]$. Then, we can write the reduced form as

$$\begin{aligned} \mathbf{y}^*_1 &= \mathbf{Z}\boldsymbol{\psi}^*_1 - \boldsymbol{\epsilon}^*_1 \\ \mathbf{y}^*_2 &= \mathbf{Z}\boldsymbol{\psi}^*_2 - \boldsymbol{\epsilon}^*_2 \end{aligned} \tag{B.2}$$

where $\boldsymbol{\psi}^*_1 = (\boldsymbol{\psi}_1 - \theta_{21}\boldsymbol{\psi}_2)/(1 - \theta_{12}\theta_{21})$ and $\boldsymbol{\psi}^*_2 = (\boldsymbol{\psi}_2 - \theta_{12}\boldsymbol{\psi}_1)/(1 - \theta_{12}\theta_{21})$; $\boldsymbol{\epsilon}^*_1 = (\boldsymbol{\epsilon}_1 - \theta_{21}\boldsymbol{\epsilon}_2)/(1 - \theta_{12}\theta_{21})$ and $\boldsymbol{\epsilon}^*_2 = (\boldsymbol{\epsilon}_2 - \theta_{12}\boldsymbol{\epsilon}_1)/(1 - \theta_{12}\theta_{21})$. And for each agent i, the reduced form model is

$$\begin{aligned} y^*_{i1} &= \mathbf{z}'_i\boldsymbol{\psi}^*_1 - \epsilon^*_{i1} \\ y^*_{i2} &= \mathbf{z}'_i\boldsymbol{\psi}^*_2 - \epsilon^*_{i2} \end{aligned} \tag{B.3}$$

where $\mathbf{z}'_i = (\mathbf{w}_i\mathbf{p}_1, \mathbf{w}_i\mathbf{p}_2, \mathbf{x}'_i)$ is the i-th row of $\mathbf{Z}$. And $\mathbf{w}_i = (w_{i1}, \cdots, w_{in})'$ is the i-th row of the network structure matrix $\mathbf{W}$. The reduced form random error vector for each agent i, i.e., $(\epsilon_{i1}, \epsilon_{i2})'$ satisfies the joint normal distribution, identically and independently, with zero means and variance covariance matrix as following

$$\mathbf{\Sigma}^* = \begin{bmatrix} \sigma_1^{*2} & \rho^*\sigma_1^* \\ \rho^*\sigma_1^* & 1 \end{bmatrix}$$

the unit variance of ϵ_{i2}^* is from our assumption with respect to binary-outcome activities' reduced-form random shock.

The reduce form parameters, $(\boldsymbol{\psi}_1^{*\prime}, \sigma_1^{*2})$ and $\boldsymbol{\psi}_2^*$ can be estimated through our Nested Pseudo Likelihood (NPL) algorithm. Let $\widehat{\boldsymbol{\psi}}_1^*$ and $\widehat{\boldsymbol{\psi}}_2^*$ represent the estimation result of $\boldsymbol{\psi}_1^*$ and $\boldsymbol{\psi}_2^*$. We denote $\widehat{\mathbf{p}}_1$ and $\widehat{\mathbf{p}}_2$ as the fixed point calculated from the NPL algorithm

$$\begin{aligned} \widehat{\mathbf{p}}_1 &= (\widehat{\mathbf{Z}}\widehat{\boldsymbol{\psi}}_1^*) \odot \Phi(\widehat{\mathbf{Z}}\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*) + \widehat{\sigma}_1^*\phi(\widehat{\mathbf{Z}}\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*) \\ \widehat{\mathbf{p}}_2 &= \Phi(\widehat{\mathbf{Z}}\widehat{\boldsymbol{\psi}}_2^*) \end{aligned} \tag{B.4}$$

where $\odot$ is the Hadamard (Schur) product discussed in section 7.5 of Horn and Johnson (2012). $\Phi(\cdot)$ and $\phi(\cdot)$ are CDF and PDF of the standard normal distribution function. And for each agent i, we have the following results

$$\begin{aligned} \widehat{p}_{i1} &= (\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_1^*)\Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*) + \widehat{\sigma}_1^*\phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*) \\ \widehat{p}_{i2} &= \Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_2^*) \end{aligned} \tag{B.5}$$

Let $\widehat{\mathbf{p}} = (\widehat{\mathbf{p}}_1', \widehat{\mathbf{p}}_2')$, the results from NPL algorithm, i.e., $(\widehat{\boldsymbol{\psi}}_1^{*\prime}, \widehat{\sigma}_1^{*2})' = \arg\max \ln L_1(\boldsymbol{\psi}_1^*, \sigma_1^{*2}; \widehat{\mathbf{p}})$ and

$$\begin{aligned} \ln L_1(\boldsymbol{\psi}_1^*, \sigma_1^{*2}; \widehat{\mathbf{p}}) &= \sum_{i=1}^{n} \Big\{ \mathbf{I}(y_{i1} > 0) \ln\left[\phi((\widehat{\mathbf{z}}_i'\boldsymbol{\psi}_1^* - y_{i1})/\sigma_1^*)/\sigma_1^*\right] + \mathbf{I}(y_{i1} = 0) \ln\left[1 - \Phi(\widehat{\mathbf{z}}_i'\boldsymbol{\psi}_1^*/\sigma_1^*)\right] \Big\} \\ \ln L_2(\boldsymbol{\psi}_2^*; \widehat{\mathbf{p}}) &= \sum_{i=1}^{n} \Big\{ y_{i2} \ln\left[\Phi(\widehat{\mathbf{z}}_i'\boldsymbol{\psi}_2^*)\right] + (1 - y_{i2}) \ln\left[1 - \Phi(\widehat{\mathbf{z}}_i'\boldsymbol{\psi}_2^*)\right] \Big\} \end{aligned} \tag{B.6}$$

We introduce matrix contains only zero or one as its element value, i.e., selection matrix – $\mathbf{J}_1$ and $\mathbf{J}_2$ where $\mathbf{Z}_1 = \mathbf{Z}\mathbf{J}_1$ and $\mathbf{Z}_2 = \mathbf{Z}\mathbf{J}_2$, then we can rewrite the structural model as

$$\begin{aligned} \mathbf{y}_1^* &= -\mathbf{y}_2^*\theta_{21} + \mathbf{Z}\mathbf{J}_1\boldsymbol{\psi}_1 - \boldsymbol{\epsilon}_1 \\ \mathbf{y}_2^* &= -\mathbf{y}_1^*\theta_{12} + \mathbf{Z}\mathbf{J}_2\boldsymbol{\psi}_2 - \boldsymbol{\epsilon}_2 \end{aligned} \tag{B.7}$$

then we combine the two equation system of the reduced-form model and the structural-form model, we can get

$$\begin{aligned}
\mathbf{y}_1 &= -(\mathbf{Z}\boldsymbol{\psi}_1^* - \boldsymbol{\epsilon}_1^*)\theta_{21} + \mathbf{Z}\mathbf{J}_1\boldsymbol{\psi}_1 - \boldsymbol{\epsilon}_1 = \mathbf{Z}(-\boldsymbol{\psi}_1^*\theta_{21} + \mathbf{J}_1\boldsymbol{\psi}_1) + \boldsymbol{\epsilon}_1^*\theta_{21} - \boldsymbol{\epsilon}_1 \\
\mathbf{y}_2 &= -(\mathbf{Z}\boldsymbol{\psi}_2^* - \boldsymbol{\epsilon}_2^*)\theta_{12} + \mathbf{Z}\mathbf{J}_2\boldsymbol{\psi}_2 - \boldsymbol{\epsilon}_2 = \mathbf{Z}(-\boldsymbol{\psi}_2^*\theta_{12} + \mathbf{J}_2\boldsymbol{\psi}_2) + \boldsymbol{\epsilon}_2^*\theta_{12} - \boldsymbol{\epsilon}_2
\end{aligned} \tag{B.8}$$

we combine this result with the reduced-form model, we can figure out the relation between the reduced-form model's parameters and the structural-form model parameters

$$\begin{aligned}
\boldsymbol{\psi}_1^* &= -\boldsymbol{\psi}_2^*\theta_{21} + \mathbf{J}_1\boldsymbol{\psi}_1 \\
\boldsymbol{\psi}_2^* &= -\boldsymbol{\psi}_1^*\theta_{12} + \mathbf{J}_2\boldsymbol{\psi}_2
\end{aligned} \tag{B.9}$$

then, we can derive the relation between the reduced-form parameters estimation results and the true values of the structural parameters as following

$$\begin{aligned}
\widehat{\boldsymbol{\psi}}_1^* &= -\widehat{\boldsymbol{\psi}}_2^*\theta_{21} + \mathbf{J}_1\boldsymbol{\psi}_1 + \mathbf{v}_1 \\
\widehat{\boldsymbol{\psi}}_2^* &= -\widehat{\boldsymbol{\psi}}_1^*\theta_{12} + \mathbf{J}_2\boldsymbol{\psi}_2 + \mathbf{v}_2
\end{aligned} \tag{B.10}$$

where

$$\begin{aligned}
\mathbf{v}_1 &= (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) + (\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*)\theta_{21} \\
\mathbf{v}_2 &= (\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*) + (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*)\theta_{12}
\end{aligned} \tag{B.11}$$

We denote $\boldsymbol{\Omega}$ as the asymptotic variance covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$. We also denote structural parameters in vector form as $\boldsymbol{\delta}_1 = (\theta_{21}, \boldsymbol{\psi}_1')'$ and $\boldsymbol{\delta}_2 = (\theta_{12}, \boldsymbol{\psi}_2')'$. According to the equation system (B.11), we can derive an SUR-type AGLS to estimate $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2')'$ as

$$\widehat{\boldsymbol{\delta}} = (\widehat{\mathbf{H}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\psi}}^* \tag{B.12}$$

where

$$\widehat{\mathbf{H}} = \begin{bmatrix} \widehat{\mathbf{H}}_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{H}}_2 \end{bmatrix} \tag{B.13}$$

in which $\widehat{\mathbf{H}}_1 = [-\widehat{\boldsymbol{\psi}}_2^*, \mathbf{J}_1]$ and $\widehat{\mathbf{H}}_2 = [-\widehat{\boldsymbol{\psi}}_1^*, \mathbf{J}_2]$. And $\widehat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$. We need the asymptotic variance-covariance matrix of $(\boldsymbol{\psi}_1^{*\prime}, \sigma_1^{*2})'$ and $\boldsymbol{\psi}_2^*$ to figure out the the expression of $\boldsymbol{\Omega}$, starting from the first-order condition of our NPL estimator

$$
\begin{aligned}
\frac{\partial \ln L_1(\widehat{\boldsymbol{\psi}}_1^*, \widehat{\sigma}_1^{*2}; \widehat{\mathbf{p}})}{\partial \boldsymbol{\psi}_1^*} &= \sum_{i=1}^n \Big\{ -\mathbf{I}(y_{i1}>0)\frac{(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_1^* - y_{i1})}{\widehat{\sigma}_1^{*2}} - \mathbf{I}(y_{i1}=0)\frac{\phi(\widehat{\mathbf{z}}_i^{*\prime}\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*)}{[1-\Phi(\widehat{\mathbf{z}}_i^{*\prime}\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*)](\widehat{\sigma}_1^{*2})^{\frac{1}{2}}} \Big\}\widehat{\mathbf{z}}_i = 0 \\
\frac{\partial \ln L_1(\widehat{\boldsymbol{\psi}}_1^*, \widehat{\sigma}_1^{*2}; \widehat{\mathbf{p}})}{\partial (\sigma_1^{*2})} &= \sum_{i=1}^n \Big\{ \mathbf{I}(y_{i1}>0)\Big[-\frac{1}{2\widehat{\sigma}_1^{*2}} + \frac{(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_1^* - y_{i1})^2}{2(\widehat{\sigma}_1^{*2})^2}\Big] \\
&\quad + \mathbf{I}(y_{i1}=0)\frac{(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_1^*)\phi(\widehat{\mathbf{z}}_i^{*\prime}\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*))}{2[1-\Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_1^*/\widehat{\sigma}_1^*)](\widehat{\sigma}_1^{*2})^{\frac{3}{2}}} \Big\} = 0 \\
\frac{\partial \ln L_2(\widehat{\boldsymbol{\psi}}_2^*; \widehat{\mathbf{p}})}{\partial \boldsymbol{\psi}_2^*} &= \sum_{i=1}^n \frac{[y_{i2} - \Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_2^*)]\phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_2^*)}{\Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_2^*)[1-\Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_2^*)]}\widehat{\mathbf{z}}_i = 0
\end{aligned}
$$

The first-order Taylor expansion around $\boldsymbol{\psi}_1^*$, σ_1^{*2}, and $\boldsymbol{\psi}_2^*$, of the previous first-order conditions for the Pseudo Maximum Likelihood estimation are

$$
\begin{aligned}
\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*} + \frac{\partial}{\partial \boldsymbol{\psi}_1^*}\Big(\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*}\Big)(\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) + \frac{\partial}{\partial \sigma_1^{*2}}\Big(\frac{\partial \ln L}{\partial \boldsymbol{\psi}_1^*}\Big)(\widehat{\sigma}_1^{*2} - \sigma_1^{*2}) &= O_p(1) \\
\frac{\partial \ln L_1}{\partial \sigma_1^{*2}} + \frac{\partial}{\partial \boldsymbol{\psi}_1^*}\Big(\frac{\partial \ln L_1}{\partial \sigma_1^{*2}}\Big)(\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) + \frac{\partial}{\partial \sigma_1^{*2}}\Big(\frac{\partial \ln L_1}{\partial \sigma_1^{*2}}\Big)(\widehat{\sigma}_1^{*2} - \sigma_1^{*2}) &= O_p(1) \qquad \text{(B.14)} \\
\frac{\partial \ln L_2}{\partial \boldsymbol{\psi}_2^*} + \frac{\partial}{\partial \boldsymbol{\psi}_2^*}\Big(\frac{\partial \ln L_2}{\partial \boldsymbol{\psi}_2^*}\Big)(\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*) &= O_p(1)
\end{aligned}
$$

we can derive the matrix form for $\boldsymbol{\psi}_1^*$ and σ_1^{*2} as

$$
\begin{pmatrix} \frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*} \\ \frac{\partial \ln L_1}{\partial \sigma_1^{*2}} \end{pmatrix} + \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\psi}_1^*}\Big(\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*}\Big) & \frac{\partial}{\partial \sigma_1^{*2}}\Big(\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*}\Big) \\ \frac{\partial}{\partial \boldsymbol{\psi}_1^*}\Big(\frac{\partial \ln L_1}{\partial \sigma_1^{*2}}\Big) & \frac{\partial}{\partial \sigma_1^{*2}}\Big(\frac{\partial \ln L_1}{\partial \sigma_1^{*2}}\Big) \end{bmatrix} \begin{bmatrix} (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) \\ (\widehat{\sigma}_1^{*2} - \sigma_1^{*2}) \end{bmatrix} = \begin{bmatrix} O_p(1) \\ O_p(1) \end{bmatrix} \qquad \text{(B.15)}
$$

then, the asymptotic results can be derived as

$$
\begin{aligned}
\begin{bmatrix} \sqrt{n}(\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) \\ \sqrt{n}(\widehat{\sigma}_1^{*2} - \sigma_1^{*2}) \end{bmatrix} &\overset{A}{=} -\begin{bmatrix} \frac{1}{n}\frac{\partial}{\partial \boldsymbol{\psi}_1^*}\Big(\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*}\Big) & \frac{1}{n}\frac{\partial}{\partial \sigma_1^{*2}}\Big(\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*}\Big) \\ \frac{1}{n}\frac{\partial}{\partial \boldsymbol{\psi}_1^*}\Big(\frac{\partial \ln L_1}{\partial \sigma_1^{*2}}\Big) & \frac{1}{n}\frac{\partial}{\partial \sigma_1^{*2}}\Big(\frac{\partial \ln L_1}{\partial \sigma_1^{*2}}\Big) \end{bmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}}\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1^*} \\ \frac{1}{\sqrt{n}}\frac{\partial \ln L}{\partial \sigma_1^{*2}} \end{pmatrix} \qquad \text{(B.16)} \\
\sqrt{n}(\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*) &\overset{A}{=} -\Big[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\psi}_2^*}\Big(\frac{\partial \ln L_2}{\partial \boldsymbol{\psi}_2^*}\Big)\Big]^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \ln L_2}{\partial \boldsymbol{\psi}_2^*}\Big) \qquad \text{(B.17)}
\end{aligned}
$$

The asymptotic form of $\sqrt{n}(\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*)$ can be reached in Liu (2019) as following

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*) \overset{A}{=} &\Big\{\frac{1}{n}\sum_{i=1}^n \frac{[\phi(\mathbf{z}_i'\boldsymbol{\psi}_2)]^2}{\Phi(\mathbf{z}_i'\boldsymbol{\psi}_2)[1-\Phi(\mathbf{z}_i'\boldsymbol{\psi}_2)]}\mathbf{z}_i\Big[\mathbf{z}_i' + \lambda_{12}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \boldsymbol{\psi}_2^*} + \lambda_{22}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \boldsymbol{\psi}_2^*}\Big]\Big\}^{-1} \\
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_2)[y_{i2} - \Phi(\mathbf{z}_i'\boldsymbol{\psi}_2)]}{\Phi(\mathbf{z}_i'\boldsymbol{\psi}_2)[1-\Phi(\mathbf{z}_i'\boldsymbol{\psi}_2)]}\mathbf{z}_i
\end{aligned} \qquad \text{(B.18)}
$$

Let $\Phi_{i1} = \Phi(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*)$ and $\phi_{i1} = \phi(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*)$, the second-order full-term derivatives are

$$\begin{aligned}\frac{\partial^2 \ln L_1}{\partial\boldsymbol{\psi}_1^*\partial\boldsymbol{\psi}_1^{*\prime}} =& \sum_{i=1}^{n} -\left[\frac{\Phi_{i1}}{\sigma_1^{*2}} + \frac{\phi_{i1}^2}{(1-\Phi_{i1})\sigma_1^{*2}}\mathbf{z}_i\mathbf{z}_i' - \frac{\phi_{i1}}{(\sigma_1^{*2})^{\frac{3}{2}}}\right]\mathbf{z}_i\mathbf{z}_i' \\ & - \left[\frac{\Phi_{i1}}{\sigma_1^{*2}} + \frac{\phi_{i1}^2}{(1-\Phi_{i1})\sigma_1^{*2}} - \frac{\phi_{i1}(\mathbf{z}_i'\boldsymbol{\psi}_1^*)}{\sigma_1^{*3}}\right]\mathbf{z}_i\left(\lambda_{11}^*\mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial\boldsymbol{\psi}_1^{*\prime}} + \lambda_{21}^*\mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial\boldsymbol{\psi}_1^{*\prime}}\right) \\ & - \left[\frac{\Phi_{i1}(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})}{\sigma_1^{*2}} + \frac{\phi_{i1}}{\sigma_1^*}\right]\begin{pmatrix}\mathbf{0}' \\ \mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial\boldsymbol{\psi}_1^{*\prime}} \\ \mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial\boldsymbol{\psi}_1^{*\prime}}\end{pmatrix}\end{aligned} \tag{B.19}$$

$$\begin{aligned}\frac{\partial^2 \ln L_1}{\partial(\sigma_1^{*2})^2} =& \sum_{i=1}^{n} -\frac{\Phi_{i1}}{2(\sigma_1^{*2})^2} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^3\phi_{i1}}{4(\sigma_1^{*2})^{\frac{7}{2}}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}}{4(\sigma_1^{*2})^{\frac{5}{2}}} - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}^2}{4(1-\Phi_{i1})(\sigma_1^{*2})^3} \\ & + \left[\frac{\Phi_{i1}(\mathbf{z}_i'\psi_1^* - y_{i1})}{(\sigma_1^{*2})^2}\frac{\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}}{2(\sigma_1^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}}{2(1-\Phi_{i1})(\sigma_1^{*2})^2}\right] \\ & \left(\lambda_{11}^*\mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial(\sigma_1^{*2})} + \lambda_{21}^*\mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial(\sigma_1^{*2})}\right)\end{aligned} \tag{B.20}$$

$$\begin{aligned}\frac{\partial}{\partial\sigma_1^{*2}}\left(\frac{\partial \ln L_1}{\partial\boldsymbol{\psi}_1^*}\right) =& \sum_{i=1}^{n} -\frac{\mathbf{z}_i\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}\mathbf{z}_i}{2(\sigma_1^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}^2\mathbf{z}_i}{2(1-\Phi_{i1})(\sigma_1^{*2})^2} \\ & + \left[-\frac{\Phi_{i1}\mathbf{z}_i}{\sigma_1^{*2}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}\mathbf{z}_i}{(\sigma_1^{*2})^{\frac{3}{2}}} - \frac{\phi_{i1}^2\mathbf{z}_i}{(1-\Phi_{i1})\sigma_1^{*2}}\right]\left(\lambda_{11}^*\mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial\sigma_1^{*2}} + \lambda_{21}^*\mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial\sigma_1^{*2}}\right) \\ & - \left[\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})\Phi_{i1}}{\sigma_1^{*2}} + \frac{\phi_{i1}}{(\sigma_1^{*2})^{\frac{1}{2}}}\right]\left(\mathbf{w}_i\frac{\partial\mathbf{p}_1}{\partial\sigma_1^{*2}}, \mathbf{w}_i\frac{\partial\mathbf{p}_2}{\partial\sigma_1^{*2}}, \mathbf{0}'\right)'\end{aligned} \tag{B.21}$$

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\psi}_1^*}\left(\frac{\partial \ln L_1}{\partial\sigma_1^{*2}}\right) =& \sum_{i=1}^{n} -\frac{\mathbf{z}_i\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}\mathbf{z}_i}{2(\sigma_1^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}^2\mathbf{z}_i}{2(1-\Phi_{i1})(\sigma_1^{*2})^2} \\ & + \left[\frac{\Phi_{i1}(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})}{(\sigma_1^{*2})^2} + \frac{\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}}{2(\sigma_1^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}^2}{2(1-\Phi_{i1})(\sigma_1^{*2})^2}\right] \\ & \left(\lambda_{11}^*\frac{\partial\mathbf{p}_1'}{\partial\boldsymbol{\psi}_1^*}\mathbf{w}_i' + \lambda_{21}^*\frac{\partial\mathbf{p}_2'}{\partial\boldsymbol{\psi}_1^*}\mathbf{w}_i'\right)\end{aligned} \tag{B.22}$$

According to $p_{i1} = (\mathbf{z}_i'\boldsymbol{\psi}_1^*)\Phi_{i1} + \sigma_1^*\phi_{i1}$ and $p_{i2} = \Phi(\mathbf{z}_i'\boldsymbol{\psi}_2^*)$ we can derive $\partial\mathbf{p}_1/\partial\boldsymbol{\psi}_1^*$, $\partial\mathbf{p}_1/\partial\boldsymbol{\psi}_2^*$,

$\partial\mathbf{p}_2/\partial\boldsymbol{\psi}_1^*$, $\partial\mathbf{p}_2/\partial\boldsymbol{\psi}_2^*$, $\partial\mathbf{p}_1/\partial\sigma_1^{*2}$, and $\partial\mathbf{p}_2/\partial\sigma_1^{*2}$.

$$\begin{aligned}
\frac{\partial\mathbf{p}_1}{\partial\psi_1^{*\prime}} &= \mathbf{K}_1^{-1}\mathbf{A}_1\mathbf{Z} \\
\frac{\partial\mathbf{p}_2}{\partial\psi_2^{*\prime}} &= \mathbf{K}_2^{-1}\mathbf{A}_2\mathbf{Z} \\
\frac{\partial\mathbf{p}_1}{\partial\psi_2^{*\prime}} &= (\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})\lambda_{21}^*\mathbf{A}_1\mathbf{W}\mathbf{K}_2^{-1}\mathbf{A}_2\mathbf{Z} \\
\frac{\partial\mathbf{p}_2}{\partial\psi_1^{*\prime}} &= (\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})\lambda_{12}^*\mathbf{A}_2\mathbf{W}\mathbf{K}_1^{-1}\mathbf{A}_1\mathbf{Z} \\
\frac{\partial\mathbf{p}_1}{\partial\sigma_1^{*2}} &= \mathbf{K}_1^{-1}\frac{\phi_1}{2(\sigma_1^{*2})^{1/2}} \\
\frac{\partial\mathbf{p}_2}{\partial\sigma_1^{*2}} &= (\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})\lambda_{12}^*\mathbf{A}_2\mathbf{W}\mathbf{K}_1^{-1}\frac{\phi_1}{2(\sigma_1^{*2})^{1/2}}
\end{aligned} \tag{B.23}$$

where

$$\begin{aligned}
\mathbf{K}_1 &= \mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{A}_1\mathbf{W}(\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})^{-1}\mathbf{A}_2\mathbf{W} \\
\mathbf{K}_2 &= \mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{A}_2\mathbf{W}(\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})^{-1}\mathbf{A}_1\mathbf{W} \\
\mathbf{A}_1 &= \mathrm{diag}(\Phi_{11}, \Phi_{21}, \cdots \Phi_{n1}) \\
\mathbf{A}_2 &= \mathrm{diag}(\phi_{12}, \phi_{22}, \cdots \phi_{n2})
\end{aligned} \tag{B.24}$$

Then we can derive the asymptotic variance covariance matrix of reduced-form parameter estimation results, $(\widehat{\boldsymbol{\psi}}_1^{*\prime}, \widehat{\sigma}_1^{*2}, \widehat{\boldsymbol{\psi}}_2^{*\prime})'$ according to previous algebra results is

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{bmatrix} \tag{B.25}$$

where

$$\begin{aligned}
\mathbf{V}_{11} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{b}_{11} \\ \mathbf{c}_{11}' & g_{11} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}_{11} & \mathbf{q}_{11} \\ \mathbf{q}_{11}' & s_{11} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{b}_{11} \\ \mathbf{c}_{11}' & g_{11} \end{bmatrix}^{\prime -1} \\
\mathbf{V}_{12} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{b}_{11} \\ \mathbf{c}_{11}' & g_{11} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}_{12} \\ \mathbf{q}_{12}' \end{bmatrix} \Big[\Big(\mathbf{Z} + \lambda_{12}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\boldsymbol{\psi}_2^{*\prime}} + \lambda_{22}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\boldsymbol{\psi}_2^{*\prime}}\Big)'\mathrm{diag}_{i=1}^n\Big(\frac{\phi_{i2}^2}{\Phi_{i2}(1-\Phi_{i2})}\Big)\mathbf{Z}\Big]^{-1} \\
\mathbf{V}_{22} &= \Big[\mathbf{Z}'\mathrm{diag}_{i=1}^n\Big(\frac{\phi_{i2}^2}{\Phi_{i2}(1-\Phi_{i2})}\Big)\Big(\mathbf{Z} + \lambda_{12}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\boldsymbol{\psi}_2^{*\prime}} + \lambda_{22}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\boldsymbol{\psi}_2^{*\prime}}\Big)\Big]^{-1} \\
&\quad \mathbf{Z}'\Big[\mathrm{diag}_{i=1}^n\Big(\frac{\phi_{i2}^2}{\Phi_{i2}(1-\Phi_{i2})}\Big)\Big]\mathbf{Z} \\
&\quad \Big[\Big(\mathbf{Z} + \lambda_{12}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\boldsymbol{\psi}_2^{*\prime}} + \lambda_{22}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\boldsymbol{\psi}_2^{*\prime}}\Big)'\mathrm{diag}_{i=1}^n\Big(\frac{\phi_{i2}^2}{\Phi_{i2}(1-\Phi_{i2})}\Big)\mathbf{Z}\Big]^{-1}
\end{aligned}$$

where

$$
\begin{aligned}
\mathbf{H}_{11} &= \mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})^2\Phi_{i1}}{(\sigma_1^{*2})^2}+\frac{\phi_{i1}^2}{(1-\Phi_{i1})\sigma_1^{*2}}\Big)\Big]\mathbf{Z} \\
\mathbf{q}_{11} &= \mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})\Phi_{i1}}{2(\sigma_1^{*2})^2}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})^3\Phi_{i1}}{2(\sigma_1^{*2})^3}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}^2}{2(1-\Phi_{i1})(\sigma_1^{*2})^2}\Big)\Big]\iota_n \\
s_{11} &= \iota_n'\Big[\text{diag}_{i=1}^{n}\Big(\frac{\Phi_{i1}}{4(\sigma_1^{*2})^4}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})^2\Phi_{i1}}{2(\sigma_1^{*2})^3}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})^4\Phi_{i1}}{4(\sigma_1^{*2})^4}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}^2}{2(1-\Phi_{i1})(\sigma_1^{*2})^3}\Big)\Big]\iota_n
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A}_{11} &= \mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}}{(\sigma_1^{*2})^{\frac{3}{2}}}-\frac{\Phi_{i1}}{\sigma_1^{*2}}-\frac{\phi_{i1}^2}{(1-\Phi_{i1})\sigma_1^{*2}}\Big)\Big]\Big(\mathbf{Z}+\lambda_{11}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\boldsymbol{\psi}_1^{*\prime}}+\lambda_{21}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\boldsymbol{\psi}_1^{*\prime}}\Big) \\
\mathbf{b}_{11} &= \mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(-\frac{\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}}{2(\sigma_1^{*2})^{\frac{5}{2}}}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}^2}{2(1-\Phi_{i1})(\sigma_1^{*2})^2}\Big)\Big]\iota_n \\
&\quad +\mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(-\frac{\Phi_{i1}}{\sigma_k^{*2}}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}}{(\sigma_1^{*2})^{\frac{3}{2}}}-\frac{\phi_{i1}^2}{(1-\Phi_{i1})\sigma_1^{*2}}\Big)\Big]\Big(\lambda_{11}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\sigma_1^{*2}}+\lambda_{21}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\sigma_1^{*2}}\Big) \\
&\quad -\Big(\mathbf{0},\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\sigma_1^{*2}},\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\sigma_1^{*2}}\Big)'\Big[\text{diag}_{i=1}^{n}\Big(\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})\Phi_{i1}}{\sigma_1^{*2}}+\frac{\phi_{i1}}{(\sigma_1^{*2})^{\frac{1}{2}}}\Big)\Big]\iota_n \\
\mathbf{c}_{11} &= \mathbf{Z}'\Big[\text{diag}_{i=1}^{n}\Big(-\frac{\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}}{2(\sigma_1^{*2})^{\frac{5}{2}}}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}^2}{2(1-\Phi_{i1})(\sigma_1^{*2})^2}\Big)\Big]\iota_n \\
&\quad +\Big(\lambda_{11}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial\boldsymbol{\psi}_1^{*\prime}}+\lambda_{21}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial\boldsymbol{\psi}_1^{*\prime}}\Big)' \\
&\quad \Big[\text{diag}_{i=1}^{n}\Big(\frac{\Phi_{i1}(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})}{(\sigma_1^{*2})^2}+\frac{\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}}{2(\sigma_1^{*2})^{\frac{5}{2}}}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}^2}{2(1-\Phi_{i1})(\sigma_1^{*2})^2}\Big)\Big]\iota_n \\
g_{11} &= \iota_n'\Big[\text{diag}_{i=1}^{n}\Big(-\frac{\Phi_{i1}}{2(\sigma_1^{*2})^2}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^3\phi_{i1}}{4(\sigma_1^{*2})^{\frac{7}{2}}}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}}{4(\sigma_1^{*2})^{\frac{5}{2}}}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}^2}{4(1-\Phi_{i1})(\sigma_1^{*2})^3}\Big)\Big]\iota_n \\
&\quad +\iota_n'\Big[\text{diag}_{i=1}^{n}\Big(\frac{\Phi_{i1}(\mathbf{z}_i'\boldsymbol{\psi}_1^*-y_{i1})}{(\sigma_1^{*2})^2}\frac{\phi_{i1}}{2(\sigma_1^{*2})^{\frac{3}{2}}}-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)^2\phi_{i1}}{2(\sigma_1^{*2})^{\frac{5}{2}}}+\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}}{2(1-\Phi_{i1})(\sigma_1^{*2})^2}\Big)\Big] \\
&\quad \Big(\lambda_{11}^*\mathbf{W}\frac{\partial\mathbf{p}_1}{\partial(\sigma_1^{*2})}+\lambda_{21}^*\mathbf{W}\frac{\partial\mathbf{p}_2}{\partial(\sigma_1^{*2})}\Big)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{H}_{12} &= \mathbf{Z}'\Big\{\text{diag}_{i=1}^{n}\Big[-\frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})\phi_{i2}}{\sigma_1^{*2}\Phi_{i2}}\Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \mathbf{z}_i'\boldsymbol{\psi}_2^*, \rho^*\Big) \\
&\quad + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})\phi_{i2}}{\sigma_1^{*2}(1-\Phi_{i2})}\Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, -\mathbf{z}_i'\boldsymbol{\psi}_2^*, -\rho^*\Big) \\
&\quad - \frac{\phi_{i1}\phi_{i2}}{\sigma_1^{*2}(1-\Phi_{i1})\Phi_{i2}}\Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \mathbf{z}_i'\boldsymbol{\psi}_2^*, -\rho^*\Big) \\
&\quad + \frac{\phi_{i1}\phi_{i2}}{\sigma_1^{*2}(1-\Phi_{i1})\Phi_{i2}}\Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, -\mathbf{z}_i'\boldsymbol{\psi}_2^*, \rho^*\Big)\Big]\Big\}\mathbf{Z} \\
\mathbf{q}_{12} &= \Big\{\text{diag}_{i=1}^{n}\Big[\Big(-\frac{1}{2\sigma_1^{*2}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})^2}{2(\sigma_2^{*2})^2}\Big)\frac{\phi_{i2}}{\Phi_{i2}}\Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \mathbf{z}_i'\boldsymbol{\psi}_2^*, \rho^*\Big) \\
&\quad - \Big(-\frac{1}{2\sigma_1^{*2}} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})^2}{2(\sigma_2^{*2})^2}\Big)\frac{\phi_{i2}}{1-\Phi_{i2}}\Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, -\mathbf{z}_i'\boldsymbol{\psi}_2^*, -\rho^*\Big) \\
&\quad + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}\phi_{i2}}{2(\sigma_1^{*2})^{\frac{3}{2}}(1-\Phi_{i1})\Phi_{i2}}\Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \mathbf{z}_i'\boldsymbol{\psi}_2^*, -\rho^*\Big) \\
&\quad - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi_{i1}\phi_{i2}}{2(\sigma_1^{*2})^{\frac{3}{2}}(1-\Phi_{i1})(1-\Phi_{i2})}\Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, -\mathbf{z}_i'\boldsymbol{\psi}_2^*, \rho^*\Big)\Big]\Big\}\mathbf{Z}
\end{aligned}
$$

as we can derive the following

$$
\begin{aligned}
\text{E}[\mathbf{I}(y_{i1} > 0)y_{i2}] &= \Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \mathbf{z}_i'\boldsymbol{\psi}_2^*, \rho^*\Big) \\
\text{E}[\mathbf{I}(y_{i1} > 0)(1-y_{i2})] &= \Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, -\mathbf{z}_i'\boldsymbol{\psi}_2^*, -\rho^*\Big) \\
\text{E}[\mathbf{I}(y_{i1} = 0)y_{i2}] &= \Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \mathbf{z}_i'\boldsymbol{\psi}_2^*, -\rho^*\Big) \\
\text{E}[\mathbf{I}(y_{i1} = 0)(1-y_{i2})] &= \Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, -\mathbf{z}_i'\boldsymbol{\psi}_2^*, \rho^*\Big)
\end{aligned}
$$

where $\Phi_2(\cdot,\cdot,\rho^*)$ is the standard bivariate normal distribution CDF with a coefficient ρ^*. Then we need to figure out the asymptotic variance-covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$. For notation convenience, we redefine the asymptotic covariance matrix of $\widehat{\psi}^* = (\widehat{\boldsymbol{\psi}}_1^{*\prime}, \widehat{\boldsymbol{\psi}}_2^{*\prime})'$ is

$$
\widetilde{\mathbf{V}} = \begin{bmatrix} \widetilde{\mathbf{V}}_{11} & \widetilde{\mathbf{V}}_{12} \\ \widetilde{\mathbf{V}}_{12}' & \mathbf{V}_{22} \end{bmatrix} \tag{B.26}
$$

where $\widetilde{\mathbf{V}}_{11}$ is the upper-left corner sub-matrix of $\mathbf{V}_{kk}$, i.e., only remove the last column and the last row from $\mathbf{V}_{11}$. And $\widetilde{\mathbf{V}}_{12}$ is the upper-left corner sub-matrix of $\mathbf{V}_{12}$, i.e., only remove the last row from $\mathbf{V}_{12}$. Therefore, from Liu (2019), the asymptotic covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$ is

$$
\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{bmatrix} \tag{B.27}
$$

where

$$\Omega_{11} = \widetilde{\mathbf{V}}_{11} + \theta_{21}^2 \mathbf{V}_{22} + \theta_{21}(\widetilde{\mathbf{V}}_{12} + \widetilde{\mathbf{V}}'_{12}) \tag{B.28}$$

$$\Omega_{22} = \mathbf{V}_{22} + \theta_{12}^2 \widetilde{\mathbf{V}}_{11} + \theta_{12}(\widetilde{\mathbf{V}}_{12} + \widetilde{\mathbf{V}}'_{12}) \tag{B.29}$$

$$\Omega_{12} = \theta_{12}\widetilde{\mathbf{V}}_{11} + \theta_{21}\mathbf{V}_{22} + \widetilde{\mathbf{V}}_{12} + \theta_{12}\theta_{21}\widetilde{\mathbf{V}}'_{12} \tag{B.30}$$

Appendix C

Appendix to Chapter 3

C.1 Steps of deriving BNE

$$
\begin{aligned}
p_{i1} =& \mathrm{E}(y_{i1}^*|y_{i1}^* > 0, y_{i2}^* > 0)\mathbf{Pr}(y_{i1}^* > 0, y_{i2}^* > 0) + \mathrm{E}(y_{i1}^*|y_{i1}^* > 0, y_{i2}^* \leq 0)\mathbf{Pr}(y_{i1}^* > 0, y_{i2}^* \leq 0) \\
=& \mathrm{E}(\mathbf{z}_i'\boldsymbol{\psi}_1^* - \epsilon_{i1}^*|\mathbf{z}_i'\boldsymbol{\psi}_1^* > \epsilon_{i1}^*, \mathbf{z}_i'\boldsymbol{\psi}_2^* > \epsilon_{i2}^*)\Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}, \rho^*\Big) \\
&+ \mathrm{E}(\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \epsilon_{i1}^{**}|\mathbf{z}_i'\boldsymbol{\psi}_1^{**} > \epsilon_{i1}^{**}, \mathbf{z}_i'\boldsymbol{\psi}_2^{**} \leq \epsilon_{i2}^{**})\Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}, -\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}, -\rho^{**}\Big) \\
=& (\mathbf{z}_i'\boldsymbol{\psi}_1^*)\Phi_2^* + (\mathbf{z}_i'\boldsymbol{\psi}_1^{**})\Phi_2^{**} \\
&- \int_{-\infty}^{\mathbf{z}_i'\boldsymbol{\psi}_1^*}\int_{-\infty}^{\mathbf{z}_i'\boldsymbol{\psi}_2^*} \epsilon_{i1}^* f^*(\epsilon_{i1}^*, \epsilon_{i2}^*)d\epsilon_{i2}^* d\epsilon_{i1}^* - \int_{-\infty}^{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}\int_{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}^{+\infty} \epsilon_{i1}^{**} f^{**}(\epsilon_{i1}^{**}, \epsilon_{i2}^{**})d\epsilon_{i2}^{**} d\epsilon_{i1}^{**} \\
=& (\mathbf{z}_i'\boldsymbol{\psi}_1^*)\Phi_2^* + (\mathbf{z}_i'\boldsymbol{\psi}_1^{**})\Phi_2^{**} \\
&+ \sigma_1^*\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) + \rho^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\Big] \\
&+ \sigma_1^{**}\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}\Big)\Phi\Big(\frac{\rho^{**}\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big) \\
&- \rho^{**}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\Phi\Big(\frac{\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\Big] \\
\leq& (\mathbf{z}_i'\boldsymbol{\psi}_1^*)\Phi_2^* + (\mathbf{z}_i'\boldsymbol{\psi}_1^{**})\Phi_2^{**} + \sigma_1^*\sqrt{\frac{2}{\pi}} + \sigma_1^{**}\sqrt{\frac{2}{\pi}} \\
\leq& \max\Big\{(\mathbf{z}_i'\boldsymbol{\psi}_1^*), (\mathbf{z}_i'\boldsymbol{\psi}_1^{**})\Big\} + \sigma_1^*\sqrt{\frac{2}{\pi}} + \sigma_1^{**}\sqrt{\frac{2}{\pi}}
\end{aligned} \tag{C.1}
$$

$$
\begin{aligned}
p_{i2} =& \mathrm{E}(y^*_{i2}|y^*_{i1}>0, y^*_{i2}>0)\mathbf{Pr}(y^*_{i1}>0, y^*_{i2}>0) + \mathrm{E}(y^*_{i2}|y^*_{i1}\leq 0, y^*_{i2}>0)\mathbf{Pr}(y^*_{i1}\leq 0, y^*_{i2}>0)\\
=& \mathrm{E}(\mathbf{z}_i'\boldsymbol{\psi}^*_2 - \epsilon^*_{i2}|\mathbf{z}_i'\boldsymbol{\psi}^*_1 > \epsilon^*_{i1}, \mathbf{z}_i'\boldsymbol{\psi}^*_2 > \epsilon^*_{i2})\Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}^*_1}{\sigma^*_1}, \frac{\mathbf{z}_i'\boldsymbol{\psi}^*_2}{\sigma^*_2}, \rho^*\Big)\\
&+ \mathrm{E}(\mathbf{z}_i'\boldsymbol{\psi}^{***}_2 - \epsilon^{**}_{i2}|\mathbf{z}_i'\boldsymbol{\psi}^{***}_1 \leq \epsilon^{***}_{i1}, \mathbf{z}_i'\boldsymbol{\psi}^{***}_2 > \epsilon^{***}_{i2})\Phi_2\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}^{***}_1}{\sigma^{***}_1}, \frac{\mathbf{z}_i'\boldsymbol{\psi}^{***}_2}{\sigma^{***}_2}, -\rho^{***}\Big)\\
=& (\mathbf{z}_i'\boldsymbol{\psi}^*_2)\Phi^*_2 + (\mathbf{z}_i'\boldsymbol{\psi}^{***}_2)\Phi^{***}_2\\
&- \int_{-\infty}^{\mathbf{z}_i'\boldsymbol{\psi}^*_2}\int_{-\infty}^{\mathbf{z}_i'\boldsymbol{\psi}^*_1} \epsilon^*_{i2} f^*(\epsilon^*_{i1}, \epsilon^*_{i2}) d\epsilon^*_{i1} d\epsilon^*_{i2}\\
&- \int_{-\infty}^{\mathbf{z}_i'\boldsymbol{\psi}^{***}_2}\int_{\mathbf{z}_i'\boldsymbol{\psi}^{***}_1}^{+\infty} \epsilon^{***}_{i2} f^{***}(\epsilon^{***}_{i1}, \epsilon^{***}_{i2}) d\epsilon^{***}_{i1} d\epsilon^{***}_{i2}\\
=& (\mathbf{z}_i'\boldsymbol{\psi}^*_2)\Phi^*_2 + (\mathbf{z}_i'\boldsymbol{\psi}^{***}_2)\boldsymbol{\Phi}^{***}_2 + \sigma^*_2\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}^*_2}{\sigma^*_2}\Big)\Phi\Big(\frac{\sigma^*_2\mathbf{z}_i'\boldsymbol{\psi}^*_1 - \rho^*\sigma^*_1\mathbf{z}_i'\boldsymbol{\psi}^*_2}{\sigma^*_1\sigma^*_2\sqrt{1-\rho^{*2}}}\Big)\\
&+ \rho^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}^*_1}{\sigma^*_1}\Big)\Phi\Big(\frac{\sigma^*_1\mathbf{z}_i'\boldsymbol{\psi}^*_2 - \rho^*\sigma^*_2\mathbf{z}_i'\boldsymbol{\psi}^*_1}{\sigma^*_1\sigma^*_2\sqrt{1-\rho^{*2}}}\Big)\Big]\\
&+ \sigma^{***}_2\Big[\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}^{***}_2}{\sigma^{***}_2}\Big)\Phi\Big(\frac{\rho^{***}\sigma^{***}_1\mathbf{z}_i'\boldsymbol{\psi}^{***}_2 - \sigma^{***}_2\mathbf{z}_i'\boldsymbol{\psi}^{***}_1}{\sigma^{***}_1\sigma^{***}_2\sqrt{1-\rho^{***2}}}\Big)\\
&- \rho^{***}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}^{***}_1}{\sigma^{***}_1}\Big)\Phi\Big(\frac{\sigma^{***}_1\mathbf{z}_i'\boldsymbol{\psi}^{***}_2 - \rho^{***}\sigma^{***}_2\mathbf{z}_i'\boldsymbol{\psi}^{***}_1}{\sigma^{***}_1\sigma^{***}_2\sqrt{1-\rho^{***2}}}\Big)\Big]\\
\leq& (\mathbf{z}_i'\boldsymbol{\psi}^*_1)\Phi^*_2 + (\mathbf{z}_i'\boldsymbol{\psi}^{***}_1)\Phi^{***}_2 + \sigma^*_2\sqrt{\frac{2}{\pi}} + \sigma^{***}_2\sqrt{\frac{2}{\pi}}\\
\leq& \max\Big\{(\mathbf{z}_i'\boldsymbol{\psi}^*_1), (\mathbf{z}_i'\boldsymbol{\psi}^{***}_1)\Big\} + \sigma^*_2\sqrt{\frac{2}{\pi}} + \sigma^{***}_2\sqrt{\frac{2}{\pi}}
\end{aligned}
\tag{C.2}
$$

C.2 Steps in Parameter Space Discussion

Following are the first-order derivative of $h(\mathbf{p})$ with respect to $\mathbf{p}$.

$$
\begin{aligned}
\frac{\partial p_{i1}}{\partial \mathbf{p}'} =&[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]\Phi_2^* + \frac{[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]}{\sigma_1^*}(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\\
&+ \frac{[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]}{\sigma_2^*}(\mathbf{z}_i'\boldsymbol{\psi}_1^*)\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\\
&+ \sigma_1^*\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\frac{[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]}{\sigma_1^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\\
&+ \sigma_1^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\Bigg[\frac{[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]}{\sigma_2^*\sqrt{1-\rho^{*2}}} - \frac{\rho^*[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]}{\sigma_1^*\sqrt{1-\rho^{*2}}}\Bigg]\\
&+ \rho^*\sigma_1^*\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\frac{[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]}{\sigma_2^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\\
&+ \rho^*\sigma_1^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\Bigg[\frac{[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]}{\sigma_1^*\sqrt{1-\rho^{*2}}} - \frac{\rho^*[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]}{\sigma_2^*\sqrt{1-\rho^{*2}}}\Bigg]\\
&+ [\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]\Phi_2^{**} + \frac{[\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]}{\sigma_1^{**}}(\mathbf{z}_i'\boldsymbol{\psi}_1^{**})\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}\Big)\Phi\Big(\frac{\rho^{**}\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\\
&+ \frac{[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}}(\mathbf{z}_i'\boldsymbol{\psi}_1^{**})\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\Phi\Big(\frac{\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\\
&+ \sigma_1^{**}\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}\Big)\frac{[\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]}{\sigma_1^{**}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}\Big)\Phi\Big(\frac{\rho^{**}\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\\
&+ \sigma_1^{**}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}\Big)\phi\Big(\frac{\rho^{**}\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\Bigg[\frac{\rho^{**}[\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]}{\sigma_1^{**}\sqrt{1-\rho^{**2}}} - \frac{[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Bigg]\\
&+ \rho^{**}\sigma_1^{**}\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\frac{[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\Phi\Big(\frac{\rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**} - \sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\\
&+ \rho^{**}\sigma_1^{**}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\phi\Big(\frac{\rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**} - \sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\Bigg[\frac{\rho^{**}[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}\sqrt{1-\rho^{**2}}} - \frac{[\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]}{\sigma_1^{**}\sqrt{1-\rho^{**2}}}\Bigg]\\
=&[\lambda_{11}^*\mathbf{w}_i, \lambda_{21}^*\mathbf{w}_i]\Phi_2^* + \Bigg(\mathbf{z}_i'\boldsymbol{\psi}_1^* - \frac{\rho^*\sigma_1^*}{\sigma_2^*}\mathbf{z}_i'\boldsymbol{\psi}_2^*\Bigg)\frac{[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]}{\sigma_2^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\\
&+ (1-\rho^{*2})\frac{\sigma_1^*[\lambda_{12}^*\mathbf{w}_i, \lambda_{22}^*\mathbf{w}_i]}{\sigma_2^*}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}, \rho^*\Big) + [\lambda_{11}^{**}\mathbf{w}_i, \lambda_{21}^{**}\mathbf{w}_i]\Phi_2^{**}\\
&- \Bigg(\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \frac{\rho^{**}\sigma_1^{**}}{\sigma_2^{**}}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}\Bigg)\frac{[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}\Big)\Phi\Big(\frac{\sigma_2^{**}\mathbf{z}_i'\boldsymbol{\psi}_1^{**} - \rho^{**}\sigma_1^{**}\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_1^{***}\sigma_2^{**}\sqrt{1-\rho^{**2}}}\Big)\\
&- (1-\rho^{**2})\frac{\sigma_1^{**}[\lambda_{12}^{**}\mathbf{w}_i, \lambda_{22}^{**}\mathbf{w}_i]}{\sigma_2^{**}}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{**}}{\sigma_1^{**}}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{**}}{\sigma_2^{**}}, \rho^{**}\Big)
\end{aligned}
\tag{C.3}
$$

$$
\begin{aligned}
\frac{\partial p_{i2}}{\partial \mathbf{p}'} =& [\lambda_{12}^* \mathbf{w}_i, \lambda_{22}^* \mathbf{w}_i]\Phi_2^* + \frac{[\lambda_{11}^* \mathbf{w}_i, \lambda_{21}^* \mathbf{w}_i]}{\sigma_1^*}(\mathbf{z}_i'\boldsymbol{\psi}_2^*)\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) \\
&+ \frac{[\lambda_{12}^* \mathbf{w}_i, \lambda_{22}^* \mathbf{w}_i]}{\sigma_2^*}(\mathbf{z}_i'\boldsymbol{\psi}_2^*)\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) \\
&+ \sigma_2^*\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\frac{[\lambda_{12}^* \mathbf{w}_i, \lambda_{22}^* \mathbf{w}_i]}{\sigma_2^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\Phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) \\
&+ \sigma_2^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big)\phi\Big(\frac{\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\Big[\frac{[\lambda_{11}^* \mathbf{w}_i, \lambda_{21}^* \mathbf{w}_i]}{\sigma_1^*\sqrt{1-\rho^{*2}}} - \frac{\rho^*[\lambda_{12}^* \mathbf{w}_i, \lambda_{22}^* \mathbf{w}_i]}{\sigma_2^*\sqrt{1-\rho^{*2}}}\Big] \\
&+ \rho^*\sigma_2^*\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\frac{[\lambda_{11}^* \mathbf{w}_i, \lambda_{21}^* \mathbf{w}_i]}{\sigma_1^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) \\
&+ \rho^*\sigma_2^*\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big)\Big[\frac{[\lambda_{12}^* \mathbf{w}_i, \lambda_{22}^* \mathbf{w}_i]}{\sigma_2^*\sqrt{1-\rho^{*2}}} - \frac{\rho^*[\lambda_{11}^* \mathbf{w}_i, \lambda_{21}^* \mathbf{w}_i]}{\sigma_1^*\sqrt{1-\rho^{*2}}}\Big] \\
&+ [\lambda_{12}^{***} \mathbf{w}_i, \lambda_{22}^{***} \mathbf{w}_i]\Phi_2^{***} \\
&- \frac{[\lambda_{11}^{***} \mathbf{w}_i, \lambda_{21}^{***} \mathbf{w}_i]}{\sigma_1^{***}}(\mathbf{z}_i'\boldsymbol{\psi}_2^{***})\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big)\Phi\Big(\frac{\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&+ \frac{[\lambda_{12}^{***} \mathbf{w}_i, \lambda_{22}^{***} \mathbf{w}_i]}{\sigma_2^{***}}(\mathbf{z}_i'\boldsymbol{\psi}_2^{***})\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}\Big)\Phi\Big(\frac{\rho^{***}\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&+ \sigma_2^{***}\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}\Big)\frac{[\lambda_{12}^{***} \mathbf{w}_i, \lambda_{22}^{***} \mathbf{w}_i]}{\sigma_2^{***}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}\Big)\Phi\Big(\frac{\rho^{***}\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&+ \sigma_2^{***}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}\Big)\phi\Big(\frac{\rho^{***}\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&\Big[\frac{\rho^{***}[\lambda_{12}^{***} \mathbf{w}_i, \lambda_{22}^{***} \mathbf{w}_i]}{\sigma_2^{***}\sqrt{1-\rho^{***2}}} - \frac{[\lambda_{11}^{***} \mathbf{w}_i, \lambda_{21}^{***} \mathbf{w}_i]}{\sigma_1^{***}\sqrt{1-\rho^{***2}}}\Big] \\
&- \rho^{***}\sigma_2^{***}\Big(-\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big)\frac{[\lambda_{11}^{***} \mathbf{w}_i, \lambda_{21}^{***} \mathbf{w}_i]}{\sigma_1^{***}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big)\Phi\Big(\frac{\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&- \rho^{***}\sigma_2^{***}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big)\phi\Big(\frac{\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&\Big[\frac{[\lambda_{12}^{***} \mathbf{w}_i, \lambda_{22}^{***} \mathbf{w}_i]}{\sigma_2^{***}\sqrt{1-\rho^{***2}}} - \frac{\rho^{***}[\lambda_{11}^{***} \mathbf{w}_i, \lambda_{21}^{***} \mathbf{w}_i]}{\sigma_1^{***}\sqrt{1-\rho^{***2}}}\Big] \\
=& [\lambda_{12}^* \mathbf{w}_i, \lambda_{22}^* \mathbf{w}_i]\Phi_2^* + \Big(\mathbf{z}_i'\boldsymbol{\psi}_2^* - \frac{\rho^*\sigma_2^*}{\sigma_1^*}\mathbf{z}_i'\boldsymbol{\psi}_1^*\Big)\frac{[\lambda_{11}^* \mathbf{w}_i, \lambda_{21}^* \mathbf{w}_i]}{\sigma_1^*}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big)\Phi\Big(\frac{\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}}\Big) \\
&+ (1-\rho^{*2})\frac{\sigma_2^*[\lambda_{11}^* \mathbf{w}_i, \lambda_{21}^* \mathbf{w}_i]}{\sigma_1^*}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}, \rho^*\Big) + [\lambda_{12}^{***} \mathbf{w}_i, \lambda_{22}^{***} \mathbf{w}_i]\Phi_2^{***} \\
&- \Big(\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \frac{\rho^{***}\sigma_2^{***}}{\sigma_1^{***}}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}\Big)\frac{[\lambda_{11}^{***} \mathbf{w}_i, \lambda_{21}^{***} \mathbf{w}_i]}{\sigma_1^{***}}\phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}\Big)\Phi\Big(\frac{\sigma_1^{***}\mathbf{z}_i'\boldsymbol{\psi}_2^{***} - \rho^{***}\sigma_2^{***}\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}\sigma_2^{***}\sqrt{1-\rho^{***2}}}\Big) \\
&- (1-\rho^{***2})\frac{\sigma_2^{***}[\lambda_{11}^{***} \mathbf{w}_i, \lambda_{21}^{***} \mathbf{w}_i]}{\sigma_1^{***}}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^{***}}{\sigma_1^{***}}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^{***}}{\sigma_2^{***}}, \rho^{***}\Big)
\end{aligned} \tag{C.4}
$$

C.3 Simulation Preparation

For a bivariate normal distributed vector $(\varepsilon_1, \varepsilon_2)$, suppose the mean vector is $(0, 0)$, and without loss of generality, the variance-covariance matrix is

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{C.5}$$

then we have the probability density function as

$$f(\varepsilon_1, \varepsilon_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)}\left[\left(\frac{\varepsilon_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\varepsilon_1}{\sigma_1}\right)\left(\frac{\varepsilon_2}{\sigma_2}\right) + \left(\frac{\varepsilon_1}{\sigma_1}\right)^2\right]\right\} \tag{C.6}$$

therefore, we have

$$\begin{aligned} \mathbf{Pr}(\varepsilon_1 < a, \varepsilon_2 < b) &= \mathbf{\Phi}_2\Big(\frac{a}{\sigma_1}, \frac{b}{\sigma_2}, \rho\Big) \\ \mathbf{Pr}(\varepsilon_1 < a, \varepsilon_2 > b) &= \mathbf{\Phi}_2\Big(\frac{a}{\sigma_1}, -\frac{b}{\sigma_2}, -\rho\Big) \end{aligned}$$

where $\mathbf{\Phi}_2(\cdot, \cdot, \rho)$ is the CDF of bivariate normal distributed vectors with correlation ρ. And we can also derive

$$\begin{aligned} \mathbf{Pr}(\varepsilon_1 = a, \varepsilon_2 \le b) &= \int_{-\infty}^{b} f(a, \varepsilon_2) d\varepsilon_2 \\ &= \int_{-\infty}^{b} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\quad \exp\left[-\frac{1}{2(1-\rho^2)}\Big(\frac{a^2}{\sigma_1^2} - \frac{2\rho a\varepsilon_2}{\sigma_1\sigma_2} + \frac{\varepsilon_2^2}{\sigma_2^2}\Big)\right] d\varepsilon_2 \\ &= \int_{-\infty}^{b} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\quad \exp\left[-\frac{1}{2(1-\rho^2)}\Big(\frac{\varepsilon_2^2}{\sigma_2^2} - \frac{2\rho a\varepsilon_2}{\sigma_1\sigma_2} + \frac{\rho^2 a^2}{\sigma_1^2} + \frac{(1-\rho^2)a^2}{\sigma_1^2}\Big)\right] d\varepsilon_2 \\ &= \frac{1}{\sigma_1}\phi\Big(\frac{a}{\sigma_1}\Big)\mathbf{\Phi}\Big(\frac{b}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho a}{\sigma_1\sqrt{1-\rho^2}}\Big) \end{aligned}$$

$$
\begin{aligned}
\mathbf{Pr}(\varepsilon_1 \le a, \varepsilon_2 = b) &= \int_{-\infty}^{a} f(\varepsilon_1, b) d\varepsilon_1 \\
&= \int_{-\infty}^{a} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
&\quad \exp\Big[-\frac{1}{2(1-\rho^2)}\Big(\frac{b^2}{\sigma_2^2} - \frac{2\rho b\varepsilon_1}{\sigma_1\sigma_2} + \frac{\varepsilon_1^2}{\sigma_1^2}\Big)\Big] d\varepsilon_1 \\
&= \int_{-\infty}^{a} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
&\quad \exp\Big[-\frac{1}{2(1-\rho^2)}\Big(\frac{\varepsilon_1^2}{\sigma_1^2} - \frac{2\rho b\varepsilon_1}{\sigma_1\sigma_2} + \frac{\rho^2 b^2}{\sigma_2^2} + \frac{(1-\rho^2)b^2}{\sigma_2^2}\Big)\Big] d\varepsilon_1 \\
&= \frac{1}{\sigma_2}\phi\Big(\frac{b}{\sigma_2}\Big)\mathbf{\Phi}\Big(\frac{a}{\sigma_1\sqrt{1-\rho^2}} - \frac{\rho b}{\sigma_2\sqrt{1-\rho^2}}\Big)
\end{aligned}
$$

And we have (according to the definition of partial derivatives)

$$
\begin{aligned}
\frac{\partial \mathbf{Pr}(\varepsilon_1 \le a, \varepsilon_2 \le b)}{\partial a} &= \mathbf{Pr}(\varepsilon_1 = a, \varepsilon_2 \le b) \\
\frac{\partial \mathbf{Pr}(\varepsilon_1 \le a, \varepsilon_2 \le b)}{\partial b} &= \mathbf{Pr}(\varepsilon_1 \le a, \varepsilon_2 = b)
\end{aligned} \tag{C.7}
$$

$$
\begin{aligned}
&\mathrm{E}(\varepsilon_1 | \varepsilon_1 < a, \varepsilon_2 < b)\mathbf{Pr}(\varepsilon_1 < a, \varepsilon_2 < b) \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} \varepsilon_1 f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1 \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{\varepsilon_1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big[-\frac{1}{2(1-\rho^2)}\Big(\frac{\varepsilon_1^2}{\sigma_1^2} - \frac{2\rho\varepsilon_1\varepsilon_2}{\sigma_1\sigma_2} + \frac{\varepsilon_2^2}{\sigma_2^2}\Big)\Big] d\varepsilon_2 d\varepsilon_1 \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{\varepsilon_1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big[-\frac{1}{2(1-\rho^2)}\Big(\frac{\varepsilon_2^2}{\sigma_2^2} - \frac{2\rho\varepsilon_1\varepsilon_2}{\sigma_1\sigma_2} + \frac{\rho^2\varepsilon_1^2}{\sigma_1^2} + \frac{(1-\rho^2)\varepsilon_1^2}{\sigma_1^2}\Big)\Big] d\varepsilon_2 d\varepsilon_1 \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\Big[-\frac{1}{2(1-\rho^2)\sigma_2^2}\Big(\varepsilon_2 - \frac{\rho\sigma_2\varepsilon_1}{\sigma_1}\Big)^2\Big] \frac{\varepsilon_1}{\sqrt{2\pi}\sigma_1} \exp\Big(-\frac{\varepsilon_1^2}{2\sigma_1^2}\Big) d\varepsilon_2 d\varepsilon_1 \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{1}{\sigma_2\sqrt{1-\rho^2}}\phi\Big(\frac{\varepsilon_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho\varepsilon_1}{\sigma_1\sqrt{1-\rho^2}}\Big)\frac{\varepsilon_1}{\sigma_1}\phi\Big(\frac{\varepsilon_1}{\sigma_1}\Big) d\varepsilon_2 d\varepsilon_1 \\
&= \int_{-\infty}^{a} \mathbf{\Phi}\Big(\frac{b}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho\varepsilon_1}{\sigma_1\sqrt{1-\rho^2}}\Big)\frac{\varepsilon_1}{\sigma_1}\phi\Big(\frac{\varepsilon_1}{\sigma_1}\Big) d\varepsilon_1 \\
&= \sigma_1 \int_{-\infty}^{\frac{a}{\sigma_1}} \mathbf{\Phi}\Big(\frac{b}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho\varepsilon_1}{\sigma_1\sqrt{1-\rho^2}}\Big)\frac{\varepsilon_1}{\sigma_1}\phi\Big(\frac{\varepsilon_1}{\sigma_1}\Big) d\Big(\frac{\varepsilon_1}{\sigma_1}\Big) \\
&= -\sigma_1\Big[\phi\Big(\frac{a}{\sigma_1}\Big)\mathbf{\Phi}\Big(\frac{b}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho a}{\sigma_1\sqrt{1-\rho^2}}\Big) + \rho\phi\Big(\frac{b}{\sigma_2}\Big)\mathbf{\Phi}\Big(\frac{a}{\sigma_1\sqrt{1-\rho^2}} - \frac{\rho b}{\sigma_2\sqrt{1-\rho^2}}\Big)\Big]
\end{aligned} \tag{C.8}
$$

$$
\begin{aligned}
&\mathrm{E}(\varepsilon_1|\varepsilon_1 < a, \varepsilon_2 > b)\mathbf{Pr}(\varepsilon_1 < a, \varepsilon_2 > b) \\
=&\int_{-\infty}^{a}\int_{b}^{+\infty} \varepsilon_1 f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1 \\
=&\int_{-\infty}^{a}\int_{-\infty}^{b} \frac{\varepsilon_1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big[-\frac{1}{2(1-\rho^2)}\Big(\frac{\varepsilon_1^2}{\sigma_1^2} - \frac{2\rho\varepsilon_1\varepsilon_2}{\sigma_1\sigma_2} + \frac{\varepsilon_2^2}{\sigma_2^2}\Big)\Big] d\varepsilon_2 d\varepsilon_1 \\
=&\int_{-\infty}^{a}\int_{b}^{+\infty} \frac{\varepsilon_1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big[-\frac{1}{2(1-\rho^2)}\Big(\frac{\varepsilon_2^2}{\sigma_2^2} - \frac{2\rho\varepsilon_1\varepsilon_2}{\sigma_1\sigma_2} + \frac{\rho^2\varepsilon_1^2}{\sigma_1^2} + \frac{(1-\rho^2)\varepsilon_1^2}{\sigma_1^2}\Big)\Big] d\varepsilon_2 d\varepsilon_1 \\
=&\int_{-\infty}^{a}\int_{b}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\Big[-\frac{1}{2(1-\rho^2)\sigma_2^2}\Big(\varepsilon_2 - \frac{\rho\sigma_2\varepsilon_1}{\sigma_1}\Big)^2\Big] \frac{\varepsilon_1}{\sqrt{2\pi}\sigma_1} \exp\Big(-\frac{\varepsilon_1^2}{2\sigma_1^2}\Big) d\varepsilon_2 d\varepsilon_1 \\
=&\int_{-\infty}^{a} \Big[1 - \mathbf{\Phi}\Big(\frac{b}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho\varepsilon_1}{\sigma_1\sqrt{1-\rho^2}}\Big)\Big] \frac{\varepsilon_1}{\sigma_1}\phi\Big(\frac{\varepsilon_1}{\sigma_1}\Big) d\varepsilon_1 \\
=&\int_{-\infty}^{a} \mathbf{\Phi}\Big(\frac{\rho\varepsilon_1}{\sigma_1\sqrt{1-\rho^2}} - \frac{b}{\sigma_2\sqrt{1-\rho^2}}\Big) \frac{\varepsilon_1}{\sigma_1}\phi\Big(\frac{\varepsilon_1}{\sigma_1}\Big) d\varepsilon_1 \\
=&\sigma_1 \int_{-\infty}^{\frac{a}{\sigma_1}} \mathbf{\Phi}\Big(\frac{\rho\varepsilon_1}{\sigma_1\sqrt{1-\rho^2}} - \frac{b}{\sigma_2\sqrt{1-\rho^2}}\Big) \frac{\varepsilon_1}{\sigma_1}\phi\Big(\frac{\varepsilon_1}{\sigma_1}\Big) d\Big(\frac{\varepsilon_1}{\sigma_1}\Big) \\
=&-\sigma_1\Big[\phi\Big(\frac{a}{\sigma_1}\Big)\mathbf{\Phi}\Big(\frac{\rho a}{\sigma_1\sqrt{1-\rho^2}} - \frac{b}{\sigma_2\sqrt{1-\rho^2}}\Big) + \rho\phi\Big(\frac{b}{\sigma_2}\Big)\mathbf{\Phi}\Big(\frac{\rho b}{\sigma_2\sqrt{1-\rho^2}} - \frac{a}{\sigma_1\sqrt{1-\rho^2}}\Big)\Big]
\end{aligned} \tag{C.9}
$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF of the standard normal distributed variable. According to Tallis (1961) and Amemiya (1974), we consider the following model suppose the probability density function of (x_1, x_2) is

$$
\begin{aligned}
f(x_1, x_2) =& \frac{\phi_2(x_1/\sigma_1, x_2/\sigma_2, \rho)}{\sigma_1\sigma_2\mathbf{Pr}(x_1 \le a_1, x_2 \le a_2)} \qquad x_1 \le a_1 \text{ and } x_2 \le a_2 \\
=&0 \qquad \text{otherwise}
\end{aligned} \tag{C.10}
$$

then we will have following two equations as with similar form to Amemiya (1974) (2.5)

$$
\begin{aligned}
\frac{\mathrm{E}(x_1^2)}{\sigma_1^2(1-\rho^2)} - \frac{\rho\mathrm{E}(x_1x_2)}{\sigma_1\sigma_2(1-\rho^2)} =& 1 + \frac{a_1\mathrm{E}(x_1)}{\sigma_1^2(1-\rho^2)} - \frac{a_1\rho\mathrm{E}(x_2)}{\sigma_1\sigma_2(1-\rho^2)} \\
\frac{\mathrm{E}(x_2^2)}{\sigma_2^2(1-\rho^2)} - \frac{\rho\mathrm{E}(x_1x_2)}{\sigma_1\sigma_2(1-\rho^2)} =& 1 + \frac{a_2\mathrm{E}(x_2)}{\sigma_2^2(1-\rho^2)} - \frac{a_2\rho\mathrm{E}(x_1)}{\sigma_1\sigma_2(1-\rho^2)}
\end{aligned} \tag{C.11}
$$

and we define

$$
\begin{aligned}
y_1 =& a_1 - x_1 \\
y_2 =& a_2 - x_2
\end{aligned} \tag{C.12}
$$

then by applying (2.11) in Amemiya (1974), we have

$$\begin{aligned} E(y_1^2) - E(y_1 a_1) &= E(x_1^2) - a_1 E(x_1) \\ E(y_1 y_2) - E(y_1 a_2) &= E(x_1 x_2) - a_1 E(x_2) \\ E(y_2^2) - E(y_2 a_2) &= E(x_2^2) - a_2 E(x_2) \\ E(y_1 y_2) - E(y_2 a_1) &= E(x_1 x_2) - a_2 E(x_1) \end{aligned} \tag{C.13}$$

then after combine C.11 and C.13, we can derive similar results to (2.13) in Amemiya (1974)

$$\begin{aligned} \frac{E(y_1^2)}{\sigma_1^2(1-\rho^2)} - \frac{\rho E(y_1 y_2)}{\sigma_1\sigma_2(1-\rho^2)} &= 1 + \frac{a_1 E(y_1)}{\sigma_1^2(1-\rho^2)} - \frac{\rho a_2 E(y_2)}{\sigma_1\sigma_2(1-\rho^2)} \\ \frac{E(y_2^2)}{\sigma_2^2(1-\rho^2)} - \frac{\rho E(y_1 y_2)}{\sigma_1\sigma_2(1-\rho^2)} &= 1 + \frac{a_2 E(y_2)}{\sigma_2^2(1-\rho^2)} - \frac{\rho a_1 E(y_1)}{\sigma_1\sigma_2(1-\rho^2)} \end{aligned}$$

After rearrangement, we have

$$\begin{aligned} E(y_1^2) &= \sigma_1^2(1-\rho^2) + a_1 E(y_1) - \rho\frac{\sigma_1}{\sigma_2} a_2 E(y_2) + \rho\frac{\sigma_1}{\sigma_2} E(y_1 y_2) \\ E(y_2^2) &= \sigma_2^2(1-\rho^2) + a_2 E(y_2) - \rho\frac{\sigma_2}{\sigma_1} a_1 E(y_1) + \rho\frac{\sigma_2}{\sigma_1} E(y_1 y_2) \end{aligned} \tag{C.14}$$

Before we conduct the simulation, let's consider the following simplified simultaneous equation model with truncated outcomes

$$\begin{aligned} y_{i1}^* &= -\theta_{21} y_{i2} + x_{i1}\beta_1 - \epsilon_{i1} \\ y_{i2}^* &= -\theta_{12} y_{i1} + x_{i2}\beta_2 - \epsilon_{i2} \end{aligned}$$

where

$$\begin{aligned} y_{i1} &= y_{i1}^* \mathbf{1}(y_{i1}^* > 0) \\ y_{i2} &= y_{i2}^* \mathbf{1}(y_{i2}^* > 0) \end{aligned}$$

and x_{i1}, x_{i2} are exogenous explanatory variables. $(\epsilon_{i1}, \epsilon_{i2})$ are i.i.d joint normally with a zero mean and variance-covariance matrix as

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

then the log-likelihood function can be written as Amemiya (1974) and Maddala (1986)

$$\begin{aligned}
\ln L &= d_{i1}d_{i2}\log\left[(1-\theta_{21}\theta_{12})\frac{1}{\sigma_1\sigma_2}\phi_2\Big(\frac{x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1}}{\sigma_1},\frac{x_{i2}\beta_2-\theta_{12}y_{i1}-y_{i2}}{\sigma_2},\rho\Big)\right] \\
&\quad +d_{i1}(1-d_{i2})\log\left[\frac{1}{\sigma_1}\phi\Big(\frac{x_{i1}\beta_1-y_{i1}}{\sigma_1}\Big)\mathbf{\Phi}\left(\frac{\rho(x_{i1}\beta_1-y_{i1})}{\sigma_1\sqrt{1-\rho^2}}-\frac{x_{i2}\beta_2-\theta_{12}y_{i1}}{\sigma_2\sqrt{1-\rho^2}}\right)\right] \\
&\quad +(1-d_{i1})d_{i2}\log\left[\frac{1}{\sigma_2}\phi\Big(\frac{x_{i2}\beta_2-y_{i2}}{\sigma_2}\Big)\mathbf{\Phi}\left(\frac{\rho(x_{i2}\beta_2-y_{i2})}{\sigma_2\sqrt{1-\rho^2}}-\frac{x_{i1}\beta_1-\theta_{21}y_{i2}}{\sigma_1\sqrt{1-\rho^2}}\right)\right] \\
&\quad +(1-d_{i1})(1-d_{i2})\log\left[\mathbf{\Phi}_2\left(-\frac{x_{i1}\beta_1}{\sigma_1},-\frac{x_{i2}\beta_2}{\sigma_2},\rho\right)\right] \\
&= d_{i1}d_{i2}\log\left[\frac{(1-\theta_{21}\theta_{12})}{\sigma_1\sigma_2}\phi_{21}\right]+d_{i1}(1-d_{i2})\log[\frac{1}{\sigma_1}\phi_{01}\mathbf{\Phi}_{11}] \\
&\quad +(1-d_{i1})d_{i2}\log[\frac{1}{\sigma_2}\phi_{02}\mathbf{\Phi}_{12}]+(1-d_{i1})(1-d_{i2})\log\mathbf{\Phi}_{24}
\end{aligned}$$

where $d_{i1}=\mathbf{1}(y_{i1}>0)$ and $d_{i2}=\mathbf{1}(y_{i2}>0)$. The first-order conditions are

$$\begin{aligned}
\frac{\partial\ln L}{\partial\beta_1} &= -d_{i1}d_{i2}\frac{1}{1-\rho^2}\left[\frac{x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1}}{\sigma_1}-\frac{\rho(x_{i2}-\theta_{12}y_{i1}-y_{i2})}{\sigma_2}\right]\frac{x_{i1}}{\sigma_1} \\
&\quad +d_{i1}(1-d_{i2})\left[-\frac{x_{i1}\beta_1-y_{i1}}{\sigma_1}+\frac{\phi_{12}}{\mathbf{\Phi}_{12}}\frac{\rho}{\sqrt{1-\rho^2}}\right]\frac{x_{i1}}{\sigma_1}-(1-d_{i1})d_{i2}\frac{x_{i1}}{\sigma_1\sqrt{1-\rho^2}} \\
&\quad -(1-d_{i1})(1-d_{i2})\frac{1}{\mathbf{\Phi}_{24}}\phi\Big(\frac{x_{i1}\beta_1}{\sigma_1}\Big)\mathbf{\Phi}\Big(\frac{\rho x_{i1}\beta_1}{\sigma_1\sqrt{1-\rho^2}}-\frac{x_{i2}\beta_2}{\sigma_2\sqrt{1-\rho^2}}\Big)\Big(\frac{x_{i1}}{\sigma_1}\Big)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\ln L}{\partial\beta_2} &= -d_{i1}d_{i2}\frac{1}{1-\rho^2}\left[\frac{x_{i2}\beta_2-\theta_{12}y_{i1}-y_{i2}}{\sigma_2}-\frac{\rho(x_{i1}\beta_2-\theta_{21}y_{i2}-y_{i1})}{\sigma_2}\right]\frac{x_{i1}}{\sigma_1} \\
&\quad -d_{i1}(1-d_{i2})\frac{x_{i2}}{\sigma_2\sqrt{1-\rho^2}}+(1-d_{i1})d_{i2}\left[-\frac{x_{i2}\beta_2-y_{i2}}{\sigma_2}+\frac{\phi_{13}}{\mathbf{\Phi}_{13}}\frac{\rho}{\sqrt{1-\rho^2}}\right]\frac{x_{i2}}{\sigma_2} \\
&\quad -(1-d_{i1})(1-d_{i2})\frac{1}{\mathbf{\Phi}_{24}}\phi\Big(\frac{x_{i2}\beta_2}{\sigma_2}\Big)\mathbf{\Phi}\Big(\frac{\rho x_{i2}\beta_2}{\sigma_2\sqrt{1-\rho^2}}-\frac{x_{i1}\beta_1}{\sigma_1\sqrt{1-\rho^2}}\Big)\Big(\frac{x_{i2}}{\sigma_2}\Big) \\
\frac{\partial\ln L}{\partial\theta_{21}} &= d_{i1}d_{i2}\Bigg[-\frac{\theta_{12}}{1-\theta_{21}\theta_{12}} \\
&\quad +\frac{1}{1-\rho^2}\left(\frac{x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1}}{\sigma_1}-\frac{\rho(x_{i2}-\theta_{12}y_{i1}-y_{i2})}{\sigma_2}\right)\frac{y_{i2}}{\sigma_1}\Bigg] \\
&\quad +(1-d_{i1})d_{i2}\frac{\phi_{13}}{\mathbf{\Phi}_{13}}\frac{y_{i2}}{\sigma_1\sqrt{1-\rho^2}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L}{\partial \theta_{12}} &= d_{i1}d_{i2}\bigg[-\frac{\theta_{12}}{1-\theta_{21}\theta_{12}} \\
&\quad +\frac{1}{1-\rho^2}\Big(\frac{x_{i2}\beta_2-\theta_{12}y_{i1}-y_{i2}}{\sigma_2}-\frac{\rho(x_{i1}\beta_2-\theta_{21}y_{i2}-y_{i1})}{\sigma_2}\Big)\frac{y_{i2}}{\sigma_1}\bigg] \\
&\quad +d_{i1}(1-d_{i2})\frac{\phi_{12}}{\mathbf{\Phi}_{12}}\frac{y_{i1}}{\sigma_2\sqrt{1-\rho^2}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L}{\partial \sigma_1} &= d_{i1}d_{i2}\bigg[-\frac{1}{\sigma_1} \\
&\quad +\frac{1}{1-\rho^2}\Big[\frac{(x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1})^2}{\sigma_1^3}-\frac{\rho(x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1})(x_{i2}-\theta_{12}y_{i1}-y_{i2})}{\sigma_1^2\sigma_2}\Big]\bigg] \\
&\quad +d_{i1}(1-d_{i2})\bigg[-\frac{1}{\sigma_1}+\frac{(x_{i1}\beta_1-y_{i1})^2}{\sigma_1^3}-\frac{\phi_{12}}{\mathbf{\Phi}_{12}}\frac{\rho(x_{i1}\beta_1-y_{i1})}{\sigma_1^2\sqrt{1-\rho^2}}\bigg] \\
&\quad +(1-d_{i1})d_{i2}\frac{\phi_{13}}{\mathbf{\Phi}_{13}}\frac{x_{i1}\beta_1-\theta_{21}y_{i2}}{\sigma_1^2\sqrt{1-\rho^2}} \\
&\quad +(1-d_{i1})(1-d_{i2})\frac{1}{\mathbf{\Phi}_{24}}\phi(\frac{x_{i1}\beta_1}{\sigma_1})\mathbf{\Phi}\bigg(-\frac{x_{i2}\beta_2}{\sigma_2\sqrt{1-\rho^2}}+\frac{\rho x_{i1}\beta_1}{\sigma_1\sqrt{1-\rho^2}}\bigg)(\frac{x_{i1}\beta_1}{\sigma_1^2})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L}{\partial \sigma_2} &= d_{i1}d_{i2}\bigg[-\frac{1}{\sigma_2} \\
&\quad +\frac{1}{1-\rho^2}\Big[\frac{(x_{i2}\beta_2-\theta_{12}y_{i1}-y_{i2})^2}{\sigma_2^3}-\frac{\rho(x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1})(x_{i2}-\theta_{12}y_{i1}-y_{i2})}{\sigma_1\sigma_2^2}\Big]\bigg] \\
&\quad +d_{i1}(1-d_{i2})\frac{\phi_{12}}{\mathbf{\Phi}_{12}}\frac{x_{i2}\beta_2-\theta_{12}y_{i1}}{\sigma_2^2\sqrt{1-\rho^2}} \\
&\quad +(1-d_{i1})d_{i2}\bigg[-\frac{1}{\sigma_2}+\frac{(x_{i2}\beta_2-y_{i2})^2}{\sigma_2^3}-\frac{\phi_{13}}{\mathbf{\Phi}_{13}}\frac{\rho(x_{i2}\beta_2-y_{i2})}{\sigma_2^2\sqrt{1-\rho^2}}\bigg] \\
&\quad +(1-d_{i1})(1-d_{i2})\frac{1}{\mathbf{\Phi}_{24}}\phi(\frac{x_{i2}\beta_2}{\sigma_2})\mathbf{\Phi}\bigg(-\frac{x_{i1}\beta_1}{\sigma_1\sqrt{1-\rho^2}}+\frac{\rho x_{i2}\beta_2}{\sigma_2\sqrt{1-\rho^2}}\bigg)(\frac{x_{i2}\beta_2}{\sigma_2^2})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L}{\partial \rho} &= d_{i1}d_{i2}\bigg[\frac{\rho}{1-\rho^2}-\frac{1}{(1-\rho^2)^2}\bigg[\rho\bigg(\frac{x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1}}{\sigma_1}\bigg)^2 \\
&\quad -(1+\rho^2)\bigg(\frac{x_{i1}\beta_1-\theta_{21}y_{i2}-y_{i1}}{\sigma_1}\bigg)\bigg(\frac{x_{i2}\beta_2-\theta_{12}y_{i1}-y_{i2}}{\sigma_2}\bigg) \\
&\quad +\rho\bigg(\frac{x_{i2}\beta_2-\theta_{12}y_{i1}-y_{i2}}{\sigma_2}\bigg)^2\bigg]\bigg] \\
&\quad +d_{i1}(1-d_{i2})\frac{\phi_{12}}{\mathbf{\Phi}_{12}}\bigg[\frac{x_{i1}\beta_1-y_{i1}}{\sigma_1(1-\rho^2)^{\frac{3}{2}}}-\frac{\rho(x_{i2}\beta_2-\theta_{12}y_{i1})}{\sigma_2(1-\rho^2)^{\frac{3}{2}}}\bigg] \\
&\quad +(1-d_{i1})d_{i2}\frac{\phi_{13}}{\mathbf{\Phi}_{13}}\bigg[\frac{x_{i2}\beta_2-y_{i2}}{\sigma_2(1-\rho^2)^{\frac{3}{2}}}-\frac{\rho(x_{i1}\beta_1-\theta_{21}y_{i2})}{\sigma_1(1-\rho^2)^{\frac{3}{2}}}\bigg]+(1-d_{i1})(1-d_{i2})\frac{\phi_{24}}{\mathbf{\Phi}_{24}}
\end{aligned}$$

C.4 Amemiya(1974) Approach

According to Amemiya (1974), we can write the revised form of the log-likelihood function by dividing all the sample agents into two groups. The first group contains all agents with a positive outcome in both activities, and the rest agents are in group 2, here we set the notation as group 2^*, then the group-1 specific reduced form parameters will be applied in the maximum likelihood estimation steps

$$\sum_{1} \log \frac{1}{\sigma_1^*\sigma_2^*}\phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1}}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i2}}{\sigma_2^*}, \rho^*\Big) + \sum_{2^*} \log\Big[1 - \Phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}, \rho^*\Big)\Big] \tag{C.15}$$

the first-order condition, with respect to the reduced form parameter, can be written as

$$\begin{aligned}
\frac{\partial \log \mathbf{L}}{\partial \boldsymbol{\psi}_1^*} =& \sum_{1}\Big[\frac{\rho^*(\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i2})}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}} - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})}{\sigma_1^{*2}\sqrt{1-\rho^{*2}}}\Big]\mathbf{z}_i \\
&- \sum_{2^*}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*)\Phi\big((\mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^* - \rho^*\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*)/\sqrt{1-\rho^{*2}}\big)}{[1-\Phi_2(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*, \mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*, \rho^*)]\sigma_1^*}\mathbf{z}_i = 0 \\
\frac{\partial \log \mathbf{L}}{\partial \boldsymbol{\psi}_2^*} =& \sum_{1}\Big[\frac{\rho^*(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})}{\sigma_1^*\sigma_2^*\sqrt{1-\rho^{*2}}} - \frac{(\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i2})}{\sigma_2^{*2}\sqrt{1-\rho^{*2}}}\Big]\mathbf{z}_i \\
&- \sum_{2^*}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*)\Phi\big((\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^* - \rho^*\mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*)/\sqrt{1-\rho^{*2}}\big)}{[1-\Phi_2(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*, \mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*, \rho^*)]\sigma_2^*}\mathbf{z}_i = 0
\end{aligned} \tag{C.16}$$

$$\begin{aligned}
\frac{\partial \log \mathbf{L}}{\partial \sigma_1^*} =& \sum_{1}\Big[-\frac{1}{\sigma_1^*} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})^2}{(1-\rho^{*2})\sigma_1^{*3}} - \frac{\rho^*(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i2})(\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i1})}{(1-\rho^{*2})\sigma_1^{*2}\sigma_2^*}\Big] \\
&+ \sum_{2^*}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*)\Phi\big((\mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^* - \rho^*\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*)/\sqrt{1-\rho^{*2}}\big)\mathbf{z}_i'\boldsymbol{\psi}_1^*}{[1-\Phi_2(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*, \mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*, \rho^*)]\sigma_1^{*2}} = 0 \\
\frac{\partial \log \mathbf{L}}{\partial \sigma_2^*} =& \sum_{1}\Big[-\frac{1}{\sigma_2^*} + \frac{(\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i2})^2}{(1-\rho^{*2})\sigma_2^{*3}} - \frac{\rho^*(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i2})(\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i1})}{(1-\rho^{*2})\sigma_1^*\sigma_2^{*2}}\Big] \\
&+ \sum_{2^*}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*)\Phi\big((\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^* - \rho^*\mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*)/\sqrt{1-\rho^{*2}}\big)\mathbf{z}_i'\boldsymbol{\psi}_2^*}{[1-\Phi_2(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*, \mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*, \rho^*)]\sigma_2^{*2}} = 0
\end{aligned} \tag{C.17}$$

$$\begin{aligned}
\frac{\partial \log \mathbf{L}}{\partial \rho^*} =& \sum_{1}\Big[\frac{\rho^*}{1-\rho^{*2}} - \frac{\rho^*(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i1})^2}{(1-\rho^{*2})^2\sigma_1^{*2}} - \frac{\rho^*(\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i2})^2}{(1-\rho^{*2})^2\sigma_2^{*2}} + \frac{(1+\rho^{*2})(\mathbf{z}_i'\boldsymbol{\psi}_1^* - y_{i2})(\mathbf{z}_i'\boldsymbol{\psi}_2^* - y_{i1})}{(1-\rho^{*2})^2\sigma_1^*\sigma_2^{*2}}\Big] \\
&+ \sum_{2^*}\frac{\phi_2(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*, \mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*, \rho^*)}{[1-\Phi_2(\mathbf{z}_i'\boldsymbol{\psi}_1^*/\sigma_1^*, \mathbf{z}_i'\boldsymbol{\psi}_2^*/\sigma_2^*, \rho^*)]} = 0
\end{aligned} \tag{C.18}$$

Amemiya (1974) introduces IV estimation based on the reduced form. As we only use agents in group 1, i.e., $y_{i1}^* > 0$ and $y_{i2}^* > 0$ which is the same as $\epsilon_{i1}^* \leq \mathbf{z}_i'\boldsymbol{\psi}_1^*$ and $\epsilon_{i2}^* \leq \mathbf{z}_i'\boldsymbol{\psi}_2^*$, which can also be rewritten as $-\epsilon_{i1}^* > \mathbf{z}_i'\boldsymbol{\psi}_1^*$ and $-\epsilon_{i2}^* > -\mathbf{z}_i'\boldsymbol{\psi}_2^*$. Then according to equation (2.14) in Amemiya (1974), we will have the similar form of equation (4.7) in Amemiya (1974) as following

$$\begin{aligned} y_{i1}^{*2} =& \rho^* \frac{\sigma_1^*}{\sigma_2^*} y_{i1}^* y_{i2}^* - \sigma_1^{*2}(1-\rho^{*2}) - \mathbf{z}_i'(\sigma_2^{*2}\boldsymbol{\psi}_1^* - \rho^*\sigma_1^*\sigma_2^*\boldsymbol{\psi}_2^*)y_{i1} + \eta_{i1} \\ y_{i2}^{*2} =& \rho^* \frac{\sigma_2^*}{\sigma_1^*} y_{i1}^* y_{i2}^* - \sigma_2^{*2}(1-\rho^{*2}) - \mathbf{z}_i'(\sigma_1^{*2}\boldsymbol{\psi}_2^* - \rho^*\sigma_1^*\sigma_2^*\boldsymbol{\psi}_1^*)y_{i2} + \eta_{i2} \end{aligned} \tag{C.19}$$

To get the instrumental variables, we can regress each y_{i1}^* and y_{i2}^* on $\mathbf{z}_i$, $\mathbf{z}_i^2$, $\cdots$. Then we can obtain the least-squares prediction of y_{i1}^* and y_{i2}^*, which can be denoted as $\hat{y}_{i1}^*$ and $\hat{y}_{i2}^*$. Then we use $(1, \hat{y}_{i1}^*\hat{y}_{i2}^*, \mathbf{z}_i'\hat{y}_{i1}^*)'$ and $(1, \hat{y}_{i1}^*\hat{y}_{i2}^*, \mathbf{z}_i'\hat{y}_{i2}^*)'$ as instrument variables to estimate σ_1^*, σ_2^*, ρ^*, $\boldsymbol{\psi}_1^*$, and $\boldsymbol{\psi}_2^*$. Then, according to (2.6) and (2.7) in Amemiya (1974), we can derive

$$\begin{aligned} P\mathrm{E}(\epsilon_{i1}^*) =& -\sigma_1^* f_1 F_2 - \rho^*\sigma_1^* f_2 F_1 \\ P\mathrm{E}(\epsilon_{i2}^*) =& -\sigma_2^* f_2 F_1 - \rho^*\sigma_2^* f_1 F_2 \\ P\mathrm{E}(\epsilon_{i1}^{*2}) =& \sigma_1^{*2}P + \rho^*(1-\rho^{*2})\sigma_1^{*2}\phi_2^* - \sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_1^* f_1 F_2 - \rho^{*2}\frac{\sigma_1^{*2}}{\sigma_2^*}\mathbf{z}_i'\boldsymbol{\psi}_2^* f_2 F_1 \\ P\mathrm{E}(\epsilon_{i2}^{*2}) =& \sigma_2^{*2}P + \rho^*(1-\rho^{*2})\sigma_2^{*2}\phi_2^* - \sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_2^* f_2 F_1 - \rho^{*2}\frac{\sigma_2^{*2}}{\sigma_1^*}\mathbf{z}_i'\boldsymbol{\psi}_1^* f_1 F_2 \\ P\mathrm{E}(\epsilon_{i1}^*\epsilon_{i2}^*) =& \rho^*\sigma_1^*\sigma_2^*P + (1-\rho^{*2})\sigma_1^*\sigma_2^*\phi_2^* - \rho^*\sigma_2^*\mathbf{z}_i'\boldsymbol{\psi}_1^* f_1 F_2 - \rho^*\sigma_1^*\mathbf{z}_i'\boldsymbol{\psi}_2^* f_2 F_1 \end{aligned} \tag{C.20}$$

where

$$\begin{aligned} P =& \boldsymbol{\Phi}_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}, \rho^*\Big) \\ f_1 =& \phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}\Big) \\ f_2 =& \phi\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}\Big) \\ F_1 =& \boldsymbol{\Phi}\Bigg(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sqrt{1-\rho^{*2}}} - \frac{\rho^*\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*\sqrt{1-\rho^{*2}}}\Bigg) \\ F_2 =& \boldsymbol{\Phi}\Bigg(\frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*\sqrt{1-\rho^{*2}}} - \frac{\rho^*\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*\sqrt{1-\rho^{*2}}}\Bigg) \\ \phi_2^* =& \phi_2\Big(\frac{\mathbf{z}_i'\boldsymbol{\psi}_1^*}{\sigma_1^*}, \frac{\mathbf{z}_i'\boldsymbol{\psi}_2^*}{\sigma_2^*}, \rho^*\Big) \end{aligned} \tag{C.21}$$

then according to (2.7) in Amemiya (1974), we can derive a similar form as (4.9) in Amemiya (1974) as following

$$\begin{aligned} y_{i1}^* =& \mathbf{z}_i'\boldsymbol{\psi}_1^* + \frac{1}{P}\sigma_1^* f_1 F_2 + \frac{1}{P}\rho^*\sigma_1^* f_2 F_1 + \xi_{i1} \\ y_{i2}^* =& \mathbf{z}_i'\boldsymbol{\psi}_2^* + \frac{1}{P}\sigma_2^* f_2 F_1 + \frac{1}{P}\rho^*\sigma_2^* f_1 F_2 + \xi_{i2} \end{aligned} \tag{C.22}$$

Here are the steps for estimation

(i) Use only group 1 observations that $y_{i1}^* > 0$ and $y_{i2}^* > 0$.

(ii) Apply the IV estimation to (C.20)

(iii) Use the results in (ii) to estimate the regressors of (C.22)

(iv) Derive the structural parameters.

C.5 Binary Case

Now, we consider the situation in which the outcome is binary. The econometric model for two actions is

$$\begin{aligned} y_{i1} &= -\theta_{21} d_{i2} + \lambda_{11} \sum_{j=1}^{n} w_{ij} p_{j1} + \lambda_{21} \sum_{j=1}^{n} w_{ij} p_{j2} + \mathbf{x}_i' \beta_1 - \epsilon_{i1} \\ y_{i2} &= -\theta_{12} d_{i1} + \lambda_{12} \sum_{j=1}^{n} w_{ij} p_{j1} + \lambda_{22} \sum_{j=1}^{n} w_{ij} p_{j2} + \mathbf{x}_i' \beta_2 - \epsilon_{i2} \end{aligned}$$

where

$$\begin{aligned} d_{i1} &= \mathbf{1}(y_{i1} > 0) \\ d_{i2} &= \mathbf{1}(y_{i2} > 0) \end{aligned}$$

and without loss of generality, $(\epsilon_{i1}, \epsilon_{i2})$ satisfy the join normal distribution with zero mean, unit variance, and correlation ρ. Then, we have the rational expectation

$$\begin{aligned} p_{i1} &= \mathrm{E}(d_{i1}) = \mathrm{E}(d_{i1}|d_{i1}=1)\mathbf{Pr}(d_{i1}=1) + \mathrm{E}(d_{i1}|d_{i1}=0)\mathbf{Pr}(d_{i1}=0) \\ &= \mathrm{E}(d_{i1}|y_{i1}>0)\mathbf{Pr}(y_{i1}>0) \\ &= \mathrm{E}(d_{i1}|y_{i1}>0, y_{i2}>0)\mathbf{Pr}(y_{i1}>0, y_{i2}>0) \\ &\quad +\mathrm{E}(d_{i1}|y_{i1}>0, y_{i2}\le 0)\mathbf{Pr}(y_{i1}>0, y_{i2}\le 0) \\ &= \mathbf{Pr}(y_{i1}>0, y_{i2}>0) + \mathbf{Pr}(y_{i1}>0, y_{i2}\le 0) \\ &= \mathbf{Pr}(\epsilon_{i1} < \mathbf{z}_i'\boldsymbol{\psi}_1 - \theta_{21}, \epsilon_{i2} < \mathbf{z}_i'\boldsymbol{\psi}_2 - \theta_{12}) + \mathbf{Pr}(\epsilon_{i1} < \mathbf{z}_i'\boldsymbol{\psi}_1, \epsilon_{i2} \ge \mathbf{z}_i'\boldsymbol{\psi}_2 - \theta_{12}) \\ &= \boldsymbol{\Phi}_2\Big(\mathbf{z}_i'\boldsymbol{\psi}_1 - \theta_{21}, \mathbf{z}_i'\boldsymbol{\psi}_2 - \theta_{12}, \rho\Big) + \boldsymbol{\Phi}_2\Big(\mathbf{z}_i'\boldsymbol{\psi}_1, -\mathbf{z}_i'\boldsymbol{\psi}_2 + \theta_{12}, -\rho\Big) \end{aligned}$$

$$
\begin{aligned}
p_{i2} &= \mathrm{E}(d_{i2}) = \mathrm{E}(d_{i2}|d_{i2}=1)\mathbf{Pr}(d_{i2}=1) + \mathrm{E}(d_{i2}|d_{i2}=0)\mathbf{Pr}(d_{i1}=0) \\
&= \mathrm{E}(d_{i1}|y_{i2}>0)\mathbf{Pr}(y_{i2}>0) \\
&= \mathrm{E}(d_{i1}|y_{i1}>0, y_{i2}>0)\mathbf{Pr}(y_{i1}>0, y_{i2}>0) \\
&\quad +\mathrm{E}(d_{i1}|y_{i1}\leq 0, y_{i2}>0)\mathbf{Pr}(y_{i1}>0, y_{i2}\leq 0) \\
&= \mathbf{Pr}(y_{i1}>0, y_{i2}>0) + \mathbf{Pr}(y_{i1}\leq 0, y_{i2}>0) \\
&= \mathbf{Pr}(\epsilon_{i1} < \mathbf{z}_i'\boldsymbol{\psi}_1 - \theta_{21}, \epsilon_{i2} < \mathbf{z}_i'\boldsymbol{\psi}_2 - \theta_{12}) + \mathbf{Pr}(\epsilon_{i1} \geq \mathbf{z}_i'\boldsymbol{\psi}_1 - \theta_{21}, \epsilon_{i2} < \mathbf{z}_i'\boldsymbol{\psi}_2) \\
&= \boldsymbol{\Phi}_2\Big(\mathbf{z}_i'\boldsymbol{\psi}_1 - \theta_{21}, \mathbf{z}_i'\boldsymbol{\psi}_2 - \theta_{12}, \rho\Big) + \boldsymbol{\Phi}_2\Big(-\mathbf{z}_i'\boldsymbol{\psi}_1 + \theta_{21}, \mathbf{z}_i'\boldsymbol{\psi}_2, -\rho\Big)
\end{aligned}
$$

The log-likelihood function is

$$
\begin{aligned}
& \sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\mathbf{Pr}(d_1=1,d_2=1) + d_{i1}(1-d_{i2})\ln\mathbf{Pr}(d_1=1,d_2=0) \\
& +(1-d_{i1})d_{i2}\ln\mathbf{Pr}(d_1=0,d_2=1) + (1-d_{i1})(1-d_{i2})\ln\mathbf{Pr}(d_1=0,d_2=0)\Big] \\
= & \sum_{i=1}^{n} d_{i1}d_{i2}\ln\boldsymbol{\Phi}_2\Big(\mathbf{z}_i'\boldsymbol{\psi}_1 - \theta_{21}, \mathbf{z}_i'\boldsymbol{\psi}_2 - \theta_{12}, \rho\Big) + d_{i1}(1-d_{i2})\ln\boldsymbol{\Phi}_2\Big(\mathbf{z}_i'\boldsymbol{\psi}_1, -\mathbf{z}_i'\boldsymbol{\psi}_2 + \theta_{12}, -\rho\Big) \\
& +(1-d_{i1})d_{i2}\ln\boldsymbol{\Phi}_2\Big(-\mathbf{z}_i'\boldsymbol{\psi}_1 + \theta_{21}, \mathbf{z}_i'\boldsymbol{\psi}_2, -\rho\Big) + (1-d_{i1})(1-d_{i2})\ln\boldsymbol{\Phi}_2\Big(-\mathbf{z}_i'\boldsymbol{\psi}_1, -\mathbf{z}_i'\boldsymbol{\psi}_2, \rho\Big)
\end{aligned}
$$

the first-order condition can be written as

$$
\begin{aligned}
\frac{\partial \ln \mathbf{L}}{\partial \boldsymbol{\psi}_1} &= \sum_{i=1}^{n}\Bigg[d_{i1}d_{i2}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})-\rho(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12},\rho)}\mathbf{z}_i\\
&+d_{i1}(1-d_{i2})\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1)\boldsymbol{\Phi}\big([-(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})+\rho(\mathbf{z}_i'\boldsymbol{\psi}_1)]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2\big(\mathbf{z}_i'\boldsymbol{\psi}_1,-\mathbf{z}_i'\boldsymbol{\psi}_2+\theta_{12},-\rho\big)}\mathbf{z}_i\\
&-(1-d_{i1})d_{i2}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_2)-\rho(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1+\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2,-\rho)}\mathbf{z}_i\\
&-(1-d_{i1})(1-d_{i2})\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1)\boldsymbol{\Phi}\big((-\mathbf{z}_i'\boldsymbol{\psi}_2+\rho\mathbf{z}_i'\boldsymbol{\psi}_1)/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1,-\mathbf{z}_i'\boldsymbol{\psi}_2,\rho)}\mathbf{z}_i\Bigg]=0\\
\frac{\partial \ln \mathbf{L}}{\partial \boldsymbol{\psi}_2} &= \sum_{i=1}^{n}\Bigg[d_{i1}d_{i2}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})-\rho(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12},\rho)}\mathbf{z}_i\\
&-d_{i1}(1-d_{i2})\frac{\phi\big((\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})\big)\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_1)-\rho(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2((\mathbf{z}_i'\boldsymbol{\psi}_1),-(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}),-\rho)}\mathbf{z}_i\\
&+d_{i1}(1-d_{i2})\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_2)\boldsymbol{\Phi}\big([-(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})+\rho(\mathbf{z}_i'\boldsymbol{\psi}_2)]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1+\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2,-\rho)}\mathbf{z}_i\\
&-(1-d_{i1})(1-d_{i2})\frac{\phi\big((\mathbf{z}_i'\boldsymbol{\psi}_2)\big)\boldsymbol{\Phi}\big([-(\mathbf{z}_i'\boldsymbol{\psi}_1)+\rho(\mathbf{z}_i'\boldsymbol{\psi}_2)]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1,-\mathbf{z}_i'\boldsymbol{\psi}_2,\rho)}\mathbf{z}_i\Bigg]=0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ln \mathbf{L}}{\partial \theta_{21}} &= \sum_{i=1}^{n}\Bigg[-d_{i1}d_{i2}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})-\rho(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12},\rho)}\\
&-d_{i1}(1-d_{i2})\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1)\boldsymbol{\Phi}\big([-(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})+\rho(\mathbf{z}_i'\boldsymbol{\psi}_1)]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2\big(\mathbf{z}_i'\boldsymbol{\psi}_1,-\mathbf{z}_i'\boldsymbol{\psi}_2+\theta_{12},-\rho\big)}\\
&+(1-d_{i1})d_{i2}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_2)-\rho(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1+\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2,-\rho)}\\
&+(1-d_{i1})(1-d_{i2})\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_1)\boldsymbol{\Phi}\big((-\mathbf{z}_i'\boldsymbol{\psi}_2+\rho\mathbf{z}_i'\boldsymbol{\psi}_1)/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1,-\mathbf{z}_i'\boldsymbol{\psi}_2,\rho)}\Bigg]=0\\
\frac{\partial \ln \mathbf{L}}{\partial \theta_{12}} &= \sum_{i=1}^{n}\Bigg[-d_{i1}d_{i2}\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})-\rho(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12},\rho)}\\
&+d_{i1}(1-d_{i2})\frac{\phi\big((\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})\big)\boldsymbol{\Phi}\big([(\mathbf{z}_i'\boldsymbol{\psi}_1)-\rho(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12})]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2((\mathbf{z}_i'\boldsymbol{\psi}_1),-(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}),-\rho)}\\
&-d_{i1}(1-d_{i2})\frac{\phi(\mathbf{z}_i'\boldsymbol{\psi}_2)\boldsymbol{\Phi}\big([-(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21})+\rho(\mathbf{z}_i'\boldsymbol{\psi}_2)]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1+\theta_{21},\mathbf{z}_i'\boldsymbol{\psi}_2,-\rho)}\\
&+(1-d_{i1})(1-d_{i2})\frac{\phi\big((\mathbf{z}_i'\boldsymbol{\psi}_2)\big)\boldsymbol{\Phi}\big([-(\mathbf{z}_i'\boldsymbol{\psi}_1)+\rho(\mathbf{z}_i'\boldsymbol{\psi}_2)]/\sqrt{1-\rho^2}\big)}{\boldsymbol{\Phi}_2(-\mathbf{z}_i'\boldsymbol{\psi}_1,-\mathbf{z}_i'\boldsymbol{\psi}_2,\rho)}\Bigg]=0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ln \mathbf{L}}{\partial \rho} &= \sum_{i=1}^{n}\Bigg[d_{i1}d_{i2}\frac{\phi_2\big((\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}),(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}),\rho\big)}{\boldsymbol{\Phi}_2((\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}),(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}),\rho)}\\
&-d_{i1}(1-d_{i2})\frac{\phi_2((\mathbf{z}_i'\boldsymbol{\psi}_1),-(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}),-\rho)}{\boldsymbol{\Phi}_2((\mathbf{z}_i'\boldsymbol{\psi}_1),-(\mathbf{z}_i'\boldsymbol{\psi}_2-\theta_{12}),-\rho)}\\
&-(1-d_{i1})d_{i2}\frac{\phi_2(-(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}),(\mathbf{z}_i'\boldsymbol{\psi}_2),-\rho)}{\boldsymbol{\Phi}_2(-(\mathbf{z}_i'\boldsymbol{\psi}_1-\theta_{21}),(\mathbf{z}_i'\boldsymbol{\psi}_2),-\rho)}\\
&+(1-d_{i1})(1-d_{i2})\frac{\phi_2(-(\mathbf{z}_i'\boldsymbol{\psi}_1),-(\mathbf{z}_i'\boldsymbol{\psi}_2),\rho)}{\boldsymbol{\Phi}_2(-(\mathbf{z}_i'\boldsymbol{\psi}_1),-(\mathbf{z}_i'\boldsymbol{\psi}_2),\rho)}\Bigg]=0
\end{aligned}
$$

C.5.1 Binary simulation preparation

Before simulating the binary model, we start with a simplified case that only contains pure exogenous regressor

$$
\begin{aligned}
y_{i1} &= -\theta_{21}d_{i2}+\mathbf{x}_{i1}'\beta_1-\epsilon_{i1}\\
y_{i2} &= -\theta_{12}d_{i1}+\mathbf{x}_{i2}'\beta_2-\epsilon_{i2}
\end{aligned}
$$

where

$$
\begin{aligned}
d_{i1} &= \mathbf{1}(y_{i1}>0)\\
d_{i2} &= \mathbf{1}(y_{i2}>0)
\end{aligned}
$$

then we can write the log-likelihood as

$$
\begin{aligned}
&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\mathbf{Pr}(d_{i1}=1,d_{i2}=1)+d_{i1}(1-d_{i2})\ln\mathbf{Pr}(d_{i1}=1,d_{i2}=0)\\
&+(1-d_{i1})d_{i2}\log\mathbf{Pr}(d_{i1}=0,d_{i2}=1)+(1-d_{i1})(1-d_{i2})\ln\mathbf{Pr}(d_{i1}=0,d_{i2}=0)\Big]\\
=\;&\sum_{i=1}^{n}\Big[d_{i1}d_{i2}\ln\boldsymbol{\Phi}_2(x_{i1}\beta_1-\theta_{21},x_{i2}\beta_2-\theta_{12},\rho)\\
&+d_{i1}(1-d_{i2})\ln\boldsymbol{\Phi}_2(x_{i1}\beta_1,-x_{i2}\beta_2+\theta_{12},-\rho)\\
&+(1-d_{i1})d_{i2}\ln\boldsymbol{\Phi}_2(-x_{i1}\beta_1+\theta_{21},x_{i2}\beta_2,-\rho)\\
&+(1-d_{i1})(1-d_{i2})\ln\boldsymbol{\Phi}_2(-x_{i1}\beta_1,-x_{i2}\beta_2,\rho)
\end{aligned}
$$

C.5.2 Preparation Simulation Results

Simulation model

$$d_{i1} = \mathbf{1}(-\theta_{21}d_{i1} + x_{i1}\beta_1 - \epsilon_{i1} > 0)$$

$$d_{i2} = \mathbf{1}(-\theta_{12}d_{i2} + x_{i2}\beta_2 - \epsilon_{i2} > 0)$$

Estimate all parameters in one step $\beta_1 = \beta_2 = 1$						
$\theta_{21} = \theta_{12}$	ρ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{12}$	$\hat{\rho}$
0.1	0.1	0.877 (0.220)	0.876 (0.220)	0.373 (0.356)	0.386 (0.359)	0.404 (0.390)
0.2	0.3	0.822 (0.199)	0.819 (0.199)	0.587 (0.285)	0.590 (0.283)	0.710 (0.282)
0.3	0.3	0.733 (0.173)	0.738 (0.173)	0.823 (0.200)	0.823 (0.203)	0.872 (0.190)

Estimate β and θ in the first step, then estimate ρ, $\beta_1 = \beta_2 = 1$						
$\theta_{21} = \theta_{12}$	ρ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{12}$	$\hat{\rho}$
0.1	0.1	1.011 (0.089)	1.007 (0.089)	0.084 (0.077)	0.087 (0.078)	0.086 (0.071)
0.2	0.3	1.022 (0.087)	1.012 (0.087)	0.089 (0.077)	0.096 (0.081)	0.201 (0.075)
0.3	0.3	1.026 (0.089)	1.026 (0.089)	0.205 (0.096)	0.205 (0.094)	0.245 (0.075)

We could find that, no matter how we choose the order of the parameter to estimate, both the accuracy and consistency perform poorly. After diving into the data-generating process, we found that such DGP exists in unstable data points. For example, if the unobserved error terms ϵ_{i1}, ϵ_{i2}

satisfied the following

$$x_{i1}\beta_1 - \theta_{21} < \quad \epsilon_{i1} < \quad x_{i1}\beta_1$$

$$x_{i2}\beta_2 - \theta_{12} < \quad \epsilon_{i2} < \quad x_{i2}\beta_2$$

suppose the DGP of (d_{i1}, d_{i2}) starting at $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$, then we have

$$\begin{aligned}
(0,0) &\Longrightarrow \quad (1,1) \Longrightarrow \quad (0,0) \Longrightarrow \quad \cdots \\
(0,1) &\Longrightarrow \quad (0,1) \Longrightarrow \quad (0,1) \Longrightarrow \quad \cdots \\
(1,0) &\Longrightarrow \quad (1,0) \Longrightarrow \quad (1,0) \Longrightarrow \quad \cdots \\
(1,1) &\Longrightarrow \quad (0,0) \Longrightarrow \quad (1,1) \Longrightarrow \quad \cdots
\end{aligned} \tag{C.23}$$

the results are uncertain between $(1,0)$ and $(0,1)$. And we can summarize the stability of data-generating process results with respect to the range of $(\epsilon_{i1}, \epsilon_{i2})$

$(\epsilon_{i1}, \epsilon_{i2})$	$(-\infty, x_{i1}\beta_1 - \theta_{21})$	$(x_{i1}\beta_1 - \theta_{21}, x_{i1}\beta_1)$	$(x_{i1}\beta_1, +\infty)$
$(-\infty, x_{i2}\beta_2 - \theta_{12})$	$(1,1)$	$(1,0)$	$(0,1)$
$(x_{i2}\beta_2 - \theta_{12}, x_{i2}\beta_2)$	$(1,0)$	$(1,0)$ or $(0,1)$	$(0,1)$
$(x_{i2}\beta_2, \infty)$	$(1,0)$	$(0,1)$	$(0,0)$

In Schmidt (1981), to make such a simultaneous equation Probit model identifiable, all principal minor of $\mathbf{\Theta}$ must equal one if $m = 2$, the econometric model is

$$y_{i1} = \mathbf{x}'_{i1}\beta_1 - \epsilon_{i1}$$

$$y_{i2} = -\theta_{12}d_{i1} + \mathbf{x}'_{i2}\beta_2 - \epsilon_{i2}$$

if $m = 3$, the econometric model is

$$y_{i1} = \mathbf{x}'_{i1}\beta_1 - \epsilon_{i1}$$

$$y_{i2} = -\theta_{12}d_{i1} + \mathbf{x}'_{i2}\beta_2 - \epsilon_{i2}$$

$$y_{i3} = -\theta_{13}d_{i1} - \theta_{23}d_{i2} + \mathbf{x}'_{i3}\beta_3 - \epsilon_{i3}$$

Given the data-generating process, we can get the following simulation results

$$y_{i1} = x_{i1}\beta_1 - \epsilon_{i1}$$

$$y_{i2} = -\theta_{12}d_{i1} + x_{i2}\beta_2 - \epsilon_{i2}$$

θ_{12}	ρ	$\hat{\theta}_{12}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
0.2	0.2	0.202 (0.104)	1.004 (0.086)	1.008 (0.086)	0.201 (0.109)
0.5	0.2	0.501 (0.113)	1.004 (0.087)	1.006 (0.087)	0.200 (0.111)
0.8	0.2	0.801 (0.115)	1.004 (0.089)	1.004 (0.089)	0.199 (0.112)
0.2	0.5	0.201 (0.101)	1.004 (0.084)	1.009 (0.084)	0.502 (0.096)
0.5	0.5	0.500 (0.104)	1.004 (0.084)	1.007 (0.084)	0.502 (0.096)
0.8	0.5	0.802 (0.104)	1.004 (0.084)	1.005 (0.084)	0.501 (0.098)
0.2	0.8	0.199 (0.092)	1.007 (0.129)	1.011 (0.129)	0.805 (0.062)
0.5	0.8	0.501 (0.090)	1.004 (0.085)	1.008 (0.085)	0.804 (0.064)
0.8	0.8	0.803 (0.089)	1.005 (0.086)	1.006 (0.086)	0.807 (0.064)

Consider the following econometric model

$$y_{i1}^* = -\theta_{21} d_{i2} + \mathbf{x}_{i1}' \beta_1 - \epsilon_{i1}$$

$$y_{i2}^* = -\theta_{12} y_{i1} + \mathbf{x}_{i2}' \beta_2 - \epsilon_{i2}$$

where

$$y_{i1} = y_{i1} \mathbf{1}(y_{i1}^* > 0)$$

$$d_{i2} = \mathbf{1}(y_{i2}^* > 0)$$

This model is a simultaneous equation Tobit-Probit model; the simultaneous effect passes through agents' decision outcomes, and the dependent variable contains both censored and binary. According to the outcome value distribution, we have the following four cases

- $y_{i1} > 0$, $d_{i2} = 1$

$$\epsilon_{i1} < x_{i1}\beta_1 - \theta_{21}$$

$$\epsilon_{i2} - \theta_{12}\epsilon_{i1} < x_{i2}\beta_2 - \theta_{12}x_{i1}\beta_1 + \theta_{12}\theta_{21}$$

- $y_{i1} > 0$, $d_{i2} = 0$

$$\epsilon_{i1} < x_{i1}\beta_1$$

$$\epsilon_{i2} - \theta_{12}\epsilon_{i1} \geq x_{i2}\beta_2 - \theta_{12}x_{i1}\beta_1$$

- $y_{i1} = 0$, $d_{i2} = 1$

$$\epsilon_{i1} \geq x_{i1}\beta_1 - \theta_{21}$$

$$\epsilon_{i2} < x_{i2}\beta_2$$

- $y_{i1} = 0$, $d_{i2} = 0$

$$\epsilon_{i1} \geq x_{i1}\beta_1$$

$$\epsilon_{i2} \geq x_{i2}\beta_2$$

However, when $(\epsilon_{i1}, \epsilon_{i2})$ satisfies the following conditions

$$\epsilon_{i1} < x_{i1}\beta_1 - \theta_{21}$$

$$x_{i2}\beta_2 - \theta_{12}x_{i1}\beta_1 \leq \epsilon_{i2} - \theta_{12}\epsilon_{i1} < x_{i2}\beta_2 - \theta_{12}x_{i1}\beta_1 + \theta_{12}\theta_{21}$$

the outcome of agent i is uncertain between $(y_{i1}, 1)$ and $(y_{i1}, 0)$.

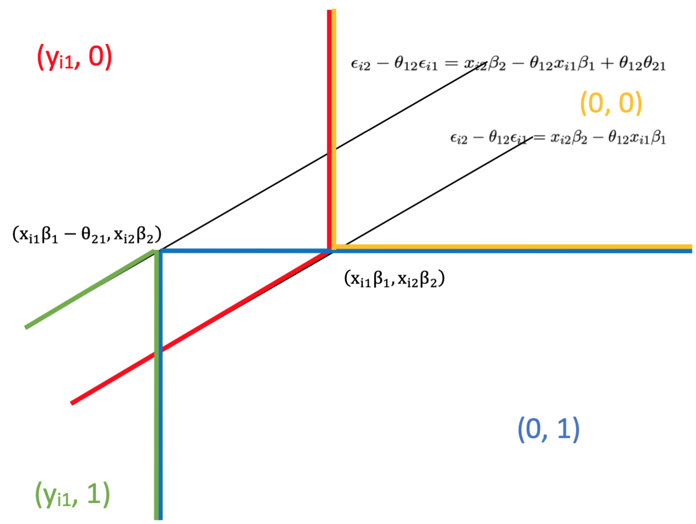
And when $(\epsilon_{i1}, \epsilon_{i2})$ satisfies the following conditions

$$x_{i1}\beta_1 - \theta_{21} \leq \epsilon_{i1} < x_{i1}\beta_1$$

$$x_{i2}\beta_2 - \theta_{12}x_{i1}\beta_1 \leq \epsilon_{i2} - \theta_{12}\epsilon_{i1}$$

$$\epsilon_{i2} < x_{i2}\beta_2$$

the outcome of agent i is uncertain between $(y_{i1}, 0)$ and $(0, 1)$.The following graph reflects the instability of this model's data-generating process.



To make the model identifiable, we need to set $\theta_{21} = 0$, and the model will be

$$y_{i1}^* = \mathbf{x}_{i1}'\beta_1 - \epsilon_{i1}$$

$$y_{i2}^* = -\theta_{12}y_{i1} + \mathbf{x}_{i2}'\beta_2 - \epsilon_{i2}$$

where

$$y_{i1} = y_{i1}\mathrm{I}(y_{i1}^* > 0)$$

$$d_{i2} = \mathrm{I}(y_{i2}^* > 0)$$