A Structural Model of Multigraph Formation: Favor Exchange and Social Networks in Villages

Jun Sung Kim†
Cornell University
Job Market Paper
November 14, 2013

Abstract

This paper proposes a structural model of multigraph formation, where 1) individuals determine two or more types of links simultaneously; 2) all networks interact with each other in the sense that the structure of one network affects an individual’s utility from the other networks; and 3) one or more networks are endogenous but not simultaneous from the econometrician’s perspective. I extend the notion of pairwise stability of a single network in Jackson and Wolinsky (1996) to a multigraph, and show that the structural model is equivalent to a multinomial choice model under pairwise stability of a multigraph. The presence of endogenous but not simultaneously determined networks is a source of an incomplete econometric model. Relying on the recent development of partially identified econometric models, I characterize the sharp identification region of utility parameters by a finite set of moment inequalities and conduct inference. I apply the model to village networks in rural India and find that friendship affects the formation of risk sharing and favor exchange networks in the same direction. On the other hand, the empirical evidence for caste homophily in risk sharing and favor exchange networks is inconclusive.

JEL Classification: C51, C14, D85, C31, O12

Keywords: Multigraph, Pairwise Stability, Moment Inequalities, Partial Identification, Random Set, Risk Sharing Network, Homophily

*I am very grateful to my advisor Francesca Molinari for her invaluable support and encouragement. I am also grateful to my committee members, Larry Blume and Jörg Stoye for their advice. I thank Allan Collard-Wexler, Panle Jia Barwick, Matt Backus, Liran Einav, Jongrim Ha, Matthew Jackson, Duk Gyoo Kim, Charles Manski, Linden McBride, Adam McCloskey, Debi Prasad Mohapatra, Priya Mukherjee, Zhuo Zhong and seminar participants at Cornell for helpful comments and conversations. All remaining errors are mine alone.

†Address: Department of Economics, Cornell University, 455 Uris Hall, Ithaca, NY 14853. Tel: +1-607-379-3489. E-mail: jk2238@cornell.edu
1 Introduction

This paper proposes a structural model of multigraph formation. A multigraph is a graph or a network where a set of nodes can have different types of links, or relations. Each type of relation can be considered as a single network. I use a multigraph to describe a structure where a set of economic agents form two or more types of links with each other. The model in this paper has three main features. First, a set of economic agents determine many but not necessarily all types of links simultaneously. Second, all networks interact with each other in the sense that the structure of one network affects an individual’s utility from the other networks. Finally, one or more networks are endogenous but not simultaneously determined from the econometrician’s perspective.

Forming a multigraph is a commonplace phenomenon among economic agents. For example, people in a village have friendship and risk sharing partnerships, exchange favors, go to temple together, etc. Also, a transportation system among cities can be described as a multigraph. A city pair can be connected by many different transportation methods such as highways, trains, flights, etc. In this paper, I consider a multigraph resulting from the strategic decisions of economic agents in a village setting, and I estimate the utility parameters of the strategic formation of that multigraph.

The main difficulties of estimating structural parameters in the strategic formation of a multigraph are twofold. First, it is well-known that network formation games often exhibit multiple equilibria. Even in single-network formation models, the number of potential equilibria grows exponentially as the number of agents grows. If I allow for multiple types of links, the number of potential equilibria is even larger. Counting and checking all possible equilibria is infeasible even with a small number of nodes. Second, multigraph formation features simultaneity in various dimensions. Such simultaneity includes the externalities generated by the formation of a given relationship as well as the endogenous determination of multiple types of relationships at once. As in a single-network case, externalities exist in the sense that the link decision of a pair affects other pairs’ decisions. Moreover, agents may determine multiple types of links at the same time. For example, if an agent were to refuse to participate in one type of exchange, say the exchange of money, this refusal may result in the severance of other types of favor exchange relationships, and vice versa. There may also exist unobserved heterogeneity that affects two or more networks even

---

1For the mathematical definition, see, for example, Chartrand, Lesniak and Zhang (2011) and many others. In the economics literature, the term multigraph has been rarely used. Chandrasekhar and Jackson (2012) use multiplexing to describe individuals’ behavior to form multiple types of links and a multigraph to denote a multiple-link structure due to multiplexing.

2When a single type of link is possible, the number of potential equilibria is $2^{N(N−1)/2}$. If I allow for multiple types of links, the number is now $2^S \times N(N−1)/2$. Note that the order of convergence is still exponential in $N^2$, although the magnitude is bigger.
when those networks are not simultaneously determined.

In order to deal with these problems, I first extend the notion of pairwise stability of a single network to the framework of a multigraph. Pairwise stability of a network, proposed by Jackson and Wolinsky (1996), is a stability notion rather than an equilibrium solution concept. In addition, it contains the assumption of myopic agents. That is, individuals do not consider future changes in the network when they deviate. As shown in Kim (2013), pairwise stability and the myopic agent assumption make other potential equilibria irrelevant when estimating utility parameters. I will explain why the other equilibria are irrelevant later in Section 5.1. The result in Kim (2013) is also applicable to the multigraph framework. When the econometrician considers a community network structure where externalities are often limited, pairwise stability and its myopic agent assumption are especially well-suited. Under pairwise stability of a multigraph, I show that the structural model of multigraph formation is equivalent to a typical multinomial choice model.

When there are no endogenous (but not simultaneous) networks, a typical multinomial choice model, e.g. multinomial probit, can be applied. However, when an endogenous network is present, it prevents point identification of utility parameters and becomes a source of an incomplete econometric model, i.e. an econometric model predicts more than one outcome for some or all values of unobservables. I employ recently developed techniques for partially identified econometric models, especially random set theory (Beresteanu, Molchanov and Molinari (2011) and Galichon and Henry (2011), BMM11 and GH hereafter, respectively) to obtain the sharp identification region of the parameter vector through a finite set of moment inequalities. The characterization of the sharp identification region in this paper does not require an excluded instrument. I conduct inference using an estimation method developed by Andrews and Soares (2010), AS henceforth. Since this paper describes very detailed procedures to implement the estimation method, it provides practical guidance to the applied econometrician.

I apply the model to village networks in rural India and use the ‘Social Networks and Microfinance’ data collected by Banerjee, Chandrasekhar, Duflo, and Jackson. In a village, individuals form many different types of relations. I focus on four networks. Two of these are favor exchange networks: (1) borrowing and lending money, or equivalently risk sharing, and (2) borrowing and lending kerosene or rice. The other two are social networks: (3) friendship and (4) kinship. Investigating the formation of favor exchange networks, I consider friendship and kinship as an underlying structure in a village.

The two main empirical questions in this paper are as follows: (1) What is the effect of friend-
ship on the formation of a risk sharing network? and (2) Do individuals have caste homophily when forming a risk sharing network? At first glance, one may think that simply running a dyadic regression of risk sharing on friendship and caste would give reasonable estimates. However, there are a few sources of bias in such an estimation procedure. First, ignoring other favor exchange networks such as borrowing or lending rice can bias the estimate if individuals diversify their partners across different favors. Second, it is very likely that the econometrician may not observe several characteristics, such as personality, that affect both friendship and risk sharing networks. This endogeneity is another source of potential bias. In many cases, including this paper’s application, the econometrician does not have access to a valid instrument for the friendship network. The structural model in this paper is suitable to both problems mentioned above since it allows for simultaneous determination of different types of networks and the endogeneity of a friendship network. Although the parameters are partially identified, I find that friendship affects the formation of risk sharing and other favor exchange networks in the same direction. However, the empirical evidence for caste homophily in risk sharing and favor exchange networks is inconclusive.4

Related Literature


In addition to the theoretical models, many researchers study empirical models of strategic network formation. Currarini, Jackson and Pin (2010) propose a search-based model of friendship formation which identifies the role of preference and bias in matching. Christakis, Fowler, Imbens and Kalyanaraman (2010) try to empirically predict what network will be formed given link-specific variables as well as observed characteristics of individuals. Their model generates a network that may not be stable. Mele (2010) establishes a dynamic game of directed network formation, where individuals form a link according to a stochastic best response dynamic (See Blume 1993). Based

4Homophily is the tendency to bond with similar individuals in a society (Lazarsfeld and Merton 1954)

4
on random matching and utility maximization, the game generates a Markov chain of networks, and he proves the existence of a unique stationary distribution. Unlike Christakis et al. (2010), Mele (2010)’s model is a directed network formation, so it may not correspond to the formation of a friendship network, where mutual consents are important. Sheng (2012) employs a simultaneous-move link announcement game and uses pairwise stability of a network as a necessary condition for equilibrium. She focuses on subnetworks to reduce the number of equilibria, and applies a partial identification approach. In Kim (2013), I use pairwise stability as a stability notion and achieve point identification of model parameters by checking pairwise stability conditions. The present paper extends the approach of Kim (2013) to the multigraph framework. To the best of my knowledge, a structural model with simultaneously determined networks has not been studied in the literature. The model in this paper is widely applicable to many different settings where a multigraph is present.

This paper also contributes to the econometrics literature on discrete choice models with endogenous explanatory variables, as well as on the practical implementation of partial identification methods. The presence of endogenous explanatory variables is a commonplace problem in practice. I focus on the situation where the econometrician has no access to a valid instrument excluded in the structural equation. I build on the recently developed use of random set theory in partial identification to characterize the sharp identification region of model parameters. The use of random sets in econometrics was first proposed by Beresteanu and Molinari (2008). They study a class of models where the sharp identification region can be represented by a transformation of the Aumann expectation of a properly defined random set. BMM11 further develop a tractable characterization of the sharp identification region for incomplete econometric models with convex moment predictions. They show various examples of such models including simultaneous move games of complete and incomplete information which admit multiple equilibria with mixed strategies. Beresteanu, Molchanov and Molinari (2012), BMM12 henceforth, revisit previous problems in the literature, e.g. best linear prediction with interval data, and illustrate the benefits of using random set theory. Chesher, Rosen and Smolinski (2011), CRS henceforth, apply random set theory to multinomial choice models where endogenous explanatory variables are present. This paper is similar but different from CRS since I consider a situation where no excluded instruments are available for an endogenous variable. The model in this paper is also similar to that of Chesher and Rosen (2012), in the sense that they cover a case that exogenous variables are included in the structural equation. However they do not consider the case with no excluded instruments distinctively. As in BMM11, BMM12, GH, and CRS, the random set approach that I take yields that the parameter vector is

\[ \text{Other recent papers in empirical models of strategic network formation include, for example, Boucher and Mourifié (2012), Koenig (2012), Miyauchi (2012) and Leung (2013) among many others.} \]
defined by a finite set of moment inequalities. To reduce the number of moment inequalities I use the notion of core determining class first proposed by GH.

Inference in partially identified econometric models recently received much attention in the literature. AS propose a new class of confidence sets and tests, which uses generalized moment selection (GMS), for models in which parameters are defined by moment inequalities and equalities. The GMS procedure has a correct asymptotic size in a uniform sense. I follow their estimation method to construct confidence sets for the sharp identification region of the model parameters. The present paper has a similar spirit to that of Ciliberto and Tamer (2009) in the sense that I take the recently developed partial identification methodology to investigate empirical questions in detail.

Finally, the empirical results in this paper contribute to the development economics literature on risk sharing and favor exchange networks. There is an extensive literature on risk sharing and favor exchange networks in developing countries. Fafchamps (1992) describes that social networks play an important role in informal risk sharing. Townsend (1994) finds that individuals cannot achieve full insurance in villages. Fafchamps and Lund (2003) show that risk sharing is limited by the extent of social network. De Weerdt and Dercon (2006) find that risk sharing occurs through a social network within one’s village. Thus, the structure of social networks is very important for individuals, poor households as well as policy makers when dealing with many types of risks such as an epidemic and famine in developing areas.

There are a few studies whose empirical questions are closely related to mine. De Weerdt (2002) investigates what determines a risk sharing network by employing a dyadic logit regression. He finds that kinship, geographical proximity, mutual friends as well as some observed characteristics affect the formation of risk sharing network. However, his model takes into account neither the potential endogeneity of friendship nor the simultaneity of other favor exchange relationships. Hence, his estimation results may suffer from endogeneity bias. Fafchamps and Gubert (2007) find that geographic proximity possibly correlated with kinship and friendship plays an important role in the formation of a risk sharing network. Recently, Kinnan and Townsend (2012) investigate the effect of kinship on a risk sharing network. This paper differs from the existing papers in the literature in the following ways. First, the model incorporates the simultaneous determination of different types of favor exchange relationships. Second, I investigate the role of friendship links in the formation of a risk sharing network by allowing for the endogeneity of friendship. Since the model in this paper allows for endogeneity and simultaneity, I provide consistent and more credible estimates for the effect of friendship as well as other observed characteristics, e.g. caste, on the formation of a risk sharing network.
Structure of the Paper

The rest of the paper is organized as follows. Section 2 describes the multigraph framework, and provides categorization of different types of networks. I introduce pairwise stability of a multigraph, and discuss the existence of a pairwise stable multigraph in section 3. Section 4 explains how pairwise stability of a multigraph reduces the structural model to a multinomial choice model. Section 5 addresses identification and estimation. Section 6 investigates the empirical application to village networks. Section 7 discusses all empirical results, and Section 8 concludes. All proofs and detailed estimation procedures are placed in the Appendix.

2 Multigraph

2.1 Categorization of Networks

A set of economic agents often form many types of networks, or a multigraph. When the econometrician analyzes the strategic formation of a network, say $A_1$, he or she needs to carefully account for the other networks. First, this is because some of the other networks are determined simultaneously with $A_1$. To see this, consider a dynamic process of a multigraph formation where a pair is chosen at each period and revises links decision between them. At each opportunity of link-revision, a pair of individuals may revise two or more types of links simultaneously. Second, there may exist networks that are not simultaneously determined but endogenous to $A_1$, in the sense that an unobserved variable is correlated with both $A_1$ and that network. Finally, some networks might be strictly exogenous. In this section, I explain the structure of a multigraph with an example of networks in villages. Note that labeling which network is exogenous, endogenous or simultaneously depends on the context, or more specifically on the network of main interest. In this section, I use $A_1$ to denote a network that the econometrician is interested in. Below I define the other networks relative to $A_1$.

**Exogenous network:** A network $A_{ex}$ is strictly exogenous if the structure of $A_1$ does not affect the formation of $A_{ex}$, and $A_{ex}$ is statistically independent with an unobserved variable. For example, suppose that the econometrician is interested in the strategic formation of a risk sharing network in a village. The underlying kinship network among the villagers is strictly exogenous.

**Endogenous (but not simultaneously determined) network:** A network $A_{en}$ is endogenous but not simultaneously determined with $A_1$ if the structure of $A_1$ does not affect $A_{en}$, but

---

$^6$Manski (1993) uses ‘correlated effects’ to denote a tendency that individuals behave similarly to others due to facing similar characteristics or environment. In my case, a network is endogenous (but not simultaneous) to $A_1$ due to unobserved characteristics. In this sense, my classification of ‘endogenous network’ and ‘simultaneous network’ is similar to his ‘correlated effects’ and ‘endogenous effects’, respectively.
there exists an unobserved variable which affects both $A_1$ and $A_{en}$. If $A_{en}$ is taken as exogenous in an econometric model, the endogeneity bias will occur. For example, consider the village example again. A friendship network among village individuals is always endogenous but might not be simultaneously determined with the risk sharing network $A_1$. One may think friendship and risk sharing network are determined simultaneously. However, the rationale to put friendship into this category is as follows. Although rejecting mutual insurance or asking risk sharing may lead to the severance of friendship, one or a couple of such events is unlikely to break up the friendship, especially in a developing area. When the econometrician considers the dynamic formation of a risk sharing network with the assumption of myopic agents, it is reasonable to assume that friendship is a fixed structure among the individuals given the short period of the time frame. However, it is not strictly exogenous, since there may exist a variable unobserved to the econometrician, which is correlated with both networks $A_1$ and $A_{en}$. Personality can be a good example of such an unobserved variable. Hence, the friendship network is endogenous to the risk sharing network in a village.

**Simultaneous network:** A network $A_{sim}$ is simultaneously determined with $A_1$ if agents (or pairs) determine their relationships on $A_{sim}$ and $A_1$ at the same time or in a negligible lag of time. Moreover, a link decision of an individual or a pair in $A_{sim}$ affect their utility of forming a link in $A_1$, and vice versa. In this case, $A_{sim}$ and $A_1$, or link decisions on those networks should be considered jointly as an outcome variable. In the village example, different types of favor exchange networks can fall into this category. Also, a network which represents relatively instant relations can be simultaneously determined. More specifically, a network of borrowing or lending kerosene or rice, and a network of providing or receiving medical help can be determined simultaneously with the risk sharing network. These types of relationships are relatively more instant or spontaneous, so the severance or formation of one such relationship may result in the severance or formation of the others immediately. Hence, it is reasonable to put such networks together as an outcome variable.

2.2 **Set-up**

Let $m = 1, \cdots, M$ be an index for communities or villages. Let $i = 1, \cdots, n_m$ be an index for individuals in village $m$. I use $N = \sum_{m=1}^{M} n_m$ as the number of all individuals in the data, and $ij = 12, 13, \cdots, (n_m - 1)n_m$ to denote unordered pairs of individuals $i$ and $j$. Let $A_{s,m}$ be a type-$s$ network among $n_m$ individuals for $s = 1, \cdots, S$, where $S$ is finite and small. With a slight abuse of notation, I also use $A_{s,m}$ as an $n_m$ by $n_m$ adjacency matrix with its $(i, j)$th element $a_{ij}^{(s)}$. That
is,

\[ a_{ij}^{(s)} = \begin{cases} 
1, & \text{if } i \text{ and } j \text{ have a type}-s \text{ relation.} \\
0, & \text{otherwise.} 
\end{cases} \]

All networks are undirected in the sense that if \( i \) nominates \( j \) as a friend for example, then \( j \) also views \( i \) as a friend. Let \( ij^s \) be the link between \( i \) and \( j \) on the \( s \)-th network, i.e. a type-\( s \)-link between \( i \) and \( j \). Let \( G_m = \{ A_{s,m}, s = 1, \cdots, S \} \) collect all networks in a village \( m \), thus it is a multigraph. Links in a multigraph may be simultaneous, endogenous and exogenous to each other as explained in Section 2.1. For simplicity, I omit the village index \( m \) until necessary. Let \( A_1 \) be the network of the econometrician’s main interest. I use \( Y = \{ A_1, \cdots, A_p \} \) to denote simultaneously determined networks with \( A_1 \), including \( A_1 \) itself, where \( p \) is the number of all simultaneously determined networks. Similarly, \( W = \{ A_{p+1}, \cdots, A_{p+q} \} \) and \( V = \{ A_{p+q+1}, \cdots, A_S \} \) are endogenous (but not simultaneous) and exogenous networks, respectively. I am interested in estimating utility parameters of forming \( Y \) given the other networks \( W \) and \( V \), and observed individual characteristics \( X \). A set of links, \( Y_{ij} = (a_{ij}^{(1)}, \cdots, a_{ij}^{(p)})' \) denotes the links decision of \( i \) and \( j \) on \( Y \). Let \( \mathcal{Y}_{ij} \) be the set of all possible \( Y_{ij} \)'s, e.g. if \( p = 2 \) (two simultaneously determined networks), \( \mathcal{Y}_{ij} = \{(0,0), (0,1), (1,0), (1,1)\} \).

Individual \( i \)'s utility from a simultaneously determined multigraph \( Y \) given the rest of the networks is

\[
U_i(Y|W, V, X, \varepsilon) = \sum_{s=1}^{p} \sum_{j=1}^{n} a_{ij}^{(s)} (u_{ij}^{(s)}(W, V, X) + f^{(s)}(Y_{ij})) + \sum_{l \neq s} \sum_{s=1}^{p} \sum_{j=1}^{n} a_{ij}^{(s)} a_{ij}^{(t)} \delta^{(s,t)}(W, V, X) + \varepsilon_{ij} (Y_{ij}). \tag{2.1}
\]

In the above expression, the base utility \( u_{ij}^{(s)}(W, V, X) \) in the first term captures \( i \)'s utility from \( j \) when \( i \) has a type-\( s \) relationship with \( j \). It may include the cost of maintaining a link. The term \( f^{(s)}(Y_{-ij}) \) captures externalities due to other pairs’ link decisions \( Y_{-ij} \), which will be specified and explained later with more detail. The term \( \delta^{(s,t)}(W, V, X) \) captures an additional utility by having two relationships \( s \) and \( t \) with a partner \( j \). All observed characteristics are collected in a \( \kappa_x \times \binom{N}{2} \) matrix \( X = (x_{i2}', \cdots, x_{n-1,n}')' \), where \( x_{ij} \) is an \( \kappa_x \times 1 \) vector of observed characteristics for \( i \) and \( j \). The term \( \varepsilon_{ij}(Y_{ij}) \) is an a match-specific unobservable which depends on the current set of links \( Y_{ij} \). It may not be independent across alternatives \( Y_{ij} \)'s.

Now, consider \( f^{(s)}(Y_{-ij}) \) and \( \delta^{(s,t)}(W, V, X) \). The function \( f^{(s)}(Y_{-ij}) \) captures network (or more precisely multigraph) externalities due to other pairs’ links in \( Y \). Let \( h_{ij}^{(s)} = \sum_k a_{ik}^{(s)} a_{jk}^{(s)} \) be the number of mutual type-\( s \) relationship partners between \( i \) and \( j \), and let \( h_{ij} \) collect all \( h_{ij}^{(s)} \)'s, i.e.
\[ h_{ij} = (h_{ij}^{(1)}, h_{ij}^{(2)}, \ldots, h_{ij}^{(p)})'. \] Then, \( f^{(s)}(Y_{-ij}) \) can be written as

\[ f^{(s)}(Y_{-ij}) = h'_{ij} \gamma^{(s)} = \sum_{t=1}^{p} \gamma_{t}^{(s)} \sum_{k} a_{ik}^{(t)} a_{jk}^{(t)}. \] (2.2)

The coefficient \( \gamma_{t}^{(s)} \) captures the effect of having mutual type-\( t \) friends with \( j \) on the utility of forming a type-\( s \) link with \( j \). The additional benefits \( \delta^{(s,t)}(W, V, X) \) of having both relationships may depend on the endogenous and/or exogenous networks as well as observed characteristics of \( i \) and \( j \).

Let \( Y'_{ij}^{(-s)} = (a_{ij}^{(1)}, \ldots, a_{ij}^{(s-1)}, a_{ij}^{(s+1)}, \ldots, a_{ij}^{(p)}) \) be the current links decision of \( i \) and \( j \) except type-\( s \) relation. Let \( mu_{i}^{(s)}(j|Y_{ij}^{(-s)}, W, V, X, \varepsilon) = mu_{i}^{(s)}(j|Y'_{ij}^{(-s)}) \) be individual \( i \)'s marginal utility of forming a type-\( s \) link with \( j \) versus not forming one given that their other relationships are \( Y_{ij}^{(-s)} \).

Then it can be written as

\[ mu_{i}^{(s)}(j|Y'_{ij}^{(-s)}) = u_{ij}^{(s)}(W, V, X) + f^{(s)}(Y_{-ij}) + \sum_{t \neq s}^{p} a_{ij}^{(t)} \delta^{(s,t)}(W, V, X) + \varepsilon_{ij}(Y'_{ij}) - \varepsilon_{ij}(Y_{ij}), \] (2.3)

where \( Y'_{ij} \) adds the type-\( s \) link to \( Y_{ij} \).

**Example.** Two simultaneous determined networks in a multigraph, i.e. \( p = 2 \).

Individual \( i \)'s utility from a multigraph \( Y \) is

\[
U_{i}(Y|W, V, X, \varepsilon) = \sum_{j=1}^{n} a_{ij}^{(1)} u_{ij}^{(1)}(W, V, X) + h'_{ij} \gamma^{(1)}
+ \sum_{j=1}^{n} a_{ij}^{(2)} u_{ij}^{(2)}(W, V, X) + h'_{ij} \gamma^{(2)}
+ \sum_{j=1}^{n} a_{ij}^{(1)} a_{ij}^{(2)} h^{(1,2)}(W, V, X) + \varepsilon_{ij}(a_{ij}^{(1)}, a_{ij}^{(2)}). \] (2.4)

Individual \( i \)'s marginal utility from having the first relationship given that the second relationship is absent is

\[ mu_{i}^{(1)}(j|0) = u_{ij}^{(1)}(W, V, X) + \gamma_{1}^{(1)} h_{ij}^{(1)} + \gamma_{2}^{(1)} h_{ij}^{(2)} + \varepsilon_{ij}(1,0) - \varepsilon_{ij}(0,0). \]

Other marginal utility terms can be written analogously. The sum of marginal utilities \( mu_{i}^{(s)}(\cdot) + \)
\( \mu_j^{(e)}(\cdot) \) can be written as, for example,

\[
\mu_i^{(1)}(j|0) + \mu_j^{(1)}(i|0) = u_{ij}^{(1)}(W, V, X) + u_{ji}^{(1)}(W, V, X) + 2h_{ij}^{(1)} + 2\varepsilon_{ij}(1,0) - 2\varepsilon_{ji}(0,0).
\] (2.5)

Note that \( \varepsilon_{ij}(1,0) \) is match-specific, i.e. \( \varepsilon_{ij}(1,0) = \varepsilon_{ji}(1,0) \). The assumption of the match-specific unobservables implies that the unobserved variables affect both \( i \) and \( j \)'s utility in the same way. For example, different personality between two individuals affects their utility in the same way.\(^7\)

### 3 Pairwise Stability of a Mutigraph

#### 3.1 Definition

In this section, I focus only on simultaneously determined networks \( Y \). Recall that \( Y_{ij} = (a_{ij}^{(1)}, \ldots, a_{ij}^{(p)})' \) is the current links decision of \( i \) and \( j \) in \( Y \). For example, consider there are two types of relationships in \( Y \), say risk sharing \((s = 1)\) and borrowing or lending rice \((s = 2)\). If individuals \( i \) and \( j \) have both risk sharing and friendship in a current multigraph \( Y \), then \( Y_{ij} = (1, 1) \). Now, let \( Y + Y_{ij}' \) be a multigraph which adds links between \( i \) and \( j \) in \( Y_{ij}' \) to \( Y \), and \( Y - Y_{ij} \) be a multigraph which deletes links between \( i \) and \( j \) in \( Y_{ij} \) from \( Y \).\(^8\) Those two multigraphs differ from \( Y \) by links involved with a pair \( ij \) only. I extend the notion of pairwise stability of a single network to the multigraph framework as follows.

**Definition 1.** (PSM with non-transferable utility) Let \( U_i(Y) \) be \( i \)'s utility from a multigraph \( Y \). The value of the a multigraph is \( V(Y) = \sum_i U_i(Y) \). Let \( Y_{ij} \) be the current link decisions of \( i \) and \( j \) in \( Y \). A multigraph \( Y \) satisfies pairwise stability of a multigraph (PSM) with non-transferable utility if the following conditions hold for all \( i \) and \( j \).

(i) For \( Y_{ij} \neq (0, \cdots, 0) \), \( U_i(Y) \geq U_i(Y - Y_{ij} + Y_{ij}') \) and \( U_j(Y) \geq U_j(Y - Y_{ij} + Y_{ij}') \) for all \( Y_{ij}' \neq Y_{ij} \) \( \in \mathcal{Y}_{ij} \).

(ii) For \( Y_{ij} = (0, \cdots, 0) \), if \( U_i(Y + Y_{ij}') > U_i(Y) \), then \( U_j(Y + Y_{ij}') < U_j(Y) \) for all \( Y_{ij}' \in \mathcal{Y}_{ij} \).

Condition (i) indicates that the set of current relations \( Y_{ij} \) between \( i \) and \( j \) is at least as beneficial as the other sets of relations for both \( i \) and \( j \). The deviation from the current set of

\(^7\)The assumption can be relaxed to that of both match- and individual-specific unobservables, i.e. \( \varepsilon_{ij}(1,0) \neq \varepsilon_{ji}(1,0) \). However, the latter assumption gives rise to an additional computational burden. Briefly speaking, the estimation procedure requires drawing twice larger dimension of unobservables, since the estimation relies on the method of simulated moments.

\(^8\)If some links in \( Y_{ij}' \) are already present in \( Y \), I add non-existing ones only.
relations \( Y_{ij} \neq (0, \cdots, 0)' \) requires that at least one of \( i \) and \( j \) strictly prefers the alternative \( Y'_{ij} \), and that the other individual is at least indifferent. When a pair \( ij \) has no relations in \( Y \), i.e. \( Y_{ij} = (0, \cdots, 0)' \), the condition (ii) provides that the formation of a relation or a set of relations between \( i \) and \( j \) requires only indifference between \( Y_{ij} = (0, \cdots, 0)' \) and \( Y'_{ij} \neq (0, \cdots, 0)' \).

Next, I provide pairwise stability of a multigraph, when utility is transferable.

**Definition 2.** (PSM with transferable utility) Let \( U_i(Y) \) be \( i \)'s utility from a multigraph \( Y \). The value of the multigraph is \( V(Y) = \sum_i U_i(Y) \). Let \( Y_{ij} \) be the current link decisions of \( i \) and \( j \) in \( Y \). A multigraph \( Y \) satisfies pairwise stability of a multigraph (PSM) with transferable utility if the following conditions hold for all \( i \) and \( j \).

(i) For \( Y_{ij} \neq (0, \cdots, 0) \), \( U_i(Y) + U_j(Y) \geq U_i(Y - Y_{ij} + Y'_{ij}) + U_j(Y - Y_{ij} + Y'_{ij}) \) for all \( Y'_{ij} \neq Y_{ij} \) \( \in Y_{ij} \).

(ii) For \( Y_{ij} = (0, \cdots, 0) \), \( U_i(Y) + U_j(Y) > U_i(Y + Y'_{ij}) + U_j(Y + Y'_{ij}) \) for all \( Y'_{ij} \neq Y_{ij} \) \( \in Y_{ij} \).

Condition (i) means that when \( i \) and \( j \) have a set of relations \( Y_{ij} \), it must be as beneficial as the other sets of relations in terms of the sum of their utilities. Condition (ii) tells that when \( i \) and \( j \) have no relations, it must provide the sum of utilities of \( i \) and \( j \) strictly higher than having at least one relation. Both PSM with non-transferable utility and PSM with transferable utility provide conditions under which no pairs of individuals want to deviate from their current set of relationships. For example, suppose there are only two simultaneously determined relationships: friendship and risk sharing. If \( i \) and \( j \) currently have friendship and risk sharing relationship, then none of the following combinations- friendship only, risk sharing relationship only, and no relationships- give higher utility than the current one.

The notion of pairwise stability of a multigraph is interesting for practice since it provides a tool to investigate interactions among different types of networks or relations. The non-deviation conditions from the current set of relations provide a preference ordering as I will show in Section 4. The inequality conditions corresponding to the preference ordering provide identification information for the utility parameters. Since the preference ordering is across different relations, it provides identification information for interactions among different relations, or networks.

If a type-\( s \) network \( A_s \) in a multigraph \( Y \) satisfies PSM, it is not only pairwise stable by itself but also pairwise stable jointly with other simultaneously determined networks \( \{A_1, \cdots, A_{s-1}, A_{s+1}, \cdots, A_p\} \). In this sense, when two or more networks are simultaneously determined, PSM is not only stronger than pairwise stability of a single network, but also stronger than pairwise stability of all the single networks considered one at a time. Although it is a stronger stability notion, conditions for the existence of a pairwise stable multigraph are not more restrictive than those of a single pairwise stable network. I will discuss the existence of a pairwise stable multigraph in the following section.
3.2 Existence

The results on the existence of a pairwise stable multigraph in this section rely on the results of Jackson and Watts (2001, 2002). First, I introduce a few definitions which extend similar ones from Jackson and Watts (2001, 2002) to the multigraph framework. Then, I establish conditions of $U$ under which a pairwise stable multigraph exists by Proposition 1 and 2.

**Definition.** (i) (a cycle) A set of multigraphs $C$ form a cycle if for any $Y \in C$ and $Y' \in C$ there exists an improving path connecting $Y$ to $Y'$.

(ii) (a closed cycle) A cycle $C$ is a closed cycle if no multigraph in $C$ lies on an improving path leading to a multigraph that is not in $C$.

(iii) (defeated) A multigraph $Y$ is defeated by $Y' = Y - Y_{ij} + Y'_{ij}$ if $U_i(Y') > U_i(Y)$ and $U_j(Y') > U_j(Y)$ in the non-transferable utility case, and $U_i(Y') + U_j(Y') > U_i(Y) + U_j(Y)$ in the transferable utility case.

**Proposition 1.** Fix $U$ and $V$. Suppose that $U_i(Y) \neq U_i(Y - Y_{ij} + Y'_{ij})$ for all $i$, $Y$, $Y_{ij}$ and $Y'_{ij} \neq Y_{ij}$. Then, there exists at least one pairwise stable multigraph or a closed cycle of multigraphs.

**Proof.** See Appendix A.1.

Consider a case that $Y'$ differs from $Y$ by a pair $ij$’s links decision only. If individuals $i$ and $j$ receive different utilities from those two multigraphs, then there is either a pairwise stable multigraph or a closed cycle of multigraph. The assumption of no ties restricts utilities between $Y$ and $Y' = Y - Y_{ij} + Y'_{ij}$, but does not restrict utilities from two arbitrary multigraphs in general.

If two multigraphs, $Y$ and $Y''$, differ by two or more pairs’ links, they can still have a tie for some individuals involving the difference in $Y$ and $Y''$.

From Proposition 1, a pairwise stable multigraph exists if and only if there are no cycles. The following proposition provides conditions under which such cycles are ruled out.

**Proposition 2.** Fix $U$ and $V$. Suppose that $U_i(Y) \neq U_i(Y - Y_{ij} + Y'_{ij})$ for all $i$, $Y_{ij}$ and $Y'_{ij} \neq Y_{ij}$. If there exists a potential function $\omega : Y \to \mathbb{R}$ such that $[Y' \text{ defeats } Y] \iff [\omega(Y') > \omega(Y)]$, and $Y$ and $Y'$ are different with respect to a single pair’s link decision only, then there are no cycles. Therefore, there exists at least one multigraph which is pairwise stable.

**Proof.** See Appendix A.2.

Proposition 2 tells that cycles are ruled out if there exists a function which represents the incentives of individuals with respect to any changes involving a single pair only. Under the utility specification (2.1) and (2.2) I can find a function $\omega(\cdot)$. 


Corollary 1. Suppose that $U_i$ is defined as (2.1) and (2.2), and that utility is transferable. Then, there exists a function $\omega(\cdot)$ that satisfies the condition in Proposition 2. Therefore, there exists at least one pairwise stable multigraph for any $\varepsilon$.

Proof. See Appendix A.2.

By Corollary 1, at least one pairwise stable multigraph exists for any values of $\varepsilon$. Note that Corollary 1 holds for any finite number of relationships in a multigraph when the following conditions are satisfied. (1) An individual’s base utility has an additive form across different links, (2) externalities can be represented by the form of the equation (2.2), and (3) the additional utility of having multiple relations does not depend on the structure of the multigraph. Note that these conditions are sufficient but not necessary.

I can find a potential function $\omega(\cdot)$ only when utility is transferable. When utility is non-transferable, the existence of a pairwise stable multigraph may depend on the sign and the magnitude of the externalities. Hellmann (2012) provides conditions for the existence of a pairwise stable network with non-transferable utility. His results may be extended to the multigraph framework, but I leave it as a future study.

In this paragraph, I discuss briefly the uniqueness of a pairwise stable multigraph. As in the single network case, the set of pairwise stable multigraphs, say $\mathcal{PSM}$, given a utility function $U$ is not a singleton in general. A particular formation process may not generate all pairwise stable multigraphs in $\mathcal{PSM}$. It is also possible that some formation processes may not generate a pairwise stable multigraph, or may not produce a particular pairwise stable multigraph of interests. To see this, consider a dynamic process of multigraph formation where a pair is chosen at each period and revise their links. The sequence of meetings among pairs is crucial to determine the configuration of a pairwise stable multigraph. Different meeting sequences can generate different pairwise stable multigraphs. However, under the same meeting sequence, agents always form the same pairwise stable multigraph. Note that if the econometrician is interested only in utility parameters, then knowledge about the history of meeting is not required. Consequently, other pairwise stable multigraphs in $\mathcal{PSM}$ are irrelevant, since those multigraphs are not outcomes of the multigraph formation process which generates the observed multigraph. I will discuss this with more details in Section 5.1.

4 A Multinomial Choice Model under PSM

From this section, I set $p = 2$ to simplify the model for ease of explanation. When $p$ is bigger than two but small, the results in this section can be easily extended with only several more tedious
steps. Also, I assume that the utility of forming a multigraph is transferable.

PSM provides no-deviation conditions for each pair. When there are a total of \( p \) simultaneous networks, the cardinality of \( Y_{ij} \) (the set of all possible links between \( i \) and \( j \)) is \( 2^p \). Hence, PSM provides \( 2^p - 1 \) no-deviation conditions for each pair. When \( p = 2 \), there are only four possible combinations of link types for each pair of agents. Once one of four combinations, say ‘type-1 link only’, is chosen by \( i \) and \( j \), PSM provides the following \( 2^2 - 1 \) conditions; \((1, 0) \succsim (0, 0)\), \((1, 0) \succsim (0, 1)\) and \((1, 0) \succsim (1, 1)\).

Recall that \( \mu_i^{(s)}(j|Y_{ij}^{(-s)}) = a_{ij}^{(t)} \) is \( i \)'s marginal utility from a type-\( s \) link with \( j \), when they have relations \( a_{ij}^{(t)} \). In addition, let \( \mu_i^{(1,2)}(j) \) be \( i \)'s marginal utility by adding both types of links at the same time. I describe how to construct moment inequalities under PSM with transferable utility.

First, consider the case that \( i \) and \( j \) choose \( Y_{ij} = (a_{ij}^{(1)}, a_{ij}^{(2)}) = (1, 1) \). It implies that \((1, 1)\) is weakly preferred to any other combinations, \((0, 1)\), \((1, 0)\), and \((0, 0)\). Hence, I have

\[
[(1, 1) \succsim (0, 1)] \Rightarrow \mu_i^{(1)}(j|a_{ij}^{(2)} = 1) + \mu_j^{(1)}(i|a_{ij}^{(2)} = 1) \geq 0, \quad (4.1)
\]

\[
[(1, 1) \succsim (1, 0)] \Rightarrow \mu_i^{(2)}(j|a_{ij}^{(1)} = 1) + \mu_j^{(2)}(i|a_{ij}^{(1)} = 1) \geq 0, \quad (4.2)
\]

and

\[
[(1, 1) \succsim (0, 0)] \Rightarrow \mu_i^{(1,2)}(j) + \mu_j^{(1,2)}(i) \geq 0. \quad (4.3)
\]

When \( i \) and \( j \) choose \( (a_{ij}^{(1)}, a_{ij}^{(2)}) = (1, 0) \), similarly there are following three conditions:

\[
[(1, 0) \succsim (1, 1)] \Rightarrow \mu_i^{(2)}(j|a_{ij}^{(1)} = 1) + \mu_j^{(2)}(i|a_{ij}^{(1)} = 1) \leq 0, \quad (4.4)
\]

\[
[(1, 0) \succsim (0, 1)] \Rightarrow \mu_i^{(1)}(j|a_{ij}^{(2)} = 0) + \mu_j^{(1)}(i|a_{ij}^{(2)} = 0) \geq \mu_i^{(2)}(j|a_{ij}^{(1)} = 0) + \mu_j^{(2)}(i|a_{ij}^{(1)} = 0), \quad (4.5)
\]

Even when \( p \) is large, the results in this section can be extended. However, practical implementation can be difficult. For example, when \( p = 20 \), there are more than a million choices for each pair.

I use the weak preference relation \( \succsim \) in the sense that utility of one choice is as great as that of the others. When \((0, 0)\) is chosen by a pair, the strong preference relation should be used based on the definition of PSM. Of course, when the distribution of \( \varepsilon \) is absolutely continuous, this discretion is unnecessary.
The case of \((a^{(1)}_{ij}, a^{(2)}_{ij}) = (0, 1)\) is symmetric to the above. That is,

\[
[(0, 1) \succ (1, 1)] \Rightarrow m_{ij}^{(1)}(j|a^{(2)}_{ij} = 1) + m_{ij}^{(1)}(i|a^{(2)}_{ij} = 1) \leq 0,
\]

and

\[
[(0, 1) \succ (0, 0)] \Rightarrow m_{ij}^{(2)}(j|a^{(1)}_{ij} = 0) + m_{ij}^{(2)}(i|a^{(1)}_{ij} = 0) \leq 0.
\]

Finally, when a pair \(ij\) chooses \((a^{(1)}_{ij}, a^{(2)}_{ij}) = (0, 0)\), I have

\[
[(0, 0) \succ (1, 1)] \Rightarrow m_{ij}^{(1,2)}(j) + m_{ij}^{(1,2)}(i) < 0,
\]

and

\[
[(0, 0) \succ (1, 0)] \Rightarrow m_{ij}^{(1)}(j|a^{(2)}_{ij} = 0) + m_{ij}^{(1)}(i|a^{(2)}_{ij} = 0) < 0,
\]

and

\[
[(0, 0) \succ (0, 1)] \Rightarrow m_{ij}^{(2)}(j|a^{(1)}_{ij} = 0) + m_{ij}^{(2)}(i|a^{(1)}_{ij} = 0) < 0.
\]

Taking the conditional expectation of the above inequalities with respect to the distribution of \(\varepsilon\) given observed variables produces the population moment inequalities. Then the parameter vector is defined by those moment inequalities. The identification region is the set of parameter values that satisfy the moment inequalities. I study the identification problem formally in the next section.

5 Identification and Estimation

5.1 Implications of PSM

In the end of Section 3.2., I briefly mentioned the benefits of using PSM, especially that other pairwise stable multigraphs are irrelevant when estimating the utility parameters. In this section, I explain this in detail through three steps. First, I demonstrate that for any values of \(\varepsilon_{ij} \in \mathbb{R}^{2p-1}\), a set of links of a pair is uniquely determined given the rest of the multigraph and observed characteristics. Second, I show that all pairs’ links decisions are separate from each other due to
the myopic agent assumption in PSM. Finally, I explain the irrelevance of other pairwise stable multigraphs.

Consider a dynamic process of multigraph formation as in Sections 2 and 3.2. In the process, a pair of individuals is chosen at each period, and the pair revises its current links given the current multigraph. When they revise, they do not consider the future changes in the multigraph, i.e. agents are myopic. The link formation (or revision) in each period occurs as a two people cooperative game. Hence, a pair chooses a set of links \( Y_{ij} \), which maximizes the sum of their utilities given \( W_{ij}, V_{ij}, X_{ij}, Y_{-ij} \), and \( \varepsilon_{ij} \). Lemma 1 shows that there is a unique prediction for each \( Y_{ij} \) given \( W_{ij}, V_{ij}, X_{ij} \) and \( Y_{-ij} \) for all realizations of \( \varepsilon_{ij} \in \mathbb{R}^{2p-1} \).

Lemma 1. For all values of \( \varepsilon_{ij} \in \mathbb{R}^{2p-1} \), \( Y_{ij} \) is uniquely determined under pairwise stability of a multigraph given \( W_{ij}, V_{ij}, X_{ij} \) and \( Y_{-ij} \).

Proof. Consider only the case with \( p = 2 \) for ease of exposition. It can be easily seen that the following four regions of \( \varepsilon \) corresponding to (4.1)-(4.3), (4.4)-(4.6), (4.7)-(4.9), and (4.10)-(4.12) are disjoint and compose \( \mathbb{R}^3 \).

Once the rest of the multigraph is fixed, each pair’s links decision \( Y_{ij} \) is uniquely predicted given \( W_{ij}, V_{ij} \) and \( X_{ij} \) for any realizations of \( \varepsilon_{ij} \). However, Lemma 1 does not mean that a pairwise stable multigraph \( Y \) is uniquely determined given \( W, V \) and \( X \) for all \( \varepsilon = (\varepsilon_{12}, \ldots, \varepsilon_{n-1,n}) \). This result is noticeably different from that of a two by two simultaneous move entry game in the literature (see for example, Tamer (2003)), where multiple equilibria, i.e. multiple predicted outcomes, occur for some regions of unobservables. In the link formation game, all \( n(n-1) \) pairwise outcomes are uniquely determined.

The remaining problem is whether those \( n(n-1) \) outcomes are separate. At this point, the benefit of the myopic agent assumption in PSM arises. Since observed multigraph is pairwise stable, each pair’s current links decision \( Y_{ij} \) is also pairwise stable given \( W_{ij}, V_{ij}, X_{ij} \) and \( Y_{-ij} \). The assumption of myopic agents makes all pairwise decisions separate across pairs. To see this, consider a pair \( ij \) with current links \( Y_{ij} \) on a multigraph \( Y \). In order to form pairwise stable relations \( Y_{ij} \), the pair \( ij \) has to check the difference between \( U_i(Y - Y_{ij} + Y_{ij}') + U_j(Y - Y_{ij} + Y_{ij}') \) and \( U_i(Y) + U_j(Y) \) for all \( Y_{ij}' \neq Y_{ij} \). This comparison does not take into account future changes in the rest of the multigraph \( Y_{-ij} \). In other words, the pair \( ij \) compares the sum of their utilities from \( Y - Y_{ij} + Y_{ij} \) with the sum of utilities from \( Y \). Thus, the rest of network \( Y_{-ij} \) is fixed when they make a decision, and there is no simultaneity among \( Y_{ij} \)’s. The utility parameters in the strategic formation of a multigraph can be obtained by checking pairwise stability conditions for each pairwise decision \( Y_{ij} \) separately, given \( W_{ij}, V_{ij}, X_{ij} \) and \( Y_{-ij} \).
Finally I discuss the irrelevance of other pairwise stable multigraphs in the set of all pairwise
stable multigraphs, say \( PSM \). Since pairwise stability of a multigraph is not an equilibrium
solution concept, the set \( PSM \) is not a set of equilibrium multigraphs. Rather, it is a set of
multigraphs that are outcomes of many different games. For the dynamic process of multigraph
formation explained in Section 3.2, a multigraph has a corresponding history of meetings. Given
a specific history, the same multigraph is always realized. By the implicit assumptions of the
dynamic formation process and the independence between meeting and preference, other pairwise
stable networks corresponding to different histories are irrelevant to the identification of the utility
parameters.\(^{11}\)

Now, if there are no endogenous variables, the identification problem is reduced to that of
multinomial choice models. If one imposes a distributional assumption on \( \varepsilon \) such as the normal or
the logistic distribution, the parameter vector \( \theta \) is point-identified under suitable location and scale
normalizations.\(^{12}\)

### 5.2 Identification with an Endogenous Explanatory Variable and No Instruments\(^{13}\)

I have shown that under the pairwise stability of a multigraph, the structural model of strategic
multigraph formation is equivalent to a multinomial choice model. This section focuses on the
identification and the estimation of the multinomial choice model. I build the analysis in this
subsection on CRS. For ease of exposition, let\(^{14}\)

\[
\tilde{N} = \sum_{m=1}^{M} \binom{n_m}{2}, \text{ the number of all pairs},
\]

\[
W_{ij} = (a_{ij}^{(p+1)}, \ldots, a_{ij}^{(p+q)})', \text{ endogenous relations between } i \text{ and } j,
\]

\[
Z_{ij} = (V_{ij}', X_{ij}', Y_{ij}')', \text{ the collection of all exogenous variables},
\]

\(^{11}\)One may consider a simultaneous move game of multigraph formation. However, the simultaneous move game
may not generate the observed multigraph, or even it may not have an equilibrium multigraph which is pairwise
stable. To see this, one may require an extension of pairwise Nash equilibrium to the simultaneous move game of
multigraph formation. The existence and the uniqueness of pairwise Nash equilibrium and corresponding multigraphs
are interesting future research areas.

\(^{12}\)For example, if \( \varepsilon_{ij} \) is i.i.d. normal across pairs and alternatives with zero-mean and variance-covariance matrix
\( \Sigma^{-1} \), \( \theta \) is point-identified.

\(^{13}\)The notations used in this section follow ones in BMM12 and CRS.

\(^{14}\)From this section, I use \( Y, W \) and \( Z \) as generic random variables or random vectors for \( Y_{ij}, W_{ij} \) and \( Z_{ij} \).
where

\[ V_{ij} = (a_{ij}^{(p+q+1)}, \ldots, a_{ij}^{(S)})', \text{ exogenous relations between } i \text{ and } j, \]

\[ X_{ij} = (x_{1,i}, \ldots, x_{\kappa_i,i}, x_{1,j}, \ldots, x_{\kappa_j,j})', \text{ the observed characteristics of } i \text{ and } j, \]

\[ Y_{-ij} = (a_{-ij}^{(1)}, \ldots, a_{-ij}^{(p)})', \text{ other pairs' links decisions in the simultaneously determined networks}. \]

Recall that \( Y_{ij} = (a_{ij}^{(1)}, a_{ij}^{(2)}, \ldots, a_{ij}^{(p)})' \). Let \( y \) be its generic (vectorized) value. The model satisfies the following assumptions.

**Assumption 1.**

(i) \((Y, W, Z, \varepsilon)\) is defined on a probability space \((\Omega, \mathcal{F}, P)\). The support of \(Y_{ij}\) is a finite set \(Y\). The support of \((W, Z, \varepsilon)\) is \(W \times Z \times \mathbb{R}^{2p-1}\).

(ii) \(F_{Y,W|Z}^0\) and \(F_{W|Z}^0\) are identified by the sampling process.

(iii) \(\varepsilon\) is a continuously distributed r.v. with everywhere positive density w.r.t. Lebesgue measure with distribution belonging to \(\mathcal{P}_\varepsilon\).

(iv) \(\varepsilon_{ij} = (\varepsilon_{ij,1}, \varepsilon_{ij,2}, \ldots, \varepsilon_{ij,2p-1})\) is i.i.d. across pairs, but not i.i.d. across the alternatives.

(v) \(\varepsilon\) and \(Z\) are statistically independent. No excluded instrumental variables are available for \(W\).

Note that Assumption (i), (ii), (iii), and (v) are not restrictive. The iid assumption in (iv) is prevalent in the literature, e.g. Christakis et al. (2010).

Define

\[ Q_\theta(y, w, z) \equiv \{\varepsilon : g(w, z, \varepsilon; \theta) = y\}, \]

where the structural equation \(g\) can be written as

\[ g(w, z, \varepsilon; \theta) = \sum_{y \in Y} y \times 1 \left[ \min_{k \in Y, k \neq y} \{U_i(y|w, z, \varepsilon; \theta) + U_j(y|w, z, \varepsilon; \theta) - U_i(k|w, z, \varepsilon; \theta) - U_j(y|w, z, \varepsilon; \theta)\} > 0 \right]. \]

That is, \(g(\cdot)\) represents a mechanism that generates an outcome \(y\) given \(w, z\) and \(\varepsilon\). In this case, given \(w, z\) and \(\varepsilon\), \(g(\cdot)\) chooses an alternative \(y \in Y\) which provides the highest utility.

Let \(\mathcal{C}(\mathbb{R}^{2p-1})\) be the collection of all closed subsets of \(\mathbb{R}^{2p-1}\). For any \(D \in \mathcal{C}(\mathbb{R}^{2p-1})\), let \(P_\varepsilon(D)\) be the probability of the event \(\{\varepsilon \in D\}\). The econometrician want to identify the duple \((\theta, P_\varepsilon)\) from the following two conditions. First, from the Assumption 1. (v), \(\varepsilon\) and \(Z\) are independently distributed. From this independence condition, the conditional distribution \(P_\varepsilon|W,Z\) must satisfy

\[ \int_{w \in W} P_\varepsilon|W,Z(D|w, z; \theta) dP_W^0|Z(w|z) = P_\varepsilon(D). \quad (5.1) \]
When $W$ is discrete as in the multigraph framework,

$$
\sum_{w \in W} P_{\varepsilon|W,Z}(D|w, z; \theta) \Pr(W = w|Z = z; F^0) = P_\varepsilon(D).
$$

(5.2)

Second, all duples $(\theta, P_\varepsilon)$ in the identification region must satisfy the observational equivalence. This condition can be written as follows: For any value of $w$ and $z$, a value of $\theta$ and the distribution of $\varepsilon$ in the identification region must yield $Y = y$ if and only if $\varepsilon \in Q_\theta(y, w, z)$. In other words, given the empirical evidence alone, the econometrician can not rule out any duples $(\theta, P_\varepsilon)$ that yield $Y = y$ when $\varepsilon \in Q_\theta(y, w, z)$. More formally, for all $y, w$ and $z$,

$$
P_{\varepsilon|W,Z}(Q_\theta(y, w, z)|w, z; \theta) = \Pr(Y = y|W = w, Z = z; F^0].
$$

(5.3)

The above two conditions- independence and observational equivalence- generate a system of inequalities which all admissible duples $(\theta, P_\varepsilon)$ must satisfy. Let $H[\theta, P_\varepsilon]$ be the sharp identification region of $(\theta, P_\varepsilon)$. Then, by combining (5.1) or (5.2) and (5.3), for any closed set $D \in C(\mathbb{R}^{2p-1})$, all admissible duples $(\theta, P_\varepsilon)$ in the sharp identification region $H[\theta, P_\varepsilon]$ satisfy the following inequality.

$$
P_{\varepsilon|W,Z}(D|w, z; \theta) \geq \sum_{y \in Y} \mathbb{1}[Q_\theta(y, w, z) \subseteq D] \Pr(Y = y|W = w, Z = z; F^0),
$$

(5.4)

for all $w$ and $z$. To see why the inequality (5.4) holds, consider an arbitrary closed set $D \in C(\mathbb{R}^{2p-1})$. The smallest value that $P_{\varepsilon|W,Z}(D|w, z; \theta)$ can obtain is equal to the sum of the probabilities $\Pr(Y = y|W = w, Z = z; F^0]$ corresponding to all sets $Q_\theta(y, w, z)$ contained entirely within $D$. If a particular duple of parameter value $\theta$ and a conditional distribution $P_{\varepsilon|W,Z}$ provides $P_{\varepsilon|W,Z}(D|w, z; \theta)$ smaller than the sum of the probabilities, then the duple violates observational equivalence. I marginalize both sides of the inequality (5.4) with respect to $W$ as in (5.1) or (5.2). By the statistical independence of $\varepsilon$ and $Z$, I obtain when $W$ is continuous,

$$
P_\varepsilon(D) \geq \int_{w \in W} \left( \sum_{y \in Y} \mathbb{1}[Q_\theta(y, w, z) \subseteq D] \Pr(Y = y|W = w, Z = z; F^0] \right) F_{W|Z}(w|z),
$$

When $W$ is discrete, I have

$$
P_\varepsilon(D) \geq \sum_{w \in W} \left( \sum_{y \in Y} \left\{ \mathbb{1}[Q_\theta(y, w, z) \subseteq D] \Pr(Y = y|W = w, Z = z; F^0] \right\} \right) \times \Pr(W = w|Z = z; F^0)\right).
$$

(5.5)
Now I define a random closed set as

\[ Q_\theta(Y, W, z) \equiv \{ \varepsilon : g(W, z, \varepsilon; \theta) = Y \}. \]

The advantage of using random sets is that the above inequality (5.5) can be written concisely with the containment functional and the capacity functional. See Appendix A.4. for the definitions of a random closed set, the containment and capacity functionals. The containment functional and the capacity functional are, respectively,

\[
\begin{align*}
\Pr[Q_\theta(Y, W, z) \subseteq D; F^0] & = \sum_{w \in W} \left( \sum_{y \in Y} \{1|Q_\theta(y, w, z) \subseteq D| \Pr[Y = y|W = w, Z = z; F^0] \right) \\
& \times \Pr(W = w|Z = z; F^0). \\
\Pr[Q_\theta(Y, W, z) \cap D \neq \emptyset; F^0] & = \sum_{w \in W} \left( \sum_{y \in Y} \{1|Q_\theta(y, w, z) \cap D \neq \emptyset| \Pr[Y = y|W = w, Z = z; F^0] \right) \\
& \times \Pr(W = w|Z = z; F^0). 
\end{align*}
\]

With the above inequalities, the sharp identification region of admissible duples \((\theta, P_\varepsilon)\) associated with \(F^0_{Y|W}\) is defined below. I characterize the sharp identification region as in BMM12 and CRS.

**Proposition 3.** Let Assumption 1 hold. Then, the sharp identification region for \((\theta, P_\varepsilon)\) associated with \(F^0_{Y|W}\) is given by

\[
H[(\theta, P_\varepsilon)] = \left\{ (\theta, P_\varepsilon) \in \Theta \times \mathcal{P} | P_\varepsilon(D) \geq \Pr[Q_\theta(Y, W, z) \subseteq D; F^0], \forall D \in \mathcal{C}(\mathbb{R}^{p-1}) \text{ a.e. } z \in \mathcal{Z} \right\}. 
\]

**Proof.** See Appendix A.4. \(\square\)

One may already notice that it is infeasible to check the inequalities in (5.6) for all closed subsets in \(\mathbb{R}^{2p-1}\). However, I only need to consider the inequality for a finite number of sets \(S\). One naive way to construct such a class of sets is as follows. Since \((Y, W, Z)\) takes only a finite number of values, I construct a class \(\mathcal{C}_1(\theta) = \{Q_\theta(y, w, z) : y \in \mathcal{Y}, w \in \mathcal{W}, \text{ given } z \in \mathcal{Z} \}.\) This class, however, does not take into account all regions of unobservables which provide the sharp identification information. Hence, I consider a bigger class that is all possible unions of
$Q_θ(y, w, z)$'s. Let this power set of $Q_θ(y, w, z)$'s be $C_2(θ)$. The sharp identification region is obtained by considering all closed sets in $C(θ)$. However, it is not the smallest class. The smallest class is defined as the core determining class by GH. The core determining class significantly reduces the number of sets for which the set of moment inequalities characterize the sharp identification region. In order to obtain the core determining class, the following sets will be eliminated among the sets in $C_2(θ)$: the entire set, the empty set, non-connected sets and duplicated sets. Proposition 4, which is similar to Theorem 2 in CRS, provides a way of finding the core determining class.

**Definition 3.** (Molchanov and Molinari; 2013) A family of closed sets $ℳ$ is said to be a core determining class for a random closed set $Q$ if any probability measure $μ$ satisfying the inequalities

$$μ(F) ≥ P(Q ⊂ F)$$

for all $F ∈ ℳ$, is the distribution of a selection of $Q$.

**Proposition 4.** Let Assumption 1 hold. Define $C_1(θ) = \{Q_θ(y, w, z) : y ∈ Y, w ∈ W, and z ∈ Z\}$. Then all connected unions of sets in $C_1(θ)$ except $R_{2p−1}$ yield the core determining class $ℳ(θ)$.

**Proof.** See Appendix A.5.

Then, the sharp identification region is written as

$$H[(θ, P_ε)] = \{(θ, P_ε) ∈ Θ × ℙ | P_ε(D) ≥ Pr[Q_θ(Y, W, z) ⊆ D; F_0], \forall D ∈ ℳ(θ) a.e. z ∈ Z\}. \quad (5.7)$$

This size reduction is important. Suppose that the cardinality of $Z$ is $K$. There are $2^{(p+q)}$ possible values of $(Y_{ij}, W_{ij})$ since $p$ and $q$ are the number of simultaneous and endogenous networks, respectively. Then, the econometrician should consider the power set of those $2^{(p+q)}$ sets, which is of cardinality $2^{2(p+q)}$. Even with rather small $p$ and $q$, the power set can be very large. For example, let $p = 2$ and $q = 1$. Then, there are a total $2^8 = 256$ sets in $C_2(θ)$. Even if I exclude the entire set and the empty set, there are still 254 sets. The cardinality of $Z$ is often very large, since it is the number of all possible values of exogenous variables. For example, if one includes five binary variables in $Z$, then $K = 32$. In that case, there are $254 × 32 = 8128$ moment inequalities, and computation becomes burdensome. Furthermore, some sets in $C_2(θ)$ are duplicated, so that the corresponding moments are perfectly collinear. This collinearity may prevent the use of existing inference methods for moment inequality models. In my empirical application, I find that the core determining class $ℳ(θ)$ has a total of 36 sets as opposed to 254. Note that it is more than a 85% reduction.
5.3 Estimation Methods

Estimation of partially identified models has been studied extensively in the recent econometrics literature, for example, Manski and Tamer (2002), Imbens and Manski (2004), Chernozhukov, Hong and Tamer (2007), Beresteanu and Molinari (2008), Romano and Shaikh (2008), Rosen (2008), Stoye (2009), AS and Andrews and Barwick (2012). I adopt the estimation method developed by AS. In this section I explain how the structural model in this paper fits into the framework of AS.

The moment inequalities in the model are written as

$$E_{F^0}[m(Y_{ij}, W_{ij}, Z_{ij}; \theta)] \geq 0,$$

where

$$m(y, w, z; \theta) = (m_{1,1}(y, w, z; \theta), \cdots, m_{1,L}(y, w, z; \theta), m_{2,L}(y, w, z; \theta), \cdots, m_{K,L}(y, w, z; \theta))^\prime$$

is the $K \times L$ dimensional vector of moments. Note that $K$ is the cardinality of $Z$, and $L$ is the cardinality of the core determining class $C(\theta)$. I impose the following assumptions.

**Assumption 2.** (i) $\theta \in \Theta \subset \mathbb{R}^d$, where $d$ is the dimension of $\theta$,  
(ii) \{$(Y_{ij}, W_{ij}, Z_{ij} : \forall ij)$\} are i.i.d. under $F^0$,  
(iii) $\sigma_{F,k,l}^2(\theta) = \text{var}_F(m_{k,l}(Y_{ij}, W_{ij}, Z_{ij}; \theta)) \in (0, \infty)$ for $k = 1, \cdots, K$ and $l = 1, \cdots, L$,  
(iv) $E_F|m_{k,l}(Y_{ij}, W_{ij}, Z_{ij}; \theta)/\sigma_{F,k,l}(\theta)|^{2+\delta} < \infty$ for some $\delta > 0$ for all $k = 1, \cdots, K$ and $l = 1, \cdots, L$.  
(v) $\varepsilon \sim N(0, \Sigma_{\varepsilon})$ and i.i.d. across pairs.

Note that Assumption 2. (i), (iii) and (iv) are not restrictive, and that Assumption 2. (ii) holds due to the myopic agent assumption in PSM. Assumption 2. (vi) is added to the zero conditional median assumption in Assumption 1. Hence, the distribution of the unobservable is known up to a finite vector of parameters, and this satisfies the framework of AS. Under Assumption 2, I employ the method of AS for estimating the structural parameter $\theta$ in the model.\(^{15}\) Their method can be applied to models where parameters are defined by moment inequalities and/or equalities. It does not require point identification of parameters. I will explain detailed estimation procedures corresponding to the empirical application in Section 7 and Appendix C. One can also find the estimation procedures and asymptotic properties of the estimator in AS.

\(^{15}\)Actually, I estimate both $\theta$ and $P_\varepsilon$, especially the variance-covariance matrix $\Sigma_\varepsilon$ of $\varepsilon$. I assume that the distribution of $\varepsilon$ belongs to a parametric family. For the sake of exposition, $\Sigma_\varepsilon$ is subsumed into $\theta$ at least in this section.
6 Empirical Application

6.1 Village Networks: Risk Sharing, Kerosene-rice, Friendship, and Kinship Networks

It has been recognized in the literature that individuals share risk within a village through an underlying social network. Fafchamps (1992) describes that social networks play an important role in informal risk sharing. Fafchamps and Lund (2003) show that risk sharing is limited by the extent of a social network. De Weerdt (2002), Fafchamps and Gubert (2007) and Kinnan and Townsend (2012) find some evidence on the positive effects of social proximity, mostly kinship ties, on the formation of risk sharing relationships. For the role of friendship, De Weerdt (2002) shows that the number of mutual friends is one of the determinants of the formation of a risk sharing network. To my knowledge, however, no papers empirically investigate the effect of a friendship tie between two individuals on the formation of their risk sharing partnership. Also, endogeneity of friendship has not been controlled for in the literature. Hence, my first goal in the empirical study is to investigate the effects of friendship on the formation of a risk sharing network by allowing for endogeneity of friendship. My conjecture is that an underlying social network plays a substantial role in the formation of a risk sharing network even after controlling for its endogeneity. This is because friendship can facilitate the formation of a risk sharing partnership by providing low costs of maintaining a risk sharing partnership. The low costs may include implicit costs such as low monitoring costs and easy punishment as well as low transaction costs.

The second empirical question focuses on interactions between risk sharing and other favor exchange relationships such as gift exchange, borrowing or lending rice, medical help, etc. One reasonable conjecture may be that villagers exchange different types of favors with the same set of people. For example individual $i$ may borrow money from $j$ who provides a medical help to $i$. On the other hand, it is also possible that individuals want to diversify their favor exchange partners by the types of favors. One may want to borrow money from older individuals but exchange gifts with individuals with similar ages. Thus, I am interested in whether villagers are more inclined to aggregate different types of favor exchange relationships, or more inclined to diversify them. Note that the proposed structural model is well-suited for this problem. The term $\delta^{(s,t)}(W, V, X)$ captures whether one of the two opposite directional incentives dominates the other, although I cannot separately identify each of them.

Finally, I investigate the role of caste in the formation of risk sharing and favor exchange networks. Estimation results from a single network formation model indicate that individuals have strong caste homophily when forming a risk sharing network. See Kim (2013). However, I conjecture that caste affects risk sharing network formation only indirectly through an underlying friendship
network. Hence, if I allow for the endogeneity of friendship, the previous result may change. Economic theory predicts that individuals want to share their risk with those who have different characteristics, especially occupation, age, wealth, etc. It is reasonable to consider caste to belong to this type of variables. On the other hand, economic theory also predicts that individuals have homophily when risk sharing, since social proximity facilitates risk sharing. In the data, friendship is correlated with caste difference, and thus without controlling for friendship, the coefficient on caste suffers from selection bias. However, the endogeneity of friendship has precluded including it into empirical models in the literature. The structural model in this paper not only controls for friendship but also allows for its endogeneity. Therefore, I can consistently estimate the role of caste in the formation of a risk sharing network. Also, a similar argument applies to a favor exchange network.

6.2 Data

I use the data set of “Social Networks and Microfinance” collected by Abhijit Banerjee, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. They collect the data from 75 villages in rural Karnataka, a state in South West India. According to the data description in Banerjee et al. (2012) and Jackson, Rodriguez-Barraquer and Tan (2012), the average population per village is about 900, and over a half of households were surveyed. Also, the eligible members and their spouse in each household are surveyed. There are a total of 14 social networks; (1) Close non-relatives, (2) Close relatives, (3) Visit-go, (4) Visit-come, (5) Borrow money from, (6) Lend money to, (7) Give advice, (8) Ask for advice, (9) Borrow kerosene or rice from, (10) Lend kerosene or rice to, (11) Temple-company, (12) Medical-help, (13) Intersection of relationships, (14) Union of relationships. In the data set both household and individual-level networks are available, but I focus on the individual-level networks only. I use the ‘close non-relative (1)’ network as a friendship network, and combine the networks (5) and (6) above to construct the risk sharing network (RS). For the other favor exchange network, I combine the networks (9) and (10) to have ‘kerosene-rice network (KR)’. ‘Combining’ means that \(i\) and \(j\) have a link in a new network if they form a link in one or both of the original networks. I use (2) as a kinship network.

Although the data set contains several individual characteristics, I only use the difference in caste between each pair. The reason is mostly computational efficiency. As I include more explanatory variables, the number of moment inequalities grows rapidly. For example, if I include one more binary explanatory variable, the number of moment inequalities doubles. After including the caste difference, I have 1296 (= 36 \(\times\) 36) moment inequalities. Note that I already have three other exogenous variables in \(Z\): kinship network, the number of mutual risk sharing partners, and the
number of mutual favor exchange partners. Since I am interested in how individuals care about the difference in social classes when forming risk sharing and favor exchange networks, I consider only villages that have the proportion of its majority caste less than 95%. 54 villages are qualified under this criterion.16

The final data set contains a total of 13,096 individuals within the 54 villages. Technically, 85,746,060 pairs are possible among the 13,096 individuals. However, I do not consider inter-village pairs. Although individuals may have a relationship with those who live in different villages, the data set does not contain information about inter-village links. Due to the data unavailability, I consider only intra-village pairs, and there are a total of 1,733,961 such pairs.

Another issue in the data is that all networks are very sparse. For example, only 1.74% of pairs of all individuals are linked as risk sharing partners. Based on the size of population of each village (900 on average), it is unrealistic to consider all non-connected pairs as an outcome of strategic decisions. Many pairs of individuals are not connected maybe because they have not met at all. However, I do not know the fraction of pairs that have met before. In order to deal with this problem, I use the geodesic distance, or the shortest path length between two individuals in network (14)- the network of union of relationships - as a measure for meeting opportunity. I use four sets of pairs, which have respectively the geodesic distance less than or equal to one, two, three, and infinity. Table 1 shows the distribution of the geodesic distance in the data. Inevitably, classifying pairs as having met before or not by their geodesic distance is somewhat arbitrary. To gauge this arbitrarity, I currently pursue a sensitivity analysis.

I provide several descriptive statistics which show the fractions of pairs that form different combinations of relationships. Tables 2-4 focus on two relationships, and Table 5 on more than three relationships. From Table 2, it is apparent that more than one third of risk sharing partnerships and favor (kerosene or rice) exchange partnerships respectively are formed without friendship. However, it may be because of kinship which has been considered as one of the most important determinant of risk sharing in the literature. After excluding those pairs who build their risk sharing and/or favor exchange on a kinship tie, there are still 14,542 pairs. These 14,542 pairs form a risk sharing or favor exchange relationships with neither friends nor relatives. See row 2, 3, and 6 in Table 5. 17

<table>
<thead>
<tr>
<th>Geodesic Distance</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Pairs</td>
<td>60,426</td>
<td>342,349</td>
<td>804,560</td>
<td>461,196</td>
<td>65,430</td>
</tr>
<tr>
<td>%</td>
<td>3.48%</td>
<td>19.74%</td>
<td>46.40%</td>
<td>26.60%</td>
<td>2.77%</td>
</tr>
</tbody>
</table>

Table 1: The Distribution of Geodesic Distance

---

16The list of villages is as follows: 3–5, 9, 14, 15, 17, 21, 24–26, 28–30, 32, 34–36, 38–40, 42, 44–55, and 58–77.

17Note that there are still many pairs that are considered as both friends and relatives although the questionnaires are designed to avoid such answers.
Table 3 shows that about 43% of pairs form only one of the two relationships—RS and KR. From these results, I conjecture that individuals may have an incentive to diversify their favor exchange partners.

<table>
<thead>
<tr>
<th>Risk Sharing \ Friendship</th>
<th>0</th>
<th>1</th>
<th>Kerosene-Rice \ Friendship</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1,695,066</td>
<td>8,718</td>
<td>0</td>
<td>1,691,909</td>
<td>9,390</td>
</tr>
<tr>
<td>1</td>
<td>10,406</td>
<td>19,771</td>
<td>1</td>
<td>13,563</td>
<td>19,099</td>
</tr>
</tbody>
</table>

Table 2: Descriptive Statistics: Number of Pairs across Two Networks (1)

<table>
<thead>
<tr>
<th>Risk Sharing \ Kerosene-Rice</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1,693,817</td>
<td>9,967</td>
</tr>
<tr>
<td>1</td>
<td>7,482</td>
<td>22,695</td>
</tr>
</tbody>
</table>

Table 3: Descriptive Statistics: Number of Pairs across Two Networks (2)

<table>
<thead>
<tr>
<th>Risk Sharing \ Kinship</th>
<th>0</th>
<th>1</th>
<th>Kerosene-Rice \ Kinship</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1,695,812</td>
<td>7,972</td>
<td>0</td>
<td>1,694,335</td>
<td>6,964</td>
</tr>
<tr>
<td>1</td>
<td>14,701</td>
<td>15,476</td>
<td>1</td>
<td>16,178</td>
<td>16,484</td>
</tr>
</tbody>
</table>

Table 4: Descriptive Statistics: Number of Pairs across Two Networks (3)

6.3 Multinomial Probit without Controlling for Endogeneity

In the empirical application, I use a parametric form for individual $i$’s base utility from $j$ by having a type-$s$ relationship, or $u_{ij}^{(s)}(W, V, X; \beta^{(s)})$. That is,

$$u_{ij}^{(s)}(W, V, X; \beta^{(s)}) = \beta_0^{(s)} + \sum_{k=1}^{\kappa_x} \beta_{1,k}^{(s)}(x_{k,i} - x_{k,j})^2 + \sum_{t=p+1}^{S} \beta_{2,t-p}^{(s)}a_{ij}^{(t)}.$$ (6.1)

Hence, the base utility $u_{ij}^{(s)}$ depends on the difference in characteristics between $i$ and $j$. The coefficient $\beta_{1,k}^{(s)}$ captures the homophily effects of the $k$th characteristics, and $\beta_{2,t-p}^{(s)}$ captures the effects of the $t$th endogenous or exogenous network.

For computational purposes, I assume that the additional benefit $\delta^{(s,t)}(W, V, X)$ of having both relationships is homogeneous across pairs and does not depend on either $W$ or $V$. That is,

$$\delta^{(s,t)}(W, V, X) = \delta^{(s,t)}.$$ (6.2)
<table>
<thead>
<tr>
<th>Types</th>
<th>Number of Pairs</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Sharing Only</td>
<td>4,267</td>
<td>0.25</td>
</tr>
<tr>
<td>Kerosene-Rice Only</td>
<td>6,415</td>
<td>0.37</td>
</tr>
<tr>
<td>Friendship Only</td>
<td>6,786</td>
<td>0.39</td>
</tr>
<tr>
<td>Kinship Only</td>
<td>6,204</td>
<td>0.36</td>
</tr>
<tr>
<td>Risk Sharing, Kerosene-Rice Only</td>
<td>3,859</td>
<td>0.22</td>
</tr>
<tr>
<td>Risk Sharing, Friendship Only</td>
<td>2,507</td>
<td>0.14</td>
</tr>
<tr>
<td>Risk Sharing, Kinship Only</td>
<td>663</td>
<td>0.04</td>
</tr>
<tr>
<td>Kerosene-Rice, Friendship Only</td>
<td>1,836</td>
<td>0.11</td>
</tr>
<tr>
<td>Kerosene-Rice, Kinship Only</td>
<td>1,672</td>
<td>0.10</td>
</tr>
<tr>
<td>Friendship, Kinship Only</td>
<td>97</td>
<td>0.01</td>
</tr>
<tr>
<td>Risk Sharing, Kerosene-Rice, Friendship Only</td>
<td>4,068</td>
<td>0.23</td>
</tr>
<tr>
<td>Risk Sharing, Kerosene-Rice, Kinship Only</td>
<td>1,617</td>
<td>0.09</td>
</tr>
<tr>
<td>Kerosene-Rice, Friendship, Kinship Only</td>
<td>44</td>
<td>0.00</td>
</tr>
<tr>
<td>Risk Sharing, Kerosene-Rice, Friendship, Kinship Only</td>
<td>13,151</td>
<td>0.76</td>
</tr>
<tr>
<td>None</td>
<td>1,680,775</td>
<td>96.93</td>
</tr>
</tbody>
</table>

Table 5: Descriptive Statistics: Number of Pairs with Multiple Relations

The number of mutual partners in type-$s$ network, $h_{ij}^{(s)}$, is simplified as

$$h_{ij}^{(s)} = \begin{cases} 
0, & \text{if } \sum_k a_{ik}^{(s)} a_{jk}^{(s)} = 0, \\
1, & \text{if } \sum_k a_{ik}^{(s)} a_{jk}^{(s)} = 1, \\
2, & \text{otherwise.} 
\end{cases}$$

By plugging equation (6.2) into equation (2.4), individual $i$’s utility from $A_1$ (risk sharing) and $A_2$ (kerosene-rice) is written as

$$U_i(Y|W, V, X; \varepsilon; \theta) = \sum_{j=1}^{N} a_{ij}^{(1)} u_{ij}^{(1)} (W, V, X; \beta^{(1)}) + h_{ij}^{(1)} \gamma^{(1)} + \sum_{j=1}^{N} a_{ij}^{(2)} u_{ij}^{(2)} (W, V, X; \beta^{(2)}) + h_{ij}^{(2)} \gamma^{(2)} + \sum_{j=1}^{N} a_{ij}^{(1)} a_{ij}^{(2)} \delta^{(1,2)} + \varepsilon_{ij} (\{a_{ij}^{(1)}, a_{ij}^{(2)}\}) \right), \tag{6.3}$$

I use the following variables from the data.

$$Y_{ij} = (a_{ij}^{(RS)}, a_{ij}^{(KR)}), \text{ risk sharing and kerosene-rice exchange partnership between } i \text{ and } j.$$  
$$W_{ij} = a_{ij}^{(FR)}, \text{ friendship tie between } i \text{ and } j.$$  
$$V_{ij} = a_{ij}^{(REL)}, \text{ kinship tie between } i \text{ and } j.$$  
$$X_{ij} = (caste_i - caste_j)^2, \text{ caste difference between } i \text{ and } j.$$
where the variable \( \text{caste}_i \) takes one if \( i \) belongs to the general caste, and zero otherwise. Since I only include caste difference between \( i \) and \( j \), the base utility \( u_{ij}^{(s)} \) from a type-\( s \) link is

\[
u_{ij}^{(s)}(W, V, X; \beta^{(s)}) = \beta_0^{(s)} + \beta_1^{(s)}(\text{caste}_i - \text{caste}_j)^2 + \beta_2^{(s)}(\text{FR})_i + \beta_2^{(s)}(\text{REL})_i.
\]

Under this utility specification, the average of marginal utilities of \( i \) and \( j \) from forming a type-1 link, when type-2 link is not present, is

\[
\frac{1}{2} \{ \mu_{i|0}^{(1)} + \mu_{j|0}^{(1)} \} = \beta_0^{(1)} + \beta_1^{(1)}(\text{caste}_i - \text{caste}_j)^2 + \beta_2^{(1)}(\text{FR})_i + \beta_2^{(1)}(\text{REL})_i + \gamma_1^{(1)} h_{ij}^{(1)} + \gamma_2^{(1)} h_{ij}^{(2)} + \varepsilon_{ij}(1,0) - \varepsilon_{ij}(0,0).
\] (6.4)

When type-2 link is present, the average of marginal utilities is

\[
\frac{1}{2} \{ \mu_{i|1}^{(1)} + \mu_{j|1}^{(1)} \} = \beta_0^{(1)} + \beta_1^{(1)}(\text{caste}_i - \text{caste}_j)^2 + \beta_2^{(1)}(\text{FR})_i + \beta_2^{(1)}(\text{REL})_i + \gamma_1^{(1)} h_{ij}^{(1)} + \gamma_2^{(1)} h_{ij}^{(2)} + \delta^{(1,2)} + \varepsilon_{ij}(1,1) - \varepsilon_{ij}(0,1).
\] (6.5)

Since only an order among the sum of marginal utilities across alternatives matters, I use the average of marginal utilities (6.4) when conducting inference.

Before implementing estimation of the structural model with endogenous friendship, I run an unordered multinomial probit model with two different specifications. In the first specification, I do not include friendship into the model. I also consider another specification that assumes strictly exogenous friendship. Later I compare those two multinomial probit results with the estimation results from the structural model with endogenous friendship. Tables 6 and 7 show the results from the multinomial probit model.

**Multinomial Probit without Friendship**

First, I run a multinomial probit model where friendship tie \( a_{ij}^{(FR)} \) is not included. Table 6 shows the results. Interestingly, the coefficient on the variable ‘caste difference’, or \( (\text{caste}_i - \text{caste}_j)^2 \) is positive for the risk sharing network, but negative for the favor (kerosene-rice) exchange network. That is, individuals have caste homophily when exchange kerosene and/or rice, but they prefer different caste for risk sharing. The coefficient on caste difference for having both relationships is negative, so individuals have caste homophily when forming both relationships together.

I find that the coefficient of the variable ‘relative’, or \( a_{ij}^{(REL)} \) is negative for having a single relationship (risk sharing only or favor exchange only), but positive for having both relationships. The former result is somewhat counter-intuitive, since one may think that kinship is one of the most
important determinants for forming each relationship. However, I conjecture that if two individuals have a kinship, then they may be willing to share many different types of favors with each other rather than only a single type of favor. The number of mutual partners between two individuals in one relation is a strong determinant of having the relation, but relatively weak for having the other relation.

**Multinomial Probit under the Assumption of Exogenous Friendship**

Next, I run a multinomial probit model which includes friendship as an exogenous variable. Although the sign and the significance of each coefficient included in the previous multinomial probit model remain the same, the magnitude has been changed. Especially, the magnitude of the coefficients on ‘caste difference’ of favor exchange network and both relationships declined about 10% and 30%, respectively. I conjecture that these results are due to the correlation between friendship and caste difference. Interestingly, friendship affects risk sharing and favor exchange networks in the same direction as kinship does. The presence of friendship between two individuals facilitates having both relationships, but not a single relationship only.

### 6.4 Analysis Accounting for Endogeneity of Friendship

The first step for estimation of the structural model with endogeneity is to find the core determining class \( \mathcal{M}(\theta) \). I describe how to obtain the core determining class in Appendix B.

Next, I construct a set of moment inequalities corresponding to the sets in \( \mathcal{M}(\theta) \). Let \( l \) be an index for sets in the core determining class \( \mathcal{M}(\theta) \), and \( k \) be an index for \( z \). For each \( z_k \) and \( S_l \), I have

\[
E(m_{k,l}(y, w, z; \theta) \mid z = z_k) = P_{\epsilon}[Q_{\theta}(y, w, z_k) \subseteq D_l] - \sum_{w' \in \mathcal{W}} \sum_{y' \in \mathcal{Y}} (1[Q_{\theta}(y', w', z_k) \subseteq D_l] \times \Pr [Y = y' \mid W = w', Z_{ij} = z_k; F_0]) \Pr [W = w' \mid Z = z_k; F_0].
\]

The unconditional moment \( E(m_{k,j}(y_{ij}, w_{ij}, z_{ij}; \theta)) \) is written as

\[
E(m_{k,j}(y, w, z; \theta)) = \left\{ P_{\epsilon}[Q_{\theta}(y, w, z_k) \subseteq D_l] - \sum_{w' \in \mathcal{W}} \sum_{y' \in \mathcal{Y}} (1[Q_{\theta}(y', w', z_k) \subseteq D_l] \times \Pr [Y_{ij} = y' \mid W_{ij} = w', Z_{ij} = z_k; F_0]) \Pr [W_{ij} = w' \mid Z_{ij} = z_k; F_0] \right\} \times \Pr [Z_{ij} = z_k; F_0].
\]
<table>
<thead>
<tr>
<th>Type of Relationships</th>
<th>Variables</th>
<th>Coef.</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Sharing Network</td>
<td>Caste Difference</td>
<td>0.1297</td>
<td>[0.0783, 0.1811]</td>
</tr>
<tr>
<td></td>
<td>Relative</td>
<td>-0.9451</td>
<td>[-0.9966, -0.8936]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Risk Sharing Partners</td>
<td>0.2585</td>
<td>[0.2291, 0.2878]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Favor Exchange Partners</td>
<td>0.0556</td>
<td>[0.0276, 0.0836]</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>-0.7673</td>
<td>[-0.7981, -0.7364]</td>
</tr>
<tr>
<td>Favor Exchange Network (Kerosene-Rice)</td>
<td>Caste Difference</td>
<td>-0.1870</td>
<td>[-0.2408, -0.1331]</td>
</tr>
<tr>
<td></td>
<td>Relative</td>
<td>-0.8507</td>
<td>[-0.8945, -0.8069]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Risk Sharing Partners</td>
<td>0.1043</td>
<td>[0.0772, 0.1313]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Favor Exchange Partners</td>
<td>0.6091</td>
<td>[0.5828, 0.6353]</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>-0.9394</td>
<td>[-0.9708, -0.9080]</td>
</tr>
<tr>
<td>Both Risk Sharing and Favor Exchange</td>
<td>Caste Difference</td>
<td>-0.1537</td>
<td>[-0.2086, -0.0987]</td>
</tr>
<tr>
<td></td>
<td>Relative</td>
<td>0.4413</td>
<td>[0.4040, 0.4787]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Risk Sharing Partners</td>
<td>0.5712</td>
<td>[0.5456, 0.5968]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Favor Exchange Partners</td>
<td>0.6473</td>
<td>[0.6219, 0.6727]</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>-1.4135</td>
<td>[-1.4455, -1.3814]</td>
</tr>
</tbody>
</table>

Table 6: Estimation Results from Multinomial Probit without Friendship
<table>
<thead>
<tr>
<th>Type of Relationships</th>
<th>Variables</th>
<th>Coef.</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risk Sharing Network</strong></td>
<td>Friendship</td>
<td>-0.0457</td>
<td>[-0.0863, -0.0051]</td>
</tr>
<tr>
<td></td>
<td>Caste Difference</td>
<td>0.1290</td>
<td>[0.0775, 0.1804]</td>
</tr>
<tr>
<td></td>
<td>Relative</td>
<td>-0.9409</td>
<td>[-0.9936, -0.8882]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Risk Sharing Partners</td>
<td>0.2321</td>
<td>[0.2027, 0.2616]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Favor Exchange Partners</td>
<td>0.0451</td>
<td>[0.0171, 0.0731]</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>-0.7327</td>
<td>[-0.7678, -0.6976]</td>
</tr>
<tr>
<td><strong>Favor Exchange Network (Kerosene-Rice)</strong></td>
<td>Friendship</td>
<td>-0.5148</td>
<td>[-0.5555, -0.4741]</td>
</tr>
<tr>
<td></td>
<td>Caste Difference</td>
<td>-0.1703</td>
<td>[-0.2250, -0.1156]</td>
</tr>
<tr>
<td></td>
<td>Relative</td>
<td>-0.8219</td>
<td>[-0.8666, -0.7771]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Risk Sharing Partners</td>
<td>0.1087</td>
<td>[0.0813, 0.1361]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Favor Exchange Partners</td>
<td>0.5852</td>
<td>[0.5588, 0.6115]</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>-0.7839</td>
<td>[-0.8185, -0.7493]</td>
</tr>
<tr>
<td><strong>Both Risk Sharing and Favor Exchange</strong></td>
<td>Friendship</td>
<td>1.183</td>
<td>[1.1474, 1.2185]</td>
</tr>
<tr>
<td></td>
<td>Caste Difference</td>
<td>-0.2255</td>
<td>[-0.2814, -0.1695]</td>
</tr>
<tr>
<td></td>
<td>Relative</td>
<td>0.4557</td>
<td>[0.4161, 0.4962]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Risk Sharing Partners</td>
<td>0.4456</td>
<td>[0.4189, 0.4723]</td>
</tr>
<tr>
<td></td>
<td>Number of Mutual Favor Exchange Partners</td>
<td>0.6628</td>
<td>[0.6366, 0.6890]</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>-1.8743</td>
<td>[-1.9105, -1.8381]</td>
</tr>
</tbody>
</table>

Table 7: Estimation Results from Multinomial Probit with Friendship
The corresponding sample moment for a pair \(ij\) is

\[
m_{ij,k,l}(y, w, z; \theta) = \left\{ \begin{array}{cl}
1[Q_\theta(y_{ij}, w_{ij}, z_{ij}) \subseteq D_1] - \sum_{w \in \mathcal{W}} \sum_{y' \in \mathcal{Y}} (1[Q_\theta(y', w', z_{ij}) \subseteq D_1]) \\
\times 1[Y = y'|W = w', Z = z_k; F^0][W = w'|Z_{ij} = z_k; F^0] \\
\times 1[Z_{ij} = z_k; F^0].
\end{array} \right.
\]

For simplicity, I denote \(m_{ij,k,l}(\theta) = m_{k,l}(y_{ij}, w_{ij}, z_{ij}; \theta)\). I compute

\[
m_{ij}(\theta) = (m_{1,1}(\theta), \ldots, m_{L,1}(\theta), m_{1,2}(\theta), \ldots, m_{L,s_{xz}}(\theta))'.
\]

Then, I take the sample average over all pairs and get the \(K \times L\) by 1 vector of sample moments,

\[
m_{ij}(\theta) = (m_{ij,1,1}(\theta), \ldots, m_{ij,L,1}(\theta), m_{ij,1,2}(\theta), \ldots, m_{ij,L,s_{xz}}(\theta))'.
\]

I also compute the estimator \(\hat{\Sigma}_N(\theta)\) of the variance-covariance matrix of \(m_{ij}(\theta)\).

For the test statistic, I use the modified method of moments (MMM) test statistic. That is,

\[
S(\sqrt{N}\hat{m}_N(\theta), \hat{\Sigma}_N(\theta)) = \sum_{k=1}^{K} \sum_{l=1}^{L} [\sqrt{N}\hat{m}_{k,l,N}(\theta)/\hat{\sigma}_{k,l}]^2,
\]

where \(\hat{\sigma}_{k,l}\) is the \(((k-1)L+k)\)th element of \(\text{Diag}(\hat{\Sigma}_N(\theta))^{1/2}\).

The estimation procedure is summarized below. (i) Draw \(R = 100\) unobservables for each observation (\(\hat{N}\) by 300 matrix) from \(N(0, I_3)\). These draws stay fixed over all repetitions. (ii) Draw the 18 by 1 parameter vector \(\theta_0\) from the uniform distribution on \([-10, 10]\)^{18}. The parameter vector includes five elements of a three by three matrix \(\text{var}(\varepsilon)\). It is not six but five due to the scale normalization, i.e. \(\text{var}(\varepsilon_1) = \sigma^2 = 1\). (iii) I pre-multiply \(\varepsilon\) by \(\Lambda\), where \(\Lambda\) is the Cholesky decomposition of \(\Sigma_\varepsilon\), i.e. \(\Sigma_\varepsilon = \Lambda\Lambda'^T\). (iv) For each value of \(\theta_0\), compute the sample moments \(\hat{m}_N\) (36 \(\times\) 36 = 1296 by 1 vector) and their sample variance-covariance matrix \(\hat{\Sigma}_N\) (1296 by 1296). (v) Compute the test statistic \(S(\sqrt{N}\hat{m}_N(\theta_0), \hat{\Sigma}_N(\theta_0))\). (vi) Draw \(B\) number of nonparametric bootstrap samples. In the empirical application, I use \(B = 100\) for computational efficiency. (vii) Recenter the bootstrap samples. Compute \(M_{N,R}(\hat{\Theta}_{N,R}(\theta_0)) = \hat{N}^{1/2}(\hat{D}_{N,R}(\theta_0))^{-1/2}(\hat{m}_N^{*, R}(\theta_0) - \hat{m}_N(\theta_0))\) and \(\hat{\Omega}_{N,R}(\hat{\Theta}_{N,R}(\theta_0)) = (\hat{D}_{N,R}(\theta_0))^{-1/2}(\hat{\Sigma}_{N,R}(\theta_0))^{-1/2}(\hat{\Omega}_{N,R}(\theta_0))\), where \(\hat{D}_{N,R}(\theta_0) = \text{Diag}(\hat{\Sigma}_{N,R}(\theta_0))\). (viii) Implement the generalized moment selection (GMS) procedure: If \(\tilde{N}^{1/2}\hat{m}_{n,j}(\theta_0)/\hat{\sigma}_{n,j}(\theta_0) > \kappa_n = (\ln \tilde{N})^{1/2}\), eliminate the corresponding moments from \((M_{N,R}(\theta_0), \hat{\Omega}_{N,R}(\theta_0))\) and denote \((M^{*, R}(\theta_0), \hat{\Omega}^{*, R}(\theta_0))\). (ix) Compute \(S(M^{*, R}(\theta_0), \hat{\Omega}^{*, R}(\theta_0))\). (x) If \(S(\sqrt{N}\hat{m}_N(\theta_0), \hat{\Sigma}_N(\theta_0)) > \hat{c}_{N,1-\alpha}\), then reject \(\theta_0\), where
\hat{c}_{N,1-\alpha} is the $1-\alpha$ quantile of \{S(M^{**}_{N,b}(\theta_0), \hat{\Omega}^{**}_{N,b}(\theta_0)), b = 1, \cdots, B\}.

To find a confidence set for the identification region, I use simulated annealing with many different starting values. For each iteration, I save the value of objective function \(Q_{\hat{N}}(\theta) = S(\hat{m}_{\hat{N}}, \hat{\Sigma}_{\hat{N}}) - \hat{c}_{\hat{N},1-\alpha}\). After many iterations, I collect all parameter values that satisfy \(Q_{\hat{N}}(\theta) \leq 0\), and these values comprise the \(1-\alpha\)% confidence set. I project the confidence set for each parameter to obtain confidence intervals.\(^{18}\)

7 Estimation Results

I provide confidence intervals for parameters in the sharp identification region in Table 8. The confidence intervals are obtained by projecting the confidence set of the sharp identification region of parameters for each parameter. Table 8 shows that all confidence intervals include zero. However, this is not surprising. As shown by Chesher and Rosen (2013), in the presence of an endogenous explanatory variable, if one does not impose a structure between an instrument and an endogenous variable, the sharp identification region is composed of two separate regions whose projection covers zero. See for example, Appendix D. Nevertheless, a few interesting results emerge from the estimation exercise. First, compared to the multinomial probit model, all of confidence intervals from the structural model include the point estimates of parameters from the multinomial probit model.

Second, when I restrict the sign of \(\beta^{(1)}_{\text{friend}}\) in the risk sharing network to be positive, \(\beta^{(2)}_{\text{friend}}\) for kerosene-rice is estimated to be positive as well. Similarly, if I restrict the sign of \(\beta^{(1)}_{\text{friend}}\) to be negative, the sign of \(\beta^{(2)}_{\text{friend}}\) again is estimated to be the same as the sign of \(\beta^{(1)}_{\text{friend}}\). Although I cannot draw a strong conclusion about the effects of friendship on the formation of risk sharing and favor exchange networks, I have strong evidence that friendship affects both networks in the

\(^{18}\)Alternatively, I rely on support vector machine (SVM) approach in Bar and Molinari (2013) to obtain a confidence set and the estimated sharp identification region. Bar and Molinari (2013) use SVM to compute the estimated identification region and corresponding confidence region. I repeat the above steps for a total \(T\) number of parameter vectors drawn uniformly. For each \(t\)th draw of the parameter vector, let \(d_t = 1\) if \(\theta_t\) is not rejected, and \(-1\) otherwise. I save the results in \{(\(d_t, \theta_t\), \(t = 1, \cdots, T\)}, and this becomes the training data for an SVM. Next, I run the SVM with the package \texttt{libsvm} in R to get the 95% confidence region \(\{\theta \in \Theta : \hat{f}(\theta) = \sum_{t=1}^{T} \alpha_t d_t K(\theta, \theta_t) + \hat{\beta}_{\text{svm}} \geq 0\}\), where \(K(\cdot, \cdot)\) is the Gaussian kernel (the default option in \texttt{libsvm}). For the confidence interval of each scalar-valued component of \(\theta\), I solve the optimization problems

\[
\begin{align*}
\min_{\theta_k} \quad & \theta_k \\
\text{s.t.} \quad & \hat{f}(\theta) \geq 0,
\end{align*}
\]

and

\[
\begin{align*}
\max_{\theta_k} \quad & \theta_k \\
\text{s.t.} \quad & \hat{f}(\theta) \geq 0.
\end{align*}
\]

The results from the SVM will be provided soon.
Table 8: Projection of 95% CIs for the Sharp Identification Regions of Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Sharing</td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>[-1.1345, 1.1060]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-0.9022, 1.9584]</td>
</tr>
<tr>
<td>caste difference</td>
<td>[-1.4220, 1.0025]</td>
</tr>
<tr>
<td>kinship</td>
<td>[-1.9960, 3.7901]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-0.8286, 1.1871]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.4085, 0.4146]</td>
</tr>
<tr>
<td>Kerosene or Rice</td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>[-1.8675, 0.8126]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-0.9284, 1.7274]</td>
</tr>
<tr>
<td>caste difference</td>
<td>[-0.9869, 1.4099]</td>
</tr>
<tr>
<td>kinship</td>
<td>[-1.0308, 3.3371]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-1.4884, 1.4435]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.2236, 1.0060]</td>
</tr>
<tr>
<td>Both</td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>[-2.7208, 0.6705]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-1.5301, 3.2381]</td>
</tr>
<tr>
<td>caste</td>
<td>[-2.4089, 2.1096]</td>
</tr>
<tr>
<td>kinship</td>
<td>[-2.9427, 4.4427]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-2.3159, 1.5129]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.6189, 1.0427]</td>
</tr>
<tr>
<td>additional utility from both</td>
<td>[-0.9195, 1.0491]</td>
</tr>
</tbody>
</table>

same way. See Table 9.

Figure 1 shows another interesting result. In the computational illustrations of the sharp identification region in Appendix D, the true parameter lies in the larger region when the sharp identification region is given by two components. In Figure 1, I find that the region with positive $\beta^{(1)}_{friend}$ and $\beta^{(2)}_{friend}$ is larger than the negative region. One may therefore conclude that the effects of friendship on both risk sharing and favor exchange networks are more likely to be positive.

The role of caste in the formation of risk sharing and favor exchange network remains indeterminate. I cannot reject the hypothesis that individuals have caste homophily when forming a risk sharing network or a favor exchange network. The results are not favorable to the hypothesis, either.

I cannot find strong evidence on interactions between two simultaneously determined networks. In particular, the number of mutual partners in one network does not affect the formation of the other type of relationship. However, the results may be due to data limitations. For example, there may exists measurement errors, and/or the questionnaires may not be enough to reflect the true network, etc. Finally, I impose restrictions $\beta^{(1)}_{rel} > 0$ and $\beta^{(2)}_{rel} > 0$, and see whether there is a change in significance. I find no parameters that are significant under the restrictions. See Table 10.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta_{friends}^{(1)} &lt; 0$</th>
<th>$\beta_{friends}^{(1)} &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Sharing intercept</td>
<td>[-0.5268, 1.1060]</td>
<td>[-1.1345, 0.8782]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-0.9022, -0.0815]</td>
<td>[0.0060, 1.9584]</td>
</tr>
<tr>
<td>caste difference</td>
<td>[-1.4220, 1.0025]</td>
<td>[-1.3880, 0.9011]</td>
</tr>
<tr>
<td>kinship</td>
<td>[-0.7369, 2.0000]</td>
<td>[-1.9960, 3.7901]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-0.4923, 0.8254]</td>
<td>[-0.8286, 1.1871]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-0.4985, 0.4146]</td>
<td>[-1.4085, 0.3115]</td>
</tr>
<tr>
<td>Kerosene or Rice intercept</td>
<td>[-1.6538, 0.8126]</td>
<td>[-1.8671, 0.3222]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-0.9284, -0.0044]</td>
<td>[0.0023, 1.7274]</td>
</tr>
<tr>
<td>caste difference</td>
<td>[-0.9869, 1.4099]</td>
<td>[-0.2759, 1.1740]</td>
</tr>
<tr>
<td>kinship</td>
<td>[0.8944, 3.3371]</td>
<td>[-1.0308, 1.3696]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-0.4904, 1.2870]</td>
<td>[-1.4884, 1.4435]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.2236, 0.4173]</td>
<td>[-0.9158, 1.0060]</td>
</tr>
<tr>
<td>Both</td>
<td>[-1.1694, 0.4639]</td>
<td>[-2.7208, 0.6705]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-1.5301, -0.1048]</td>
<td>[0.1189, 3.2381]</td>
</tr>
<tr>
<td>caste</td>
<td>[-2.4089, 2.1096]</td>
<td>[-1.1835, 1.5939]</td>
</tr>
<tr>
<td>kinship</td>
<td>[0.5686, 4.4427]</td>
<td>[-2.9427, 4.2320]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-0.3088, 1.1074]</td>
<td>[-2.3159, 1.5129]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.6056, 0.5784]</td>
<td>[-1.6189, 1.0427]</td>
</tr>
</tbody>
</table>

Table 9: Projection of 95% CIs for Parameters when $\beta_{friends}^{(1)} < 0$ and $\beta_{friends}^{(1)} > 0$

Figure 1: The Projected Confidence Region for $(\beta_{friend}^{(1)}, \beta_{friend}^{(2)})$
95% CI

<table>
<thead>
<tr>
<th>Parameter</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Sharing intercept</td>
<td>[-1.1345, 1.1060]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-0.9022, 1.2465]</td>
</tr>
<tr>
<td>caste difference</td>
<td>[-1.4220, 1.0018]</td>
</tr>
<tr>
<td>relative</td>
<td>[0.5163, 3.7901]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-0.5279, 1.1871]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.1061, 0.2883]</td>
</tr>
<tr>
<td>Kerosene or Rice intercept</td>
<td>[-1.7254, 0.8126]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-0.7536, 1.4957]</td>
</tr>
<tr>
<td>caste difference</td>
<td>[-0.9869, 1.4099]</td>
</tr>
<tr>
<td>relative</td>
<td>[0.1708, 3.2353]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-0.5607, 1.2870]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.2236, 1.0060]</td>
</tr>
<tr>
<td>Both intercept</td>
<td>[-2.7208, 0.6705]</td>
</tr>
<tr>
<td>friendship</td>
<td>[-1.5301, 2.4607]</td>
</tr>
<tr>
<td>caste</td>
<td>[-2.4089, 2.1096]</td>
</tr>
<tr>
<td>relative</td>
<td>[1.0808, 4.4427]</td>
</tr>
<tr>
<td>mutual risk sharing partners</td>
<td>[-0.7800, 1.3078]</td>
</tr>
<tr>
<td>mutual favor exchange partners</td>
<td>[-1.6189, 0.9036]</td>
</tr>
</tbody>
</table>

Table 10: Projection of 95% CIs for Parameters when $\beta_{rel}^{(1)} > 0$ and $\beta_{rel}^{(2)} > 0$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>risk sharing kerosene or rice exchange</td>
<td>[-1.9516, 1.2448]</td>
</tr>
<tr>
<td>risk sharing both</td>
<td>[-1.3954, 1.2480]</td>
</tr>
<tr>
<td>kerosene or rice</td>
<td>[-1.2907, 2.8282]</td>
</tr>
<tr>
<td>both</td>
<td>[-1.5430, 6.7667]</td>
</tr>
</tbody>
</table>

Table 11: 95% Confidence Interval for Covariance Matrix of $\varepsilon$

I also estimate the variance-covariance matrix of the unobservables. See Table 11. The confidence intervals are relatively wide especially for variances. I cannot rule out independence of $\varepsilon$ across alternatives. However, I am able to reject the hypothesis that $\varepsilon$ from different alternatives have an identical distribution since $\text{var}(\varepsilon_3) = 1$ does not belong to the 95% confidence interval. The estimate for $\text{var}(\varepsilon_{ij}(1, 1))$ seems larger than $\text{var}(\varepsilon_{ij}(1, 0))$ and $\text{var}(\varepsilon_{ij}(0, 1))$.

8 Conclusion

In this paper I have studied a structural model of multigraph formation. A multigraph can be decomposed into simultaneous, endogenous and exogenous networks. I propose the notion of pairwise stability of a multigraph and show that the structural model of multigraph formation is equivalent to a multinomial choice model under PSM with myopic agents. PSM provides a relatively sim-
ple way to identify the structural parameters in the strategic formation of a multigraph. When endogenous networks are present, however, the model parameters are not point-identified. I therefore builds on the recently developed partial identification methods to characterize the parameters’ sharp identification region and conduct inference. In my empirical application, I find that friendship affects the formation of risk sharing and favor exchange networks in the same direction. However, the empirical evidence for caste homophily is inconclusive.

There are a few interesting extensions for further research. First, the structural model proposed in this paper is widely applicable to many other settings in social interaction models. For example, a network of a risky behavior, e.g. crime, sexual contact, etc., and the transformation of a network from one period to the other period fit into the framework of the model in this paper. Furthermore, the analysis of a discrete choice model with an endogenous variable and no instruments can provide a new approach for many economic applications. Second, it may be interesting to extend the analysis to the linear-in-means social network model, where peer effects are present through an underlying social network. By applying the econometric method in this paper, one can solve potential bias due to the endogeneity of a social network. Finally, the model in this paper allows the econometrician to recover utility parameters of multigraph formation, but it does not predict an entire multigraph configuration for different realizations of exogenous variables and/or networks. It is interesting to incorporate a particular multigraph formation process to the model and predict what multigraph will be formed as a counterfactual analysis.

References


**Appendix A. Proofs**

**A.1. Proof of Proposition 1**

In order to prove Lemma 1, I introduce the following definitions: adjacency, an improving path, and a maximal cycle. The following definitions extend corresponding definitions (under the same name) in Jackson and Watts (2002) to the multigraph framework.

**Definition.** (Extended definitions from Jackson and Wolinsky (1996) and Jackson and Watts (2002))

(i) (adjacent) If two multigraphs $Y$ and $Y'$ differ by only one pair’s decision $Y_{ij}$ and $Y'_{ij}$, they are adjacent.

(ii) (an improving path) An improving path from a multigraph $Y$ to a multigraph $Y'$ is a finite sequence of adjacent multigraphs $Y = Y_1, \ldots, Y_K = Y'$ such that for any $k \in 1, \ldots, K - 1$, $Y_{k+1} = Y_k - Y_{k,ij} + Y'_{k,ij}$ for some $ij$ such that $U_i(Y_{k+1}) + U_j(Y_{k+1}) > U_i(Y_k) + U_j(Y_k)$ in the case of transferable utility, and $U_i(Y_{k+1}) > U_i(Y_k)$ and $U_j(Y_{k+1}) > U_j(Y_k)$ in the case of nontransferable utility.

(iii) (a maximal cycle) A cycle $C$ is a maximal cycle if it is not a proper subset of a cycle.
Proof. (Proposition 1) Consider an arbitrary multigraph $Y_0$. If $Y_0$ is pairwise stable, the result is established. So, suppose not. Now, since $Y_0$ is not pairwise stable, it lies on an improving path. So, there is a pair that wants to deviate from $Y_0$ to its adjacent multigraph $Y_1$. Note that each pair’s deviation is uniquely determined for each $Y_0$ since there is no tie between adjacent multigraphs in terms of utility. This unique determination is a key to make the case of a multigraph similar to that of a single network. If the improving path ends at a multigraph $Y_T$ (i.e. no pairs want to deviate from $Y_T$), then $Y_T$ is a pairwise stable multigraph. Hence, the result is established. If the path does not end, it must hit the original multigraph $Y_0$ since the number of possible multigraph configurations is finite. Therefore, there exists at least one pairwise stable multigraph or a closed cycle of multigraphs. □

A.2. Proof of Proposition 2

Proof. I use the contra-positive to prove the proposition. Suppose that there is a cycle, \{ $Y = Y_0, Y_1, \ldots, Y_K = Y$ \}. For the sake of contradiction, suppose that there exists a potential function $\omega(\cdot)$. Without loss of generality, $Y_0$ is a multigraph that lies on a cycle. Then, $\omega(Y) = \omega(Y_0) < \omega(Y_1) < \cdots < \omega(Y_K) = \omega(Y)$. This makes a contradiction. Therefore, if there exists $\omega(\cdot)$, then there are no cycles, and at least one pairwise stable multigraph exists. □

A.3. Proof of Corollary 1

Proof. I prove the case of two types of links in $Y$. When there are more than two networks, I just need to check more cases. When there are two networks, I have three cases to consider: (1) $Y_{ij} = (0,0)$ and $Y_{ij}' = (1,0)$, (2) $Y_{ij} = (1,0)$ and $Y_{ij}' = (0,1)$, and (3) $Y_{ij} = (0,0)$ and $Y_{ij}' = (1,1)$, other cases are symmetric with one of those three. Define

$$\omega(Y) = \sum_{i} \sum_{j \neq i} a_{ij}^{(1)} u_{ij}^{(1)} + \sum_{i} \sum_{j} a_{ij}^{(1)} \left( \frac{1}{2} \gamma_1^{(1)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} + \gamma_2^{(1)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} \right) +$$

$$+ \sum_{i} \sum_{j \neq i} a_{ij}^{(2)} u_{ij}^{(2)} + \sum_{i} \sum_{j} a_{ij}^{(2)} \left( \frac{1}{2} \gamma_1^{(2)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} + \gamma_2^{(2)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} \right) +$$

$$+ \sum_{i=1}^{n} a_{ij}^{(1)} a_{ij}' a_{ij}^{(1,2)} (W, V, X) + \sum_{i}^{N} \varepsilon_{ij}(Y_{ij}).$$

Case 1. $Y_{ij} = (0,0)$ and $Y_{ij}' = (1,0)$

43
Individual $i$’s marginal utility of forming the first type relationship with $j$ is

$$U_i(Y_{ij} = (1,0)) - U_i(Y_{ij} = (0,0)) = u_{ij}^{(1)} + \gamma_1 a_{ik}^{(1)} a_{jk}^{(1)} + \gamma_2 a_{ik}^{(2)} a_{jk}^{(2)} + \varepsilon_{ij}(1,0) - \varepsilon_{ij}(0,0).$$

The sum of $i$ and $j$’s marginal utilities, say $mu_{ij}$, is

$$mu_{ij} = u_{ij}^{(1)} + u_{ji}^{(1)} + 2\gamma_1 a_{ik}^{(1)} a_{jk}^{(1)} + 2\gamma_2 a_{ik}^{(2)} a_{jk}^{(2)} + 2\varepsilon_{ij}(1,0) - 2\varepsilon_{ij}(0,0).$$

Now consider $\omega(Y') - \omega(Y)$.

$$\omega(Y') - \omega(Y) = u_{ij}^{(1)} + u_{ji}^{(1)} + 2 \times \frac{1}{2} \gamma_1 a_{ik}^{(1)} a_{jk}^{(1)} + 2 \gamma_2 a_{ik}^{(2)} a_{jk}^{(2)} + 2\varepsilon_{ij}(1,0) - 2\varepsilon_{ij}(0,0).$$

Note that the first two lines are the sum of the marginal utilities of $i$ and $j$ except multiplying $\frac{1}{2}$ to $\gamma_1 a_{ik}^{(1)} a_{jk}^{(1)}$. The last term is the sum of the additional utility (or disutility) that the rest of individuals receive due to the change in $i$ and $j$’s relationships. Hence, the difference in the sum of $i$ and $j$’s marginal utilities is the same as the difference in $\omega(Y)$.

Case 2. $Y_{ij} = (0,1)$ and $Y'_{ij} = (1,0)$

First,

$$U_i(Y_{ij} = (1,0)) - U_i(Y_{ij} = (0,1)) = u_{ij}^{(1)} - u_{ij}^{(2)} + \gamma_1 a_{ik}^{(1)} a_{jk}^{(1)} + \gamma_2 a_{ik}^{(2)} a_{jk}^{(2)} - \gamma_1 a_{ik}^{(1)} a_{jk}^{(1)} - \gamma_2 a_{ik}^{(2)} a_{jk}^{(2)} + \varepsilon_{ij}(1,0) - \varepsilon_{ij}(0,1).$$

$$mu_{ij} = u_{ij}^{(1)} - u_{ij}^{(2)} + u_{ji}^{(1)} - u_{ji}^{(2)} + 2\gamma_1 a_{ik}^{(1)} a_{jk}^{(1)} + 2\gamma_2 a_{ik}^{(2)} a_{jk}^{(2)} - 2\gamma_1 a_{ik}^{(1)} a_{jk}^{(1)} - 2\gamma_2 a_{ik}^{(2)} a_{jk}^{(2)} + 2\varepsilon_{ij}(1,0) - 2\varepsilon_{ij}(0,1).$$

44
Consider $\omega(Y') - \omega(Y)$.

\[
\omega(Y') - \omega(Y) = u_{ij}^{(1)} + u_{ji}^{(1)} + 2 \times \frac{1}{2} \gamma_1 \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} + 2 \gamma_2 \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} - 2 \gamma_2 \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} - 2 \times \frac{1}{2} \gamma_2 \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} + 2 \times 2 \gamma_2 \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} - 2 \gamma_2 \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} - 2 \times 2 \gamma_2 \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} + 2 \varepsilon_{ij}(1,0) - 2 \varepsilon_{ij}(0,0).
\]

Again, $\omega(Y') - \omega(Y)$ is equal to the sum of the marginal utilities of $i$ and $j$.

Case 3. $Y_{ij} = (0,0)$ and $Y'_{ij} = (1,1)$

In this case, I consider two separate cases: $Y_{ij} = (0,0)$ and $Y'_{ij} = (1,0)$, and $Y_{ij} = (0,0)$ and $Y'_{ij} = (0,1)$. By Case 1 and symmetry, $\omega(Y') - \omega(Y)$ is the same as the sum of marginal utilities, since the term $\sum_{i=1}^{n} a_{ij}^{(1)} a_{ij}^{(2)} \delta_{ij}^{(1,2)}(W, V, X)$ only affects $i$ and $j$’s utilities. The difference $\omega(Y') - \omega(Y)$ is the sum of marginal utilities of $i$ and $j$. Therefore, by Proposition 2, there exists at least one pairwise stable multigraph.

A.4. Definitions

Definitions in this appendix follow from GH, BMM12, Chesher and Rosen (2013) and Molchanov and Molinari (2013). Let $C$ be the collection of all closed sets. A random closed set is a measurable map $Q : \Omega \rightarrow C$, from the probability space to the collection of closed sets $\{F \in C : F \cap D \neq \emptyset\}$ for all $D \in K$.

**Definition.** A map $Q$ from a probability space $(\Omega, F, P)$ to $C$ is called a random closed set if

\[
Q^{-}(D) = \{\omega : Q(\omega) \cap D \neq \emptyset\} \in F
\]

for each compact set $D \subset \mathbb{R}^d$.

Let $K$ be the collection of all compact sets. The capacity functional and the containment functional are defined as follows.

**Definition.** (i) A functional $T_X(D) : K \rightarrow [0,1]$ given by

\[
T_X(D) = P(X \cap D \neq \emptyset), D \in K,
\]

is called the capacity functional of $X$. 

45
(ii) A functional \( C_X(F) : \mathcal{C} \to [0,1] \) given by

\[
C_X(F) = P(X \subset F), \quad F \in \mathcal{C},
\]

is called the containment functional of \( X \).

### A.5. Proof of Proposition 3

**Proof.** Since all admissible duples \((\theta, P_\varepsilon)\) are in \( H[(\theta, P_\varepsilon)] \), I only need to prove that a duple \((\theta', P_\varepsilon')\) which is not admissible by the model is not in \( H[(\theta, P_\varepsilon)] \). First, from Theorem 2.1. in BMM12, I obtain the equivalence between containment functional and capacity functional. Hence, I can write the sharp identification region as

\[
H[(\theta, P_\varepsilon)] = \left\{ (\theta, P_\varepsilon) \in \Theta \times \mathcal{P} | P_\varepsilon(D) \leq \Pr[Q_\theta(Y, W, z) \cap D \neq \emptyset; F_0^0], \forall D \in \mathcal{C}(\mathbb{R}^{2p-1}) \ a.e. \ z \in \mathcal{Z} \right\}.
\]

Then, the sharpness is due to the Arstein’s inequality. see Artstein (1983), Norberg (1992), or Molchanov (2005). \( \square \)

### A.6. Proof of Proposition 4

**Proof.** Recall the sharp identification region (5.6)

\[
H[(\theta, P_\varepsilon)] = \left\{ (\theta, P_\varepsilon) \in \Theta \times \mathcal{P} | P_\varepsilon(D) \geq \Pr[Q_\theta(Y, W, z) \subseteq D; F_0^0], \forall D \in \mathcal{C}(\mathbb{R}^{2p-1}) \ a.e. \ z \in \mathcal{Z} \right\}.
\]

First, I want to rule out sets that are not the union of sets in \( \mathcal{C}_1(\theta) = \{Q_\theta(y, w, z) : y \in \mathcal{Y}, w \in \mathcal{W}, \text{and } z \in \mathcal{Z}\} \). Let \( T_1 \) be an arbitrary set which is not a union of sets in \( \mathcal{C}_1(\theta) \). I can find the largest possible union of sets in \( \mathcal{C}_1(\theta) \), which is a subset of \( T_1 \). Denote the largest possible union \( D_1 \).

From Assumption 1. (iii), the distribution of \( \varepsilon \) is continuous, and \( \varepsilon \) has everywhere positive density. Hence, \( P_\varepsilon(D_1) \leq P_\varepsilon(T_1) \). Now consider \( \Pr[Q_\theta(Y, W, z) \subseteq T_1; F_0^0] \) and \( \Pr[Q_\theta(Y, W, z) \subseteq D_1; F_0^0] \).

If \( Q_\theta(y, w, z) \) is a subset of \( T_1 \), then it is also a subset of \( D_1 \) since \( D_1 \) is the largest possible union. So, I have

\[
P_\varepsilon(T_1) \geq P_\varepsilon(D_1) \geq \Pr[Q_\theta(Y, W, z) \subseteq D_1; F_0^0] = \Pr[Q_\theta(Y, W, z) \subseteq T_1; F_0^0],
\]

Thus, \( P_\varepsilon(D_1) \geq \Pr[Q_\theta(Y, W, z) \subseteq D_1; F_0^0] \) is not informative. Therefore I can rule out sets that are not the unions of sets in \( \mathcal{C}_1(\theta) \).

Next, I want to rule out non-connected sets among the unions of sets in \( \mathcal{C}_1(\theta) \). Pick an arbitrary non-connected union \( T_2 \). Without loss of generality, suppose \( Q_1 \) and \( Q_2 \) are non-connected, and
\( T_2 = Q_1 \cup Q_2 \). For \( Q_1 \) and \( Q_2 \), I know

\[
\begin{align*}
P_\varepsilon(Q_1) & \geq \Pr[Q_\theta(Y, W, z) \subseteq Q_1; F^0], \\
P_\varepsilon(Q_2) & \geq \Pr[Q_\theta(Y, W, z) \subseteq Q_2; F^0].
\end{align*}
\]

Also, I have

\[
P_\varepsilon(T_2) \geq \Pr[Q_\theta(Y, W, z) \subseteq T_2; F^0]
\]

\[
= \Pr[Q_\theta(Y, W, z) \subseteq Q_1; F^0] + \Pr[Q_\theta(Y, W, z) \subseteq Q_2; F^0],
\]

by construction of the containment functional, i.e.

\[
\Pr[Q_\theta(Y, w, z) \subseteq D; F^0] = \sum_{w \in W} \left\{ \sum_{y \in Y} \left( \sum_{y \in Y} 1[Q_\theta(y, w, z) \subseteq D] \Pr[Y = y | W = w, Z = z; F^0] \right) \right\}. 
\]

Since \( Q_1 \) and \( Q_2 \) are disjoint, \( P_\varepsilon(T_2) = P_\varepsilon(Q_1) + P_\varepsilon(Q_2) \). Hence, I have

\[
P_\varepsilon(T_2) \geq \Pr[Q_\theta(Y, W, z) \subseteq T_2; F^0]
\]

\[
\Leftrightarrow
\]

\[
P_\varepsilon(Q_1) + P_\varepsilon(Q_2) \geq \Pr[Q_\theta(Y, W, z) \subseteq Q_1; F^0] + \Pr[Q_\theta(Y, W, z) \subseteq Q_2; F^0].
\]

The last inequality is not informative.

Finally, it is trivial that the union of sets in \( C_1(\theta) \), which constitute \( R^{2^{p-1}} \) is not informative. Therefore, all connected unions of sets in \( C_1(\theta) \) except \( R^{p-1} \) consist the core determining class.

**Appendix B. Core Determining Class**

In this appendix, I explain how to obtain a core determining class in practice. With the utility function (6.3), \( C_2(\theta) \), or the power set of \( C_1(\theta) = \{ Q_\theta(y, w, z) : y \in Y, w \in W, z \in Z \} \) has cardinality equal to \( 2^8 = 256 \). The cardinality of \( Z \) is \( K = 2^2 \times 3^2 = 36 \). Without using the core determining class, the econometrician would have to compute \( 36 \times 256 = 9216 \) moments. I find sets in the core determining class as follows. First, I denote the eight \( Q_\theta(y, w, z) \)'s as \( \{ Q_1, \cdots, Q_8 \} \). I construct the adjacency matrix \( G_\theta \) among \( \{ Q_1, \cdots, Q_8 \} \) based on the connectedness of each pair \( (Q_i, Q_j) \). That is, \( G_{\theta,ij} = 1 \) if \( Q_i \) and \( Q_j \) are connected, and zero otherwise. Then, among those \( 2^8 \) sets, I delete non-connected sets and duplicated sets based on \( G_\theta \). Finally, I delete the entire
set \( \mathbb{R}^3 \) and the empty set. Note that this process should be done with respect to the value of \( \theta \), or more specifically \( \beta_3^{(1)} \) and \( \beta_3^{(2)} \).

The core determining class differs by the value of the parameter vector \( \theta \). More precisely, it depends on the coefficients on the endogenous explanatory variable, i.e. \( \beta_3^{(1)} \) and \( \beta_3^{(2)} \). There are 8 possible cases with respect to the values of \( \beta_3^{(1)} \) and \( \beta_3^{(2)} \). For a given value of \( Z = z \), I first denote \( Q_1, \cdots, Q_8 \) as follows.

\[
Q_1 = Q_\theta((0, 0), 0, z), \ Q_2 = Q_\theta((0, 0), 1, z), \ Q_3 = Q_\theta((1, 0), 0, z), \\
Q_4 = Q_\theta((1, 0), 1, z), \ Q_5 = Q_\theta((0, 1), 0, z), \ Q_6 = Q_\theta((0, 1), 1, z), \\
Q_7 = Q_\theta((1, 1), 0, z), \text{ and } Q_8 = Q_\theta((1, 1), 1, z).
\]

- Case 1. \( \beta_3^{(2)} < \beta_3^{(1)} < 0 \).

Figure 2 shows how the area of unobservables is divided into 8 different regions. Table 12 shows the adjacency matrix \( G_\theta \) for Case 1, which characterizes the connectedness of \( Q_i \)'s. The core determining class \( M(\theta) \) is described in Table 13.

<table>
<thead>
<tr>
<th></th>
<th>( Q_1 )</th>
<th>( Q_2 )</th>
<th>( Q_3 )</th>
<th>( Q_4 )</th>
<th>( Q_5 )</th>
<th>( Q_6 )</th>
<th>( Q_7 )</th>
<th>( Q_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q_1 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( Q_2 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( Q_3 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( Q_4 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( Q_5 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( Q_6 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( Q_7 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( Q_8 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 12: The adjacency matrix \( G(\theta) \) when \( \beta_3^{(2)} < \beta_3^{(1)} < 0 \).

There are other 7 more cases corresponding to \( (\beta_3^{(1)}, \beta_3^{(2)}) \). They are as follows.

- Case 2: \( \beta_3^{(1)} < \beta_3^{(2)} < 0 \),
- Case 3: \( \beta_3^{(2)} > \beta_3^{(1)} > 0 \),
- Case 4: \( \beta_3^{(1)} > \beta_3^{(2)} > 0 \),
- Case 5: \( \beta_3^{(1)} < 0 < \beta_3^{(2)} \), and \( \beta_3^{(1)} + \beta_3^{(2)} > 0 \),
- Case 6: \( \beta_3^{(1)} < 0 < \beta_3^{(2)} \), and \( \beta_3^{(1)} + \beta_3^{(2)} < 0 \),
- Case 7: \( \beta_3^{(2)} < 0 < \beta_3^{(1)} \), and \( \beta_3^{(1)} + \beta_3^{(2)} > 0 \),

and Case 8: \( \beta_3^{(2)} < 0 < \beta_3^{(1)} \), and \( \beta_3^{(1)} + \beta_3^{(2)} < 0 \).
Figure 2: Regions of Unobservables Corresponding to $Q(y, w, z; \theta)$ given $z$ when $\beta_3^{(2)} < \beta_3^{(1)} < 0$.

Appendix C. Detailed Estimation Procedures

In this appendix, I explain the estimation procedure with more details.

C.1. Drawing Parameters

For the simulated annealing procedure, I draw more than 500 initial values of $\theta$ from $Unif(-1, 1)$ in order to make sure that all regions in $\Theta$ are visited by simulated annealing. I set the lower and upper bounds for simulated annealing to be -5 and 5, respectively for each parameter. For the SVM procedures, I first draw $T$ values of the 18 dimensional parameter vector from $U(-5, 5)$. The parameter vector has 18 elements since it includes five parameters in $\Lambda$, or the Cholesky decomposition of $\Sigma_e$, i.e. $\Sigma_e = \Lambda \Lambda'$. Note that I have only five parameters from $\Sigma_e$, since scale normalization is done by setting $var(\varepsilon_1) = 1$. 

49
After drawing parameters, a core determining class will be computed for each parameter draw. Before I start the estimation procedures, I construct 8 core determining classes corresponding to the 8 different cases shown in Appendix B.

C.2. Simulation of Unobservables

Since I do not observe \( \varepsilon \), I need to simulate \( R = 100 \) sets of \( \varepsilon \). I draw \( \varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \) from \( N(0, I_3) \). Then I multiply the Cholesky decomposition matrix \( \Lambda \). This type of simulation is based on the method of simulated moments (MSM) proposed by McFadden (1989) and Pakes and Pollard (1989). I fix \( \varepsilon \) during all estimation procedure.

Next, I compute a total of 1296 moments \( m_{ij}^{(r)} \) for each observation \( ij \) and \( r \)th simulation. That is,

\[
m_{ij}^{(r)}(\theta) = (m_{ij,1,1}^{(r)}(\theta), m_{ij,1,2}^{(r)}(\theta), \ldots, m_{ij,1,36}^{(r)}(\theta), \ldots, m_{ij,k,1}^{(r)}(\theta), \ldots, m_{ij,36,36}^{(r)}(\theta))',
\]

where

\[
m_{ij,k,1}^{(r)}(\theta) = 1[\varepsilon_i^{(r)} \in D_i] - \sum_{w \in W} \sum_{y \in Y} 1[Q_0(y, w, z) \subseteq D_i]1[Y_{ij} = y|W_{ij} = w, Z_{ij} = z_k]1[W_{ij} = w|Z_{ij} = z_k].
\]

Table 13: The core determining class \( C(\theta) \) when \( \beta_3^{(2)} < \beta_3^{(1)} < 0 \).
I take the sample average of \( \{m_{ij}^{(r)}(\theta), r = 1, \cdots, R\} \) over the unobservable draws for each \( ij \). That is,
\[
m_{ij}(\theta) = (m_{ij,1,1}(\theta), m_{ij,1,2}(\theta), \cdots, m_{ij,1,36}(\theta), \cdots, m_{ij,k,l}(\theta), \cdots, m_{ij,36,36}(\theta))',
\]
where
\[
m_{ij,k,l}(\theta) = \frac{1}{R} \sum_{r=1}^{R} \{1[\varepsilon_{i}^{(r)} \in D_i] - \sum_{w \in W} \sum_{y \in Y} 1[Q_{\theta}(y, w, z) \subseteq D_i]1[Y_{ij} = y|W_{ij} = w, Z_{ij} = z_k]1[W_{ij} = w|Z_{ij} = z_k]\}.
\]
I still do not use a notation \( \bar{m} \) at this stage, since I have not taken the sample average over pairs.

C.3. Computing Sample Moments and Test Statistic

I take the sample average over pairs as
\[
\bar{m}_N(\theta) = (\bar{m}_{N,1,1}(\theta), \bar{m}_{N,1,2}(\theta), \cdots, \bar{m}_{N,1,36}(\theta), \cdots, \bar{m}_{N,k,l}(\theta), \cdots, \bar{m}_{N,36,36}(\theta))',
\]

\[
\bar{m}_{N,k,l}(\theta) = \tilde{N}^{-1} \sum_{m=1}^{M} \sum_{j > i}^{n_m} \sum_{i=1}^{n_m-1} \frac{1}{R} \sum_{r=1}^{R} \{1[\varepsilon_{i}^{(r)} \in D_i] - \sum_{w \in W} \sum_{y \in Y} 1[Q_{\theta}(y, w, z) \subseteq D_i]1[Y_{ij} = y|W_{ij} = w, Z_{ij} = z_k]1[W_{ij} = w|Z_{i} = z_k]\},
\]
where \( \tilde{N} = \sum_{m=1}^{M} \binom{n_m}{2} \). I also compute the variance-covariance matrix \( \hat{\Sigma}_N(\theta) \) of \( m_{ij}(\theta) \). In practice, I only need to compute the variance of each \( m_{ij,k,l}(\theta) \).
\[
\hat{\Sigma}_N(\theta) = \tilde{N}^{-1} \sum_{m=1}^{M} \sum_{j > i}^{n_m} \sum_{i=1}^{n_m-1} (m_{ij}(\theta) - \bar{m}_N(\theta))(m_{ij}(\theta) - \bar{m}_N(\theta))'.
\]
From AS, I have several choices of functions for computing the test statistic. For computational efficiency, I choose \( S_3 \) in AS, which is
\[
S_3(m, \Sigma) = \sum_{j=1}^{J} [m_j/\sigma_j]^2,
\]
\[ [x]^- = \begin{cases} x, & \text{if } x \leq 0 \\ 0, & \text{otherwise} \end{cases} \]

Then, the test statistic \( T_S(\theta_0) \) is defined as

\[
T_S(\theta_0) = S_3(\sqrt{N}\tilde{m}_S(\theta), \tilde{\Sigma}_S(\theta)) = \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{\sqrt{N}\tilde{m}_{S,k,l}(\theta)/\hat{\sigma}_{k,l}}{(\ln \tilde{N})^{1/2}},
\]

where \( K = 36 \) and \( L = 36 \) in practice.

### C.4. Bootstrap and Moment Selection

I determine the generalized moment selection (GMS) critical value \( \hat{c}_N(\theta_0, 1-\alpha) \) by bootstrapping.

First, I simulate \( B = 100 \) (nonparametric iid) bootstrap samples. \( \{Y_{ij,b}, W_{ij,b}, Z_{ij,b}, \varepsilon_{ijb}, \forall ij \leq \tilde{N}\} \), where \( \varepsilon_{ij} \) is the \( R \) set of unobservables drawn at the first stage.

Compute

\[
M_{N,b}^*(\theta_0) = \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{\sqrt{N}\tilde{m}_{N,k,l}(\theta)/\hat{\sigma}_{k,l}}{SD_{N,b}^*(\theta_0) \tilde{\Sigma}_{N,b}^*(\theta_0) \hat{D}_{N,b}^{-1/2}(\theta_0)},
\]

for all \( b = 1, \ldots, B \), where \( \hat{D}_{N,b}(\theta_0) \) is the diagonal matrix of \( \hat{\Sigma}_{N,b}^*(\theta_0) \).

Next, I determine whether \( \tilde{N}^{1/2}\tilde{m}_{N,k,l}(\theta_0)/\hat{\sigma}_{N,k,l}(\theta_0) > \kappa_N = (\ln \tilde{N})^{1/2} \) for each \( (k, l) \)th sample moment. Eliminate the elements in \( (M_{N,b}^*(\theta_0), \tilde{\Sigma}_{N,b}^*(\theta_0)) \) for all \( b = 1, \ldots, B \) that correspond to the moments that satisfy the above condition. Then the resulting statistics are denoted by \( (M_{N,b}^{**}(\theta_0), \tilde{\Sigma}_{N,b}^{**}(\theta_0)) \) for \( b = 1, \ldots, B \).

Finally, I take the critical value \( \hat{c}_N(\theta_0, 1-\alpha) \) to be the \( 1-\alpha \) sample quantile of \( \{S(M_{N,b}^{**}(\theta_0), \tilde{\Sigma}_{N,b}^{**}(\theta_0)) \} : b = 1, \ldots, B \}. \) If \( T_n(\theta_0) \leq \hat{c}_N(\theta_0, 1-\alpha) \), then \( \theta_0 \) should be included in the \( 1-\alpha \) confidence set.

I repeat the above procedures for all values of parameters drawn in Appendix C.1., when using SVM. For simulated annealing, these procedures are repeated until the optimization process ends.
Appendix D. Computational Examples of the Identification Region

In this appendix, I illustrate the computation of the sharp identification region of parameter vector with the similar settings of CRS. Since the purpose of this appendix is to see how the sharp identification region looks like, I just focus on a binary outcome variable: \( y \in \{0, 1\} \). I consider two cases of binary choice models with \( \kappa_w = 2 \) and \( \kappa_w = 4 \). Let \( U(y = 1|w, z; \theta) = \alpha + \beta w + \gamma z + v \), where \( \theta = (\alpha, \beta, \gamma)' \). The joint distribution of \( Y \) and \( W \) given \( Z = z \) is specified as ordered probit for \( W \) given \( Z = z \) and multinomial logit for \( Y \) given \( W = w_k \) and \( Z = z \). Then,

\[
F^0(Y = 1 \text{ and } W = w_k | Z = z) = \frac{\exp(\alpha + \beta w_k + \gamma z)}{1 + \exp(\alpha + \beta w_k + \gamma z)} \left( \Phi\left(\frac{c_k - d_1 z}{d_2}\right) - \Phi\left(\frac{c_k - 1 - d_1 z}{d_2}\right)\right),
\]

and

\[
F^0(Y = 0 \text{ and } W = w_k | Z = z) = \frac{1}{1 + \exp(\alpha + \beta w_k + \gamma z)} \left( \Phi\left(\frac{c_k - d_1 z}{d_2}\right) - \Phi\left(\frac{c_k - 1 - d_1 z}{d_2}\right)\right),
\]

where \( c_0 = -\infty \) and \( c_{\kappa_w} = \infty \). The unobservable \( \varepsilon \) has the iid Type 1 extreme distribution. I set \( d_1 = d_2 = 1 \), and \( \alpha = 0, \beta = 1, \) and \( \gamma = -0.5 \).

Example. \( \kappa_w = 2 \).

The supports of \( w \) and \( z \) are as follows: \( w \in \{-1, 1\} \), and \( z \in \{-1, 1\} \). For the three parameters, I construct a grid of approximately 930,000 values and plot the sharp identification region. When \( \beta \) approaches to zero, i.e. the model has no endogenous explanatory variable, the sharp identification region of \( \gamma \) becomes the empty set. Figure 3 shows the sharp identification region.

Example. \( \kappa_w = 4 \).

The supports of \( w \) and \( z \) are as follows: \( w \in \{-1, -1/2, 1/2, 1\} \), and \( z \in \{-1, 1\} \). \( c = (-\infty, -1/2, 0, 1/2, \infty) \). Again, for the three parameters, I construct a grid of approximately 1.2 million values and plot the sharp identification region. As \( \beta \) approaches to zero, i.e. the model has no endogenous explanatory variable, the sharp identification region of \( \gamma \) shrinks to the empty set. Figure 4 shows the sharp identification region.

As Figures 3 and 4 indicate, the sharp identification regions include both negative and positive values for each parameter. From this perspective, it may not be strange that the confidence intervals in the estimation results of the empirical application include both negative and positive values for all parameters.
Figure 3: The sharp identification region of $\theta$ when $\kappa_w = 2$

Figure 4: The sharp identification region of $\theta$ when $\kappa_w = 4$