DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 23-10

Simultaneous Equations with Censored Outcomes and Social Interactions

Sutianjie Zhou
University of Colorado Boulder

October 31, 2023

Department of Economics

University of Colorado Boulder
Boulder, Colorado 80309

Simultaneous Equations with Censored Outcomes and Social Interactions

Sutianjie Zhou*

Oct 31, 2023

Abstract

This paper introduces a censored-outcome simultaneous-equation model with social interactions. The construction of the microeconomics foundation for this model is from the equilibrium in a large-network-based game with incomplete information, in which each agent conducts multiple actions and interacts with other agents through a network through a linear quadratic utility function. The sufficient condition of the unique Bayesian Nash Equilibrium (BNE) existence is characterized. We also discuss the identification of the econometric model. We propose a two-stage method to estimate the model in which we apply the nested pseudo-likelihood (NPL) to estimate the reduced parameters and then derive the structural form parameters by Amemiya Generalized Least Square estimator (AGLS). Monte Carlo simulation shows that the estimation performs well in finite samples. The estimation also shows the feasibility of the computation when the network size is large.

JEL classification: C31, C34

Key words: Censored Model, Limited Dependent, Networks, Spatial Autoregressive Models, Simultaneous Equation Models, Social Interaction Models, Tobit model

1 Introduction

In many real economic situations, the outcomes of agents' activities are censored. For example, the spending amount of household annually traveling should be zero or some positive value. Researchers of teenagers' behaviors are interested in how many cigarettes a teenager smokes and how many alcoholic beverages a teenager drinks. Both results are also zeros or positive values. Such result of each agent's certain activity can be influenced by the same agent's other activities' results and also can be influenced by other agents' activities'. From the household consumption example, a household's annual traveling spending will influence its annual entertainment spending and vice

*Department of Economics, University of Colorado at Boulder

versa. And both spending can be impacted by their friends' spending. As for the teenager behavior research example, the number of cigarettes a teenager smokes also impacts how many alcoholic beverages he/she drinks and vice versa. Both the number of cigarettes he/she smokes and the number of alcoholic beverages he/she drinks are influenced by his/her friends. The multi-action censored outcomes of different actions determined by the same agent inter-depend, and agents also impact each other. This paper proposes a simultaneous equation model with censored outcomes and social interactions to characterize the interdependence of an agent's economic decision outcomes both across agents and economic activities.

From the seminal literature, Amemiya (1974) proposed a multivariate Tobit model, which is an extension of Tobin (1958) and Amemiya (1973), and build a simplified version of consistent estimation by grouping agents into two types based on the outcomes of there activities. Nelson and Olson (1978) generate a different multivariate Tobit model with latent endogenous variables under a more general truncation structure, i.e., not all zeros, and also develop an estimator. The model they propose has one completely observed endogenous variable and one truncated. They derived a reduced-form-based two-stage estimation containing the application of the least square method and likelihood method. Then, Amemiya (1979) extends the simultaneous-equation Tobit model more generally, which combines truncated and non-truncated endogenous outcomes and develops a GLS-based estimator. According to the variance-covariance matrix developing, the estimation process Amemiya (1979) is more efficient than that in Nelson and Olson (1978). Our econometric model is similar to the traditional simultaneous equation model, but the endogenous variables of our structure model are dependent not on the observed outcomes but on the unobserved reservation values.

Our paper is based on the intuition in which each agent conducts various activities at the same time and shows the results, and each agent cannot observe others' outcomes before making decisions, however, their expectation of other agents' outcomes will influence their behaviors. That means the econometricians and economists need to figure out the Bayesian Nash Equilibrium in the incomplete information network game. Liu (2019) discuss a simultaneous-equation binary model with social interactions under incomplete information and conduct the estimation based on maximum likelihood and nested pseudo-likelihood, which is introduced in Aguirregabiria and Mira (2007). Yang et al. (2018) discuss the incomplete information situation of the single-equation Tobit model with social interactions and compare it to the complete information case, which is an extension of Xu and Lee (2015), in which a maximum likelihood estimator is developed to estimate the single-equation SAR Tobit model. Other related literature, such as Lee et al. (2014), Lin and Xu (2017), and Yang and Lee (2017) discuss the existence of Bayesian Nash Equilibrium under incomplete information social network in which individuals share rational expectations on others' outcomes.

The rest of this paper is organized as follows. Section 2 clarifies the microeconomic foundation of this paper, a network multi-activity game under incomplete information case. Section 3 is the generation of the econometric model of this paper. Section 4 is the estimation. Section 5 is the Monte Carlo simulation. Section 6 concludes.

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 network matrix $W = [w_{ij}]$ for i and j from 1 to n, where $w_{ii} = 0$ for all i, and w_{ij} is known non-negative constant, 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. And $y_{ik} = y^*_{ik} I(y^*_{ik} > 0)$ where $I(\cdot)$ 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 introduce a linear-quadratic form utility function for agent i as

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

where $\vartheta_{kl} = \vartheta_{lk}$, and $\vartheta_{kk} \neq 0$ for all k from 1 to m. 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 second item, $\frac{1}{2} \sum_{k=1}^{m} \sum_{l=1}^{m} \vartheta_{lk} y^*_{ik} y^*_{il}$, represents the cost generated by $\{y^*_{ik}\}_{k=1}^{m}$.

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) 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, agents' activities' outcomes are censored instead of perfect observable values. Second, there is an unobserved random error term (shock) for each activity of each agent i.e., ε_{ik}.

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

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

where

$$p_{jl} = \mathrm{E}(y_{jl}|\{\varpi_{ik}\}_{i=1,k=1}^{n,m})$$

From the first-order condition in maximizing conditional expected utility, we have

$$\sum_{l=1}^m \frac{\vartheta_{lk}}{\vartheta_{kk}} y_{il}^* = \sum_{l=1}^m \frac{\varrho_{lk}}{\vartheta_{kk}} \sum_{j=1}^n w_{ij} p_{jl} + \frac{\varpi_{ik} - \varepsilon_{ik}}{\vartheta_{kk}} \tag{3}$$

without loss of generality, we denote $\theta_{lk} = \frac{\vartheta_{lk}}{\vartheta_{kk}}$, $\lambda_{lk} = \frac{\varrho_{lk}}{\vartheta_{kk}}$, $\pi_{ik} = \frac{\varpi_{ik}}{\vartheta_{kk}}$, and $\epsilon_{ik} = \frac{\varepsilon_{ik}}{\vartheta_{kk}}$, then we can rewrite the first-order condition of maximizing conditional utility in the following form

$$\sum_{l=1}^m \theta_{lk} y_{il}^* = \sum_{l=1}^m \lambda_{lk} \sum_{j=1}^n w_{ij} p_{jl} + \pi_{ik} - \epsilon_{ik} \tag{4}$$

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

$$\begin{aligned} \mathbf{y}_l^* = & \quad (y_{1l}^*, y_{2l}^*, ..., y_{nl}^*)' \\ \mathbf{p}_l = & \quad (p_{1l}, p_{2l}, ..., p_{nl})' \\ \pi_k = & \quad (\pi_{1k}, \pi_{2k}, ..., \pi_{nk})' \\ \epsilon_k = & \quad (\epsilon_{1k}, \epsilon_{2k}, ..., \epsilon_{nk})' \end{aligned} \tag{5}$$

the vector form can be written as

$$\sum_{l=1}^{m} \theta_{lk} \mathbf{y}_l^* = \sum_{l=1}^{m} \lambda_{lk} \mathbf{W} \mathbf{p}_l + \pi_k - \epsilon_k \tag{6}$$

In the vector form equation (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 inner-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 inter-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

$$\begin{aligned} \mathbf{Y} = & \quad [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m] \\ \mathbf{P} = & \quad [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m] \\ \mathbf{\Pi} = & \quad [\pi_1, \pi_2, ..., \pi_m] \\ \mathbf{E} = & \quad [\epsilon_1, \epsilon_2, ..., \epsilon_m] \end{aligned} \tag{7}$$

then the matrix form can be written as

$$\mathbf{Y}^* \mathbf{\Theta} = \mathbf{W} \mathbf{P} \mathbf{\Lambda} + \mathbf{\Pi} - \mathbf{E} \tag{8}$$

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

$$\mathbf{Y}^* = \mathbf{W} \mathbf{P} \mathbf{\Lambda}^* + \mathbf{\Pi}^* - \mathbf{E}^* \tag{9}$$

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

$$y^*_{ik} = \sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik} - \epsilon^*_{ik} \tag{10}$$

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

$$\begin{aligned} p_{ik} =& \mathrm{E}(y_{ik}|\{\varpi_{jl}\}^{n,m}_{j=1,l=1}) \\ =& \mathrm{E}(y_{ik}|\{\varpi_{jl}\}^{n,m}_{j=1,l=1}, y_{ik} > 0)\mathbf{Pr}(y_{ik} > 0|\{\varpi_{jl}\}^{n,m}_{j=1,l=1}) \\ &+ \mathrm{E}(y_{ik}|\{\varpi_{jl}\}^{n,m}_{j=1,l=1}, y_{ik} = 0)\mathbf{Pr}(y_{ik} = 0|\{\varpi_{jl}\}^{n,m}_{j=1,l=1}) \\ =& \mathrm{E}(y_{ik}|\{\varpi_{jl}\}^{n,m}_{j=1,l=1}, y_{ik} > 0)\mathbf{Pr}(y_{ik} > 0|\{\varpi_{jl}\}^{n,m}_{j=1,l=1}) \end{aligned} \tag{11}$$

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. *The simultaneous effect (interdependence effect) matrix* Θ *is nonsingular and has a unit diagonal.*

Assumption 2. *For each agent* i *in our model, the structural form random shock vector* $(\epsilon_{i1}, \cdots, \epsilon_{im})'$ *satisfies the multivariate normal distribution with zeros mean and* $\mathbf{\Sigma} = [\rho_{kl}\sigma_k\sigma_l]^{m,m}_{k=1,l=1}$ *(*$\rho_{kk} = 1$ *for* $k = 1, \cdots, m$*) variance-covariance matrix, and are independently among agents.*

With Assumption 1 and Assumption 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 $\mathbf{\Sigma}^* = \mathbf{\Theta}'^{-1}\mathbf{\Sigma}^*\mathbf{\Theta}^{-1} = [\sigma^*_{kl}]^{m,m}_{k=1,l=1}$, and all the diagonal elements of the matrix $\mathbf{\Sigma}^*$ are finite. This can interpreted by the following proposition.

Proposition 2.1. *If Assumption 1 and 2 hold, the reduced form random shock vector is jointly normally distributed with zero means and finite value diagonal elements variance-covariance matrix.*

Then equation 11 can be derived as

$$p_{ik} = \Big(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}\Big)\mathbf{\Phi}\Big(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}}{\sigma^*_{kk}}\Big) + \sigma^*_{kk}\phi\Big(\frac{\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1}^{n} w_{ij} p_{jl} + \pi^*_{ik}}{\sigma^*_{kk}}\Big) \tag{12}$$

According to the expression of $\{p_{ik}\}_{i=1,k=1}^{n,m}$, 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 $\mathbf{p} = (\mathbf{p}_1', \cdots, \mathbf{p}_m')'$

$$\mathbf{p}_k = \left(\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \pi_k\right) \odot \mathbf{\Phi}\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \pi_k}{\sigma_{kk}^*}\right) + \sigma_{kk}^* \phi\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \pi_k}{\sigma_{kk}^*}\right) \tag{13}$$

then, we set the notation

$$\vec{h}(\mathbf{p}) = [\vec{h}_1(\mathbf{p})', \cdots, \vec{h}_m(\mathbf{p})']' \tag{14}$$

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

$$\begin{aligned} \vec{h}_k(\mathbf{p}) =&(u_{1k}\mathbf{\Phi}_{1k} + \sigma_{kk}^*\phi_{1k}, \cdots, u_{nk}\mathbf{\Phi}_{nk} + \sigma_{kk}^*\phi_{nk})' \\ =&\left(\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \pi_k\right) \odot \mathbf{\Phi}\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \pi_k}{\sigma_{kk}^*}\right) + \sigma_{kk}^* \phi\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \mathbf{W}\mathbf{p}_l + \pi_k}{\sigma_{kk}^*}\right) \end{aligned} \tag{15}$$

where

$$\begin{aligned} u_{ik} =& \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^* \\ \mathbf{\Phi}_{ik} =& \mathbf{\Phi}\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^*}{\sigma_{kk}^*}\right) \\ \Phi_{ik} =& \mathbf{\Phi}\left(\frac{\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^*}{\sigma_{kk}^*}\right) \end{aligned} \tag{16}$$

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

$$||\mathbf{\Lambda}^*||_1 ||\mathbf{W}||_\infty < 1 \text{ or } ||\mathbf{\Lambda}^*||_\infty ||\mathbf{W}||_1 < 1$$

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

$$\begin{aligned} ||\mathbf{A}||_\infty &= \max_{i=1,2,\ldots,n} \sum_{j=1}^{m} |a_{ij}| \\ ||\mathbf{A}||_1 &= \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} |a_{ij}| \end{aligned} \tag{17}$$

Suppose we denote $\mathbf{p} = \text{vec}(\mathbf{P})$, $\mathbf{y}^\star = \text{vec}(\mathbf{Y}^\star)$, and $\pi^* = \text{vec}(\mathbf{X}\mathbf{B}^*)$, $\epsilon^* = \text{vec}(\mathbf{E}^*)$, then we have

Proposition 2.2. *If Assumption 1, 2, and 3 holds, then the incomplete information network game has a unique pure strategy BNE, given the equilibrium strategy* $\mathbf{y}^\star$ *as*

$$\mathbf{y}^\star = (\mathbf{\Lambda}^{*'} \otimes \mathbf{W})\mathbf{p}^\star + \pi^* - \epsilon^* \tag{18}$$

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

$$\mathbf{p} = \vec{h}(\mathbf{p}) \tag{19}$$

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

$$F_k(u) = u\Phi(\frac{u}{\sigma_k^*}) + \sigma_k^* \phi(\frac{u}{\sigma_k^*})$$

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, therefore, according to Brouwer fixed-point theorem, there at least exist one solution to $p = \vec{h}(p)$. 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

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

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 all the previous assumptions hold, we can derive the uniqueness of the solution from proposition 2.2 according to the completeness and coherency of the model in Tamer (2003). The contraction mapping (19) with a fixed point will guarantee the convergence to a consistent estimator in Kasahara and Shimotsu (2012), hence suggesting the Nested Pseudo Likelihood estimation algorithm. We will propose the econometric model based on the incomplete social network game and discuss the identification and estimation in the next section.

3 Econometric Model

Suppose $\pi_k = \mathbf{X}\beta_k$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]'$ is an $n \times q$ matrix, representing exogenous variables; and β_k is unknown q-dimension parameter vector reflect the direct effect in the model, then we can consider a simultaneous-equation model with a censored outcome variable $y_{ik} = y^*_{ik} \cdot \mathbf{1}(y^*_{ik} > 0)$, where

$$
y^*_{ik} = -\sum_{l=1,l\neq k}^{m}\theta_{lk}y^*_{il} + \sum_{l=1}^{m}\lambda_{lk}\sum_{j=1,j\neq i}^{n}w_{ij}p_{jl} + \mathbf{x}_i\beta_k - \epsilon_{ik}, \tag{20}
$$

for $i = 1, \cdots, n$ and $k = 1, \cdots, m$. Let $\mathbf{Y}^* = [\mathbf{y}^*_1, \mathbf{y}^*_2, \cdots, \mathbf{y}^*_m]$, $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_m]$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]'$, $\mathbf{E} = [\epsilon_1, \epsilon_2, \cdots, \epsilon_m]$, and $W = [w_{ij}]$, where $\mathbf{y}^*_k = (y^*_{1k}, y^*_{2k}, \cdots, y^*_{nk})'$, $\mathbf{p}_k = (p_{1k}, p_{2k}, \cdots, p_{nk})'$, and $\epsilon_k = (\epsilon_{1k}, \epsilon_{2k}, \cdots, \epsilon_{nk})'$. Let $\mathbf{\Theta} = [\theta_{kl}]$, an $(m \times m)$-dimension matrix that reflects the inner effect cross activities of the same agent, where $\theta_{kk} = 1$ for all k, $\mathbf{\Lambda} = [\lambda_{kl}]$, an $(m \times m)$-dimension matrix that reflects the peer effects among agents and activities; and $\mathbf{B} = [\beta_1, \cdots, \beta_m]$, which reflects the direct effects. In matrix form, Equation (20) can be written as

$$\mathbf{Y}^*\mathbf{\Theta} = \mathbf{WP\Lambda} + \mathbf{XB} - \mathbf{E}. \tag{21}$$

Based on our Assumption 2, we can extend our assumption that $\text{vec}(\mathbf{E})|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n)$. The (k,l)-th element of $\mathbf{\Sigma}$ is $\rho_{kl}\sigma_k\sigma_l$, with $\rho_{kk} = 1$ for all $k = 1, \cdots, m$. If $\mathbf{\Theta}$ is nonsingular, the reduced form of Equation (21) is

$$\mathbf{Y}^* = \mathbf{WP\Lambda}^* + \mathbf{XB}^* - \mathbf{E}^*, \tag{22}$$

where $\mathbf{\Lambda}^* = \mathbf{\Lambda\Theta}^{-1}$, $\mathbf{B}^* = \mathbf{B\Theta}^{-1}$, and $\mathbf{E}^* = \mathbf{E\Theta}^{-1}$. As $\text{vec}(\mathbf{E})|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n)$, we have $\text{vec}(\mathbf{E}^*)|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma}^* \otimes \mathbf{I}_n)$ where $\mathbf{\Sigma}^* = \mathbf{\Theta}'^{-1}\mathbf{\Sigma\Theta}^{-1}$. The (k,l)-th element of $\mathbf{\Sigma}^*$ is $\rho^*_{kl}\sigma^*_k\sigma^*_l$, with $\rho^*_{kk} = 1$ for all $k = 1, \cdots, m$. From equation (22), we can derive the scalar reduced form of the econometric model as

$$y^*_{ik} = \sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1, j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i\beta^*_k - \epsilon^*_{ik}. \tag{23}$$

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

$$\Pr(d_{ik} = 1) = \Pr(y^*_{ik} > 0) = \Pr(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1, j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i\beta^*_k > \epsilon^*_{ik}) = \Phi_{ik},$$

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

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

where $\phi_{ik} = \phi[(\sum_{l=1}^{m} \lambda^*_{lk} \sum_{j=1,j\neq i}^{n} w_{ij}p_{jl} + \mathbf{x}_i\beta^*_k)/\sigma^*_k]$. Let us consider the identification of the reduced form parameter first, i.e., $\mathbf{\Lambda}^* = [\lambda^*_{kl}]^{m,m}_{k=1,l=1}$, $\mathbf{B}^* = [\beta^*_k]^m_{k=1}$, and $\{\sigma^*_k\}^m_{k=1}$. 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 $\left(\mathbf{\Lambda}^*, \mathbf{B}^*, \{\sigma^*_k\}^m_{k=1}\right)$ and $\left(\widetilde{\mathbf{\Lambda}}^*, \widetilde{\mathbf{B}}^*, \{\widetilde{\sigma}^*_k\}^m_{k=1}\right)$

are observational equivalent if

$$(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}\widetilde{p}_{jl}+\mathbf{x}_i\widetilde{\beta}_k^{*})\widetilde{\Phi}_{ik}+\widetilde{\sigma}_k^{*}\widetilde{\phi}_{ik}=(\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}p_{jl}+\mathbf{x}_i\beta_k^{*})\Phi_{ik}+\sigma_k^{*}\phi_{ik} \tag{24}$$

for all $i=1,\cdots,n$ and $k=1,\cdots,m$. Based on our Assumption 1, 2, and 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

$$\begin{aligned}&[\mathbf{WP},\mathbf{X}][\mathbf{\Lambda}^{*\prime},\mathbf{B}^{*\prime}]'\odot\mathbf{\Phi}([\mathbf{WP},\mathbf{X}][\mathbf{\Lambda}^{*\prime},\mathbf{B}^{*\prime}]'\mathbf{D}_m^{*-1})+\phi([\mathbf{WP},\mathbf{X}][\mathbf{\Lambda}^{*\prime},\mathbf{B}^{*\prime}]'\mathbf{D}_m^{*-1})\mathbf{D}_m^{*}\\=&[\mathbf{WP},\mathbf{X}][\widetilde{\mathbf{\Lambda}}^{*\prime},\widetilde{\mathbf{B}}^{*\prime}]'\odot\mathbf{\Phi}([\mathbf{WP},\mathbf{X}][\widetilde{\mathbf{\Lambda}}^{*\prime},\widetilde{\mathbf{B}}^{*\prime}]'\widetilde{\mathbf{D}}_m^{*-1})+\phi([\mathbf{WP},\mathbf{X}][\widetilde{\mathbf{\Lambda}}^{*\prime},\widetilde{\mathbf{B}}^{*\prime}]'\widetilde{\mathbf{D}}_m^{*-1})\widetilde{\mathbf{D}}_m^{*}\end{aligned} \tag{25}$$

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

$$(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}\widetilde{p}_{jl}+\mathbf{x}_i\widetilde{\beta}_k^{*})+\widetilde{\sigma}_k^{*}\frac{\widetilde{\phi}_{ik}}{\widetilde{\Phi}_{ik}}=(\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}p_{jl}+\mathbf{x}_i\beta_k^{*})+\sigma_k^{*}\frac{\phi_{ik}}{\Phi_{ik}} \tag{26}$$

Suppose we denote $\mathbf{L}=[\phi_{ik}/\Phi_{ik}]_{i=1,k=1}^{n,m}$, 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

$$[\mathbf{WP},\mathbf{X},\mathbf{L}]\Big([\mathbf{\Lambda}^*,\mathbf{B}^*,\mathbf{D}_m^*]-[\widetilde{\mathbf{\Lambda}}^*,\widetilde{\mathbf{B}}^*,\widetilde{\mathbf{D}}_m^*]\Big)=\mathbf{0} \tag{27}$$

the scalar form results are

$$\begin{aligned}(\sum_{l=1}^{m}\lambda_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}p_{jl}+\mathbf{x}_i\beta_k^{*})&=(\sum_{l=1}^{m}\widetilde{\lambda}_{lk}^{*}\sum_{j=1,j\neq i}^{n}w_{ij}\widetilde{p}_{jl}+\mathbf{x}_i\widetilde{\beta}_k^{*})\\\sigma_k^{*2}&=\widetilde{\sigma}_k^{*2}\end{aligned} \tag{28}$$

And in matrix form, it will be

$$\begin{aligned}\mathbf{WP\Lambda}^*+\mathbf{XB}^*&=\mathbf{WP}\widetilde{\mathbf{\Lambda}}^*+\mathbf{X}\widetilde{\mathbf{B}}^*\\\mathbf{D}_m^*&=\widetilde{\mathbf{D}}_m^*\end{aligned} \tag{29}$$

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

$$\left[\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}, \mathbf{D}_m^*\right]' = \left[\widetilde{\mathbf{\Lambda}}^{*\prime}, \widetilde{\mathbf{B}}^{*\prime}, \widetilde{\mathbf{D}}_m^*\right]' \tag{30}$$

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

Assumption 4. $[\mathbf{WP}, \mathbf{X}, \mathbf{L}]$ *has full column rank.*

When the Assumption 4 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 $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{\Theta}^{-1}$, $\mathbf{B} = \mathbf{B}\mathbf{\Theta}^{-1}$, and $\mathbf{D}_m^*$ is an $m \times m$ diagonal matrix whose diagonal elements are $\mathbf{\Theta}'^{-1}\mathbf{\Sigma}\mathbf{\Theta}^{-1}$ 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 $\boldsymbol{R_k \gamma_k = 0}$ and rank($\boldsymbol{R_k}\mathbf{\Gamma}$)$= m-1$ for $k = 1, \cdots, m-1$, which is the sufficient rank condition to identify structural parameters from reduced form parameters by Schmidt (1976).

Assumption 5. *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* $\mathbf{R}_k\gamma_k = \mathbf{0}$, *and rank*$(\mathbf{R}_k\mathbf{\Gamma}) = m-1$ *for* $k = 1, \cdots, m$.

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

4 Estimation

We will derive the estimation of the model based on the identification process in the previous section. There are two steps of the estimation. The first step is to estimate the reduced form parameters $\mathbf{\Lambda}^*$, $\mathbf{B}^*$, and $(\sigma_1^{*2}, \cdots, \sigma_m^{*2})$ by Nested Pseudo Likelihood (NPL) algorithm, 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 $\Psi^* = [\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}, \mathbf{L}_m]'$ and at $t = 0$ the NPL starts from an initial vector $\mathrm{p}^{(0)} \in [0, +\infty)^{mn}$ and conduct the following iterative steps:

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

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

where

$$u_{ik}^{(t-1)} = \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j \neq i}^{n} w_{ij} p_{jl}^{(t-1)} + \mathbf{x}_i \beta_k^*$$

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

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

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

with

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

for $k \in \{1, 2, \cdots, m\}$. Update $\mathbf{p}^{(t-1)}$ in **Step 1** to $\mathbf{p}^{(t)}$. Repeat these two steps until convergence.

As we have proved, the convergence of the contraction mapping in Bayesian Nash Equilibrium and the unique fixed point exists. The convergence of the NPL algorithm will be guaranteed by Kasahara and Shimotsu (2012). Given the convergence of the NPL algorithm, the estimator $\widehat{\mathbf{\Psi}}^* = [\hat{\psi}_1^*, \cdots, \hat{\psi}_m^*]$ is the result that satisfies $\arg\max \ln L(\psi_k^*, \widehat{\mathbf{p}})$ for $k = 1, \cdots, m$. And $\widehat{\mathbf{p}}$ is calculated as $\widehat{\mathbf{p}} = \overrightarrow{h}(\widehat{\mathbf{p}}; \widehat{\mathbf{\Psi}}^*)$. According to our previous assumptions and propositions, we can draw a conclusion that our NPL algorithm estimation result is square-root n consistent and asymptotically normal. There are similar arguments and deriving processes in Aguirregabiria and Mira (2007), Lin and

Xu (2017), and Liu (2019). The detailed steps of deriving the asymptotic distribution of the NPL algorithm estimator are in 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

$$\mathbf{Y}^* = \mathbf{Z}[\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]' - \mathbf{E}^* = \mathbf{Z}\mathbf{\Psi} - \mathbf{E}^* \tag{31}$$

After the estimation of reduced form parameters $\mathbf{\Lambda}^*$, $\mathbf{B}^*$, and based on the rank condition of constraints of the structural form parameters, we can estimate the reduced form parameters $\mathbf{\Theta}$, $\mathbf{\Lambda}$, 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 $\mathbf{\Theta}$, 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

$$\mathbf{y}_1^* = -\mathbf{Y}_1^*\boldsymbol{\theta}_1 + \mathbf{Z}_1\boldsymbol{\psi}_\mathbf{1} - \boldsymbol{\epsilon}_\mathbf{1} \tag{32}$$

where $\mathbf{\Psi} = [\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]'$, and $\boldsymbol{\psi}_1$ is the first column of $\mathbf{\Psi}$. $\boldsymbol{\theta}_1$ is the first column of $\mathbf{\Theta}$. 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

$$\mathbf{y}_1^* = -\mathbf{Y}^*\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 + \mathbf{Z}\mathbf{J}_{Z_1}\boldsymbol{\psi}_\mathbf{1} - \boldsymbol{\epsilon}_\mathbf{1} \tag{33}$$

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

$$\begin{aligned} \mathbf{y}_1^* &= -(\mathbf{Z}\mathbf{\Psi} - \mathbf{E}^*)\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 + \mathbf{Z}\mathbf{J}_{Z_1}\boldsymbol{\psi}_\mathbf{1} - \boldsymbol{\epsilon}_\mathbf{1} \\ &= -\mathbf{Z}(\mathbf{\Psi}\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 - \mathbf{J}_{Z_1}\boldsymbol{\psi}_\mathbf{1}) + \mathbf{E}^*\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 - \boldsymbol{\epsilon}_\mathbf{1} \end{aligned} \tag{34}$$

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

$$\boldsymbol{\psi}_1^* = -\mathbf{\Psi}^*\mathbf{J}_{Y_1}\boldsymbol{\theta}_1 + \mathbf{J}_{Z_1}\boldsymbol{\psi}_\mathbf{1} \tag{35}$$

and the regression equation is

$$\widehat{\boldsymbol{\psi}}_1^* = \widehat{\boldsymbol{\Psi}}^* \mathbf{J}_{Y_1} \boldsymbol{\theta}_1 + \mathbf{J}_{Z_1} \boldsymbol{\psi}_{\mathbf{1}} + \mathbf{v}_1 \tag{36}$$

and the regression-error item is

$$\mathbf{v}_1 = (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) + (\widehat{\boldsymbol{\Psi}}^* - \boldsymbol{\Psi}^*) \mathbf{J}_{Y_1} \boldsymbol{\theta}_1 \tag{37}$$

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

$$(\widehat{\boldsymbol{\theta}}_1', \widehat{\boldsymbol{\psi}}_1^*) = (\widehat{\mathbf{H}}_1' \widehat{\boldsymbol{\Omega}}_{11}^{-1} \widehat{\mathbf{H}}_1)^{-1} \widehat{\mathbf{H}}_1' \widehat{\boldsymbol{\Omega}}_{11}^{-1} \widehat{\boldsymbol{\psi}}_1^* \tag{38}$$

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.

5 Monte Carlo Simulation

5.1 Simulation Setup

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

$$\begin{aligned} \mathbf{y}_1^* &= -\theta_{21} \mathbf{y}_2^* + \lambda_{11} \mathbf{W} \mathbf{p}_1 + \lambda_{21} \mathbf{W} \mathbf{p}_2 + \mathbf{X}_1 \beta_1 - \epsilon_1 \\ \mathbf{y}_2^* &= -\theta_{12} \mathbf{y}_1^* + \lambda_{12} \mathbf{W} \mathbf{p}_1 + \lambda_{22} \mathbf{W} \mathbf{p}_2 + \mathbf{X}_2 \beta_2 - \epsilon_2 \end{aligned} \tag{39}$$

Then the matrix form of structural parameters are

$$\boldsymbol{\Theta} = \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix} \qquad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix} \tag{40}$$

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 $\mathbf{E} = [\epsilon_1, \epsilon_2]$, we can derive the matrix form of the two-equation model as

$$\mathbf{Y}^* \boldsymbol{\Theta} = \mathbf{W} \mathbf{P} \boldsymbol{\Lambda} + \mathbf{X} \mathbf{B} - \mathbf{E}. \tag{41}$$

then we can derive the reduced form

$$\mathbf{Y}^* = \mathbf{W}\mathbf{P}\mathbf{\Lambda}^* + \mathbf{X}\mathbf{B}^* - \mathbf{E}^*. \tag{42}$$

the matrix form of reduced-form parameters are

$$\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{\Theta}^{-1} = \begin{bmatrix} \lambda_{11}^* & \lambda_{12}^* \\ \lambda_{21}^* & \lambda_{22}^* \end{bmatrix} \qquad \mathbf{B}^* = \mathbf{B}\mathbf{\Theta}^{-1} = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix} \tag{43}$$

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

$$\begin{aligned} &\theta_{21} = -\frac{\beta_{12}^*}{\beta_{11}^*} \qquad \theta_{12} = -\frac{\beta_{21}^*}{\beta_{22}^*} \qquad \beta_1 = \beta_{11}^* - \frac{\beta_{12}^*\beta_{21}^*}{\beta_{22}^*} \qquad \beta_2 = \beta_{22}^* - \frac{\beta_{12}^*\beta_{21}^*}{\beta_{11}^*} \\ &\lambda_{11} = \lambda_{11}^* - \frac{\beta_{21}^*}{\beta_{22}^*}\lambda_{12}^* \qquad \lambda_{12} = \lambda_{12}^* - \frac{\beta_{12}^*}{\beta_{11}^*}\lambda_{11}^* \qquad \lambda_{21} = \lambda_{21}^* - \frac{\beta_{21}^*}{\beta_{22}^*}\lambda_{22}^* \qquad \lambda_{22} = \lambda_{22}^* - \frac{\beta_{12}^*}{\beta_{11}^*}\lambda_{21}^* \end{aligned} \tag{44}$$

the equation (44) 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 $vec(\mathbf{E}^*)|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma}^* \otimes \mathbf{I_n})$ where

$$\mathbf{\Sigma}^* = \begin{bmatrix} \sigma_1^{*2} & \rho^*\sigma_1^*\sigma_2^* \\ \rho^*\sigma_1^*\sigma_2^* & \sigma_2^{*2} \end{bmatrix} \tag{45}$$

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

$$\begin{aligned} \mathbf{y}_1^* =& \lambda_{11}^*\mathbf{W}\mathbf{p}_1 + \lambda_{21}^*\mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_{11}^* + \mathbf{X}_2\beta_{21}^* - \epsilon_1^* \\ \mathbf{y}_2^* =& \lambda_{12}^*\mathbf{W}\mathbf{p}_1 + \lambda_{22}^*\mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\beta_{12}^* + \mathbf{X}_2\beta_{22}^* - \epsilon_2^* \end{aligned} \tag{46}$$

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

$$\sigma_1^* = f(a_1) = e^{a_1} \qquad \sigma_2^* = f(a_2) = e^{a_2} \qquad \rho^* = g(a_3) = 1 - \frac{2}{1 + e^{a_3}} \tag{47}$$

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^*.

$$\begin{aligned}
\frac{\partial \ln L}{\partial a_1} &= \frac{\partial \ln L}{\partial \sigma_1^*}\frac{\partial \sigma_1^*}{\partial a_1} = \sigma_1^* \frac{\partial \ln L}{\partial \sigma_1^*} \\
\frac{\partial \ln L}{\partial a_2} &= \frac{\partial \ln L}{\partial \sigma_2^*}\frac{\partial \sigma_2^*}{\partial a_2} = \sigma_2^* \frac{\partial \ln L}{\partial \sigma_2^*} \\
\frac{\partial \ln L}{\partial a_3} &= \frac{\partial \ln L}{\partial \rho^*}\frac{\partial \rho^*}{\partial a_3} = \rho^* \frac{\partial \ln L}{\partial \rho^*}
\end{aligned} \tag{48}$$

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 $d_{i1} = \mathbf{1}(y_{i1} > 0)$, $d_{i2} = \mathbf{1}(y_{i2} > 0)$, then we can write the log-likelihood as

$$\begin{aligned}
\ln L_1 = \sum_{i=1}^n \Big[& d_{i1} \ln \Big[\frac{1}{\sigma_1^*}\phi\Big(\frac{\lambda_{11}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{21}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^* + x_{i2}\beta_{21}^* - y_{i1}}{\sigma_1^*}\Big)\Big] \\
& + (1 - d_{i1}) \ln \Big[1 - \Phi\Big(\frac{\lambda_{11}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{21}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^* + x_{i2}\beta_{21}^*}{\sigma_1^*}\Big)\Big]\Big] \\
\ln L_2 = \sum_{i=1}^n \Big[& d_{i2} \ln [\frac{1}{\sigma_2^*}\phi\Big(\frac{\lambda_{12}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{22}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^* + x_{i2}\beta_{22}^* - y_{i2}}{\sigma_2^*}\Big)] \\
& + (1 - d_{i2}) \ln \Big[1 - \Phi\Big(\frac{\lambda_{12}^* \sum_{j=1}^n w_{ij} p_{i1}^{(t)} + \lambda_{22}^* \sum_{j=1}^n w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^* + x_{i2}\beta_{22}^*}{\sigma_2^*}\Big)\Big]\Big]
\end{aligned} \tag{49}$$

the results are

$$\begin{aligned}
(\lambda_{11}^{*(t)}, \lambda_{21}^{*(t)}, \beta_{11}^{*(t)}, \beta_{21}^{*(t)}, \sigma_1^{*2(t)}) &= \arg\max \ln L_1(\lambda_{11}^*, \lambda_{21}^*, \beta_{11}^*, \beta_{21}^*, \sigma_1^*) \\
(\lambda_{12}^{*(t)}, \lambda_{22}^{*(t)}, \beta_{12}^{*(t)}, \beta_{22}^{*(t)}, \sigma_2^{*2(t)}) &= \arg\max \ln L_2(\lambda_{12}^*, \lambda_{22}^*, \beta_{12}^*, \beta_{22}^*, \sigma_2^*)
\end{aligned} \tag{50}$$

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

$$\begin{aligned} p_{i1}^{(t+1)} =& (\lambda_{11}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{21}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)})\Phi_{i1}^{(t)} + \sigma_1^{*(t)}\phi_{i1}^{(t)} \\ p_{i2}^{(t+1)} =& (\lambda_{12}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{22}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^{*(t)} + x_{i2}\beta_{22}^{*(t)})\Phi_{i2}^{(t)} + \sigma_2^{*(t)}\phi_{i2}^{(t)} \end{aligned} \tag{51}$$

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

$$\begin{aligned} \Phi_{i1}^{(t)} =& \Phi\Big(\frac{\lambda_{11}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{21}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)}}{\sigma_1^*}\Big) \\ \phi_{i1}^{(t)} =& \phi\Big(\frac{\lambda_{11}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{21}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{11}^{*(t)} + x_{i2}\beta_{21}^{*(t)}}{\sigma_1^*}\Big) \\ \Phi_{i2}^{(t)} =& \Phi\Big(\frac{\lambda_{12}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{22}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^{*(t)} + x_{i2}\beta_{22}^{*(t)}}{\sigma_2^*}\Big) \\ \phi_{i2}^{(t)} =& \phi\Big(\frac{\lambda_{12}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i1}^{(t)} + \lambda_{22}^{*(t)} \sum_{j=1}^{n} w_{ij} p_{i2}^{(t)} + x_{i1}\beta_{12}^{*(t)} + x_{i2}\beta_{22}^{*(t)}}{\sigma_2^*}\Big) \end{aligned} \tag{52}$$

the same as 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

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

$$-\frac{\hat{\lambda}^*_{11}\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}^*_{21}\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}^*_{11}+x_{i2}\hat{\beta}^*_{21}}{\hat{\sigma}^*_1\sqrt{1-\rho^{*2}}}\Big)$$

$$\phi\Big(\frac{\hat{\lambda}^*_{12}\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}^*_{22}\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}^*_{12}+x_{i2}\hat{\beta}^*_{22}-y_{i2}}{\hat{\sigma}^*_2}\Big)\Big]$$

$$+(1-d_{i1})(1-d_{i2})\ln\Phi_2\Big(\frac{\hat{\lambda}^*_{11}\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}^*_{21}\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}^*_{11}+x_{i2}\hat{\beta}^*_{21}}{\hat{\sigma}^*_1},$$

$$\frac{\hat{\lambda}^*_{12}\sum_{j=1}^n w_{ij}\hat{p}_{i1}+\hat{\lambda}^*_{22}\sum_{j=1}^n w_{ij}\hat{p}_{i2}+x_{i1}\hat{\beta}^*_{12}+x_{i2}\hat{\beta}^*_{22}}{\hat{\sigma}^*_2},\hat{\sigma}^*_1,\hat{\sigma}^*_2,\rho^*\Big)\Big]$$

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

$$\begin{bmatrix}\hat{\sigma}^{*2}_1 & \rho^*\hat{\sigma}^*_1\hat{\sigma}^*_2\\ \rho^*\hat{\sigma}^*_1\hat{\sigma}^*_2 & \hat{\sigma}^{*2}_2\end{bmatrix} \tag{53}$$

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

$$\hat{\rho}^* = \arg\max \ln L(\rho^*) \tag{54}$$

Suppose we denote $\mathbf{Z} = [\mathbf{Wp_1}, \mathbf{Wp_2}, \mathbf{X}]$, and $\psi^*_1 = (\lambda^*_{11}, \lambda^*_{21}, \beta^*_{11}, \beta^*_{21})'$ and $\psi^*_2 = (\lambda^*_{12}, \lambda^*_{22}, \beta^*_{12}, \beta^*_{22})'$, then we have

$$\begin{aligned}\mathbf{y}_1 =& \mathbf{Z}\psi^*_1 - \epsilon^*_1\\ \mathbf{y}_2 =& \mathbf{Z}\psi^*_2 - \epsilon^*_2\end{aligned} \tag{55}$$

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

$$\begin{aligned}(\hat{\psi}^*_1, \hat{\sigma}^{*2}_1) =& \arg\max \ln L(\psi^*_1, \sigma^{*2}_1; \hat{\mathbf{p}}_1)\\ (\hat{\psi}^*_2, \hat{\sigma}^{*2}_2) =& \arg\max \ln L(\psi^*_2, \sigma^{*2}_2; \hat{\mathbf{p}}_2)\end{aligned} \tag{56}$$

where

$$\begin{aligned}\ln L(\psi^*_1, \sigma^*_1; \hat{\mathbf{p}}_1) =& \sum_{i=1}^n d_{i1}\ln\Big[\frac{1}{\sigma^*_1}\phi\Big(\frac{\hat{\mathbf{z}}'_i\psi^*_1 - y_{i1}}{\sigma^*_1}\Big)\Big] + (1-d_{i1})\ln\Big[1-\Phi\Big(\frac{\hat{\mathbf{z}}'_i\psi^*_1}{\sigma^*_1}\Big)\Big]\\ \ln L(\psi^*_2, \sigma^*_2; \hat{\mathbf{p}}_2) =& \sum_{i=1}^n d_{i2}\ln\Big[\frac{1}{\sigma^*_2}\phi\Big(\frac{\hat{\mathbf{z}}'_i\psi^*_2 - y_{i2}}{\sigma^*_2}\Big)\Big] + (1-d_{i2})\ln\Big[1-\Phi\Big(\frac{\hat{\mathbf{z}}'_i\psi^*_2}{\sigma^*_2}\Big)\Big]\end{aligned} \tag{57}$$

the first-order conditions are

$$
\begin{aligned}
\frac{\partial L}{\partial \psi_k^*} &= \sum_{i=1}^{n} -d_{ik}\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\mathbf{z}_i}{\sigma_k^{*2}} - (1-d_{ik})\frac{\phi_{ik}\mathbf{z}_i}{(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{1}{2}}} = 0 \\
\frac{\partial L}{\partial (\sigma_k^{*2})} &= \sum_{i=1}^{n} d_{ik}\Big[-\frac{1}{2\sigma_k^{*2}} + \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{2(\sigma_k^{*2})^2}\Big] + (1-d_{ik})\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} = 0
\end{aligned}
\tag{58}
$$

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.,

$$
\begin{aligned}
\mathbf{E}\big[y_{ik} - (\mathbf{z}_i'\psi_k^*)\Phi_{ik} - \sigma_k^*\phi_{ik}\big] &= 0 \\
\mathbf{E}\big[d_{ik} - \Phi_{ik}\big] &= 0
\end{aligned}
\tag{59}
$$

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 reduce-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 1. (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

$$
\frac{\partial \ln L(\hat{\psi}_k^*; \hat{\mathbf{p}})}{\partial \psi_k^*} = \sum_{i=1}^{n} \frac{[d_{ik} - \Phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)]\phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)}{\Phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)[1-\Phi(\hat{\mathbf{z}}_i'\hat{\psi}_k^*)]}\hat{\mathbf{z}}_i = 0 \tag{60}
$$

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

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

$$= O_p(1)$$

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

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_{11}) performs better than those in circular networks. In a random network, the estimator average accuracy of simultaneous effect matrix parameters (θ_{12} and θ_{11}) 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.

6 Conclusion

This paper derives a simultaneous equation model with peer effects and rational expectations under an incomplete information network. The econometric model and the microeconomic foundation are discussed. The sufficient condition of the existence of a Bayesian Nash Equilibrium under an incomplete information network game is derived. A nested pseudo-likelihood estimation process and the estimator's asymptotic distribution properties are developed. Monte Carlo simulation shows the consistency of the estimation results when the sample size is finite and the network size is large. In this current paper, the network is predetermined and has no relation with agents' random shock and attributes. However, in a real economic situation, an agent's connection with other agents in the network can partially reflect his/her observed and unobserved characteristics. Therefore, the endogenous network structure can be a potential future extension. Another extension in our current work is that our econometric model proposes the simultaneous effect from the intention of different activities of each agent. However, it is very possible that the simultaneous effect is the relation between the activity's intention and other activities' outcomes. This means the condition of an option model is necessary to discuss, and the potential model-selection problem needs to be handled in the future.

References

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.

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

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

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.

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

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.

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.

Appendices

A. AGLS Estimator

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

$$\begin{aligned} \mathbf{y}_1^* &= -\mathbf{y}_2^*\theta_{21} + \mathbf{Z}_1\psi_1 - \epsilon_1 \\ \mathbf{y}_2^* &= -\mathbf{y}_1^*\theta_{12} + \mathbf{Z}_2\psi_2 - \epsilon_2 \end{aligned} \tag{61}$$

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

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{Z}\psi_1^* - \epsilon_1^* \\ \mathbf{y}_2 &= \mathbf{Z}\psi_2^* - \epsilon_2^* \end{aligned} \tag{62}$$

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 (62), 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

$$\mathbf{p}_k = (\widehat{\mathbf{Z}}\widehat{\psi}_k^*) \odot \mathbf{\Phi}(\widehat{\mathbf{Z}}\widehat{\psi}_k^*/\widehat{\sigma}_k^*) + \widehat{\sigma}_k^*\phi(\widehat{\mathbf{Z}}\widehat{\psi}_k^*/\widehat{\sigma}_k^*) \tag{63}$$

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

$$\ln L(\psi_k^*, \sigma_k^*; \widehat{\mathbf{p}}) = \sum_{i=1}^{n} \left\{ d_{ik} \ln\left[\phi((\widehat{\mathbf{z}}_i'\psi_k^* - y_{ik})/\sigma_k^*)/\sigma_k^*\right] + (1 - d_{ik}) \ln\left[1 - \mathbf{\Phi}(\widehat{\mathbf{z}}_i'\psi_k^*/\sigma_k^*)\right] \right\} \tag{64}$$

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

$$\begin{aligned} \mathbf{y}_1 &= -\mathbf{y}_2\theta_{21} + \mathbf{Z}\mathbf{J}_1\psi_1 - \epsilon_1 \\ \mathbf{y}_2 &= -\mathbf{y}_1\theta_{12} + \mathbf{Z}\mathbf{J}_2\psi_2 - \epsilon_2 \end{aligned} \tag{65}$$

then we put the reduced-form equations in the model (62) into the reorganized structural model (65), we can get

$$\begin{aligned} \mathbf{y}_1 &= -(\mathbf{Z}\psi_1^* - \epsilon_1^*)\theta_{21} + \mathbf{Z}\mathbf{J}_1\psi_1 - \epsilon_1 = \mathbf{Z}(-\psi_1^*\theta_{21} + \mathbf{J}_1\psi_1) + \epsilon_1^*\theta_{21} - \epsilon_1 \\ \mathbf{y}_2 &= -(\mathbf{Z}\psi_2^* - \epsilon_2^*)\theta_{12} + \mathbf{Z}\mathbf{J}_2\psi_2 - \epsilon_2 = \mathbf{Z}(-\psi_2^*\theta_{12} + \mathbf{J}_2\psi_2) + \epsilon_2^*\theta_{12} - \epsilon_2 \end{aligned} \tag{66}$$

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

$$\begin{aligned} \psi_1^* &= -\psi_2^*\theta_{21} + \mathbf{J}_1\psi_1 \\ \psi_2^* &= -\psi_1^*\theta_{12} + \mathbf{J}_2\psi_2 \end{aligned} \tag{67}$$

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

$$\begin{aligned} \widehat{\psi}_1^* &= -\widehat{\psi}_2^*\theta_{21} + \mathbf{J}_1\psi_1 + \mathbf{v}_1 \\ \widehat{\psi}_2^* &= -\widehat{\psi}_1^*\theta_{12} + \mathbf{J}_2\psi_2 + \mathbf{v}_2 \end{aligned} \tag{68}$$

where

$$\begin{aligned} \mathbf{v}_1 &= (\widehat{\psi}_1^* - \psi_1^*) + (\widehat{\psi}_2^* - \psi_2^*)\theta_{21} \\ \mathbf{v}_2 &= (\widehat{\psi}_2^* - \psi_2^*) + (\widehat{\psi}_1^* - \psi_1^*)\theta_{12} \end{aligned} \tag{69}$$

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

$$\widehat{\delta} \;=\; (\widehat{\mathbf{H}}'\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\widehat{\mathbf{\Omega}}^{-1}\widehat{\psi}^* \tag{70}$$

where

$$\widehat{\mathbf{H}} = \begin{bmatrix} \widehat{\mathbf{H}}_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{H}}_2 \end{bmatrix} \tag{71}$$

in which $\widehat{\mathbf{H}}_1 = [-\widehat{\psi}_2^*, \mathbf{J}_1]$ and $\widehat{\mathbf{H}}_2 = [-\widehat{\psi}_1^*, \mathbf{J}_2]$. And $\widehat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$. To derive the detailed form of $\boldsymbol{\Omega}$, 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

$$\begin{aligned} \frac{\partial \ln L}{\partial \psi_k^*} + \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) (\widehat{\psi}_k^* - \psi_k^*) + \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \psi_k^*} \right) (\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) = O_p(1) \\ \frac{\partial \ln L}{\partial \sigma_k^{*2}} + \frac{\partial}{\partial \psi_k^*} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) (\widehat{\psi}_k^* - \psi_k^*) + \frac{\partial}{\partial \sigma_k^{*2}} \left(\frac{\partial \ln L}{\partial \sigma_k^{*2}} \right) (\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) = O_p(1) \end{aligned} \tag{72}$$

the matrix form is

$$\begin{pmatrix} \frac{\partial \ln L}{\partial \psi_k^*} \\ \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} + \begin{bmatrix} \frac{\partial}{\partial \psi_k^*} \begin{pmatrix} \frac{\partial \ln L}{\partial \psi_k^*} \end{pmatrix} & \frac{\partial}{\partial \sigma_k^{*2}} \begin{pmatrix} \frac{\partial \ln L}{\partial \psi_k^*} \end{pmatrix} \\ \frac{\partial}{\partial \psi_k^*} \begin{pmatrix} \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} & \frac{\partial}{\partial \sigma_k^{*2}} \begin{pmatrix} \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} \end{bmatrix} \begin{bmatrix} (\widehat{\psi}_k^* - \psi_k^*) \\ (\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) \end{bmatrix} = \begin{bmatrix} O_p(1) \\ O_p(1) \end{bmatrix} \tag{73}$$

then the asymptotic results can be written as

$$\begin{bmatrix} \sqrt{n}(\widehat{\psi}_k^* - \psi_k^*) \\ \sqrt{n}(\widehat{\sigma}_k^{*2} - \sigma_k^{*2}) \end{bmatrix} \overset{A}{=} - \begin{bmatrix} \frac{1}{n} \frac{\partial}{\partial \psi_k^*} \begin{pmatrix} \frac{\partial \ln L}{\partial \psi_k^*} \end{pmatrix} & \frac{1}{n} \frac{\partial}{\partial \sigma_k^{*2}} \begin{pmatrix} \frac{\partial \ln L}{\partial \psi_k^*} \end{pmatrix} \\ \frac{1}{n} \frac{\partial}{\partial \psi_k^*} \begin{pmatrix} \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} & \frac{1}{n} \frac{\partial}{\partial \sigma_k^{*2}} \begin{pmatrix} \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} \end{bmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} \frac{\partial \ln L}{\partial \psi_k^*} \\ \frac{1}{\sqrt{n}} \frac{\partial \ln L}{\partial \sigma_k^{*2}} \end{pmatrix} \tag{74}$$

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{75}
$$

$$
\begin{aligned}
\frac{\partial^2 L}{\partial(\sigma_k^{*2})^2} =& \sum_{i=1}^{n} d_{ik}\left[\frac{1}{2(\sigma_k^{*2})^2} - \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2}{(\sigma_k^{*2})^3}\right.\\
&\left. + \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2}\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial(\sigma_k^{*2})} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial(\sigma_k^{*2})}\right)\right]\\
&+ (1-d_{ik})\left\{\left[\frac{(\mathbf{z}_i'\psi_k^*)^3\phi_{ik}}{4(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{7}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{4(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{4(1-\Phi_{ik})^2(\sigma_k^{*2})^3}\right]\right.\\
&+ \left[\frac{\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})^2(\sigma_k^{*2})^2}\right]\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial(\sigma_k^{*2})}\right.\\
&\left.\left. + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial(\sigma_k^{*2})}\right)\right\}\\
=& \sum_{i=1}^{n} -\frac{\Phi_{ik}}{2(\sigma_k^{*2})^2} + \frac{(\mathbf{z}_i'\psi_k^*)^3\phi_{ik}}{4(\sigma_k^{*2})^{\frac{7}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{4(\sigma_k^{*2})^{\frac{5}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{4(1-\Phi_{ik})(\sigma_k^{*2})^3}\\
&+ \left[\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2}\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\right]\\
&\left(\lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial(\sigma_k^{*2})} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial(\sigma_k^{*2})}\right)
\end{aligned}
\tag{76}
$$

$$
\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}}}\left[\frac{(\mathbf{z}_i'\psi_k^*)^2}{2(\sigma_k^{*2})^2} - \frac{(\mathbf{z}_i'\psi_k^*)}{\sigma_k^{*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)\right] \\
&\quad - (1-d_{ik})\frac{\mathbf{z}_i\phi_{ik}^2}{(1-\Phi_{ik})^2(\sigma_k^{*2})^{\frac{1}{2}}}\left[-\frac{(\mathbf{z}_i'\psi_k^*)}{2(\sigma_k^{*2})^{\frac{3}{2}}} + \frac{1}{(\sigma_k^{*2})^{\frac{1}{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)\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}}} \\
&\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})\left[\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}}\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) \\
&= \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 + \left[-\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}}\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) \\
&\quad - \left[\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\Phi_{ik}}{\sigma_k^{*2}} + \frac{\phi_{ik}}{(\sigma_k^{*2})^{\frac{1}{2}}}\right]\left(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}}\right)'
\end{aligned}
\tag{77}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \psi_k^*}\left(\frac{\partial L}{\partial \sigma_k^{*2}}\right) = &\sum_{i=1}^{n}\left[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}}}\right.\\
&\left. - (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}\right]\\
&\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}\\
& + \left[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}}}\right.\\
&\left. - (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}\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)\\
= &\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{78}
$$

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 $l = 1, 2$)

$$
\begin{aligned}
\frac{\partial \mathbf{p}_1}{\partial \psi_1^{*\prime}} &= \mathbf{K}_1^{-1}\mathbf{A}_1\mathbf{Z} \\
\frac{\partial \mathbf{p}_2}{\partial \psi_2^{*\prime}} &= \mathbf{K}_2^{-1}\mathbf{A}_2\mathbf{Z} \\
\frac{\partial \mathbf{p}_1}{\partial \psi_2^{*\prime}} &= (\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})\lambda_{21}^*\mathbf{A}_1\mathbf{W}\mathbf{K}_2^{-1}\mathbf{A}_2\mathbf{Z} \\
\frac{\partial \mathbf{p}_2}{\partial \psi_1^{*\prime}} &= (\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})\lambda_{12}^*\mathbf{A}_2\mathbf{W}\mathbf{K}_1^{-1}\mathbf{A}_1\mathbf{Z} \\
\frac{\partial \mathbf{p}_1}{\partial \sigma_1^{*2}} &= \mathbf{K}_1^{-1}\frac{\phi_1}{2(\sigma_1^{*2})^{1/2}} \\
\frac{\partial \mathbf{p}_2}{\partial \sigma_2^{*2}} &= \mathbf{K}_2^{-1}\frac{\phi_2}{2(\sigma_2^{*2})^{1/2}} \\
\frac{\partial \mathbf{p}_1}{\partial \sigma_2^{*2}} &= (\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})\lambda_{21}^*\mathbf{A}_1\mathbf{W}\mathbf{K}_2^{-1}\frac{\phi_2}{2(\sigma_2^{*2})^{1/2}} \\
\frac{\partial \mathbf{p}_2}{\partial \sigma_1^{*2}} &= (\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})\lambda_{12}^*\mathbf{A}_2\mathbf{W}\mathbf{K}_1^{-1}\frac{\phi_1}{2(\sigma_1^{*2})^{1/2}}
\end{aligned} \tag{79}
$$

where

$$
\begin{aligned}
\mathbf{K}_1 &= \mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{A}_1\mathbf{W}(\mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W})^{-1}\mathbf{A}_2\mathbf{W} \\
\mathbf{K}_2 &= \mathbf{I}_n - \lambda_{22}^*\mathbf{A}_2\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{A}_2\mathbf{W}(\mathbf{I}_n - \lambda_{11}^*\mathbf{A}_1\mathbf{W})^{-1}\mathbf{A}_1\mathbf{W} \\
\mathbf{A}_1 &= \text{diag}(\Phi_{11}, \Phi_{21}, \cdots \Phi_{n1}) \\
\mathbf{A}_2 &= \text{diag}(\Phi_{12}, \Phi_{22}, \cdots \Phi_{n2})
\end{aligned} \tag{80}
$$

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

$$
\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{bmatrix} \tag{81}
$$

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

$$
\mathbf{V}_{kk} = \begin{bmatrix} \mathbf{A}_{kk} & \mathbf{b}_{kk} \\ \mathbf{c}_{kk}' & g_{kk} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}_{kk} & \mathbf{p}_{kk} \\ \mathbf{q}_{kk}' & s_{kk} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{kk} & \mathbf{b}_{kk} \\ \mathbf{c}_{kk}' & g_{kk} \end{bmatrix}^{\prime -1} \tag{82}
$$

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

$$\mathbf{A}_{kk} = \mathbf{Z}'\Bigg[\text{diag}_{i=1}^{n}\Bigg(\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{\Phi_{ik}}{\sigma_k^{*2}} - \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}}\Bigg)\Bigg]\Bigg(\mathbf{Z} + \lambda_{11}^*\mathbf{W}\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{21}^*\mathbf{W}\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\Bigg) \tag{83}$$

$$\begin{aligned}\mathbf{b}_{kk} =& \mathbf{Z}'\Bigg[\text{diag}_{i=1}^{n}\Bigg(-\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\Bigg)\Bigg]\iota_n \\ &+ \mathbf{Z}'\Bigg[\text{diag}_{i=1}^{n}\Bigg(-\frac{\Phi_{ik}}{\sigma_k^{*2}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}}\Bigg)\Bigg]\Bigg(\lambda_{1k}^*\mathbf{W}\frac{\partial \mathbf{p}_1}{\partial \sigma_k^{*2}} + \lambda_{2k}^*\mathbf{W}\frac{\partial \mathbf{p}_2}{\partial \sigma_k^{*2}}\Bigg) \\ &- \Bigg(\mathbf{0}, \mathbf{W}\frac{\partial \mathbf{p}_1}{\partial \sigma_k^{*2}}, \mathbf{W}\frac{\partial \mathbf{p}_2}{\partial \sigma_k^{*2}}\Bigg)'\Bigg[\text{diag}_{i=1}^{n}\Bigg(\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\Phi_{ik}}{\sigma_k^{*2}} + \frac{\phi_{ik}}{(\sigma_k^{*2})^{\frac{1}{2}}}\Bigg)\Bigg]\iota_n\end{aligned} \tag{84}$$

$$\begin{aligned}\mathbf{c}_{kk} =& \mathbf{Z}'\Bigg[\text{diag}_{i=1}^{n}\Bigg(-\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\Bigg)\Bigg]\iota_n \\ &+ \Bigg(\lambda_{1k}^*\mathbf{W}\frac{\partial \mathbf{p}_1}{\partial \psi_k^{*\prime}} + \lambda_{2k}^*\mathbf{W}\frac{\partial \mathbf{p}_2}{\partial \psi_k^{*\prime}}\Bigg)' \\ &\Bigg[\text{diag}_{i=1}^{n}\Bigg(\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2} + \frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\Bigg)\Bigg]\iota_n\end{aligned} \tag{85}$$

$$\begin{aligned}g_{kk} =& \iota_n'\Bigg[\text{diag}_{i=1}^{1}\Bigg(-\frac{\Phi_{ik}}{2(\sigma_k^{*2})^2} + \frac{(\mathbf{z}_i'\psi_k^*)^3\phi_{ik}}{4(\sigma_k^{*2})^{\frac{7}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{4(\sigma_k^{*2})^{\frac{5}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{4(1-\Phi_{ik})(\sigma_k^{*2})^3}\Bigg)\Bigg]\iota_n \\ &+ \iota_n'\Bigg[\text{diag}_{i=1}^{n}\Bigg(\frac{\Phi_{ik}(\mathbf{z}_i'\psi_k^* - y_{ik})}{(\sigma_k^{*2})^2}\frac{\phi_{ik}}{2(\sigma_k^{*2})^{\frac{3}{2}}} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}}{2(\sigma_k^{*2})^{\frac{5}{2}}} + \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\Bigg)\Bigg] \\ &\Bigg(\lambda_{1k}^*\mathbf{W}\frac{\partial \mathbf{p}_1}{\partial (\sigma_k^{*2})} + \lambda_{2k}^*\mathbf{W}\frac{\partial \mathbf{p}_2}{\partial (\sigma_k^{*2})}\Bigg)\end{aligned} \tag{86}$$

$$\mathbf{H}_{kk} = \mathbf{Z}'\Bigg[\text{diag}_{i=1}^{n}\Bigg(\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2\Phi_{ik}}{(\sigma_k^{*2})^2} + \frac{\phi_{ik}^2}{(1-\Phi_{ik})\sigma_k^{*2}}\Bigg)\Bigg]\mathbf{Z} \tag{87}$$

$$\mathbf{p}_{kk} = \mathbf{q}_{kk} = \mathbf{Z}'\Bigg[\text{diag}_{i=1}^{n}\Bigg(\frac{(\mathbf{z}_i'\psi_k^* - y_{ik})\Phi_{ik}}{2(\sigma_k^{*2})^2} - \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^3\Phi_{ik}}{2(\sigma_k^{*2})^3} - \frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^2}\Bigg)\Bigg]\iota_n \tag{88}$$

$$s_{kk} = \iota_n'\Bigg[\text{diag}_{i=1}^{n}\Bigg(\frac{\Phi_{ik}}{4(\sigma_k^{*2})^4} - \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^2\Phi_{ik}}{2(\sigma_k^{*2})^3} + \frac{(\mathbf{z}_i'\psi_k^* - y_{ik})^4\Phi_{ik}}{4(\sigma_k^{*2})^4} - \frac{(\mathbf{z}_i'\psi_k^*)^2\phi_{ik}^2}{2(1-\Phi_{ik})(\sigma_k^{*2})^3}\Bigg)\Bigg]\iota_n \tag{89}$$

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

$$\mathbf{V}_{kl} = \begin{bmatrix}\mathbf{A}_{kk} & \mathbf{b}_{kk} \\ \mathbf{c}_{kk}' & g_{kk}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{H}_{kl} & \mathbf{p}_{kl} \\ \mathbf{q}_{kl}' & s_{kl}\end{bmatrix}\begin{bmatrix}\mathbf{A}_{ll} & \mathbf{b}_{ll} \\ \mathbf{c}_{ll}' & g_{ll}\end{bmatrix}^{\prime -1} \tag{90}$$

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

$$\begin{aligned}\mathbf{H}_{kl} =&\mathbf{Z}'\Big\{\text{diag}_{i=1}^n\Big[\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})(\mathbf{z}_i'\psi_l^*-y_{il})}{\sigma_k^{*2}\sigma_l^{*2}}\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\\&+\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})\phi_{il}}{(1-\Phi_{il})\sigma_k^{*2}\sigma_l^*}\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&+\frac{(\mathbf{z}_i'\psi_l^*-y_{il})\phi_{ik}}{(1-\Phi_{ik})\sigma_k^*\sigma_l^{*2}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&+\frac{\phi_{ik}\phi_{il}}{(1-\Phi_{ik})(1-\Phi_{il})\sigma_k^*\sigma_l^*}\Big]\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\Big\}\mathbf{Z}\end{aligned}\tag{91}$$

$$\begin{aligned}\mathbf{p}_{kl} =&-\mathbf{Z}'\text{diag}_{i=1}^n\Big\{\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})}{\sigma_k^{*2}}\Big[\frac{(\mathbf{z}_i'\psi_l^*-y_{il})^2}{2(\sigma_l^{*2})^2}-\frac{1}{2\sigma_l^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\\&+\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})(\mathbf{z}_i'\psi_l^*)\phi_{il}}{2(1-\Phi_{il})\sigma_k^{*2}(\sigma_l^{*2})^{\frac{3}{2}}}\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&+\frac{\phi_{ik}}{(1-\Phi_{ik})\sigma_k^*}\Big[\frac{(\mathbf{z}_i'\psi_l^*-y_{il})^2}{2(\sigma_l^{*2})^2}-\frac{1}{2\sigma_l^{*2}}\Big]\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&+\frac{(\mathbf{z}_i'\psi_l^*)\phi_{ik}\phi_{il}}{2(1-\Phi_{ik})(1-\Phi_{il})\sigma_k^*(\sigma_l^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\Big\}\iota_n\end{aligned}\tag{92}$$

$$\begin{aligned}\mathbf{q}_{kl} =&-\mathbf{Z}'\text{diag}_{i=1}^n\Big\{\frac{(\mathbf{z}_i'\psi_l^*-y_{il})}{\sigma_l^{*2}}\Big[\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})^2}{2(\sigma_k^{*2})^2}-\frac{1}{2\sigma_k^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\\&+\frac{\phi_{il}}{(1-\Phi_{il})\sigma_l^*}\Big[\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})^2}{2(\sigma_k^{*2})^2}-\frac{1}{2\sigma_k^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&+\frac{(\mathbf{z}_i'\psi_l^*-y_{il})(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})\sigma_l^{*2}(\sigma_k^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&+\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}\phi_{il}}{2(1-\Phi_{ik})(1-\Phi_{il})\sigma_l^*(\sigma_k^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\Big\}\iota_n\end{aligned}\tag{93}$$

$$\begin{aligned}s_{kl} =&\iota_n'\text{diag}_{i=1}^n\Big\{\Big[\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})^2}{2(\sigma_k^{*2})^2}-\frac{1}{2\sigma_k^{*2}}\Big]\Big[\frac{(\mathbf{z}_i'\psi_l^*-y_{il})^2}{2(\sigma_l^{*2})^2}-\frac{1}{2\sigma_l^{*2}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\\&+\Big[\frac{(\mathbf{z}_i'\psi_k^*-y_{ik})^2}{2(\sigma_k^{*2})^2}-\frac{1}{2\sigma_k^{*2}}\Big]\Big[\frac{(\mathbf{z}_i'\psi_l^*)\phi_{il}}{2(1-\Phi_{il})(\sigma_l^{*2})^{\frac{3}{2}}}\Big]\Phi_2\Big(\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&+\Big[\frac{(\mathbf{z}_i'\psi_l^*-y_{il})^2}{2(\sigma_l^{*2})^2}-\frac{1}{2\sigma_l^{*2}}\Big]\Big[\frac{(\mathbf{z}_i'\psi_k^*)\phi_{ik}}{2(1-\Phi_{ik})(\sigma_k^{*2})^{\frac{3}{2}}}\Big]\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},-\rho^*\Big)\\&\frac{(\mathbf{z}_i'\psi_k^*)(\mathbf{z}_i'\psi_l^*)\phi_{ik}\phi_{il}}{4(1-\Phi_{ik})(1-\Phi_{il})(\sigma_k^{*2})^{\frac{3}{2}}(\sigma_l^{*2})^{\frac{3}{2}}}\Phi_2\Big(-\frac{\mathbf{z}_i'\psi_k^*}{\sigma_k^*},-\frac{\mathbf{z}_i'\psi_l^*}{\sigma_l^*},\rho^*\Big)\Big\}\iota_n\end{aligned}\tag{94}$$

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

$$\widetilde{\mathbf{V}} = \begin{bmatrix} \widetilde{\mathbf{V}}_{11} & \widetilde{\mathbf{V}}_{12} \\ \widetilde{\mathbf{V}}'_{12} & \widetilde{\mathbf{V}}_{22} \end{bmatrix} \tag{95}$$

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

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega'_{12} & \Omega_{22} \end{bmatrix} \tag{96}$$

where

$$\Omega_{11} = \widetilde{\mathbf{V}}_{11} + \theta_{21}^2 \widetilde{\mathbf{V}}_{22} + \theta_{21}(\widetilde{\mathbf{V}}_{12} + \widetilde{\mathbf{V}}'_{12}) \tag{97}$$

$$\Omega_{22} = \widetilde{\mathbf{V}}_{22} + \theta_{12}^2 \widetilde{\mathbf{V}}_{11} + \theta_{12}(\widetilde{\mathbf{V}}_{12} + \widetilde{\mathbf{V}}'_{12}) \tag{98}$$

$$\Omega_{22} = \theta_{12} \widetilde{\mathbf{V}}_{11} + \theta_{21} \widetilde{\mathbf{V}}_{22} + \widetilde{\mathbf{V}}_{12} + \theta_{12}\theta_{21} \widetilde{\mathbf{V}}'_{12} \tag{99}$$

B. Simulation Results

Random Network $\theta_{12} = \theta_{21} = 0.5$, $\beta_1 = \beta_2 = 1$, $n = 2000$, rep $= 1000$, $\sigma_1^* = \sigma_2^* = 1$, $\rho_{12}^* = 0.1$.												
$\lambda_{11} = \lambda_{22}$	$\lambda_{12} = \lambda_{21}$	$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1^*$	$\hat{\sigma}_2^*$	$\hat{\rho}_{12}^*$
0.1	0.1	0.500	0.502	0.117	0.115	0.120	0.120	1.000	0.999	0.997	0.998	0.100
		(0.032)	(0.031)	(0.067)	(0.065)	(0.067)	(0.066)	(0.046)	(0.047)	(0.029)	(0.028)	(0.033)
0.2	0.1	0.500	0.502	0.226	0.121	0.126	0.230	1.000	0.999	0.997	0.998	0.100
		(0.031)	(0.031)	(0.065)	(0.064)	(0.066)	(0.064)	(0.046)	(0.047)	(0.029)	(0.028)	(0.033)
0.2	0.2	0.500	0.501	0.229	0.228	0.234	0.234	1.000	0.998	0.997	0.998	0.099
		(0.031)	(0.032)	(0.063)	(0.061)	(0.062)	(0.061)	(0.047)	(0.047)	(0.028)	(0.029)	(0.032)
0.3	0.2	0.501	0.501	0.336	0.230	0.235	0.340	1.000	0.998	0.998	0.998	0.099
		(0.031)	(0.031)	(0.061)	(0.060)	(0.061)	(0.058)	(0.047)	(0.047)	(0.029)	(0.029)	(0.031)
0.3	0.3	0.500	0.501	0.336	0.336	0.339	0.340	1.001	0.999	0.998	0.997	0.100
		(0.031)	(0.031)	(0.059)	(0.058)	(0.059)	(0.058)	(0.047)	(0.047)	(0.028)	(0.029)	(0.031)
0.4	0.2	0.500	0.501	0.443	0.231	0.233	0.444	1.001	0.998	0.998	0.998	0.099
		(0.031)	(0.031)	(0.058)	(0.059)	(0.061)	(0.057)	(0.046)	(0.047)	(0.028)	(0.029)	(0.030)
0.4	0.3	0.500	0.501	0.441	0.335	0.336	0.442	1.001	0.999	0.998	0.997	0.100
		(0.031)	(0.031)	(0.057)	(0.057)	(0.060)	(0.057)	(0.046)	(0.046)	(0.028)	(0.029)	(0.030)
0.4	0.4	0.500	0.502	0.436	0.436	0.438	0.438	1.000	0.998	0.998	0.997	0.100
		(0.031)	(0.031)	(0.056)	(0.056)	(0.058)	(0.057)	(0.047)	(0.047)	(0.028)	(0.029)	(0.029)
0.5	0.3	0.500	0.502	0.542	0.329	0.332	0.546	1.000	0.999	0.998	0.996	0.099
		(0.031)	(0.031)	(0.053)	(0.056)	(0.059)	(0.053)	(0.047)	(0.047)	(0.029)	(0.029)	(0.029)
0.5	0.4	0.500	0.501	0.536	0.428	0.432	0.539	1.001	0.999	0.998	0.996	0.100
		(0.031)	(0.031)	(0.053)	(0.054)	(0.057)	(0.053)	(0.047)	(0.048)	(0.028)	(0.029)	(0.028)
0.5	0.5	0.500	0.501	0.526	0.526	0.529	0.530	1.001	0.999	0.998	0.997	0.101
		(0.032)	(0.031)	(0.053)	(0.056)	(0.057)	(0.054)	(0.046)	(0.047)	(0.028)	(0.029)	(0.028)
0.6	0.3	0.500	0.502	0.643	0.321	0.324	0.647	1.000	0.998	0.998	0.996	0.099
		(0.032)	(0.032)	(0.051)	(0.056)	(0.059)	(0.051)	(0.047)	(0.048)	(0.029)	(0.030)	(0.029)

Random Network $\theta_{12}=\theta_{21}=0.5$, $\beta_1=\beta_2=1$, $n=2000$, rep $=1000$, $\sigma_1^*=\sigma_2^*=1$, $\rho_{12}^*=0.1$ (Cont.)												
$\lambda_{11}=\lambda_{22}$	$\lambda_{12}=\lambda_{21}$	$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1^*$	$\hat{\sigma}_2^*$	$\hat{\rho}_{12}^*$
0.6	0.4	0.500 (0.032)	0.502 (0.031)	0.633 (0.050)	0.418 (0.054)	0.422 (0.057)	0.638 (0.050)	1.000 (0.047)	0.999 (0.048)	0.999 (0.028)	0.996 (0.029)	0.100 (0.028)
0.6	0.5	0.500 (0.032)	0.502 (0.031)	0.623 (0.050)	0.514 (0.054)	0.517 (0.057)	0.626 (0.051)	1.000 (0.046)	0.998 (0.048)	0.999 (0.029)	0.997 (0.029)	0.101 (0.027)
0.6	0.6	0.501 (0.033)	0.501 (0.032)	0.608 (0.050)	0.608 (0.053)	0.612 (0.055)	0.613 (0.050)	1.000 (0.048)	0.998 (0.050)	0.997 (0.031)	0.996 (0.030)	0.101 (0.026)
0.7	0.4	0.501 (0.033)	0.501 (0.031)	0.730 (0.047)	0.406 (0.054)	0.409 (0.057)	0.734 (0.047)	1.000 (0.047)	0.999 (0.049)	0.998 (0.029)	0.996 (0.030)	0.100 (0.027)
0.7	0.5	0.501 (0.034)	0.502 (0.031)	0.717 (0.048)	0.500 (0.055)	0.503 (0.057)	0.721 (0.049)	1.000 (0.047)	0.998 (0.050)	0.998 (0.030)	0.996 (0.031)	0.100 (0.026)
0.7	0.6	0.501 (0.033)	0.501 (0.032)	0.702 (0.049)	0.593 (0.054)	0.597 (0.057)	0.706 (0.049)	1.000 (0.048)	0.998 (0.052)	0.996 (0.031)	0.996 (0.032)	0.100 (0.025)
0.7	0.7	0.502 (0.036)	0.500 (0.035)	0.688 (0.050)	0.686 (0.053)	0.691 (0.055)	0.690 (0.050)	1.001 (0.052)	0.998 (0.055)	0.994 (0.034)	0.996 (0.033)	0.100 (0.024)
0.8	0.4	0.501 (0.034)	0.501 (0.032)	0.825 (0.045)	0.391 (0.054)	0.394 (0.058)	0.828 (0.043)	0.999 (0.047)	0.999 (0.051)	0.999 (0.030)	0.996 (0.032)	0.100 (0.026)
0.8	0.5	0.502 (0.034)	0.503 (0.034)	0.813 (0.045)	0.488 (0.057)	0.484 (0.057)	0.810 (0.045)	0.998 (0.051)	0.997 (0.052)	0.998 (0.032)	0.998 (0.031)	0.098 (0.024)
0.8	0.6	0.500 (0.036)	0.501 (0.036)	0.799 (0.046)	0.582 (0.057)	0.576 (0.056)	0.794 (0.046)	0.999 (0.054)	1.000 (0.055)	1.000 (0.033)	0.998 (0.034)	0.099 (0.025)
0.9	0.5	0.503 (0.037)	0.500 (0.035)	0.904 (0.042)	0.471 (0.057)	0.473 (0.060)	0.903 (0.042)	1.000 (0.054)	0.998 (0.057)	0.996 (0.034)	0.998 (0.035)	0.099 (0.025)

Circular Network $\theta_{12} = \theta_{21} = 0.5$, $\beta_1 = \beta_2 = 1$, $n = 2000$, rep $= 1000$, $\sigma_1^* = \sigma_2^* = 1$, $\rho_{12}^* = 0.1$.

$\lambda_{11} = \lambda_{22}$	$\lambda_{12} = \lambda_{21}$	$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1^*$	$\hat{\sigma}_2^*$	$\hat{\rho}_{12}^*$
0.1	0.1	0.500	0.501	0.118	0.117	0.117	0.119	1.001	0.999	0.997	0.997	0.100
		(0.032)	(0.031)	(0.045)	(0.045)	(0.046)	(0.045)	(0.046)	(0.048)	(0.028)	(0.029)	(0.033)
0.2	0.1	0.500	0.500	0.225	0.122	0.123	0.227	1.001	1.001	0.997	0.997	0.100
		(0.031)	(0.032)	(0.041)	(0.044)	(0.044)	(0.042)	(0.047)	(0.047)	(0.028)	(0.028)	(0.032)
0.2	0.2	0.502	0.502	0.230	0.231	0.229	0.229	0.998	0.997	0.999	0.997	0.100
		(0.032)	(0.030)	(0.042)	(0.044)	(0.043)	(0.041)	(0.046)	(0.047)	(0.030)	(0.029)	(0.031)
0.3	0.2	0.503	0.503	0.338	0.235	0.232	0.336	0.995	0.994	0.999	0.996	0.101
		(0.032)	(0.033)	(0.039)	(0.044)	(0.044)	(0.040)	(0.049)	(0.049)	(0.029)	(0.028)	(0.031)
0.3	0.3	0.506	0.505	0.337	0.337	0.340	0.339	0.987	0.989	0.998	0.997	0.100
		(0.032)	(0.033)	(0.038)	(0.043)	(0.041)	(0.038)	(0.049)	(0.047)	(0.028)	(0.028)	(0.030)
0.4	0.2	0.508	0.506	0.441	0.233	0.235	0.442	0.989	0.988	0.996	0.998	0.100
		(0.033)	(0.035)	(0.037)	(0.044)	(0.045)	(0.037)	(0.049)	(0.049)	(0.031)	(0.030)	(0.031)
0.4	0.3	0.508	0.510	0.440	0.338	0.338	0.440	0.983	0.983	0.997	0.997	0.101
		(0.032)	(0.031)	(0.034)	(0.040)	(0.041)	(0.033)	(0.047)	(0.047)	(0.029)	(0.029)	(0.031)
0.4	0.4	0.512	0.511	0.435	0.436	0.440	0.439	0.979	0.977	0.997	0.999	0.101
		(0.033)	(0.032)	(0.033)	(0.039)	(0.039)	(0.033)	(0.047)	(0.048)	(0.030)	(0.029)	(0.027)
0.5	0.3	0.515	0.513	0.540	0.333	0.338	0.544	0.975	0.974	0.994	0.995	0.101
		(0.034)	(0.035)	(0.030)	(0.042)	(0.042)	(0.032)	(0.048)	(0.046)	(0.030)	(0.030)	(0.029)
0.5	0.4	0.517	0.515	0.537	0.431	0.434	0.538	0.970	0.970	0.995	0.998	0.103
		(0.033)	(0.033)	(0.030)	(0.040)	(0.041)	(0.031)	(0.047)	(0.047)	(0.029)	(0.030)	(0.029)
0.5	0.5	0.519	0.521	0.529	0.531	0.530	0.531	0.964	0.962	0.998	0.997	0.102
		(0.033)	(0.033)	(0.031)	(0.040)	(0.039)	(0.030)	(0.048)	(0.048)	(0.030)	(0.031)	(0.026)
0.6	0.3	0.517	0.517	0.640	0.331	0.330	0.639	0.965	0.966	0.990	0.993	0.098
		(0.038)	(0.038)	(0.028)	(0.047)	(0.046)	(0.028)	(0.049)	(0.050)	(0.033)	(0.032)	(0.029)

Circular Network $\theta_{12} = \theta_{21} = 0.5$, $\beta_1 = \beta_2 = 1$, $n = 2000$, rep = 1000, $\sigma_1^* = \sigma_2^* = 1$, $\rho_{12}^* = 0.1$. (Cont.)

$\lambda_{11} = \lambda_{22}$	$\lambda_{12} = \lambda_{21}$	$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1^*$	$\hat{\sigma}_2^*$	$\hat{\rho}_{12}^*$
0.6	0.4	0.520	0.522	0.633	0.429	0.428	0.633	0.960	0.961	0.994	0.995	0.102
		(0.035)	(0.037)	(0.027)	(0.046)	(0.043)	(0.028)	(0.050)	(0.048)	(0.032)	(0.032)	(0.027)
0.6	0.5	0.524	0.525	0.628	0.525	0.522	0.626	0.953	0.955	0.995	0.995	0.102
		(0.034)	(0.035)	(0.026)	(0.041)	(0.041)	(0.027)	(0.048)	(0.047)	(0.032)	(0.030)	(0.026)
0.6	0.6	0.529	0.528	0.617	0.617	0.619	0.618	0.948	0.948	0.997	0.995	0.104
		(0.035)	(0.033)	(0.027)	(0.039)	(0.041)	(0.028)	(0.045)	(0.047)	(0.033)	(0.032)	(0.025)
0.7	0.4	0.527	0.527	0.727	0.422	0.421	0.727	0.943	0.944	0.988	0.987	0.100
		(0.039)	(0.040)	(0.023)	(0.047)	(0.046)	(0.023)	(0.049)	(0.050)	(0.035)	(0.036)	(0.025)
0.7	0.5	0.530	0.530	0.721	0.515	0.516	0.721	0.941	0.941	0.992	0.991	0.103
		(0.038)	(0.040)	(0.022)	(0.045)	(0.046)	(0.023)	(0.049)	(0.050)	(0.034)	(0.034)	(0.025)
0.7	0.6	0.534	0.532	0.713	0.609	0.611	0.714	0.940	0.940	0.994	0.993	0.106
		(0.038)	(0.037)	(0.024)	(0.043)	(0.046)	(0.023)	(0.046)	(0.046)	(0.035)	(0.035)	(0.024)
0.7	0.7	0.530	0.533	0.705	0.707	0.706	0.705	0.967	0.967	1.015	1.015	0.128
		(0.039)	(0.040)	(0.025)	(0.046)	(0.045)	(0.025)	(0.046)	(0.045)	(0.039)	(0.038)	(0.027)
0.8	0.4	0.534	0.532	0.813	0.416	0.419	0.813	0.921	0.921	0.969	0.970	0.088
		(0.049)	(0.047)	(0.018)	(0.052)	(0.054)	(0.018)	(0.054)	(0.055)	(0.042)	(0.042)	(0.026)
0.8	0.5	0.533	0.529	0.811	0.503	0.507	0.812	0.933	0.929	0.980	0.980	0.099
		(0.044)	(0.044)	(0.017)	(0.050)	(0.049)	(0.017)	(0.049)	(0.051)	(0.039)	(0.037)	(0.025)
0.8	0.6	0.530	0.526	0.808	0.595	0.599	0.808	0.964	0.964	1.001	1.002	0.125
		(0.046)	(0.046)	(0.018)	(0.052)	(0.052)	(0.018)	(0.049)	(0.049)	(0.043)	(0.043)	(0.026)
0.9	0.5	0.517	0.517	0.898	0.490	0.491	0.899	0.956	0.954	0.975	0.974	0.106
		(0.053)	(0.055)	(0.011)	(0.058)	(0.056)	(0.011)	(0.057)	(0.058)	(0.050)	(0.049)	(0.026)