IDENTIFICATION AND ESTIMATION OF NETWORK FORMATION MODELS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jun Sung Kim

May 2014

IDENTIFICATION AND ESTIMATION OF NETWORK FORMATION MODELS

Jun Sung Kim, Ph.D.

Cornell University 2014

My thesis studies identification and estimation in network formation models. First, I study what can be learned from pairwise stable networks. Pairwise stability of a network gives strong identification power when I consider the probability that the observed network is pairwise stable. I propose a semiparametric maximum score estimator which is simple and computationally feasible. I apply the empirical model to social and economic networks in rural India, and find homophily patterns in village networks.

Second, I propose a structural model of multigraph formation, where 1) individuals determine multiple types of links simultaneously; 2) all networks interact with each other; and 3) one or more networks are endogenous but not simultaneous. I extend the notion of pairwise stability to a multigraph, and show that the structural model is equivalent to a multinomial choice model. The presence of endogenous but not simultaneous networks is a source of an incomplete econometric model. Relying on partially identified econometric models, I characterize the sharp identification region of parameters by a finite set of moment inequalities. I apply the model to village networks and find that friendship affects risk sharing and favor exchange networks in the same direction.

The last chapter studies an empirical model of network formation in the U.S. airline industry and investigates the size of network externalities. I assume that each airline builds a network that satisfies a weak notion of stability. That is, no

airlines want to deviate from their current networks by a single route change. In this framework, I can use an entry game to investigate the airline industry and include network measures in the profit function to estimate network externalities. I find that when I control for the number of one-stop flights the effect of hub-size is larger than the case without considering one-stop flights.

BIOGRAPHICAL SKETCH

Jun Sung Kim attended Korea University for his undergraduate studies, completing the bachelor's degree in business administration. He earned his master's degrees in statistics from University of California, San Diego, and in economics from Cornell. He plans on graduating from Cornell University with Ph.D. in economics in may 2014.

To My Family

and

In Memory of My Grandmother Kyoungjo Lee

ACKNOWLEDGEMENTS

I am deeply indebted to the members of my dissertation committee, Francesca Molinari, Larry Blume, and Jörg Stoye for their invaluable comments, suggestions and support. Especially, I am grateful to Francesca Molinari for being an outstanding mentor. During my doctoral studies, I always try to follow her in all aspects; wise and humble, strong and kind, enthusiastic and patient.

I thank my friends at Cornell for helpful conversations. I do not name each of them since there are too many! I also thank my old friends in Korea, Kyoungil Kim and Grimilgi, Pastor Hyungjin Ko and the members of DongsanY, and friends whom I have met in various places. I am very lucky to have strong ties with you all in my social network!

I want to thank my lovely family, uncles and aunts. My wife Elu So Young Yoon has always been with me even when we face difficulties in our lives. Now, it is my turn to support her with all my efforts. I want to thank my brother Youngjae Kim and his wife Younjin Sung, and my brother-in-law Sung Dae Yoon. They are the reason I could keep pursuing my academic career abroad. I am grateful to my father Jung Soo Kim, my mother Jung Kyoung Choi, and my parents-in-law Chang Kook Yoon and Hae Hee Kwon for their continuous support. Without their support, I could not make all this happen. I want to say 'thank you and I miss you' to my grandmother Kyoungjo Lee who would be the happiest person in heaven. All glory belongs to God and the Lord Jesus Christ.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Economic agents form many different types of networks and make important decisions based on those networks. Economists have been interested in the theoretical analysis of strategic network formation. However, the literature on the empirical models of strategic network formation has not been fully established. My thesis contributes to the literature by studying identification and estimation problems in the models of strategic network formation. Furthermore, I propose a structural model of multigraph formation which can capture the richness of our social interactions. I focus on stable networks and multigraphs to solve the well-known problem of the curse of dimensionality in the network formation models. I employ recently developed econometric techniques such as partial identification and simulation methods and apply them to the data on social and economic networks.

As a starting point, I first study what can be learned from a single pairwise stable network in the second chapter. Recent literature on empirical models of strategic network formation confronts problems such as the curse of dimensionality and multiple equilibria. To solve these problems, I consider the probability that the observed network is pairwise stable, instead of the probability that a certain equilibrium outcome is observed. Pairwise stability of a network and the assumption of myopic agents contained in it give strong identification power when I consider the probability that the observed network is pairwise stable. I propose a semiparametric maximum score estimator which is simple and computationally feasible. I apply the empirical model to social and economic networks in rural India, and find that individuals have caste homophily in all types of village networks.

The third chapter 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.

The last chapter studies an empirical model of network formation in the U.S. airline industry and investigates the size of network externalities. I define the entry of an airline carrier in a market as an operation of a direct flight in that market. With this definition of entry, being an incumbent is equivalent to the formation of a link on the airline network. I assume that airline carriers build a network that satisfies a weak notion of stability which is similar to pairwise stability in the usual social network settings. An airline's decision in each market, which is derived from an entry game against other airline carriers, provides positive profits given the rest of the network. That is, no airlines want to deviate from their current networks by a single route change at a time. In this framework, the econometrician can use a typical entry game in the literature to investigate the airline industry. In addition, under this assumption of stable networks, I include various network measures in post-entry profit function and estimate network externalities. From the recent Airline Origin and Destination Survey data, I find that when I control for the number of one-stop flights the effect of hub-size is larger than the case without considering one-stop flights.

CHAPTER 2

STRUCTURAL ESTIMATION OF PAIRWISE STABLE NETWORKS: AN APPLICATION TO SOCIAL NETWORKS IN RURAL INDIA

2.1 Introduction

Individuals form a social network based on their socioeconomic characteristics. While economists have proposed many theoretical models of strategic network formation, their empirical counterparts have not been fully established. The retarded progress of the development in empirical network formation models is due to the following reasons. First of all, it is well-known that games of strategic network formation often exhibit multiple equilibria. Second, externalities naturally appear in the network formation models. That is, the link decision of a pair of individuals affects other pairs' link decisions and vice versa. Finally, network formation models suffer from the curse of dimensionality. As the number of individuals grows, the number of all possible network configurations grows exponentially. Since these three difficulties arise simultaneously, identification and estimation of network formation models are intimidating. For example, counting all possible equilibria and corresponding network configurations with large number of players is often impossible even in a very simple network formation game. In order to solve these problems, this chapter considers pairwise stability of a network. I propose an empirical model of strategic network formation, and show that pairwise stability of a network introduced by Jackson and Wolinsky (1996) provides strong identification power. Also, pairwise stability makes estimation procedures very simple. Finding these two advantages of pairwise stability is the main contribution of this chapter. I apply the model to village data in rural India, and find that individuals have strong caste homophily.

Recent literature on the structural econometrics of games has been dealing with identification problems due to the presence of multiple equilibria. Bjorn and Vuong (1984) first describe the problem of multiple equilibria in a labor participation game between husband and wife. Bresnahan and Reiss (1990, 1991b) investigate an entry game in which the presence of multiple equilibria is commonplace. They obtain point identification by taking the number of firms in a market as an outcome variable with an assumption of a homogeneous product across firms. Tamer (2003) distinguishes incomplete econometric models with incoherent economic models. The entry game where multiple equilibria are present is an example of incomplete econometric models. He shows that there is still identification information in incomplete models and that point identification of parameters can be achieved with additional conditions such as exclusion restrictions. He also employs a partial identification approach to construct bounds for parameters in case that no exclusion restrictions are available. After Tamer (2003), many authors consider the identification problem due to multiple equilibria. See Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2010), Beresteanu, Molchanov, and Molinari (2011), and Kline (2012) among others. Although it is ideal to apply a partial identification approach to network formation games, the use of such approaches is limited due to the curse of dimensionality. More specifically, it is often infeasible to compute all possible equilibria in a network formation game. Hence, I suggest a similar approach to that of Bresnahan and Reiss to obtain the point identification of structural parameters. I focus on the probability that an observed network is pairwise stable, instead of the probability that a certain equilibrium network is observed. The necessary conditions for pairwise stability of a network are strong enough to point identify the utility parameters in the model. Since I am interested in maximizing the probability that the observed network is pairwise stable, potential other pairwise stable networks are irrelevant to the empirical work, which will be explained later with more details.

This chapter also contributes to the growing literature on the empirical models of a strategic network formation. Currarini, Jackson, and Pin (2009) propose various indices to measure the degree of homophily within a network. They also develop a search-based model of friendship formation that explains segregation patterns in a social network. They find that the type sensitivity of preferences and bias in matching together explain the observed homophily patterns among U.S. adolescents. Christakis, Fowler, Imbens, and Kalyanaraman (2010) empirically predict what network will be formed based on observed characteristics of individuals as well as link-specific variables. They propose a sequential game of strategic network formation among myopic agents, and use two-step Bayesian MCMC methods to estimate structural parameters. Their model is limited since the resulting network is not necessarily stable. Also, it requires the augmentation of the history of meetings since the meeting history among individuals is not observed. Mele (2010) employs a directed friendship formation and uses the stochastic best response dynamics (see Blume 1993) to achieve a unique prediction of the network formation game. However, it is more reasonable to consider friendship as a relationship under mutual consents. Sheng (2012) employs a simultaneous-move link announcement game first proposed by Myerson (1991, p448). She focuses on subnetworks, and uses a partial identification approach to obtain an outer region of parameters. However, adopting pairwise stability as necessary conditions for an equilibrium in the simultaneous-move link announcement game may lead a misspecification problem.[1] To my knowledge, no previous papers in the literature fully utilize pairwise stability of a network as a stability notion to obtain the point identification of the utility parameters.

While the existing papers which impose a particular type of games, the approach in this chapter is immune to the different types of games as long as observed networks are pairwise stable. The econometrician do not have enough knowledge about the type of a network formation game (static, dynamic, or combined). I impose neither a potentially unrealistic game nor its equilibrium concept. Nevertheless, I provide a consistent estimator for the utility parameters under pairwise stability of a network. The estimation procedure uses a semiparametric maximum score estimator first proposed by Manski (1975, 1985). The estimator is semiparametric in the sense that no distributional assumption is imposed on the unobservables. In the literature on the empirical models of strategic network formation, the distribution of unobservables is often assumed to be logistic. The semiparametric approach has an advantage since it is robust to the distributional assumption prevalent in the literature. To the best of my knowledge, a maximum score estimator is first proposed in this chapter for the empirical models of strategic network formation. [2]

In an empirical application, I apply the model to social networks in rural India. I use the "Social Networks and Microfinance" data collected by Banerjee, Chandrasekhar, Duflo, and Jackson.[3] The results in this chapter show that

[1]For the simultaneous-move link announcement game, an appropriate equilibrium concept is pairwise Nash equilibrium (PNE). Pairwise stability is a necessary condition for PNE. However, the existence of PNE is not guaranteed in general. This non-existence may need a misspecification problem.

[2]Maximum score estimator has been proposed in the empirical matching models. See Fox (2008).

[3]Abhijit Banerjee; Arun Chandrasekar; Esther Duflo; Matthew Jackson,

individuals have strong homophily in many different characteristics. Lazarsfeld and Merton (1954) first introduced the term homophily which defines the tendency of individuals in a society to bond with similar others. The structural model in this chapter allows a researcher to find different segregation patterns by the types of social networks. Individuals may borrow money from those who are similar in a certain characteristic, while they may ask help to others that are different in the characteristic. I am mostly interested in how a social class, or more specifically caste, affects the formation of different types of networks: friendship, borrowing money, give-advice, etc. I find that village individuals have strong caste homophily. I also show that different characteristics are important in the formation of different types networks.

The rest of the chapter is organized as follows. Section 2 describes a strategic network formation model. Section 3 provides discussion on network formation games and pairwise stability of a network. Section 4 solves the identification problem by using pairwise stability of a network. Section 5 explains the estimation method and how the structural model is connected to typical parametric models such as probit. Section 6 collects data explanation, estimation results and economic interpretations. Section 7 concludes.

2.2 Set Up

A network formation process consists of a meeting technology, preferences, and a network formation game. While a meeting technology governs the opportunity of forming a link, the network formation game and the preferences of indi-

2011-08, Social Networks and Microfinance, http://hdl.handle.net/1902.1/16559 UNF:5:4EmgOYAQGaoQugFowckNfA== Jameel Poverty Action Lab [Distributor] V5 [Version] viduals determine actual link formation decisions given a set of opportunities. This section describes a meeting technology, preferences and their relation.

2.2.1 A Network

Let $m = 1, \cdots, M$ be an index for villages. Individuals in village m are indexed by $i = 1, \cdots, N_m$. The number of individuals in each village is fixed. Let $\mathcal{N}_m$ be the set of individuals in village m. For ease of exposition, I will omit the village index m unless necessary. A social network formed by the individuals is represented by an adjacency matrix A, where the ith row and jth column element of A is written as;

$$a_{ij} = \begin{cases} 1 & \text{, if } i \text{ and } j \text{ are friends} \\ 0 & \text{, otherwise.} \end{cases} \tag{2.1}$$

I set $a_{ii} = 0$ for all i. A link between two individuals is undirected, so if i nominates j as a friend, then j also thinks i as a friend. Alternatively, a network can be represented by a set of links as one can see in graph theory. With a slight abuse of notations, I use the same A to denote the set of links. For example, if a network is composed of a single link between i and j, $A = \{ij\}$. In terms of matrix notation, this network A has $a_{ij} = a_{ji} = 1$, and zeroes in all other positions. The alternative notation from graph theory makes it easy to introduce pairwise stability of a network.

2.2.2 Meeting Technology

Consider how to distinguish a meeting from a linking. First, I distinguish an initial meeting between a pair from meetings which occur after the first meeting. I assume that individuals in a village know each other. Thus, the probability of the initial meeting between i and j is one for all i and j. The assumption is not harmful when a village is not very large.

Now consider a meeting technology given that all individuals in a village know each other. If two individuals meet very often due to some common activities or environments, they are more likely to be friends. In that case, meeting technology is the major factor of forming a link. However, it is hard to distinguish whether some contextual variables (e.g. age) play a role in meeting technology or a role in preferences. For example, consider a high school student who has all friends with the same age. There are two possible explanations. She may choose her friends based on preferences to age, or she has few chances to meet people with different ages. In order to rule out such problems, I limit the meeting technology in the model to a completely random meeting. Thus, an individual i's probability of meeting j is identical across i and j, conditional on contextual variables. Under the random meeting technology, I conclude that meeting opportunities do not have any prediction power for network formation. Although I admit that variables such as age and gender give rise to differences in meeting, it is hard to identify the roles of such variables in both the meeting technology and preferences simultaneously.

2.2.3 Preferences

After imposing a process of completely random meetings, I define individual i's utility function. An individual's utility of forming a network in this chapter has a similar form to ones in the existing literature. Let x_i be an $L \times 1$ vector of observed characteristics of i, and X be an $L \times K$ matrix of the observed characteristics of all individuals. The utility function of agent i by forming a network A is written as[4]

$$U_i(A|X, \varepsilon; \theta) = \sum_{j=1}^{N} a_{ij}(u_{ij} + \varepsilon_{ij}) + \sum_{j} a_{ij} \sum_{k} \Delta_{ik} a_{ik} a_{jk}. \tag{2.2}$$

The first term in (2.2) represents the sum of intrinsic values of i's friends to i. In particular, $u_{ij} + \varepsilon_{ij}$ is j's intrinsic value to i, where u_{ij} is observed to the econometrician, but ε_{ij} is not. Assume that ε_{ij} is independent and identically distributed across all individuals and pairs. The intrinsic value u_{ij} depends on i's characteristics as well as j's. That is,

$$u_{ij} = \alpha_0 + \alpha_1 x_{j,l} + \cdots + \alpha_L x_{j,L} + \sum_{l}^{L} \beta_l (x_{i,l} - x_{j,l})^2. \tag{2.3}$$

As in Christakis et al. (2010), the utility function (2.3) captures homophily effects with respect to the lth individual characteristic by β_l. If it is negative, individuals prefer similar types of friends over different ones, and vice versa. The term Δ_{ik} in the right hand side of (2.2) is an additional utility from having a mutual friend k. There may exist heterogeneity in Δ_{ik}, but I assume $\Delta_{ik} = \Delta$ for all i and k

[4]The previous version of this chapter employed the distance-based utility which can be found in the connections model of Jackson and Wolinsky (1996). The distance-based utility may be more realistic, but it leaves the problem of the existence of a pairwise stable network unsolved without a strong assumption. See Appendix B for more detail.

when estimating it.[5] I also assume $\Delta_{ik} \geq 0$. The non-negative utility from having mutual friends is intuitive, and also guarantees the existence of a pairwise stable network. I will show the existence result later. I use $\theta = (\alpha', \beta', \Delta)$ to denote the collection of all parameters.

I denote $A - ij$ as a network obtained by severing a link between i and j from A, if $a_{ij} = 1$ for A. Likewise, $A + ij$ denotes a network obtained by adding a link ij to A. Then, i's marginal utility of j is[6]

$$U_i(A + ij|X, \varepsilon; \theta) - U_i(A - ij|X, \varepsilon; \theta) = u_{ij} + \varepsilon_{ij} + \Delta \sum_k a_{ik} a_{jk}. \tag{2.4}$$

When it is not confusing, I sometimes use mu_{ij} and mu_{ij}^d to denote the marginal utility and its deterministic parts, respectively.

2.3 Network Formation Games

2.3.1 Games and Pairwise Stability

Before I discuss pairwise stability (PS) of a network, I discuss a network formation game first. Although the notion of pairwise stability does not depend on a particular network formation process, a researcher may want to consider one. Here, I give an example of network formation games, which is a simultaneous-move link announcement game of complete information proposed by Myerson (1991). In this game, players announce link formation decision with the

[5]One would consider Δ_{ik} as a random coefficient to capture unobserved heterogeneity. It may be an interesting way to extend the model, but I leave it as a future study.

[6]Note that if $ij \in A$, $A + ij = A$. Also, if $ij \notin A$, $A - ij = A$. Thus, (2.4) is true regardless of a_{ij}.

other $N-1$ players. Let $Y_i = \{0,1\}^{N-1}$ denote the set of actions of i, and $y_i = (y_{i1}, \cdots, y_{i,i-1}, y_{i,i+1}, \cdots, y_{iN}) \in Y_i$ be the generic vector of an action for player i. Let $y = (y_1, \cdots, y_N) \in Y = \times Y_i$ and $y_{-i} = (y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_N) \in Y_{-i} = \times_{j \neq i} Y_j$ be the set of actions of all players and that of all other players except i, respectively. Since the link formation is based on mutual agreement, the resulting network A is such that for all i and j,

$$a_{ij} = \begin{cases} 1 & \text{, if } y_{ij} = y_{ji} = 1 \\ 0 & \text{, otherwise.} \end{cases} \tag{2.5}$$

The Nash equilibrium of the game is an equilibrium strategy profile y^{NE} that satisfies$U_i(A^{NE}|X, \varepsilon; \theta) = U_i(y_i^{NE}|y_{-i}^{NE}, X, \varepsilon; \theta) \geq U_i(y_i|y_{-i}^{NE}, X, \varepsilon; \theta)$ for all y_i and i. Nash equilibrium networks are network configurations, or graphs A^{NE} corresponding to the NE strategies. Nash equilibrium of the network formation game may not be unique even in a two-player game. When the number of players becomes larger, counting the number of all Nash equilibria and corresponding networks is often impossible. Thus, the econometrician cannot make a prediction by using NE conditions. In this context, the advantages of pairwise stability arise.

Pairwise stability of a network is first introduced by Jackson and Wolinsky (1996). Pairwise stability of a network is a **stability notion** rather than a solution concept of network formation games. This is important since a stability notion does not require imposing a particular game, while an equilibrium solution concept does. The set of all pairwise stable networks contains all pairwise stable network configurations that are outcomes of possibly many different games. Hence, different pairwise stable network configurations may be outcomes of different games. Imposing a particular game may not generate all sets of pairwise stable networks. Bloch and Jackson (2006) point out that pairwise stability is not based on a non-cooperative game. Suppose that the econometrician imposes a noncooperative game and tries to identify parameters with pairwise stability. In this case, the observed network may not correspond to the equilibrium network of the game. This may result in a misspecification problem.

Now I introduce pairwise stability. As I mentioned above, I use A to refer both an adjacency matrix and a graph (i.e. $A = \{ij | a_{ij} = 1, \text{and } j > i\}$). The value of graph A is simply the aggregation of individual utilities, or $U = \sum_i U_i(A|X, \varepsilon; \theta)$. For simplicity, I use $U_i(A)$ to abbreviate $U_i(A|X, \varepsilon; \theta)$, whenever there is no confusion. The formal definition of pairwise stability of a network is as follows.

Definition (Jackson and Wolinsky (1996)) *The graph A is pairwise stable with respect to U if*
(i) for all $ij \in A$, $U_i(A) \geq U_i(A - ij)$ and $U_j(A) \geq U_j(A - ij)$, and
(ii) for all $ij \notin A$, if $U_i(A) > U_i(A + ij)$ then $U_j(A) < U_j(A + ij)$.

Now, when observing a network A with $a_{ij} = 1$, I can infer that

$$\begin{aligned} U_i(A|X, \varepsilon; \theta) &\geq U_i(A - ij|X, \varepsilon; \theta) \\ &\text{and} \\ U_j(A|X, \varepsilon; \theta) &\geq U_j(A - ij|X, \varepsilon; \theta). \end{aligned} \tag{2.6}$$

Also, I can infer from some $a_{ij} = 0$ that

$$\begin{aligned} \text{if } U_i(A + ij|X, \varepsilon; \theta) &> U_i(A|X, \varepsilon; \theta), \\ \text{then } U_j(A + ij|X, \varepsilon; \theta) &< U_j(A|X, \varepsilon; \theta). \end{aligned} \tag{2.7}$$

The above conditions are exactly the same conditions as pairwise stability defined in Jackson and Wolinsky (1996).

I point out a few things. First, the benefit of employing pairwise stability of a network is that it takes into account both link formation and severance without imposing a strong assumption as explained in Jackson and Wolinsky (1996). These link formation and severance conditions give a prediction power in empirical applications. Second, pairwise stability contains the assumption of myopic agents. When a pair of individuals tries to deviate from the current network, they do not consider future changes in the network. Finally, I recognize that my approach has a limitation. If the econometrician is interested in predicting a network given different set of individuals, I cannot tell what network configuration would be realized without imposing a network formation game. However, when the econometrician tries to impose a particular game, he or she has to be careful because pairwise stability notion is not an equilibrium solution concept. In this case, one may want to consider a stronger solution concept than NE, e.g. pairwise Nash equilibrium (PNE, see Bloch and Jackson 2006). I leave this for future studies.

2.3.2 The Existence of Pairwise Stable Networks

The existence of pairwise stable network is not guaranteed in general. Jackson and Watts (2001, 2002a) prove that there exists either a closed cycle of networks or a pairwise stable network. In this subsection, I summarize the theoretical results in the literature on pairwise stability of a network, and discuss which results are applicable to the case in this chapter.

Lemma 2.3.1. *(Jackson and Watts (2002a)) There exists at least one pairwise stable network or closed cycle of networks. Consequently, if there are no cycles, then there exists at least one pairwise stable network.*

Proof. See Jackson and Watts (2002a).

Based on the lemma, if one can rule out a cycle of networks, a pairwise stable network must exist. Jackson and Watts (2001, 2002a) also give conditions under which no closed cycles of networks exist. In order to discuss the conditions, it is necessary to introduce the concepts of adjacency and no indifference in Jackson and Watts (2002a). Two networks A and A' are called *adjacent* if they differ by one link. The utility function U exhibits *no indifference* if for any two adjacent matrices A and A' either A defeats A' or A' defeats A. Suppose that $ij \in A$. Network A is said to *defeat* $A - ij$ if $U_i(A) > U_i(A - ij)$ and $U_j(A) > U_j(A - ij)$. Likewise, $A - ij$ defeats A if either $U_i(A - ij) \geq U_i(A)$ or $U_j(A - ij) \geq U_j(A)$.

Let $\mathcal{A}_N$ be the set of all graphs with N agents. If a potential function can be found, then there are no cycles. Theorem 1 in Jackson and Watts (2002a) formally states it.

Lemma 2.3.2. *(Theorem 1 in Jackson and Watts (2002a)) Fix U. If there exists a function $W : \mathcal{A}_N \to \mathbb{R}$, such that "$A$ defeats $A' \Leftrightarrow W(A) > W(A')$, where A and A' are adjacent.", then there are no cycles. Conversely, if U exhibits no indifference, then there are no cycles only if there exists a function $W : \mathcal{A}_N \to \mathbb{R}$, such that "$A$ defeats $A' \Leftrightarrow W(A) > W(A')$, where A and A' are adjacent.".*

Proof. See Jackson and Watts (2002a).

It is difficult to find a function $W(\cdot)$ under the current utility specification of (2.2). There are a couple of papers that provide different models to find a potential function and the existence of equilibria. Mele (2010) employs directed link formation and shows that there exists a potential function with similar utility function to (2.2). The existence of a potential function relates to the existence and the uniqueness of a stationary distribution. Accordingly, the Markov process of the stochastic best response dynamics converges to a unique stationary distribution. Although directed network formation can happen in some applications, it is more realistic to consider the network formation to be undirected in friendship networks. If one considers the other person as a friend but not vice versa, their friendship would not last long. As a consequence, the network will change, and the stability is not guaranteed. Under undirected link formation it is difficult to find a potential function without allowing for utility transfer between individuals. Sheng (2012) uses a transferable utility model of an undirected link formation and shows the existence of a potential function.

I employ a slightly simple utility specification in this chapter, and apply theorems in Hellmann (2009) and Hellmann (2012) to prove the existence of pairwise stable networks. Hellmann (2012) shows that there exists at least one pairwise stable network if the utility function is ordinal convex and exhibits strategic complementarity. For the definitions of ordinal convexity and strategic complementarity, see Appendix A.1. Also, if the utility function is ordinal concave and satisfies the ordinal substitute property, there exists a unique pairwise stable network. If $\Delta_{ik} \geq 0$, I show the existence of a pairwise stable network. Sheng (2012) also has similar results with a slightly different utility function.

Proposition 2.3.3. *Consider the utility function (2.3). If Δ_{ik} is non-negative for all i and k, then there exists a pairwise stable network for all ε.*

Proof. See Appendix A.1.1.

2.4 Identification

2.4.1 Identification Power from Pairwise Stability

In this subsection, I describe how I obtain the identification of utility parameters in the model under pairwise stability of a network. The identification procedure is carried out in three steps. First, I show that the link decision a_{ij} of a pair i and j is uniquely determined given the rest of network A_{-ij}. Second, I show how pairwise stability of a network rules out simultaneity between a_{ij} and A_{-ij}. Finally, I provide parametric identification results in the next subsection.

I recall that pairwise stability of a network is a **stability notion** rather than a solution concept. Pairwise stability of a network implies that no pairs of individuals have an incentive to deviate from the current network configuration. For example, consider an arbitrary network A. Let A_{-ij} be the network of all pairs of individuals except the pair i and j. If the network A satisfies pairwise stability, then all pairs do not want to deviate from their current link decisions. Thus, by fixing the rest of the network A_{-ij}, pairwise stability of a network gives conditions (2.6)-(2.7) for all $N(N-1)/2$ pairs. Note that the conditions for each pair are satisfied given that other pairs' link decisions are fixed.

Recall the utility function and the marginal utility:

$$U_i(A|X, \varepsilon; \theta) = \sum_{j=1}^{N} a_{ij}(u_{ij} + \varepsilon_{ij}) + \Delta \sum_j a_{ij} \sum_k a_{ik} a_{jk}, \tag{2.8}$$

and

$$U_i(A + ij|X, \varepsilon; \theta) - U_i(A - ij|X, \varepsilon; \theta) = u_{ij} + \varepsilon_{ij} + \Delta \sum_k a_{ik} a_{jk}. \tag{2.9}$$

For notational simplicity, I denote $t_{ij} = \sum_k a_{ik} a_{jk}$. Then, $a_{ij} = 1$ is equivalent to

$$\begin{aligned} u_{ij} + \varepsilon_{ij} + \Delta t_{ij} &> 0 \\ &\text{and} \\ u_{ji} + \varepsilon_{ji} + \Delta t_{ij} &> 0. \end{aligned} \tag{2.10}$$

Putting the unobserved variable to the left hand side yields

$$\begin{aligned} \varepsilon_{ij} &> -u_{ij} + \Delta t_{ij} \\ &\text{and} \\ \varepsilon_{ji} &> -u_{ji} + \Delta t_{ij}. \end{aligned} \tag{2.11}$$

Likewise $a_{ij} = 0$ provides

$$\begin{aligned} \varepsilon_{ij} &\leq -u_{ij} - \Delta t_{ij} \\ &\text{or} \\ \varepsilon_{ji} &\leq -u_{ji} - \Delta t_{ij}. \end{aligned} \tag{2.12}$$

Since there are $N(N-1)/2$ pairs of individuals, I have $N(N-1)/2$ sets of inequalities. Let y_i be

$$y_i = \begin{cases} 1 & \text{, if } \varepsilon_{ij} > -u_{ij} - \Delta t_{ij} \\ 0 & \text{, otherwise.} \end{cases} \tag{2.13}$$

Then $a_{ij} = 1$ if $(y_i, y_j) = (1, 1)$ and zero otherwise. Figure 2.1 shows that pairwise stability provides a unique prediction for the link decision of two individuals regardless of the regions of ε_{ij} and ε_{ji}. Figure 2.1 also compares pairwise stable outcomes with Nash equilibrium outcomes.[7] A link is formed by a mutual agreement, while the severance of a link can be done unilaterally. If j never wants to form a link, then i's decision does not matter for a_{ij}. Second, in the center region of ε, NE predicts $(y_i, y_j) = (1, 1)$ or $(0, 0)$. In this case, pairwise stability implies that i and j always choose $(1, 1)$, or $a_{ij} = 1$, since they are better off by deviating from (0,0). Hence, for any values of $\varepsilon = (\varepsilon_{ij}, \varepsilon_{ji})$, pairwise stability predicts a unique outcome for a single pair's link decision a_{ij} given the rest of network A_{-ij} and X.

Next, I have to deal with the problem of simultaneity between a_{ij} and A_{-ij}. Although I have shown that pairwise stability provides a unique prediction for the two-player link formation game, i's marginal utility of j still depends on decisions A_{-ij} of other pairs. The main advantage of pairwise stability comes at

[7] In the figure, for ease of exposition I simplify the model such that $\Delta t_{ij} > 0$.

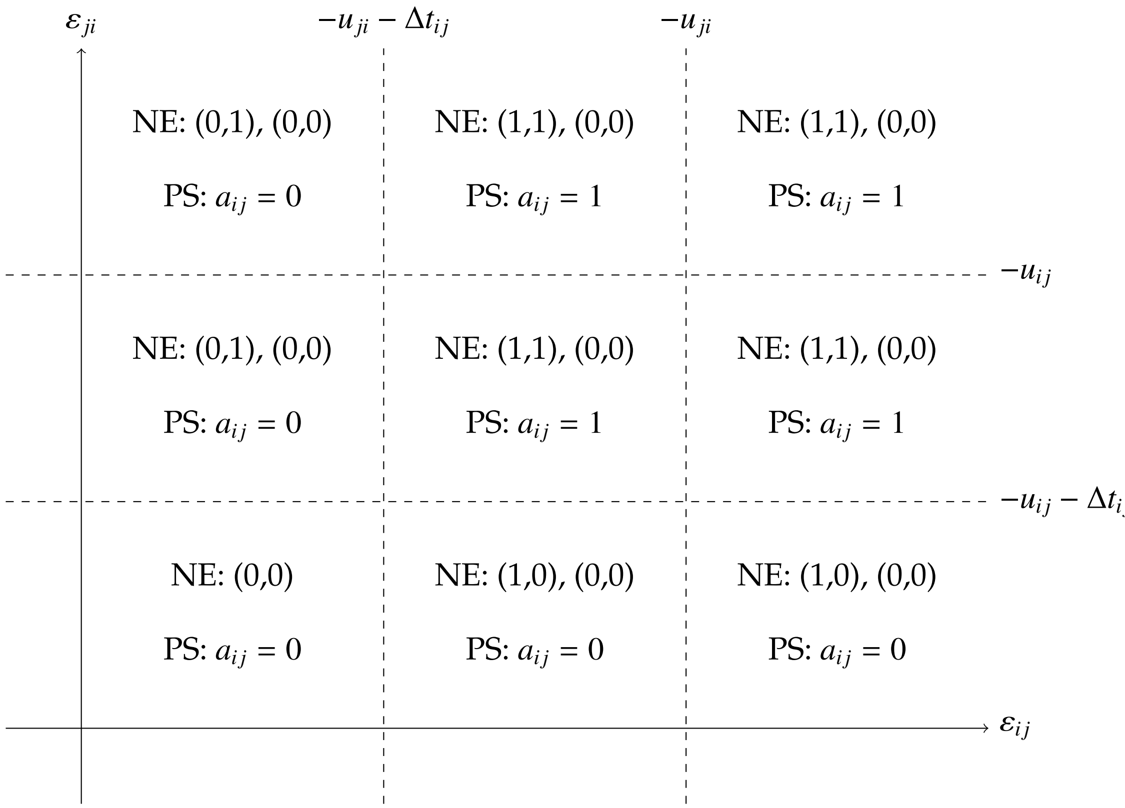

Figure 2.1: NE vs. pairwise stable outcome in two by two link formation game

this point. At the end of subsection 2.3.2, I have briefly mentioned the implicit assumption of myopic agents contained in pairwise stability. From the next paragraph, I will explain it with more detail.

According to Jackson and Watts (2001), a network is pairwise stable if and only if it has no improving paths emanating from it.

Definition (Jackson and Watts (2001)) *An improving path from a network A to a network A' is a finite sequence of networks $A_1, A_2, \cdots, A_K$ with $A_1 = A$ and $A_K = A'$ such that for any $k = 1, \cdots, K$ either*

(i) $A_{k+1} = A_k - ij$ for some ij such that $U_i(A_k - ij) \geq U_i(A_k)$, or

(ii) $A_{k+1} = A_k + ij$ *for some* ij *such that* $U_i(A_k + ij) > U_i(A_k)$ *and* $U_j(A_k + ij) > U_j(A_k)$

They explained that *'an improving path is a sequence of networks that might be observed in a dynamic process where agents are myopically adding and deleting a link'*. Consider a dynamic game of network formation. If an observed network is pairwise stable, it does not have an improving path. Equivalently, players do not have an incentive to add and/or delete a link myopically under the pairwise stable network. In addition, when a pair ij makes a deviation decision, the pair compares utility from the current network A with utility from $A + ij$ (or $A - ij$) only. The rest of the network A_{-ij} is given when they make a decision. In other words, rather than taking into account A_{-ij} as other pairs' decisions, i and j consider A_{-ij} as a fixed structure, since they are myopic. Hence, there is no simultaneity between a_{ij} and A_{-ij}.

I formally state the above argument below. In the literature, the econometrician is often interested in the probability that the outcome A is observed given X and θ, or $\Pr(A|X, \theta)$. In network formation games this probability is

$$\Pr(A|X;\theta) = \Pr(A \text{ is chosen}|A \text{ is PS}, X;\theta)\Pr(A \text{ is PS}|X;\theta) \tag{2.14}$$

$$\int_{\varepsilon} \{\mathbf{1}[A \text{ is chosen}|A \text{ is PS}, X, \varepsilon;\theta] \times \mathbf{1}[A \text{ is PS}|X, \varepsilon;\theta]\}\, dF(\varepsilon). \tag{2.15}$$

In this chapter, I am interested in probability of a different type of an outcome; the likelihood that A is pairwise stable. Since I focus on a different outcome variable, this idea is similar to one in Bresnahan and Reiss (1990) where they are interested in the number of firms rather than an equilibrium outcome in each market. I am interested in maximizing the following likelihood.

$$\Pr(A \text{ is PS}|X;\, \theta) = \int_{\varepsilon} \mathbf{1}[A \text{ is pairwise stable}|X,\, \varepsilon] dF(\varepsilon). \tag{2.16}$$

According to the assumption of myopic agents in pairwise stability, I rewrite the indicator term in equation (2.16) as

$$\mathbf{1}[A \text{ is PS}|X,\, \varepsilon;\, \theta] = \mathbf{1}[(a_{12}, \cdots, a_{N-1,N}) \text{ is PS}|X,\, \varepsilon;\, \theta] \tag{2.17}$$

$$= \prod_{ij} \mathbf{1}[a_{ij} \text{ is PS}|A_{-ij},\, X,\, \varepsilon;\, \theta]. \tag{2.18}$$

Equation (2.18) is derived from (2.17) due to the assumption of myopic agents. If the econometrician considers pairwise stability as a solution concept of a game (which is not correct), deviation of a link decision of a pair, say a'_{12} would change the decisions of other pairs $a_{-12} = (a_{13,} \cdots, a_{N-1,N})$. However, as I explained above, pairwise stability and the myopic agents assumption make other pairs decisions A_{-ij} fixed. It is a very important implication of pairwise stability of a network for empirical applications.

Note that when maximizing the likelihood $\Pr(A \text{ is PS}|X;\, \theta)$, other pairwise stable networks are *irrelevant*. Regardless of the number of pairwise stable networks and the selection probability $\Pr(A \text{ is chosen}|A \text{ is PS},\, X;\, \theta)$, I can obtain point identification from $\Pr(A \text{ is PS}|X;\, \theta)$. To be more specific, consider a dynamic process of network formation. A particular network is always observed given a history of meetings which may not be observed. Hence, other pairwise stable networks that cannot be realized under the current formation process are all irrelevant.

2.4.2 Identification

In this subsection, I prove the identification of the structural parameters in (2.2) and (2.3). Let $\theta \in \Theta$ collect all parameters. i.e. $\theta = (\alpha_0, \cdots, \alpha_L, \beta_1, \cdots, \beta_L, \Delta)$.

$$\begin{aligned}
\Pr(A \text{ is PS}|X;\, \theta) &= \int_{\varepsilon} \mathbf{1}[A \text{ is PS}|X,\, \varepsilon] dF(\varepsilon) \\
& \int_{\varepsilon} \prod_{ij} \mathbf{1}[a_{ij} \text{ is PS}|A_{-ij},\, X,\, \varepsilon;\, \theta] dF(\varepsilon) \\
&= \int_{\varepsilon} \prod_{ij} \left\{ (p_{ij})^{a_{ij}} (1-p_{ij})^{1-a_{ij}} \right\} dF(\varepsilon), \qquad (2.19)
\end{aligned}$$

where $p_{ij} = \Pr(mu_{ij} > 0 \,\text{and}\, mu_{ji} > 0|X,\, A_{-ij},\, \varepsilon)$. Also let the $(2L+2) \times 1$ vector x_{ij} of explanatory variables in (3.2) be

$$x_{ij} = (1,\, x_{j,1},\, \cdots,\, x_{j,L},\, (x_{i,1} - x_{j,1})^2,\, \cdots,\, (x_{i,L} - x_{j,L})^2,\, t_{ij})', \qquad (2.20)$$

and I use $\mathcal{X}$ to denote the space of X

Let θ_0 be the utility parameter. Define

$$\begin{aligned}
X_\theta &= \{X \in \mathcal{X} :\, \Pr(A \text{ is PS}|X,\, \theta) \\
&\neq \Pr(A \text{ is PS}|X,\, \theta_0)\}\,. \qquad (2.21)
\end{aligned}$$

If $\Pr(X \in X_\theta) > 0$, then there is no observationally equivalent θ to θ_0. Thus, θ_0 is point identified.

Definition θ_0 is point identified relative to Θ if $\Pr(X \in X_\theta) > 0$

Assumption (i) (Support conditions) At least two elements of x_{ij}, say x_{1j} and

x_{2j} have full support, and corresponding parameters are non zero. Furthermore $f_{x_1|x_2=x}$ is strictly positive for all x.

(ii) (Scale normalization) θ is normalized to have $||\theta|| = 1$.

(ii) (Location normalization) ε_{ij} is i.i.d. and has a continuous and strictly increasing distribution function F_ε with $med(\varepsilon|x) = 0$.

(iv) (Linear independence) There exists no proper linear subspace of $\mathbb{R}^{(2L+2)}$ having probability 1 under F_x.

Unlike typical binary choice models, I need at least two explanatory variables with full support. Briefly speaking, when $a_{ij} = 1$, it requires two marginal utilities to be positive simultaneously. In order to satisfy this condition, I need two explanatory variables with full support. Also, the support of two variables must be independent. Note that they are not necessary condition, and the conditional full support assumption is testable in practice. From the location normalization assumption, I have

$$\begin{aligned}\Pr(A \text{ is PS}|X;\ \theta) &= \prod_{j>i}\prod_{i}\left\{\int\int (p_{ij})^{a_{ij}}(1-p_{ij})^{1-a_{ij}}dF(\varepsilon_{ij})dF(\varepsilon_{ji})\right\} \\ &= \prod_{j>i}\prod_{i}\left[\left\{\Pr(mu^d_{ij} > 0|X, A_{-ij})\Pr(mu^d_{ji} > 0)|X, A_{-ij}\right\}^{a_{ij}}\right. \\ &\quad \left.\times\left\{1-\Pr(mu^d_{ij} > 0|X, A_{-ij})\Pr(mu^d_{ji} > 0)|X, A_{-ij}\right\}^{1-a_{ij}}\right]. \qquad (2.22)\end{aligned}$$

From (2.22), I can rewrite X_θ as

$$\begin{aligned}\tilde{X}_\theta &= \Big\{(x_{ij},\ x_{ji}) \in \widetilde{\mathcal{X}} \times \widetilde{\mathcal{X}} :\ \Pr(mu_{ij} > 0 \text{ and } mu_{ji} > 0|x_{ij},\ x_{ji};\ \theta) \\ &\qquad \neq \Pr(mu_{ij} > 0 \text{ and } mu_{ji} > 0|x_{ij},\ x_{ji};\ \theta_0)\Big\}, \qquad (2.23)\end{aligned}$$

where $\widetilde{\mathcal{X}}$ is a space for x_{ij}. Showing $\Pr[(x_{ij}, x_{ji}) \in \tilde{X}_\theta] > 0$ is enough for point identification.

Proposition 2.4.1. *Let the utility function satisfy the form of (2.3), (2.8) and (2.18). Suppose that Assumption 1 holds. Then, the parameter vector θ is point identified.*

Proof. See Appendix A.1.3.

Identification of parameters crucially depends on the existence of two full support explanatory variables. This condition may not meet in practice. When the econometrician does not have such explanatory variables, he or she is still able to obtain partial identification of parameters. For instance, Komarova (2012) provides a recursive method to obtain an outer region of parameters. I leave this as a future study.

2.5 Estimation Methods

2.5.1 Maximum Score Estimator

The estimation procedure is semiparametric since I do not impose parametric distributional assumptions on the unobserved characteristics. Maximum score estimator is first proposed by Manski (1975, 1985). Let $mu^d_{ij}(\theta)$ be the deterministic component of $U_i(A_{-ij} + ij|X, \varepsilon; \theta) - U_i(A_{-ij}|X, \varepsilon; \theta)$, and θ_0 be the true parameter vector. Then, $mu^d_{ij}(\theta_0)$ represents the true value of marginal utility of i if i form a link to j given X and A_{-ij};

$$mu_{ij}^d(\theta_0) = u_{ij,0} + \Delta_0 \sum_{k \notin \mathbf{N}_i(A)} a_{ik} a_{jk}. \tag{2.24}$$

Define the population score function as

$$\begin{aligned} S(\theta_0) \quad = \quad & E\Big[a_{ij} \times \mathbf{1}\{mu_{ij}^d(\theta_0) > 0\} \times \mathbf{1}\{mu_{ji}^d(\theta_0) > 0\} \\ & +(1 - a_{ij}) \times \Big(1 - \mathbf{1}\{mu_{ij}^d(\theta_0) > 0\} \times \mathbf{1}\{mu_{ji}^d(\theta_0) > 0\}\Big)\Big] \\ & = \Pr(a_{ij} = 1 \text{ and } \{mu_{ij}^d(\theta_0) > 0\} \cap \{mu_{ji}^d(\theta_0) > 0\}) \\ & + \Pr(a_{ij} = 0 \text{ and } \{mu_{ij}^d(\theta_0) \leq 0\} \cup \{mu_{ji}^d(\theta_0) \leq 0\}). \end{aligned} \tag{2.25}$$

and the sample score function as

$$\begin{aligned} S_M(\theta) \quad = \quad & \frac{1}{M} \sum_m \binom{N_m}{2}^{-1} \sum_{i=1}^{N_m} \sum_{j>i}^{N_m} \Big[a_{ij} \times \mathbf{1}\{mu_{ij}^d(\theta) > 0\} \times \mathbf{1}\{mu_{ji}^d(\theta) > 0\} \\ & +(1 - a_{ij}) \times \Big(1 - \mathbf{1}\{mu_{ij}^d(\theta) > 0\} \times \mathbf{1}\{mu_{ji}^d(\theta) > 0\}\Big)\Big]. \end{aligned} \tag{2.26}$$

The maximum score estimator is

$$\hat{\theta}_{MSE} = \arg\max_{\theta} S_M(\theta). \tag{2.27}$$

Equivalently, I can rewrite (2.25) as

$$\begin{aligned}
\tilde{S}(\theta_0) &= E\left[\left|a_{ij} - \mathbf{1}\{mu^d_{ij}(\theta_0) > 0\} \times \mathbf{1}\{mu^d_{ji}(\theta_0) > 0\}\right|\right] \\
&= \Pr(a_{ij} = 0 \text{ and } \{mu^d_{ij}(\theta_0) > 0\} \cap \{mu^d_{ji}(\theta_0) > 0\}) \\
&\quad + \Pr(a_{ij} = 1 \text{ and } \{mu^d_{ij}(\theta_0) \leq 0\} \cup \{mu^d_{ji}(\theta_0) \leq 0\}) \\
&= 1 - S(\theta_0).
\end{aligned}$$

The maximum score estimator minimizes the sample analogue of $\tilde{S}(\theta_0)$.

2.5.2 Asymptotics

The asymptotic argument on this version of maximum score estimator is the number of villages M. The number of pairs, $\binom{N_m}{2} = N_m(N_m - 1)/2$ and the number of individuals N_m remain fixed. I do not use the number of pairs (or individuals) in a village as an asymptotic argument due to following reasons. First, the model in this chapter implicitly assumes that people in a village had a chance to know each other. Hence, I utilize non connected pairs ($a_{ij} = 0$) as well as existing links (ij such that $a_{ij} = 1$) for estimation. If village population becomes very large, it may be true that some pairs of individuals are not connected due to no chance of a meeting. Furthermore, as the number of people goes to infinity, I may not guarantee the existence of a pairwise stable network. For these reasons, I use M as an asymptotic argument. Consistency of the estimator is a simple extension of Manski (1985), and I do not pursue it here.

2.5.3 Connection to Parametric Models

One may notice that the model in this chapter is similar to parametric binary choice models. e.g. probit or logit models. Indeed, equation (2.22) is similar to the likelihood function of parametric binary choice models except: 1) it has the product of two probabilities in each term, and 2) I do not specify the distribution of unobservables. I can rule out the former difference by considering transferable utility models. Employing pairwise stability with transferable utility gives the following conditions. $a_{ij} = 1$ if $mu_i(A + ij) + mu_j(A + ij) > 0$, and $a_{ij} = 0$ if $mu_i(A + ij) + mu_j(A + ij) \leq 0$. In addition, one may assume $\varepsilon_{ij} = \varepsilon_{ji}$. Then, p_{ij} in (2.22) can be written as

$$\Pr(\varepsilon_{ij} \geq -\alpha_0 - \alpha' \bar{x}_{ij} - \sum \beta_l (x_{il} - x_{jl})^2 - \Delta t_{ij} | A_{-ij}; \theta), \tag{2.28}$$

where $\bar{x}_{ij}$ is the average characteristics of i and j. Finally, if one imposes a distributional assumption (e.g. Normal), the model is equivalent a parametric binary choice model.[8] In this sense, pairwise stability a network gives a rationale to the applied econometrician to use logit or probit model for link level predictions. However, one must be careful on imposing these assumptions (transferable utility, $\varepsilon_{ij} = \varepsilon_{ji}$, and a distributional assumption). For example, it may not be natural to assume transferable utility in friendship network formation.

[8]Khan (2010) shows the observational equivalence between semiparametric binary choice model and heteroskedastic probit/logit models.

2.6 Data and Empirical Results

2.6.1 Data

I use the "Social Networks and Microfinance" data set collected by Abhijit Banerjee, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. They collect the data from 75 villages in rural areas in Karnataka, which is an area of southern India. According to Banerjee, Chandrasekhar, Duflo, and Jackson (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. They construct 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 (MH), 13) Intersection of relationships, 14) Union of relationships (ALL). I use the individual-level network data.

One of my main interests is how individuals care about social class (caste or sub caste) of others when forming a network. For this reason, I consider only villages where the proportion of its majority caste is less than 95% of the total population. 22 villages are qualified under this criterion. I first construct four social networks among those 8 networks. Visit network (VISIT), Money network (MONEY), Advice network (ADV) and Kerosene-Rice network (KERICE) are the unions of 3) and 4), 5) and 6), 7) and 8), and 9) and 10), respectively. Then, I use a total of 7 networks: close non-relatives (or friendship network, FR henceforth), MONEY, VISIT, ADV, KERICE, MH, and ALL networks. I do not use network 11 and 13, since they exhibit relatively small number of links. Table 2.1 shows descriptive statistics.

	Friend	Money	Kero-rice	Visit	Advice	Medic	All
# of links	13603 1.80%	14234 1.89%	15472 2.05%	17342 2.30%	11256 1.49%	11630 1.54%	29593 3.92%
same caste	12033	12781	13987	15381	10325	10755	25468
different	1570	1453	1485	1961	931	875	4125

Table 2.1: Descriptive Statistics for Individual Networks

2.6.2 Estimation Results

I use the maximum score estimator, (2.27). Since the objective function is discontinuous, I use a global optimization method. Fox (2008) and Fox and Santiago (2008) suggest the differential evolution algorithm for maximum score estimation. The algorithm is proposed by Storn and Price (1997) and appropriate for this type of problems. They provide a MATLAB code, `devec3.m` for implementing the algorithm. Even with this algorithm, I had to try a lot of population members to make sure convergence, since the algorithm has a stochastic feature. Tables 2.2-2.6 show the results.

	Friendship	95% CI	Money	95% CI
constant	-0.1735**	(-0.2183, -0.0883)	-0.3824**	(-0.4813, -0.3659)
α's for x_j				
Gender	0.3171**	(0.2285, 0.3686)	0.6452**	(0.6412, 0.7933)
Age	0.0011**	(0.0010, 0.0013)	0.0006**	(0.0002, 0.0007)
Religion	-0.1150**	(-0.1646, -0.0781)	0.0014	(-0.0158, 0.0927)
Education	0.0071**	(0.0048, 0.0116)	0.0055**	(0.0025, 0.0095)
Village Native	-0.0238	(-0.0532, 0.0357)	-0.1282**	(-0.1849, -0.0730)
Work	-0.0789**	(-0.1353, -0.0330)	-0.1908**	(-0.2806, -0.1782)
Caste	-0.0093	(-0.0506, 0.0108)	0.0312	(-0.0117, 0.0782)
β's on $(x_i - x_j)^2$				
Gender	0.0154	(-0.0208, 0.1914)	0.0066	(-0.0382, 0.1654)
Age	-0.1365**	(-0.1630, -0.0209)	-0.2835**	(-0.3500, -0.2093)
Religion	-0.6102**	(-0.8064, -0.5816)	-0.1894**	(-0.2722, -0.0424)
Education	-0.1148**	(-0.1338, -0.0186)	-0.2358**	(-0.2857, -0.1485)
Village Native	-0.5943**	(-0.7549, -0.5494)	-0.1761**	(-0.2265, -0.0140)
Work	-0.1450	(-0.1765, 0.0337)	-0.2865**	(-0.3556, -0.1478)
Caste	-0.2138**	(-0.2498, -0.0385)	-0.2446**	(-0.2993, -0.0911)
Δ	0.1575**	(0.1066, 0.1878)	0.2020**	(0.1694, 0.2492)

Table 2.2: Estimation Results for Friendship and Money Networks.

(**: significant at 5%, *: significant at 10%)

	Kerosene-Rice	95% CI	Visit	95% CI
constant	-0.3592**	(-0.4483, -0.3063)	-0.4344**	(-0.5493, -0.3961)
α's for x_j				
Gender	0.5460**	(0.5053, 0.6677)	0.3899**	(0.2975, 0.4683)
Age	0.0007**	(0.0004, 0.0008)	0.0008**	(0.0005, 0.0009)
Religion	-0.0690**	(-0.1179, -0.0022)	0.0178**	(0.0025, 0.1056)
Education	0.0096**	(0.0071, 0.0164)	0.0086**	(0.0063, 0.0139)
Village Native	-0.0999**	(-0.1831, -0.0557)	-0.0629**	(-0.1252, -0.0077)
Work	-0.0830**	(-0.1529, -0.0288)	0.0434**	(0.0272, 0.1216)
Caste	0.0476**	(0.0127, 0.1023)	-0.0035*	(-0.0035, 0.0838)
β's on $(x_i - x_j)^2$				
Gender	-0.0030	(-0.0639, 0.1546)	-0.0328	(-0.0896, 0.1178)
Age	-0.2452**	(-0.3134, -0.1823)	-0.4581**	(-0.5785, -0.4437)
Religion	-0.3286**	(-0.4563, -0.2354)	-0.3984**	(-0.5369, -0.3131)
Education	-0.1528**	(-1919, -0.0620)	-0.2565**	(-0.3095, -0.1718)
Village Native	-0.2670**	(-0.3374, -0.1559)	-0.3815**	(-0.4978, -0.2846)
Work	-0.3869**	(-0.4992, -0.2928)	-0.1523	(-0.1825, 0.0410)
Caste	-0.2741**	(-0.3387, -0.1394)	-0.1646**	(-0.1988, -0.0106)
Δ	0.2388**	(0.1878, 0.2939)	0.1435**	(0.0979, 0.1749)

Table 2.3: Estimation Results for Kerosene-Rice and Visit Networks

	Advice	95% CI	Medical Help	95% CI
constant	-0.1480**	(-0.1775, -0.0427)	-0.2663**	(-0.3369, -0.1957)
α's for x_j				
Gender	0.3609**	(0.2894, 0.4232)	0.4681**	(0.4166, 0.5721)
Age	0.00055**	(0.00026, 0.00063)	0.0008**	(0.0006, 0.0010)
Religion	-0.1292**	(-0.1841, -0.0807)	-0.0812**	(-0.1229, -0.0330)
Education	0.0078**	(0.0059, 0.0124)	0.0046**	(0.0011, 0.0082)
Village Native	-0.1425**	(-0.2410, -0.1019)	-0.0192	(-0.0657, 0.0468)
Work	0.0222	(-0.0203, 0.1190)	-0.1509**	(-0.2466, -0.1172)
Caste	-0.0176	(-0.0685, 0.0051)	0.0450**	(0.0188, 0.0886)
β's on $(x_i - x_j)^2$				
Gender	-0.3845**	(-0.5464, -0.3146)	0.1212**	(0.1071, 0.3405)
Age	-0.3690**	(-0.4725, -0.3036)	-0.1599**	(-0.1914, -0.0538)
Religion	-0.1014	(-0.1782, 0.0615)	-0.5375**	(-0.7287, -0.5147)
Education	-0.3059**	(-0.3871, -0.2452)	-0.2006**	(-0.2457, -0.1061)
Village Native	-0.2813**	(-0.3630, -0.1560)	-0.2870**	(-0.3578, -0.1589)
Work	-0.4120**	(-0.5257,-0.3269)	-0.0781	(-0.0949, 0.1045)
Caste	-0.3747**	(-0.4662, -0.2698)	-0.4207**	(-0.5272, -0.3115)
Δ	0.1812**	(0.0989, 0.2132)	0.2082**	(0.1541, 0.2544)

Table 2.4: Estimation Results for Advice and Medical-help Networks

	All	95% CI
constant	-0.1335	(-0.1476, 0.0684)
α's for x_j		
Gender (male=1)	0.2356**	(0.1368, 0.3004)
Age	0.0009**	(0.0006, 0.0011)
Religion	-0.1198*	(-0.1676, 0.0083)
Education	-0.0016	(-0.0051, 0.0089)
Village Native	-0.0492	(-0.1146, 0.0173)
Work	-0.0236	(-0.0756, 0.0663)
Caste	0.0111	(-0.0449, 0.0818)
β's on $(x_i - x_j)^2$		
Gender)	-0.1116**	(-0.2313, -0.0202)
Age	-0.2650**	(-0.3413, -0.1999)
Religion	-0.1289	(-0.2190, 0.0550)
Education	-0.5499**	(-0.7080, -0.5834)
Village Native	-0.2244**	(-0.3311, -0.1320)
Work	-0.4813**	(-0.6374, -0.4189)
Caste	-0.4659**	(-0.6019, -0.3840)
Δ	0.0920*	(-0.0011, 0.0946)

Table 2.5: Estimation Results for All Network

I also report the 95% confidence interval for each parameter. The confidence interval is obtained by subsampling. I draw 200 subsamples in which 10% of pairs are randomly chosen from the original data, and run the maximum score estimation for each subsample. The 95% confidence intervals are constructed according to Politis and Romano (1994) and Politis, Romano, and Wolf (1999).

Consider the estimates of α's. The constant α_0 is negative and significant in most types of networks except ALL network (insignificant). Among the parameters on the opponents' characteristics, individuals prefer having male individuals in all types of networks.Those who have more education are more attractive in most networks, but the magnitude is relatively small. Individuals prefer friends who are not currently working, and also tend to borrow or give money to those who are not working. The latter may seem counter intuitive.

	Friendship	Money	Kerosene-Rice	Visit	Advice	Medical Help	All
constant	-0.1735**	-0.3824**	-0.3592**	-0.4344**	-0.1480**	-0.2663**	-0.1335
α's for x_j							
Gender	0.3171**	0.6452**	0.5460**	0.3899**	0.3609**	0.4681**	0.2356**
Age	0.0011**	0.0006**	0.0007**	0.0008**	0.00055**	0.0008**	0.0009**
Religion	-0.1150**	0.0014	-0.0690**	0.0178**	-0.1292**	-0.0812**	-0.1198*
Education	0.0071**	0.0055**	0.0096**	0.0086**	0.0078**	0.0046**	-0.0016
Village Native	-0.0238	-0.1282**	-0.0999**	-0.0629**	-0.1425**	-0.0192	-0.0492
Work	-0.0789**	-0.1908**	-0.0830**	0.0434**	0.0222	-0.1509**	-0.0236
Caste	-0.0093	0.0312	0.0476**	-0.0035*	-0.0176	0.0450**	0.0111
β's on $(x_i - x_j)^2$							
Gender	0.0154	0.0066	-0.0030	-0.0328	-0.3845**	0.1212**	-0.1116**
Age	-0.1365**	-0.2835**	-0.2452**	-0.4581**	-0.3690**	-0.1599**	-0.2650**
Religion	-0.6102**	-0.1894**	-0.3286**	-0.3984**	-0.1014	-0.5375**	-0.1289
Education	-0.1148**	-0.2358**	-0.1528**	-0.2565**	-0.3059**	-0.2006**	-0.5499**
Village Native	-0.5943**	-0.1761**	-0.2670**	-0.3815**	-0.2813**	-0.2870**	-0.2244**
Work	-0.1450	-0.2865**	-0.3869**	-0.1523	-0.4120**	-0.0781	-0.4813**
Caste	-0.2138**	-0.2446**	-0.2741**	-0.1646**	-0.3747**	-0.4207**	-0.4659**
Δ	0.1575**	0.2020**	0.2388**	0.1435**	0.1812**	0.2082**	0.0920*

Table 2.6: Comparison of Parameters in Different Types of Networks

Caste is significant in KERICE, VISIT, and MH networks, but its sign differs by networks. Interestingly, α's are less significant in ALL network. I may conclude that aggregation of all types of links may reduce the effects of a single characteristic of opponent individuals.

I focus more on β which captures preferences to homophily. The coefficient β_{caste} is negative and significant in all types of social networks. The results indicate that individuals have strong preference to homophily in caste. However, caste does not seem to be the strongest factor for the segregation patterns of the networks. Indeed, religion is the strongest for FR, and VISIT networks, and work status is the strongest for MONEY, KERICE, and ADVICE networks. It makes a sense that individuals strongly prefer ones with similar work status in favor exchange networks. Overall, most β's are negative and significant, but β_{gender} is insignificant in most networks. Also, I find that β_{gender} is positive in MH (medical help) network. I conjecture that female members may play an important role in medical treatments in rural areas. Unlike α, β is significant in ALL network. Aggregation of different types of links does not reduce homophily.

The additional utility Δ from having mutual friends is large and significant in all types of networks. The effect of one additional mutual friend is almost as strong as that of β_{caste} in most networks, and at least about a half of β_{caste} in all types of networks (except ALL network).

2.7 Conclusion

I use the notion of pairwise stability to identify and estimate the preferences on network formation. Pairwise stability not only facilitates the identification of the model but also delivers a computationally feasible way to estimate the structural parameters. I propose a semiparametric maximum score estimator for strategic formation of a network. The results from the semiparametric maximum score estimation show that individuals have strong caste homophily for all types of networks. The size of caste homophily slightly differs by the types of networks.

CHAPTER 3

A STRUCTURAL MODEL OF MULTIGRAPH FORMATION: FAVOR EXCHANGE AND SOCIAL NETWORKS IN VILLAGES

3.1 Introduction

This chapter 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.[1] 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 chapter 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 chapter, I consider a multigraph resulting from the strategic decisions of economic agents in a village setting, and I estimate the utility parameters of the

[1]For 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.

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.[2] 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 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 the previous chapter, pairwise stabil-

[2]When 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.

ity 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 the previous chapter 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 chapter does not require an excluded instrument. I conduct inference using an estimation method developed by Andrews and Soares (2010), AS henceforth. Since this chapter 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.[3] 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 chapter are as follows: (1) What is the effect of friendship 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 chapter's application, the econometrician does not have access to a valid instrument for the friendship network. The structural model in this chapter 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

[3]Abhijit Banerjee; Arun Chandrasekar; Esther Duflo; Matthew Jackson, 2011-08, Social Networks and Microfinance, http://hdl.handle.net/1902.1/16559 UNF:5:4EmgOYAQGaoQugFowckNfA== Jameel Poverty Action Lab [Distributor] V5 [Version]

sharing and favor exchange networks is inconclusive.[4]

Related Literature

First, this chapter is related to the literature on strategic network formation. There are many types of theoretical models of strategic network formation in the literature. Myerson (1991) proposes a simultaneous link announcement game. Jackson and Wolinsky (1996) study pairwise stability and efficiency of networks. Bala and Goyal (2000) present a one-sided and non-cooperative link-formation model. Watts (2001) and Jackson and Watts (2002b) examine a dynamic network formation and stochastic evolution of networks. Bloch and Jackson (2006) and Calvó-Armengol and İlkılıç (2009) investigate relations between stability and equilibrium concepts in network formation. This chapter contributes to the theory literature by proposing pairwise stability of a multigraph (PSM), which will be defined in Section 3. It is a simple extension of the notion of pairwise stability of a single network in Jackson and Wolinsky (1996) to a multigraph setting.

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

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

a stochastic best response dynamic (See Blume 1993). Based 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.[5] In the previous chapter, I use pairwise stability as a stability notion and achieve point identification of model parameters by checking pairwise stability conditions. This chapter extends the approach of the previous chapter 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 chapter is widely applicable to many different settings where a multigraph is present.

This chapter 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

[5] 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.

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 chapter 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 chapter 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 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 chapter 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 chapter 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 chapter 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 chapter 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 Chapter

The rest of the chapter 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 B.

3.2 Multigraph

3.2.1 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.[6] 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 simultaneous 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

[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.

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 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.

3.2.2 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 sth 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 multigraph is

$$\begin{aligned} U_i(Y|W, V, X, \varepsilon) &= \sum_{j=1}^{n} \Bigg[\sum_{s=1}^{p} a_{ij}^{(s)} u_{ij}^{(s)} + \sum_{t \neq s}^{p} \sum_{s=1}^{p} a_{ij}^{(s)} a_{ij}^{(t)} \delta_{ij}^{(s,t)} \\ & \sum_{r \neq s,t}^{p} \sum_{t \neq s}^{p} \sum_{s=1}^{p} a_{ij}^{(s)} a_{ij}^{(t)} a_{ij}^{(r)} \delta_{ij}^{(s,t,r)} + \cdots + \prod_{s=1}^{p} a_{ij}^{(s)} \delta_{ij}^{(1,\cdots,p)} \Bigg], \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} u_{ij}^{(s)} &= u^{(s)}(W_{ij}, V_{ij}, X_{ij}, Y_{-ij}, \varepsilon_{ij}^{(s)}), \\ \delta_{ij}^{(s,t)} &= \delta^{(s,t)}(W_{ij}, V_{ij}, X_{ij}, Y_{-ij}, \varepsilon_{ij}^{(s,t)}), \\ \delta_{ij}^{(s,t,r)} &= \delta^{(s,t,r)}(W_{ij}, V_{ij}, X_{ij}, Y_{-ij}, \varepsilon_{ij}^{(s,t,r)}) \\ &\vdots \\ \delta_{ij}^{(1,\cdots,p)} &= \delta^{(1,\cdots,p)}(W_{ij}, V_{ij}, X_{ij}, Y_{-ij}, \varepsilon_{ij}^{(1,\cdots,p)}) \end{aligned}$$

In the above expression, the base utility $u_{ij}^{(s)}$, $s = 1, \cdots, p$ captures i's utility from j when i has a type-s relationship with j. It may include the cost of maintaining a link. Additional utilities, $\delta_{ij}^{(s,t)}, \delta_{ij}^{(s,t)}, \cdots, \delta_{ij}^{(1,\cdots,p)}$ capture the effects of having multiple relationships $(s, t), (s, t, r), \cdots, (1, \cdots, p)$, respectively with a partner j. Both the base and additional utilities may depend on the following components: a pair ij's current relations in the underlying multigraph structure (W, V), observed characteristics and externalities due to other pairs' decisions Y_{-ij}. All observed characteristics are collected in a $\kappa_x \times \binom{N}{2}$ matrix $X = (X'_{12}, \cdots, X'_{n-1,n})'$, where X_{ij} is an $\kappa_x \times 1$ vector of observed characteristics for i and j. The term $\varepsilon_{ij}^{(s)}$ is a match-specific unobserved variable for type-s relation, and $\varepsilon_{ij}^{(s,t)}$ is for both type-s and type-t relations. These variables are public information to individuals but unobserved to the econometrician. These unobserved variables can be viewed as a personality match between i and j. This personality match differs across relations, and there are additional unobserved components when they are linked in multiple relations. Note that ε_{ij}'s are match-specific, i.e. $\varepsilon_{ij} = \varepsilon_{ji}$. 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]

Note that the utility function (3.1) is very flexible and has very few restrictions. Utilities from different individuals and base utilities across different networks are additively separable. However, the utility function is able to capture a rich set of non-additive structures across individuals as well as different networks, since the base utility $u_{ij}^{(s)}$ depends on Y_{-ij} that contains other pairs' decisions in the type-s network as well as other types of networks. Thus, it can cap-

[7]The assumption can be relaxed to that of both match- and individual- specific unobservables, i.e. $\varepsilon_{ij} \neq \varepsilon_{ji}$. 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.

ture utility from having a three people triad or a four people rectangle within the same type or across different types of relations. The additional utility $\delta_{ij}^{(s,t)}$ from multiple relations allows non-additive separability across networks. Finally, the base utility and the additional utilities allow for non-additively separable unobservables.

Let $Y_{ij}^{(-s)} = (a_{ij}^{(1)}, \cdots, a_{ij}^{(s-1)}, a_{ij}^{(s+1)}, \cdots, a_{ij}^{(p)})$ 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 $Y_{ij}^{(-s)}$.

Example (Two simultaneous determined networks in a multigraph, i.e. $p = 2$) Individual i's utility from a multigraph Y is

$$U_i(Y|W, V, X, \varepsilon) = \sum_{j=1}^{n} \left[a_{ij}^{(1)}u_{ij}^{(1)} + a_{ij}^{(2)}u_{ij}^{(2)} + a_{ij}^{(1)}a_{ij}^{(2)}\delta_{ij}^{(1,2)}\right] \tag{3.2}$$

Individual i's marginal utility from having the first relationship given that they already have the second relationship is

$$mu_i^{(1)}(j|0) := mu_i^{(1)}(j|a_{ij}^{(2)} = 0) = u_{ij}^{(1)}$$

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

$$mu_i^{(1)}(j|0) + mu_j^{(1)}(i|0) = u_{ij}^{(1)} + u_{ji}^{(1)}. \tag{3.3}$$

3.3 Pairwise Stability of a Mutigraph

3.3.1 Definition

In this section, I focus only on simultaneously determined networks Y. Recall that $Y_{ij} = (a_{ij}^{(1)}, \cdots, a_{ij}^{(p)})'$ 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.

I provide pairwise stability of a multigraph with transferable utility (PSMt) in this section and pairwise stability of a multigraph without transferable utility (PSMnt) in Appendix B.2.

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

$$U_i(Y) + U_j(Y) \geq U_i(Y - Y_{ij} + Y'_{ij}) + U_j(Y - Y_{ij} + Y'_{ij}) \text{ for all } Y_{ij},\ Y'_{ij}(\neq Y_{ij}) \in \mathcal{Y}_{ij}.$$

Condition (i) means that when i and j have a set of relations Y_{ij}, it must be

[8]If some links in Y'_{ij} are already present in Y, I add non-existing ones only.

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 PSMt and PSMnt 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 relations: 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 has a benefit in 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.

3.3.2 Existence

The results on the existence of a pairwise stable multigraph in this section rely on the results of Jackson and Watts (2001, 2002b). First, I introduce a few definitions which extend similar ones from Jackson and Watts (2001, 2002b) 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 3.3.1. *Fix U. 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 B.1.1.

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 3.3.2. *Fix U. 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 function $\omega : \mathcal{Y} \to \mathbb{R}$ such that [Y' defeats Y] $\Leftrightarrow$[$\omega(Y') > \omega(Y)$, 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 B.1.2.

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. When utility is transferable, by restricting externalities in $u_{ij}^{(s)}$ and $u_{ij}^{(s,t)}$, I can find a function $\omega(\cdot)$ under the utility specification (3.1).

Assumption 3.1. (i) Externalities from Y_{-ij} in $u_{ij}^{(s)}$ are additively separable and affect the base utility $u_{ij}^{(s)}$, for all $s = 1, \cdots, p$, only linearly through $\sum_k a_{ik}^{(t)} a_{jk}^{(t)}$, $t = 1, \cdots, p$. That is,

$$u_{ij}^{(s)} := u^{(s)}(W_{ij}, V_{ij}, X_{ij}, \varepsilon_{ij}^{(s)}) + \sum_{t=1}^{p} \gamma^{(t,s)} \sum_k a_{ik}^{(t)} a_{jk}^{(t)}, \; s = 1, \cdots, p \tag{3.4}$$

(ii) The additional utilities from multiple relations do not depend on Y_{-ij}.

Assumption 3.1.(*i*) restricts that the base utility depends only on the number of mutual partners between i and j across relations. For example, the number of mutual risk sharing partners between i and j affects not only i's base utility from having j as a risk sharing partner but also i's base utility from having j as a partner in relations other than risk sharing. Assumption 1. (*ii*) implies that other pairs' decisions in Y are already considered in the base utility and do not have an additional impact on having multiple relations.

Proposition 3.3.3. *Let U_i be defined as (3.1). Suppose that Assumption 1 holds. Then, there exists at least one pairwise stable multigraph with transferable utility.*

Proof. See Appendix B.1.3.

By Proposition 3.3.3, at least one pairwise stable multigraph exists. Note that Proposition 3.3.3 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 $\sum_k a_{ik}^{(s)} a_{jk}^{(s)}$, and (3) the additional utility of having multiple relations does not depend on the rest of the multigraph Y_{-ij} . 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.

Before moving on to the next section, 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 3.5.1.

3.4 A Multinomial Choice Model under PSMt

3.4.1 A Multinomial Structure under PSMt

From this section, I set $p = 2$ to simplify the model for ease of explanation. When p is bigger than two but still small, the results in this section can be easily extended with only several more steps.[9] I also 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 $\mathcal{Y}_{ij}$ (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 $(1, 0) \succsim (1, 1)$.[10]

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

$$[(1,\ 1) \succsim (0,\ 1)] \quad \Rightarrow \quad mu_i^{(1)}(j|a_{ij}^{(2)} = 1) + mu_j^{(1)}(i|a_{ij}^{(2)} = 1) \geq 0, \tag{3.5}$$

$$[(1,\ 1) \succsim (1,\ 0)] \quad \Rightarrow \quad mu_i^{(2)}(j|a_{ij}^{(1)} = 1) + mu_j^{(2)}(i|a_{ij}^{(1)} = 1) \geq 0, \tag{3.6}$$

$$[(1,\ 1) \succsim (0,\ 0)] \quad \Rightarrow \quad mu_i^{(1,2)}(j) + mu_j^{(1,2)}(i) \geq 0 \tag{3.7}$$

[9] 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.

[10] I use the weak preference relation $\succsim$ in the sense that utility of one choice is as great as that of the others. When the distribution of ε is absolutely continuous, this discretion is unnecessary.

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

$$[(1, 0) \succsim (1, 1)] \Rightarrow mu_i^{(2)}(j|a_{ij}^{(1)} = 1) + mu_j^{(2)}(i|a_{ij}^{(1)} = 1) \leq 0, \tag{3.8}$$

$$\begin{aligned}[(1, 0) \succsim (0, 1)] \Rightarrow\ & mu_i^{(1)}(j|a_{ij}^{(2)} = 0) + mu_j^{(1)}(i|a_{ij}^{(2)} = 0) \\ & \geq mu_i^{(2)}(j|a_{ij}^{(1)} = 0) + mu_j^{(2)}(i|a_{ij}^{(1)} = 0),\end{aligned} \tag{3.9}$$

$$[(1, 0) \succsim (0, 0)] \Rightarrow mu_i^{(1)}(j|a_{ij}^{(2)} = 0) + mu_j^{(1)}(i|a_{ij}^{(2)} = 0) \geq 0. \tag{3.10}$$

The case of $(a_{ij}^{(1)}, a_{ij}^{(2)}) = (0, 1)$ is symmetric to the above. That is,

$$[(0, 1) \succsim (1, 1)] \Rightarrow mu_i^{(1)}(j|a_{ij}^{(2)} = 1) + mu_j^{(1)}(i|a_{ij}^{(2)} = 1) \leq 0, \tag{3.11}$$

$$\begin{aligned}[(0, 1) \succsim (1, 0)] \Rightarrow\ & mu_i^{(1)}(j|a_{ij}^{(2)} = 0) + mu_j^{(1)}(i|a_{ij}^{(2)} = 0) \\ & \leq mu_i^{(2)}(j|a_{ij}^{(1)} = 0) + mu_j^{(2)}(i|a_{ij}^{(1)} = 0),\end{aligned} \tag{3.12}$$

$$[(0, 1) \succsim (0, 0)] \Rightarrow mu_i^{(2)}(j|a_{ij}^{(1)} = 0) + mu_j^{(2)}(i|a_{ij}^{(1)} = 0) \leq 0. \tag{3.13}$$

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

$$[(0, 0) \succsim (1, 1)] \Rightarrow mu_i^{(1,2)}(j) + mu_j^{(1,2)}(i) \leq 0, \tag{3.14}$$

$$[(0, 0) \succsim (1, 0)] \Rightarrow mu_i^{(1)}(j|a_{ij}^{(2)} = 0) + mu_j^{(1)}(i|a_{ij}^{(2)} = 0) \leq 0, \tag{3.15}$$

$$[(0, 0) \succsim (0, 1)] \Rightarrow mu_i^{(2)}(j|a_{ij}^{(1)} = 0) + mu_j^{(2)}(i|a_{ij}^{(1)} = 0) \leq 0. \tag{3.16}$$

Since each pair makes a decision among multiple (2^p) alternatives, and the set of relations chosen gives the maximum sum of utilities among all alternatives, the model has a multinomial structure. However, the model is not exactly a typical multinomial choice model, e.g. multinomial probit, because the marginal utilities depend on on Y_{-ij} and may have non-additively separable ε_{ij}.

3.4.2 A Multinomial Choice Model under PSMt

In the end of Section 3.3.2., I briefly mentioned the benefits of using pairwise stability of a multigraph. In this section, I will show that the structural model is equivalent to a typical multinomial choice model under PSMt and additional assumptions. It will be done through three steps. First, I impose additive separability of ε_{ij} and demonstrate that for any values of $\varepsilon_{ij} \in \mathbb{R}^{2^p-1}$, 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 additional assumption of myopic agents. Finally, I explain the irrelevance of other pairwise stable multigraphs.

Assumption 3.2. (i) ε_{ij} is additively separable.

(ii) ε is a continuously distributed r.v. with everywhere positive density w.r.t. Lebesgue measure.

(iii) $\varepsilon_{ij} = (\varepsilon_{ij,1}, \varepsilon_{ij,2}, \cdots, \varepsilon_{ij,2^p-1})$ is i.i.d. across pairs, but not i.i.d. across the alternatives.

Assumptions 3.2.(i) and 3.2.(ii) are not strong assumptions. The i.i.d. assumption rules out the role of unobserved individual personality in multigraph formation. It is a strong assumption, but prevalent in the literature of empirical network formation, e.g. Christakis et al. (2010). From Assumption 3.2.(i), the base utility and the additional utility are written as

$$u_{ij}^{(s)} := u^{(s)}(W_{ij}, V_{ij}, X_{ij}) + \sum_{t=1}^{p} \gamma^{(t,s)} \sum_{k} a_{ik}^{(t)} a_{jk}^{(t)} + \varepsilon_{ij}^{(s)},$$

and

$$\delta_{ij}^{(s,t)} := \delta^{(s,t)}(W_{ij}, V_{ij}, X_{ij}) + \varepsilon_{ij}^{(s,t)}.$$

Then, the sum of marginal utilities in (3.5)-(3.16) are written as, for example,

$$\begin{aligned} mu_i^{(1)}(j|a_{ij}^{(2)} = 1) + mu_j^{(1)}(i|a_{ij}^{(2)} = 1) \quad = \quad & u^{(1)}(W_{ij}, V_{ij}, X_{ij}) + u^{(1)}(W_{ij}, V_{ij}, X_{ji}) \\ & +2 \sum_{t=1}^{2} \gamma^{(t,s)} \sum_{k} a_{ik}^{(t)} a_{jk}^{(t)} + \delta^{(s,t)}(W_{ij}, V_{ij}, X_{ij}) \\ & +\delta^{(s,t)}(W_{ij}, V_{ij}, X_{ji}) + 2\varepsilon_{ij}^{(1)} + 2\varepsilon_{ij}^{(1,2)}. \end{aligned}$$

For notational simplicity, I use $\varepsilon_{ij}(Y_{ij})$ to denote the sum of unobservables which corresponds to Y_{ij}, for example, $\varepsilon_{ij}(1, 1) = \varepsilon_{ij}^{(1)} + \varepsilon_{ij}^{(2)} + \varepsilon_{ij}^{(1,2)}$. Then, I have

$$\begin{aligned} mu_i^{(1)}(j|a_{ij}^{(2)} = 1) + mu_j^{(1)}(i|a_{ij}^{(2)} = 1) \quad = \quad & u^{(1)}(W_{ij}, V_{ij}, X_{ij}) + u^{(1)}(W_{ij}, V_{ij}, X_{ji}) \\ & +2 \sum_{t=1}^{2} \gamma^{(t,s)} \sum_{k} a_{ik}^{(t)} a_{jk}^{(t)} + \delta^{(s,t)}(W_{ij}, V_{ij}, X_{ij}) \\ & +\delta^{(s,t)}(W_{ij}, V_{ij}, X_{ji}) \\ & +2\varepsilon_{ij}(1, 1) - 2\varepsilon_{ij}(0, 1). \end{aligned} \tag{3.17}$$

The right hand side of (3.17) is the same form as that of a typical multinomial choice model.

Now consider a dynamic process of multigraph formation as in Sections 3.2 and 3.3.2. In the process, a pair of individuals is chosen at each period, and the pair revises its current links given the current multigraph. The link formation (or revision) in each period occurs as a two people cooperative game. When they revise, they do not consider the future changes in the multigraph, i.e. agents are myopic. 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 ε_{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 $\varepsilon_{ij} \in \mathbb{R}^{2^p-1}$.

Lemma 3.4.1. *The utility function is defined as (3.1). Let Assumption 2 hold. Then, given W_{ij}, V_{ij}, X_{ij} and Y_{-ij}, each pair's links decision Y_{ij} is uniquely determined under PSMt for all values of $\varepsilon_{ij} \in \mathbb{R}^{2^p-1}$,*

Proof. Consider only the case with $p = 2$ for ease of exposition. It can be easily seen that the following four regions of ε corresponding to (3.5)-(3.7), (3.8)-(3.10), (3.11)-(3.13), and (3.14)-(3.16) 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 ε_{ij}. However, Lemma 1 does not mean that a pairwise stable multigraph Y is uniquely determined given W, V and X for all $\varepsilon = (\varepsilon_{12}, \cdots, \varepsilon_{n-1,n})$. 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. It is mainly because of the nature of the game. In the link formation game of a pair, two individuals compare the sum of their utilities across alternatives and choose one that gives the maximum utility. In the entry game, on the other hand, two firms separately make an entry decision in a market without cooperation.

The remaining problem is the simultaneity among al pairs' links decisions, $(Y_{12}, Y_{13}, \cdots, Y_{n-1,n})$. At this point the benefit of pairwise stability of a multigraph arises. When checking pairwise stability of a multigraph, I do not need to consider no-deviation conditions for a set of individuals larger than a pair. Checking pairwise stability of each pair's current links decision Y_{ij} only requires comparison of utilities between Y and $Y - Y_{ij} + Y'_{ij}$. Hence, if a multigraph is pairwise stable, it is possible to consider the multigraph as if it is formed by myopic agents. This 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 $\mathcal{PSM}$. Since pairwise stability of a multigraph is not an equilibrium solution concept, the set $\mathcal{PSM}$ 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 3.3.2, 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 ε such as the normal or the logistic distribution, the parameter vector θ is point-identified under suitable location and scale normalizations.[12] I consider the case with an endogenous variable formally in the next section.

3.5 Identification and Estimation

3.5.1 Identification with an Endogenous Explanatory Variable and No Instruments

I have shown that under pairwise stability of a multigraph, the structural model of strategic multigraph formation is equivalent to a multinomial choice model.[13] This section focuses on the identification and the estimation of the multinomial

[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 ε_{ij} is i.i.d. normal across pairs and alternatives with zero-mean and variance-covariance matrix $I_{2^{p-1}}$, θ is point-identified.

[13]The notations used in this section follow ones in BMM12 and CRS.

choice model. I build the analysis in this subsection on CRS. For ease of exposition, let[14]

$$
\begin{aligned}
\tilde{N} &= \sum_{m=1}^{M} \binom{n_m}{2}, \text{ the number of all pairs,} \\
W_{ij} &= (a_{ij}^{(p+1)}, \cdots, a_{ij}^{(p+q)})', \text{ endogenous relations between } i \text{ and } j, \\
Z_{ij} &= (V_{ij}', X_{ij}', Y_{-ij}')', \text{ the collection of all exogenous variables,}
\end{aligned}
$$

where

$$
\begin{aligned}
V_{ij} &= (a_{ij}^{(p+q+1)}, \cdots, a_{ij}^{(S)})', \text{ exogenous relations between } i \text{ and } j, \\
X_{ij} &= (x_{1,i}, \cdots, x_{\kappa_x,i}, x_{1,j}, \cdots, x_{\kappa_x,j})', \text{ the observed characteristics of } i \text{ and } j, \\
Y_{-ij} &= (a_{-ij}^{(1)}, \cdots, a_{-ij}^{(p)})', \text{ other pairs' links decisions in } Y.
\end{aligned}
$$

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

Assumption 3.3. (i) (Y, W, Z, ε) is defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The support of Y_{ij} is a finite set $\mathcal{Y}$. The support of (W, Z, ε) is $\mathcal{W} \times \mathcal{Z} \times \mathbb{R}^{2^p-1}$.

(ii) The true data generating processes, $F_{Y,W|Z}^0$ and $F_{W|Z}^0$, are identified by the sample.

(iii) ε and Z are statistically independent. No excluded instrumental variables are available for W.

[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}.

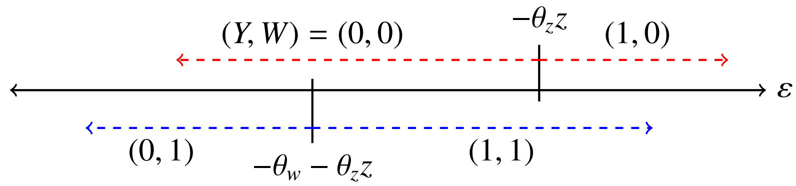

Figure 3.1: An Example of a Binary Choice Model When $\theta_w > 0$

Define the structural equation g as

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

That is, $g(\cdot)$ generates a pairwise stable outcome y for each pair given w, z and ε, which provides the maximum sum of utilities to a pair of individuals, compared to other alternatives y'. Now, define a random set

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

In other words, it is a set-valued random variable which represents model predictions. Let the inverse function of $Q_\theta(\varepsilon, z)$ be

$$\mathcal{E}_\theta(Y, W, z) \equiv \{\varepsilon : \; g(W, z, \varepsilon; \theta) = Y\}.$$

Since Y and W are random variables, the inverse function is also a random set. For example, consider a simple case where $\mathcal{Y} = \{0, 1\}$ and $\mathcal{W} = \{0, 1\}$. The sum of utilities of a pair is simplified as $U_i(Y_{ij} = 1|w, z, \varepsilon) + U_j(Y_{ij} = 1|w, z, \varepsilon) = \theta_w w + \theta_z z + \varepsilon$. Hence, $y = 1\{\theta_w w + \theta_z z + \varepsilon > 0\}$. Due to the endogeneity of W, the model predicts multiple outcomes for all regions of ε. For example, in Figure

3.1, when θ_w is positive, the random set $Q_\theta(\varepsilon, z)$ takes multiple values of (Y, W), e.g. $Q_\theta(\varepsilon, z) = \{(0,0), (1,1)\}$ when $\varepsilon \in [-\theta_w - \theta_z z, -\theta_z z]$.

Let $\text{Sel}(Q_\theta)$ and $\text{Sel}(\mathcal{E}_\theta)$ be all selections of Q_θ and $\mathcal{E}_\theta$, respectively. Then, a parameter vector θ is in the identification region if and only if the model prediction associated with a parameter vector θ contains the observed outcome. That is, $(Y, W) \in \text{Sel}(Q_\theta)$. From the inverse relation between $Q_\theta(\varepsilon|z)$ and $\mathcal{E}_\theta(Y, W, z)$, I have

$$(Y, W) \in \text{Sel}(Q_\theta) \quad \Leftrightarrow \quad \varepsilon \in \text{Sel}(\mathcal{E}_\theta).$$

Let $C(\mathbb{R}^{2^p-1})$ be the collection of all closed subsets of $\mathbb{R}^{2^p-1}$. For any $F \in C(\mathbb{R}^{2^p-1})$, let $P_\varepsilon(F)$ be the probability of the event $\{\varepsilon \in D\}$ corresponding to a probability distribution P_ε. By the Artstein's inequality Artstein (1983), a candidate distribution of ε, P_ε is the distribution of $\text{Sel}(\mathcal{E}_\theta)$ if and only if

$$P_\varepsilon(F|Z = z) \geq \Pr[\mathcal{E}_\theta(Y, W, z) \subseteq F;\ F^0].$$

That is, the probability distribution of the unobservables ε dominates the (lower) probability distribution of the random set $\mathcal{E}_\theta(Y, W, z)$. From the statistical independence between ε and Z, I have

$$P_\varepsilon(F) \geq \Pr[\mathcal{E}_\theta(Y, W, z) \subseteq F;\ F^0]. \tag{3.18}$$

The right hand side probability is equal to the sum of the probabilities $\Pr[Y = y|W = w, Z = z; F^0]$ corresponding to all sets $\mathcal{E}_\theta(Y, W, z)$ contained **entirely** within

F. When W is continuous, it can be written as

$$\Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq F;\ F^0] \;=\; \int\limits_{w\in\mathcal{W}} \left(\sum_{y\in\mathcal{Y}} 1[\mathcal{E}_\theta(y,\ w,\ z) \subseteq F] \Pr[Y = y|W = w,\ Z = z; F^0] \right) dF^0_{W|Z}(w|z).$$

When W is discrete,

$$\Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq F;\ F^0] \;=\; \sum_{w\in\mathcal{W}} \left(\sum_{y\in\mathcal{Y}} \left\{ 1[\mathcal{E}_\theta(y,\ w,\ z) \subseteq F] \Pr[Y = y|W = w,\ Z = z; F^0] \right\} \times \Pr(W = w|Z = z;\ F^0) \right), \tag{3.19}$$

All duples $(\theta,\ P_\varepsilon)$ in the identification region must satisfy the inequality (3.18) for all z. The advantage of using random sets is that the random set theory provides the sharpness of the identification region, which will be explained in Proposition 3.5.1. In addition, the above inequality (3.19) can be written concisely with the containment functional (or the capacity functional) as in (3.18). See Appendix B.1.4., for the definitions of a random closed set, the containment and capacity functionals. I characterize the sharp identification region of admissible duples $(\theta,\ P_\varepsilon)$ as in BMM12 and CRS.

Proposition 3.5.1. *Let Assumptions 3.1-3.3 hold. Then, the sharp identification region for $(\theta,\ P_\varepsilon)$ associated with $F^0_{Y|W}$ is given by*

$$\mathrm{H}[(\theta, P_\varepsilon)] \;=\; \Big\{ (\theta, P_\varepsilon) \in \Theta \times \mathcal{P} |\, P_\varepsilon(F) \geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq F;\ F^0],\ \forall F \in \mathcal{C}(\mathbb{R}^{2^p-1})\, a.e.\, z \in \mathcal{Z} \Big\}. \tag{3.20}$$

Proof. See Appendix B.1.5.

One may already notice that it is infeasible to check the inequalities in (3.20) for all closed subsets in $\mathbb{R}^{2^p-1}$. However, I only need to consider the inequality for a finite number of sets F. 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 $C_1(\theta) = \{\mathcal{E}_\theta(Y, W, z) : y \in \mathcal{Y}, w \in \mathcal{W}$, 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 $\mathcal{E}_\theta(Y, W, z)$'s. Let this power set of $\mathcal{E}_\theta(Y, W, z)$'s be $C_2(\theta)$. The sharp identification region is obtained by considering all closed sets in $C(\theta)$. 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 $C_2(\theta)$: the entire set, the empty set, non-connected sets and duplicated sets. Proposition 3.5.1, which is similar to Theorem 2 in CRS, provides a way of finding the core determining class.

Definition (Molchanov and Molinari, 2013) A family of closed sets $\mathcal{M}$ is said to be a core determining class for a random closed set $\mathcal{E}$ if any probability measure μ satisfying the inequalities

$$\mu(F) \geq P(\mathcal{E} \subset F)$$

for all $F \in \mathcal{M}$, is the distribution of a selection of $\mathcal{E}$.

Proposition 3.5.2. *Let Assumptions 3.1-3.3 hold. Define* $C_1(\theta) = \{\mathcal{E}_\theta(Y, W, z) : y \in$

$\mathcal{Y}$, $w \in \mathcal{W}$, *and* $z \in \mathcal{Z}\}$. *Then all connected unions of sets in* $C_1(\theta)$ *except* $\mathbb{R}^{2^p-1}$ *yield the core determining class* $\mathcal{M}(\theta)$.

Proof. See Appendix B.1.5.

Then, the sharp identification region is written as

$$\begin{aligned} \mathrm{H}[(\theta, P_\varepsilon)] &= \Big\{(\theta, P_\varepsilon) \in \Theta \times \mathcal{P} | P_\varepsilon(D) \geq \Pr[\mathcal{E}_\theta(Y, W, z) \subseteq F;\ F^0], \\ &\quad \forall F \in \mathcal{M}(\theta)\, a.e.\, z \in \mathcal{Z}\Big\}\,. \end{aligned} \tag{3.21}$$

This size reduction is important. Suppose that the cardinality of $\mathcal{Z}$ 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 $C_2(\theta)$. 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 $254 \times 32 = 8128$ moment inequalities, and computation becomes burdensome. Furthermore, some sets in $C_2(\theta)$ 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 $\mathcal{M}(\theta)$ has a total of 36 sets as opposed to 254. Note that it is more than a 85% reduction.

3.5.2 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 chapter fits into the framework of AS.

The moment inequalities in the model are written as

$$E_{F^0}[m(Y, W, Z; \theta)] \geq 0,$$

where

$$\begin{aligned} m(Y, W, Z; \theta) &= (m_{1,1}(Y, W, z_1; \theta), \cdots, m_{1,L}(Y, W, z_1; \theta), \\ &\quad m_{2,L}(Y, W, z_2; \theta), \cdots, m_{K,L}(Y, W, z_K; \theta))' \end{aligned}$$

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

Assumption Assumption 3.4. (i) $\theta \in \Theta \subset \mathbb{R}^d$, where d is the dimension of θ,

(ii) $\{(Y_{ij}, W_{ij}, Z_{ij} : \forall ij\}$ are i.i.d. under F^0,

(iii) $\sigma^2_{F,k,l}(\theta) = var_F(m_{k,l}(Y, W, Z; \theta)) \in (0, \infty)$ for $k = 1, \cdots, K$ and $l = 1, \cdots, L$,

(iv) $E_F|m_{k,l}(Y, W, Z; \theta)/\sigma_{F,k,l}(\theta)|^{2+\delta} < \infty$ for some $\delta > 0$ for all $k = 1, \cdots, K$ and $l = 1, \cdots, L$.

(v) $\varepsilon \sim N(0, \Sigma_\varepsilon)$.

Note that Assumption 3.4. (i), (iii) and (iv) are not restrictive, and that Assumption 3.4. (ii) holds due to the myopic agent assumption. From Assumption 3.4. (v), the distribution of the unobservable is known up to a finite vector of parameters, and this satisfies the framework of AS. Under Assumption 3.4, I employ the method of AS for estimating the structural parameter θ 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 3.7 and Appendix B.4. One can also find the estimation procedures and asymptotic properties of the estimator in AS.

3.6 Empirical Application

3.6.1 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

[15] Actually, I estimate both θ and P_ε, especially the variance-covariance matrix Σ_ε of ε. I assume that the distribution of ε belongs to a parametric family. For the sake of exposition, Σ_ε is subsumed into θ at least in this section.

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 the previous chapter. 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 chapter 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.

3.6.2 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 (FR) 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 (KIN) 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 (= 36×36) 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 3.1 shows the distribution of the geodesic distance in the data. In-

[16]The 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.

Geodesic Distance	1	2	3	4	5+
Number of Pairs	60,426	342,349	804,560	461,196	65,430
%	3.48%	19.74%	46.40%	26.60%	2.77%

Table 3.1: The Distribution of Geodesic Distance

evitably, 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 3.2-3.4 focus on two relationships, and Table 5 on more than three relationships. From Table 3.2, 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 rows 2, 3, and 6 in Table 3.5. [17] Table 3.3 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.

RS\ FR	0	1	KR \ FR	0	1
0	1,695,066	8,718	0	1,691,909	9,390
1	10,406	19,771	1	13,563	19,099

Table 3.2: Descriptive Statistics: Number of Pairs across Two Networks (1)

[17] Note that there are still many pairs that are considered as both friends and relatives although the questionnaires are designed to avoid such answers.

RS\ KR	0	1
0	1,693,817	9,967
1	7,482	22,695

Table 3.3: Descriptive Statistics: Number of Pairs across Two Networks (2)

RS\ KIN	0	1	KR \ KIN	0	1
0	1,695,812	7,972	0	1,694,335	6,964
1	14,701	15,476	1	16,178	16,484

Table 3.4: Descriptive Statistics: Number of Pairs across Two Networks (3)

3.7 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 $u^{(s)}(W_{ij}, V_{ij}, X_{ij}; \boldsymbol{\beta}^{(s)})$. That is,

$$u^{(s)}(W_{ij}, V_{ij}, X_{ij}; \boldsymbol{\beta}^{(s)}) = \beta_0^{(s)} + \sum_{k=1}^{\kappa_x} \beta_{1,k}^{(s)}(x_{k,i} - x_{k,j})^2 + \sum_{t=p+1}^{S} \beta_{2,t-p}^{(s)} a_{ij}^{(t)}. \tag{3.22}$$

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 kth characteristics, and $\beta_{2,t-p}^{(s)}$ captures the effects of the tth endogenous or exogenous network.

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

$$\delta^{(s,t)}(W_{ij}, V_{ij}, X_{ij}) = \delta^{(s,t)}. \tag{3.23}$$

Types	Number of Pairs	Percent
Risk Sharing Only	4,267	0.25
Kerosene-Rice Only	6,415	0.37
Friendship Only	6,786	0.39
Kinship Only	6,204	0.36
Risk Sharing, Kerosene-Rice Only	3,859	0.22
Risk Sharing, Friendship Only	2,507	0.14
Risk Sharing, Kinship Only	663	0.04
Kerosene-Rice, Friendship Only	1,836	0.11
Kerosene-Rice, Kinship Only	1,672	0.10
Friendship, Kinship Only	97	0.01
RS, KR, FR Only	4,068	0.23
RS, KR, KIN Only	1,617	0.09
KR, FR, KIN Only	44	0.00
RS, KR, FR, KIN	13,151	0.76
None	1,680,775	96.93
Total	1,733,961	100

Table 3.5: Descriptive Statistics: Number of Pairs with Multiple Relations

The number of mutual partners in type-s network 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 (3.23) into the utility function (3.1), individual i's utility from A_1 (risk sharing) and A_2 (kerosene-rice) is written as

$$\begin{aligned} U_i(Y|W, V, X, \varepsilon; \theta) &= \sum_{j=1}^{N} a_{ij}^{(1)} (u^{(1)}(W_{ij}, V_{ij}, X_{ij}; \boldsymbol{\beta}^{(1)}) + h'_{ij}\gamma^{(1)}) \\ &\quad + \sum_{j=1}^{N} a_{ij}^{(2)} (u^{(2)}(W_{ij}, V_{ij}, X_{ij}; \boldsymbol{\beta}^{(2)}) + h'_{ij}\gamma^{(2)}) \\ &\quad + \sum_{j=1}^{N} a_{ij}^{(1)} a_{ij}^{(2)} \delta^{(1,2)} + \varepsilon_{ij}\{(a_{ij}^{(1)}, a_{ij}^{(2)})\}, \end{aligned} \tag{3.24}$$

I use the following variables from the data.

$$
\begin{aligned}
Y_{ij} &= (a_{-ij}^{(RS)}, a_{-ij}^{(KR)})', \text{ RS and KR exchange between } i \text{ and } j. \\
W_{ij} &= a_{ij}^{(FR)}, \text{ friendship tie between } i \text{ and } j. \\
V_{ij} &= a_{ij}^{(REL)}, \text{ kinship tie between } i \text{ and } j. \\
X_{ij} &= (caste_i - caste_j)^2, \text{ caste difference between } i \text{ and } j,
\end{aligned}
$$

where the variable $caste_i$ 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

$$
u^{(s)}(W_{ij}, V_{ij}, X_{ij}; \boldsymbol{\beta}^{(s)}) = \beta_0^{(s)} + \beta_1^{(s)}(caste_i - caste_j)^2 + \beta_{2,1}^{(s)} a_{ij}^{(FR)} + \beta_{2,2}^{(s)} a_{ij}^{(REL)}.
$$

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

$$
\begin{aligned}
\frac{1}{2}\{mu_i^{(1)}(j|0) + mu_j^{(1)}(i|0)\} &= \beta_0^{(1)} + \beta_1^{(1)}(caste_i - caste_j)^2 + \beta_{2,1}^{(1)} a_{ij}^{(FR)} + \beta_{2,2}^{(1)} a_{ij}^{(REL)} \\
&\quad + \gamma_1^{(1)} h_{ij}^{(1)} + \gamma_2^{(1)} h_{ij}^{(2)} + \varepsilon_{ij}(1,0) - \varepsilon_{ij}(0,0). \qquad (3.25)
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{1}{2}\{mu_i^{(1)}(j|1) + mu_j^{(1)}(j|1)\} &= \beta_0^{(1)} + \beta_1^{(1)}(caste_i - caste_j)^2 + \beta_{2,1}^{(1)} a_{ij}^{(FR)} + \beta_{2,2}^{(1)} a_{ij}^{(REL)} \\
&\quad + \gamma_1^{(1)} h_{ij}^{(1)} + \gamma_2^{(1)} h_{ij}^{(2)} + \delta^{(1,2)} + \varepsilon_{ij}(1,1) - \varepsilon_{ij}(0,1). \qquad (3.26)
\end{aligned}
$$

Since only an order among the sum of marginal utilities across alternatives matters, I use the average of marginal utilities (3.25) when conducting inference.

Before implementing estimation of the structural model with endogenous friendship, I run an independent multinomial probit model using `mprobit` in STATA, which has *i.i.d.* errors across alternatives, i.e. $\Sigma_\varepsilon = I_3$. I run the model with four different specifications. For the first two specifications, I restrict the additional benefit of having both relationships $\delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij})$ to be homogeneous across pairs, i.e. $\delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij}) = \delta^{(1,2)}$. The next two specifications allow for heterogeneity in $\delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij})$. That is,

$$\delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij}) = \delta_0 + \delta_1(caste_i - caste_j)^2 + \delta_{2,1}a_{ij}^{(FR)} + \delta_{2,2}a_{ij}^{(REL)} + \delta_{3,1}h_{ij}^{(1)} + \delta_{3,2}h_{ij}^{(2)}. \tag{3.27}$$

For each case, one specification does not include friendship $a_{ij}^{(FR)}$, while the other includes it. Table 3.6 shows the estimation results.

Friendship: With the specifications that restrict $\delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij})$, friendship gives positive utility to both risk sharing and favor exchange relationships. On the other hand, if I do not impose the restriction on $\delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij})$, friendship gives negative utility to both risk sharing and favor exchange relationships when each pair forms only one of those relationships. Friendship provides positive utility for risk sharing and favor exchange networks only when a pair builds both relationships together. In sum, if two individuals are friends, they are less likely to form only one of two relationships, risk sharing and favor exchange.

Caste: Interestingly, the coefficient on the variable 'caste difference', or $(caste_i - caste_j)^2$ is positive and significant for the risk sharing network, but negative and significant for the favor exchange network in all four specifications. That is, individuals have caste homophily when they exchange kerosene

Type of Relationships		(1)	(2)	(3)	(4)
Risk Sharing	Friendship		0.9119***		-0.0457*
	Caste Difference	0.1398***	0.1035***	0.1297***	0.1290***
	Relative	0.3486***	0.3621***	-0.9451***	-0.9409***
	Mutual RS Partners	0.4035***	0.3272***	0.2585***	0.2321***
	Mutual FV Partners	0.0294**	0.0472***	0.0556***	0.0451**
	_cons	-1.1705***	-1.4639***	-0.7673***	-0.7327***
Favor Exchange	Friendship		0.3603***		-0.5148***
(Kerosene-Rice)	Caste Difference	-0.2448***	-0.2619***	-0.1870***	-0.1703***
	Relative	0.1502***	0.1908***	-0.8507***	-0.8219***
	Mutual RS Partners	0.2069***	0.1790***	0.1043***	0.1087***
	Mutual FV Partners	0.6093***	0.6166***	0.6091***	0.5852***
	_cons	-1.3038***	-1.4153***	-0.9394***	-0.7839***
Both Relationships	Additional Benefit ($\delta^{(1,2)}$)	1.0925***	1.0080***		
	Friendship				1.1830***
	Caste Difference			-0.1537***	-0.2255***
	Relative			0.4413***	0.4557***
	Mutual RS Partners			0.5712***	0.4456***
	Mutual FV Partners			0.6473***	0.6628***
	_cons			-1.4135***	-1.8743***
Number of Pairs= 60, 426					

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ $^{***}p < 0.001$

(1) Independent multinomial probit with $\delta(X_{ij}, Y_{-ij}, \varepsilon_{ij}) = \delta$.
(2) Independent multinomial probit with $\delta(X_{ij}, A^{(FR)}, Y_{-ij}, \varepsilon_{ij}) = \delta$.
(3) Independent multinomial probit with $\delta(X_{ij}, Y_{-ij}, \varepsilon_{ij})$ as in equation (3.27).
(4) Independent multinomial probit with $\delta(X_{ij}, A^{(FR)}, Y_{-ij}, \varepsilon_{ij})$ as in equation (3.27).

Table 3.6: Estimation Results from Multinomial Probit when $\Sigma_{\varepsilon} = I$

and/or rice, but they prefer different caste for money. The coefficient on caste difference for having both relationships is negative, so individuals have caste homophily when forming both relationships together.

Relative: The variable 'relative' has very similar effects as friendship. It gives negative utility when a pair has a single relationship but it gives positive utility when they have both relationships. The negative effect of being a relative on having only a single relationship is very strong for both risk sharing and favor exchange relationships. From this result, I conjecture that if two individuals have a kinship, then they have a strong incentive to have many different types of mutual agreements with each other rather than only a single one.

Other variables: The number of mutual partners in risk sharing and favor exchange respectively give positive utility to risk sharing, favor exchange and both relationships. Mutual partners in one relation is a strong determinant of having the same relation, but a relatively weak determinant for having the other relation. The additional benefits of having multiple relationship under the homogeneity restriction is positive and significant.

3.8 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.3.

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 F_l, I have

$$\begin{aligned} E(m_{k,l}(Y, W, z; \theta)|z = z_k) &= P_\varepsilon(F_l) - \sum_{w\in\mathcal{W}}\sum_{y\in\mathcal{Y}}(1[\mathcal{E}_\theta(y, w, z_k) \subseteq F_l] \\ &\quad \times \Pr[Y = y|W = w, Z_{ij} = z_k; F^0]\Pr[W = w|Z = z_k; F^0]\Big). \end{aligned}$$

The unconditional moment $E(m_{k,l}(Y, W, z; \theta))$ is written as

$$\begin{aligned} E(m_{k,l}(Y, W, z; \theta)) &= \Big\{P_\varepsilon(F_l) - \sum_{w\in\mathcal{W}}\sum_{y\in\mathcal{Y}}(1[\mathcal{E}_\theta(y, w, z_k) \subseteq F_l] \\ &\quad \times \Pr[Y_{ij} = y|W_{ij} = w, Z_{ij} = z_k; F^0]\Pr[W_{ij} = w|Z_{ij} = z_k; F^0]\Big)\Big\} \\ &\quad \times \Pr[Z_{ij} = z_k; F^0]. \end{aligned}$$

The corresponding sample moment for a pair ij is

$$\begin{aligned} m_{ij,k,l}(Y, W, z; \theta) &= \Big\{1[\varepsilon_{ij} \in F_l] - \sum_{w\in\mathcal{W}}\sum_{y\in\mathcal{Y}}\Big(1[\mathcal{E}_\theta(y, w, z_{ij}) \subseteq F_l] \\ &\quad \times 1[Y_{ij} = y|W_{ij} = w, Z_{ij} = z_k; F^0]1[W_{ij} = w|Z_{ij} = z_k; F^0]\Big)\Big\} \\ &\quad \times 1[Z_{ij} = z_k; F^0]. \end{aligned}$$

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

$$m_{ij}(\theta) = (m_{1,1}(\theta), \cdots, m_{L,1}(\theta), m_{1,2}(\theta), \cdots, m_{L,\kappa_Z}(\theta))'.$$

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

$$m_{ij}(\theta) = (m_{ij,1,1}(\theta), \cdots, m_{ij,L,1}(\theta), m_{ij,1,2}(\theta), \cdots, m_{ij,L,\kappa_Z}(\theta))'.$$

I also compute the estimator $\hat{\Sigma}_{\tilde{N}}(\theta)$ 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,

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

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

The estimation procedure is summarized below. (i) Draw $R = 100$ unobservables for each observation ($\tilde{N}$ by 300 matrix) from $N(0, I_3)$. These draws stay fixed over all repetitions. (ii) Draw the 18 by 1 parameter vector θ_0 from the uniform distribution on $[-10, 10]^{18}$. The parameter vector includes five elements of a three by three matrix $var(\varepsilon)$. It is not six but five due to the scale normalization, i.e. $var(\varepsilon_1) = \sigma_1^2 = 1$. (iii) I pre-multiply ε by Λ, where Λ is the Cholesky decomposition of Σ_ε, i.e. $\Sigma_\varepsilon = \Lambda\Lambda^T$. (iv) For each value of θ_0, compute the sample moments $\bar{m}_{\widetilde{N}}$ ($36 \times 36 = 1296$ by 1 vector) and their sample variance-covariance matrix $\hat{\Sigma}_{\tilde{N}}$ (1296 by 1296). (v) Compute the test statistic $S(\sqrt{\tilde{N}}\bar{m}_{\widetilde{N}}(\theta_0), \hat{\Sigma}_{\tilde{N}}(\theta_0))$. (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 $M^*_{\widetilde{N},b} = \tilde{N}^{1/2}(\hat{D}^*_{\tilde{N},b}(\theta_0))^{-1/2}(\bar{m}^*_{\tilde{N},b}(\theta_0) - \bar{m}_{\widetilde{N}}(\theta_0))$ and $\hat{\Omega}^*_{\tilde{N},b} = (\hat{D}^*_{\tilde{N},b}(\theta_0))^{-1/2}\hat{\Sigma}^*_{\tilde{N},b}(\theta_0)(\hat{D}^*_{\tilde{N},b}(\theta_0))^{-1/2}$, where $\hat{D}^*_{\tilde{N},b}(\theta_0) = \text{Diag}(\hat{\Sigma}^*_{\tilde{N},b}(\theta_0))$. (viii) Implement the generalized moment selection (GMS) procedure: If $\tilde{N}^{1/2}\bar{m}_{n,j}(\theta_0)/\hat{\sigma}_{n,j}(\theta_0) > \kappa_n = (\ln\tilde{N})^{1/2}$, eliminate the corresponding moments from $(M^{*}_{\widetilde{N},b}(\theta_0), \hat{\Omega}^{*}_{\widetilde{N},b}(\theta_0))$ and denote $(M^{**}_{\widetilde{N},b}(\theta_0), \hat{\Omega}^{**}_{\widetilde{N},b}(\theta_0))$. (ix) Compute $S(M^{**}_{\widetilde{N},b}(\theta_0), \hat{\Omega}^{**}_{\widetilde{N},b}(\theta_0))$. (x) If $S(\sqrt{\tilde{N}}\bar{m}_{\widetilde{N}}(\theta_0), \hat{\Sigma}_{\tilde{N}}(\theta_0)) > \hat{c}_{\tilde{N},1-\alpha}$, then reject θ_0, where $\hat{c}_{\tilde{N},1-\alpha}$ is the $1-\alpha$ quantile of $\{S(M^{**}_{\widetilde{N},b}(\theta_0), \hat{\Omega}^{**}_{\widetilde{N},b}(\theta_0)), b = 1, \cdots, B\}$.

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 $Q_{\tilde{N}}(\theta) = S(\bar{m}_{\widetilde{N}}, \hat{\Sigma}_{\tilde{N}}) - \hat{c}_{\tilde{N},1-\alpha}$. After many iterations, I collect all parameter values that satisfy $Q_{\tilde{N}}(\theta) \leq 0$, and these values comprise the $1-\alpha\%$ confidence set. I project the confidence set for each parameter to obtain confidence intervals.[18]

Estimation Results

I provide confidence intervals for parameters in the sharp identification region in Table 3.7. The confidence intervals are obtained by projecting the confidence

[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 tth draw of the parameter vector, let $d_t = 1$ if θ_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 `libsvm` in R to get the 95% confidence region $\{\theta \in \Theta : \hat{f}(\theta) = \sum_{t=1}^{T} \hat{\alpha}_t d_t K(\theta, \theta_t) + \hat{\beta}_{svm} \geq 0\}$, where $K(\cdot,\cdot)$ is the Gaussian kernel (the default option in `libsvm`). For the confidence interval of each scalar-valued component of θ, I solve the optimization problems

$$\begin{aligned} \min \quad & \theta_k \\ \text{s.t.} \quad & \hat{f}(\theta) \geq 0, \end{aligned}$$

and

$$\begin{aligned} \max \quad & \theta_k \\ \text{s.t.} \quad & \hat{f}(\theta) \geq 0. \end{aligned}$$

The results from the SVM will be provided soon.

set of the sharp identification region of parameters for each parameter. Table 3.7 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 B.5. 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.

		95% CI
Risk Sharing	intercept	[-1.1345, 1.1060]
	friendship	[-0.9022, 1.9584]
	caste difference	[-1.4220, 1.0025]
	kinship	[-1.9960, 3.7901]
	mutual RS partners	[-0.8286, 1.1871]
	mutual FV partners	[-1.4085, 0.4146]
Kerosene or Rice	intercept	[-1.8675, 0.8126]
	friendship	[-0.9284, 1.7274]
	caste difference	[-0.9869, 1.4099]
	kinship	[-1.0308, 3.3371]
	mutual RS partners	[-1.4884, 1.4435]
	mutual FV partners	[-1.2236, 1.0060]
Both	intercept	[-2.7208, 0.6705]
	friendship	[-1.5301, 3.2381]
	caste	[-2.4089, 2.1096]
	kinship	[-2.9427, 4.4427]
	mutual RS partners	[-2.3159, 1.5129]
	mutual FV partners	[-1.6189, 1.0427]
	additional utility from both	[-0.9195, 1.0491]

Table 3.7: Projection of 95% CIs for the Sharp Identification Regions of Parameters

Second, when I restrict the sign of $\beta^{(1)}_{friend}$ in the risk sharing network to be positive, $\beta^{(2)}_{friend}$ for kerosene-rice is estimated to be positive as well. Similarly, if I

		$\beta^{(1)}_{friends} < 0$	$\beta^{(1)}_{friends} > 0$
Risk Sharing	intercept	[-0.5268, 1.1060]	[-1.1345, 0.8782]
	friendship	[-0.9022, -0.0815]	[0.0060, 1.9584]
	caste difference	[-1.4220, 1.0025]	[-1.3880, 0.9011]
	kinship	[-0.7369, 2.0000]	[-1.9960, 3.7901]
	mutual RS partners	[-0.4923, 0.8254]	[-0.8286, 1.1871]
	mutual FV partners	[-0.4985, 0.4146]	[-1.4085, 0.3115]
Kerosene or Rice	intercept	[-1.6538, 0.8126]	[-1.8671, 0.3222]
	friendship	[-0.9284, -0.0044]	[0.0023, 1.7274]
	caste difference	[-0.9869, 1.4099]	[-0.2759, 1.1740]
	kinship	[0.8944, 3.3371]	[-1.0308, 1.3696]
	mutual RS partners	[-0.4904, 1.2870]	[-1.4884, 1.4435]
	mutual FV partners	[-1.2236, 0.4173]	[-0.9158, 1.0060]
Both	intercept	[-1.1694, 0.4639]	[-2.7208, 0.6705]
	friendship	[-1.5301, -0.1048]	[0.1189, 3.2381]
	caste	[-2.4089, 2.1096]	[-1.1835, 1.5939]
	kinship	[0.5686, 4.4427]	[-2.9427, 4.2320]
	mutual RS partners	[-0.3088, 1.1074]	[-2.3159, 1.5129]
	mutual FV partners	[-1.6056, 0.5784]	[-1.6189, 1.0427]

Table 3.8: Projection of 95% CIs for Parameters when $\beta^{(1)}_{friends} < 0$ and $\beta^{(1)}_{friends} > 0$

restrict the sign of $\beta^{(1)}_{friend}$ to be negative, the sign of $\beta^{(2)}_{friend}$ again is estimated to be the same as the sign of $\beta^{(1)}_{friend}$. 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 same way. See Table 3.8.

Figure 3.2 shows another interesting result. In the computational illustrations of the sharp identification region in Appendix E, the true parameter lies in the larger region when the sharp identification region is given by two components. In Figure 3.2, 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.

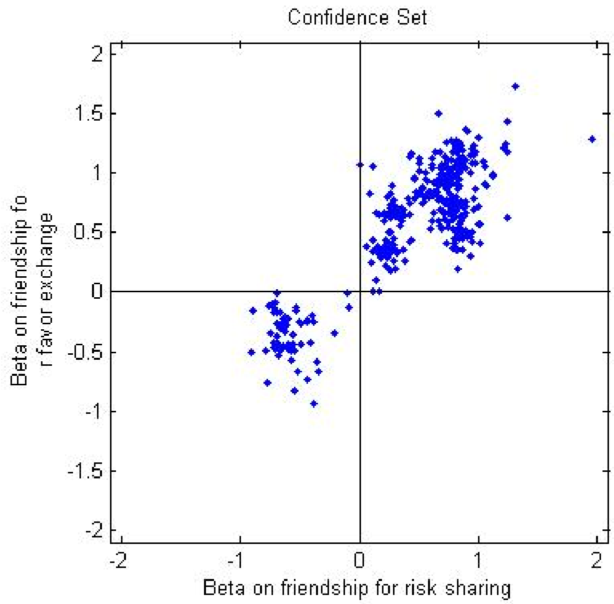

Figure 3.2: The Projected Confidence Region for $(\beta^{(1)}_{friend}, \beta^{(2)}_{friend})$

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 3.9.

		95% CI
Risk Sharing	intercept	[-1.1345, 1.1060]
	friendship	[-0.9022, 1.2465]
	caste difference	[-1.4220, 1.0018]
	relative	[0.5163, 3.7901]
	mutual RS partners	[-0.5279, 1.1871]
	mutual FV partners	[-1.1061, 0.2883]
Kerosene or Rice	intercept	[-1.7254, 0.8126]
	friendship	[-0.7536, 1.4957]
	caste difference	[-0.9869, 1.4099]
	relative	[0.1708, 3.2353]
	mutual RS partners	[-0.5607, 1.2870]
	mutual FV partners	[-1.2236, 1.0060]
Both	intercept	[-2.7208, 0.6705]
	friendship	[-1.5301, 2.4607]
	caste	[-2.4089, 2.1096]
	relative	[1.0808, 4.4427]
	mutual RS partners	[-0.7800, 1.3078]
	mutual FV partners	[-1.6189, 0.9036]

Table 3.9: Projection of 95% CIs for Parameters when $\beta_{rel}^{(1)} > 0$ and $\beta_{rel}^{(2)} > 0$

	Risk Sharing	Kerosene or Rice	Both
Risk Sharing	1	[-1.9516, 1.2448]	[-1.3954, 1.2480]
Kerosene or Rice		[0.6740, 4.5173]	[-1.2907, 2.8282]
Both			[1.5430, 6.7667]

Table 3.10: 95% Confidence Interval for Covariance Matrix of ε

I also estimate the variance-covariance matrix of the unobservables. See Table 3.10. The confidence intervals are relatively wide especially for variances. I cannot rule out independence of ε across alternatives. However, I am able to reject the hypothesis that ε from different alternatives have an identical distribution since $var(\varepsilon_3) = 1$ does not belong to the 95% confidence interval. The estimate for $var(\varepsilon_{ij}(1,1))$ seems larger than $var(\varepsilon_{ij}(1,0))$ and $var(\varepsilon_{ij}(0,1))$.

3.9 Conclusion

In this chapter 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 simple 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 build 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 chapter 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 chapter. 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 chapter, one can solve potential bias due to the endogeneity of a social network. Finally, the model in this chapter 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 may be interesting to incorporate a particular multigraph formation process into the model and predict what multigraph will be formed as a counterfactual analysis.

CHAPTER 4

ESTIMATING NETWORK EXTERNALITIES IN THE U.S. AIRLINE INDUSTRY

4.1 Introduction

This chapter studies network formation in the U.S. airline industry, and investigates the size of network externalities. The U.S. airline industry has experienced dramatic changes since the Airline Deregulation Act was enacted in October 1978. Major airlines have transformed their networks into *hub-and-spoke networks* after the deregulation act allowed them to choose their own routes and fares. The hub-and-spoke network is 'a system of connection arranged like a chariot wheel in which all transportation moves along spoke routes connected to a hub airports at the center'.[1] Unlike the legacy airline carriers such as United Airlines and Delta Airlines, Southwest Airlines has built a point-to-point network in which air travelers can fly directly from their origin to destination. Southwest Airlines made a huge success with its point-to-point network in combination of other strategies such as using a single type of aircrafts. More recently, the industry has experienced airline alliances, the advent of low cost carriers (LCCs), mergers among airlines, etc. Delta Airlines has once filed for bankruptcy in 2005, and American Airlines in November 2011. Many other airline carriers have also been suffered from a huge amount of loss in recent years.

I conjecture that airline carriers have been facing these difficulties due to negative network externalities. For example, operating a route between two cities

[1]This definition follows one in Wikipedia. See http://en.wikipedia.org/wiki/Spoke-hub_distribution_paradigm

may provide positive profits for the market, but it may not for other markets that share one of those two cities.[2] Regarding airline networks, economists have investigated advantages and disadvantages of employing a hub-and-spoke network for the last two decades. They measure 'airport presence' or 'hub-size' as the number of direct flights departing from or arriving at an airport and use it as a profit shifter in empirical models. However, hub-size does not perfectly represent each airline's network topology. Indeed, the effects of hub-size are a mixture of positive and negative network externalities. To my knowledge, no papers in the literature consider those two effects of opposite directions.

As a first step of estimating network externalities in the U.S. airline industry, I assume that each airline builds a network which satisfies a weak notion of stability. That is, no airlines want to deviate from their current networks by a single route change at a time. By maintaining this stability, airlines form each link (route) in their network by playing an entry game. Consequently, the definition of entry in this chapter is an operation of a direct flight in a market which is defined as a city pair. Then, I use an entry game as a link (route) formation framework for the network formation of airline carriers.

The stability notion gives different and more reasonable interpretation on why the econometrician can include network measures that capture competitiveness of an airline's network in the post-entry market-level profit function which will be defined later. Each airline considers just one market at a time and maintain stable network. Hence, many network measures derived from the rest of its network can be taken into account. I include various types of network variables including the number of one-stop flights and average hub-size, into the post-entry profit function. Since these various network variables are present

[2] I use city, area and airport interchangeably.

in the profit function, the econometrician can estimate the size of network externalities by investigating an entry game across markets.

In terms of an empirical model for an entry game, I revisit the empirical model of Berry (1992) and estimate the size of network externalities with various network. I use the Airline Origin and Destination Survey (DB1B) for the second quarter of 2013. I focus on the five largest airline carriers: American, Delta, Southwest, United and US Airways. From the recent data, I find that when I control for the number of one-stop flights the effect of hub-size is larger than the case without considering one-stop flights.

This chapter is related to the literature on airline markets and airline network formation. Many authors have targeted the U.S. airline industry since it has not only several interesting features such as government regulation and deregulation, but also publicly available data sets through the Bureau of Transportation Statistics. Borenstein (1989) first studies the importance of airport presence in the U.S. airline industry. Berry (1992) proposes an empirical model of entry that incorporates firm competitiveness. His model is a close benchmark of this chapter. Applying the network formation theory to the airline industry has been also interesting direction in the existing literature. In the literature on network formation, Jackson and Wolinsky (1996) define a notion of pairwise stability a network. The stability notion that I impose for an airline network is similar (but not the same) to pairwise stability of a network. A series of papers by Hendricks, Piccione, and Tan (1995, 1997, and 1999) propose a basic analytic model of network competition in the airline industry, and characterize equilibrium conditions under which a certain network topology such as a hub-and-spoke network is realized. In the empirical models, however, to the best of my knowledge, there are very few attempts to include the entire network structure into an empirical model, although the networks of airlines are observed by the econometrician.

This chapter also stems from the large literature on entry games and applications.[3] Bresnahan and Reiss (1990) assume homogeneity in the interaction effects and focus on the number of firms in a market as a dependent variable to obtain a point identified model. Berry (1992) imposes two ordered-entry assumptions to find an unique equilibrium and estimate a complicated model with the method of simulated moments. Jia (2008) proposes an empirical model which allows for correlation across different markets. Ciliberto and Tamer (2009) employ a partial application approach and allow for heterogeneity in interaction effects, and apply their model to the airline industry. Bajari et al. (2010) estimate an equilibrium selection mechanism in discrete games including entry games. Beresteanu, Molchanov, and Molinari (2011) provide an approach to sharp inference for models with convex moment inequalities including an entry game. Recently, dynamic entry games have been studied by many authors. See Bajari et al. (2010) and Aguirregabiria and Mira (2010) for a survey. Aguirregabiria and Ho (2010, 2012) propose a dynamic model and apply to the airline industry. They include a variable hub size in both demand and production side of the model to capture network externalities. In addition to air. In addition to airline industries, there are a lot of applications to various industries. For example, Jia (2008) and Grieco (2010) study the discount stores industry.

This chapter is organized as follows. Section 2 discusses network structures in the U.S. airline industry. Section 3 sets up the empirical model. Data expla-

[3]The literature on entry games and applications to the airline industry is strongly related to each other.

nation and descriptive statistics are provided in Section 4. Section 5 explains the detailed estimation methods. Section 6 collects all the estimation results and Section 7 concludes.

4.2 Network Structures in the U.S. Airline Industry

4.2.1 Industry Overview

There are several interesting features in the industry, such as the dominance of hub-and-spoke networks, the emergence of so-called low cost carriers. In addition, the industry faces frequent mergers and airline carriers build alliances with each other. Economists have also been interested in government subsidies and regulations. In this section, among many issues I focus on network structures of the U.S. airline carriers.

Hub-and-spoke networks are common characteristics of the airline industry not only in the North American markets but also in other regions of the world. In the U.S. markets, Delta Airlines first started the system at its first hub of Hartsfield-Jackson Atlanta International Airport (ATL) in 1955. After its invention of the new paradigm, combining with the deregulation of the U.S. airline industry in 1978, the new system has been adopted by many other carriers. Currently, major airlines in the United States typically have 3 to 8 hubs except Southwest Airlines, which intends to provide 'point-to-point' networks although it has hub-like airports.[4] Delta has a network which is more concen-

[4]Indeed, Southwest Airlines has more than forty destinations from airports such as Chicago-Midway, Las Vegas, Baltimore-Washington, Phoenix and Denver. Some authors (e.g. Aguirregabiria and Ho (2012)) classified these airports as its hubs.

trated on several hubs such as Detroit, Atlanta, Salt Lake City, Minneapolis, etc. On the other hand, the network of Southwest has a lot of direct links among their destinations. Virgin America, which is relatively new in the industry, builds a hub-and-spoke network with only two hubs (SFO and LAX). Airlines build hub-and-spoke networks since the system is considered to give airlines benefits of lower entry costs and variable costs. On the other hand, it has drawbacks on demand side since passengers have to transfer at a hub airport in order to go to their final destinations, and it obviously increases traveling time. If there is an unexpected delay due to weather or maintenance of airplanes, the increase of traveling time could be extreme.

Although most existing studies indicate that the hub-and-spoke network gives benefits of lower sunk costs, or an irreversible part of entry costs, and the entry deterrence advantage, several U.S. airlines have shown their behaviors which do not seem to be supportive to the literature. The most remarkable example is the huge success of Southwest Airlines with its point-to-point network.[5] Unlike to other airlines in the industry, Southwest Airlines has been trying to connect its operating cities with direct flights instead of stopping once at a hub airport.

In most existing literature, hub-size has been used as a variable in order to identify the effects of airport presence of an airline on its and others' profits. If an airline have a pure hub-and-spoke network, the bigger hub size would deter other airlines' entry decision. Additionally, all city pairs, or markets do not have a substitute from its own. For example, consider an airline that uses Chicago as a hub and operates a direct flight in Boston-Chicago market. Because

[5] A point-to-point transportation system is a system where a plane travels directly to a destination, rather than going through a central hub.(from Wikipedia)

the airline constructs a pure hub-and-spoke network, there is no substitute with one-stop flights of its own to the route Boston-Chicago. If passengers in Boston want to travel to other cities except Chicago with the airline, they have to stop at Chicago with no other choices. In turn, the large hub-size at Chicago brings only positive externalities through complementarity built in one-stop flights. In this case, the econometrician can precisely identify the effects of hub-size of the airline on its profits. By contrast, let us imagine a network on which an airline has more than one hubs. For example, United Airlines uses San Francisco and Washington DC as its hubs. Consider a passenger who wants to travel from Hartford, CT to San Francisco, CA. She might contemplate between buying a direct flight or one-stop flight which stops at Washington DC. There may also exist other airports such as Denver or Chicago at which she can change her flight without changing airlines. She can choose where to transfer because United has large airport presence at not only San Francisco International airport but also many different airports. Then, the large hub size (at San Francisco) has another effect that it increases the number of substitutes, and in turn, the substitutability may decrease profits from its direct flight between San Francisco and Washington DC.

To summarize, if airlines employ different network architectures, then the marginal effect of airport presence, or hub-size could have different economic implications. Mostly, the marginal effects include both positive (complementarity, lower entry costs, entry deterrence effects, etc) and negative (more substitutes from its own) sides. I incorporate each airline's network structure through the adjacency matrix, and separately estimate the negative effects of large airport presence and its positive effects. This approach brings more precise interpretation of the effects of hub-size as well as the effects of different network

# of Markets Entered	AA	DL	UA	US	WN
Entry (direct flight only)	491	449	1074	1406	812
Entry (regardless of transfer)	2112	3039	2861	2807	1829

Table 4.1: Number of Markets Served by Airlines with Different Entry Definitions

architectures.

4.2.2 Entry Decision and Network Formation

In this chapter, I define entry in a market as operating a direct flight in that market. In the literature, Berry (1992) and Ciliberto and Tamer (2009) define entry as an operation of a flight between two areas (or airports) regardless of the number of transfers. With their definition, an airline is classified as an incumbent in a market where the airline serves two endpoint areas. From Table 4.1, I find substantial differences in the number of entry for all airlines. The number of markets served by airlines in their definition is about two to seven times larger than when using only direct flights. For example, the previous definition counts 3039 markets for Delta, which makes Delta as an incumbent in about 86% of total markets, which is not reasonable. In the literature, the recent papers, e.g. Aguirregabiria and Ho (2010, 2012), define entry same as in this chapter.

There are a few benefits to define entry in this chapter. First, when the econometrician wants to include network variables as a firm-specific profit shifter, the previous definition may cause an simultaneity problem. For example, hub-size and the dependent variable of an entry decision are simultaneously determined. When I define entry as operating a direct flight, I can rule out such simultaneity. Second, the econometrician can estimate the size of network externalities precisely. With the entry definition in this chapter, the effects of network variables are the exact size of spillovers from a network measure to the other part (e.g. a link, or a route) of the network. Once the inclusion of such network variables in the profit function has a ground, the econometrician can estimate the size of network externalities consistently.

Finally, I impose a restriction on each airline's network. In order to estimate the effects of network variables such as hub-size, previous papers in the literature employ a couple of different approaches. One approach is a partial equilibrium proposed in Berry (1992). He argues that it is reasonable to consider each market separately due to computational difficulties. The other approach is that airlines consider their network in the previous period as given, and make their entry decisions for each market.

In this chapter, I suggest a new approach for this problem. I assume that each airline builds a network which satisfies a weak notion of stability. That is, no airlines want to deviate from their current networks by a single route change at a time. This notion of stability is similar to pairwise stability introduced by Jackson and Wolinsky (1996) in the sense that the strength of stability. However, I do not call it pairwise stability, since the deviation is not based on the incentive of a pair of nodes (here, areas). A network is formed by one agent, or an airline, so it is not the same as pairwise stability. I formally state the assumption of stable airline networks.

Assumption Each airline builds a stable network in the sense that no airlines want to deviate from their current networks by a single route change at a time.

Airline carriers build stable networks, and the formation of each route is based on an entry game. With the assumption, I can include network variables as a measure of firm's competitiveness in a market. It is because each airline does not consider a deviation which involve two or more routes at the same time. This assumption is reasonable since the number of possible deviations with two routes is already very large. In addition, when an airline tries to deviate, it has to take into account how its competitors will respond. Given the large number of competitors in the market, to make optimal decision in this setting is infeasible.

An alternative approach is a revealed preference approach. The approach takes the current network configurations of firms as a NE equilibrium, and then any deviation from the current network would give smaller or equal profits to the firm which makes the deviation. I do not pursue the approach in this chapter and leave it as a future study. See Ellickson, Houghton, and Timmins (2013) for the application of this approach in the discount store industry.

4.3 Model

Let $n = 1, \ldots, N$ be an index for airlines. Let $Y_n = (V, E_n)$ be a network of airline n, where $V = \{1, \cdots, C\}$ is the set of all cities, and E_n is the set of airline n's routes. With a slight abuse of notations, the network Y_n can be viewed as an adjacency matrix. That is, Y_n can be written as a C by C matrix, where its ith row and jth column element $y_{n,ij}$ takes value one if airline n operates a direct flight in the city pair (i, j). There are a total of C cities, so the number of city pair is $M = C(C-1)/2$. Recall that this chapter only considers operating a direct flight as an entry in a market. We denote $y_n = (y_{n,12}, y_{n,13}, \cdots, y_{n,C-1,C})'$ as a vector of firm n's decisions across all markets (city pairs).

In terms of modeling firm n's profit function, I follow Berry (1992) and Ciliberto and Tamer (2009). Let the post-entry profit function for firm n in market ij be $\pi_{n,ij}(y_{-n,ij};\theta)$, where $y_{-n,ij}$ is a vector of entry decisions of other firms except n and θ is a finite vector of parameters.

Each market ij is defined by a vector of its characteristics X_{ij} such as average population of a city pair, and the the rest of all networks except $y_{n,ij}$ and $y_{-n,ij}$. The profit function $\pi_{n,ij}$ is written as

$$\pi_{n,ij} = X_{ij}'\beta + \sum_{q\neq n} \Delta_q y_{q,ij} + \gamma' f(y_{n,-ij}) + \eta_{n,ij}. \tag{4.1}$$

The parameter Δ_l captures the interaction effects, and it is often considered as negative. That is, if other airlines are operating a direct flight, firm n's profit is more likely to decrease. The (vector valued) function $f(\cdot)$ measures each firm's competitiveness in terms of its rest of the network. For example, $f(\cdot)$ can be a vector of the average number of routes from and to the endpoint airports i and j, and the number of one-stop flights between i and j. The last term $\eta_{n,ij}$ is a profit shifter of firm n in market ij, which is unobserved by the econometrician. This unobserved variable is decomposed into two components: the market-specific unobservable $\varepsilon_{0,ij}$ and the firm- and market- specific unobserved heterogeneity $\varepsilon_{n,ij}$. These unobservables are distributed with i.i.d standard normal distribution. As in Berry (1992) , I impose a traditional constraint on $\eta_{n,ij}$ such that $var(\eta_{n,ij}) = 1$, and $corr(\eta_{n,ij}, \eta_{l,ij}) = \rho$ for all n, q and ij such that $n \neq q$. Thus, I can decompose $\eta_{n,ij}$ as

$$\eta_{n,ij} = \sqrt{1-\rho^2}\varepsilon_{n,ij} + \rho\varepsilon_{0,ij}.$$

In the literature, it has been shown that the model is not point-identified and has a computational burden even in a two players entry game. See Bresnahan and Reiss (1991a), Berry (1992) and Tamer (2003).[6] Hence, as a starting point of investigating the size of network externalities, I employ the following two assumptions. First, I assume that the interaction parameter Δ_q is homogeneous across firms. Then the post-entry profit depends only on the number of incumbents rather than the identity of incumbents. Bresnahan and Reiss (1990) first impose this homogeneity. Under this assumption the profit function becomes

$$\pi_{n,ij} = X_{ij}'\beta + \Delta \sum_{q \neq n} y_{q,ij} + \gamma' f(y_{n,-ij}) + \sqrt{1-\rho^2}\varepsilon_{n,ij} + \rho\varepsilon_{0,ij}. \tag{4.2}$$

With a slight modification, I follow Berry (1992) to make the post-entry profit depend on the number of incumbents in market ij, or N_{ij} through $\ln(N_{ij})$. That is,

$$\pi_{n,ij} = X_{ij}'\beta + \Delta \ln(N_{ij}) + \gamma' f(y_{n,-ij}) + \sqrt{1-\rho^2}\varepsilon_{n,ij} + \rho\varepsilon_{0,ij}. \tag{4.3}$$

Second, in terms of the order of entry, I assume that the most profitable firms move first. Berry (1992) uses the same assumption as one of two types of ordered entry assumptions in his paper. These two assumptions allow the econometrician easy to compute a sequential-move equilibrium and to estimate the model.

The main interest of this chapter is the set of parameter $\gamma = (\gamma_1, \cdots, \gamma_K)'$. The parameter γ_k captures the effect of the kth characteristics of airline n's network, or network externalities corresponding to the kth network measure. As I mentioned earlier, examples include the average degree of i and j in a network, or equivalently the average hub size, the number of airline n's one-stop flights,

[6] It is an interesting direction for further studies to employ partial identification method as in Ciliberto and Tamer (2009) and Beresteanu et al. (2011) to obtain heterogeneous competitive effects.

eigenvector centrality and so on. By doing so, I can separately estimate the effect of airport presence (hub size) conjectured as positive and the potentially negative effects of having one-stop flights.

4.4 Data

4.4.1 Market Data

The data mainly comes from the Airline Origin and Destination Survey (DB1B). Since I want to investigate the most recent trend of the U.S. airline industry, I use the data of the second quarter of 2013, which is the latest data available. The DB1B data provides a 10% random sample of flight tickets with variables such as distance flown, non-stop distance, fare, the number of passengers, operating, reporting and ticketing carriers, origin, connecting and destination cities. The DB1B data for the second quarter of 2013 has 6,018,517 observations corresponding to 45 carriers that sold tickets in the period. More detailed description of each variable used for estimation is as follows.

- Origin and Destination Airports: Each market is defined as a non-directional city pair. From the Annual Estimates of the Population of Metropolitan and Micropolitan Statistical Areas: April 1, 2010 to July 1, 2011 (CBSA-EST2011-01), I narrow down the number of areas. I rank the top 100 metropolitan statistical areas (MSAs) by population size. Then, I match 100 largest population areas with their corresponding airports. Among these 100 areas, 16 areas do not have a large enough airport, or they are very close to other larger metropolitan areas. I use the criterion of the number of enplanement for 'large enough'. If one area has only one airport and the airport has less than 300,000 annual enplanement, I do not count it in. Also, some areas share airports with their neighboring areas. For example, Springfield, MA has Westover Metropolitan Airport, but its number of operations is very small to exceed a certain threshold to be considered as one separate area. It is more likely that people in the area share Bradley International Airport with those who live in Hartford-West Hartford-East Hartford, CT. Hence I ruled out such areas from the list. However, when I consider the population size for the corresponding airports, I do not add the number of population of excluded area to neighboring areas. I also excluded Honolulu, HI due to its special geographic characteristics.

After these arrangements, a total of 84 areas remain separately. Tables 5-6 describe those 84 areas and their population and corresponding airports with IATA (the International Air Transport Association) abbreviations. A total of 105 airports belong to the selected areas. I sometimes use the largest city in each area to indicate the area that the city belongs to. For example, I use New York or NY for the New York-Northern New Jersey-Long Island area.

- Markets: As I mentioned earlier, markets are defined as a non-directional city pair. Since there are 84 areas, the number of city (area) pairs, or markets is $84 \times (84 - 1)/2 = 3486$.

- Airline Carriers: There is another issue in classifying firms. It is the large number of incumbents and potential entrants of the industry in the second quarter of 2013. There are a total of 45 ticketing carriers in that quarter. I first pick five major airlines: American Airlines, Delta Airlines, Southwest

Airlines, United Airlines and US Airways. There are several middle- and small- sized carriers such as Alaska Airlines. I pick five middle-sized carriers: Alaska Airlines, Jet Blue, Virgin America, Frontier Airlines and Spirit Airlines. AirTran Airways and Hawaiian Airlines are also in this category. However, I consider tickets sold by AirTran as if those tickets were sold by Southwest due to their recent merger. Hawaiian Airlines rarely sells a ticket in those markets. The portion of tickets sold by the rest of the airline carriers is less than 5%, so I do not take into account those small airlines. Since I study the networks of airline carriers, I focus on those five major airlines. I combine other five firms and call it mid-sized airlines. That is, for each market, I find one of those five airlines, which has the largest airport presence, as a competitor for the other five major airlines. Hence, in each market there are a total of six potential entrants and incumbents.

I use the following airline codes: American Airlines (AA), Delta Airlines (DL), United Airlines (UA), US Airways (US) and Southwest Airlines (NW). I use MD for the mid-sized airlines.

- Entry Decisions of Airlines: An incumbent of a market is an airline which is operating non-stop flights between two cities, regardless of directions, frequency and the size of aircrafts. There are three possible ways to count incumbent firms from DB1B data: operating, reporting, and ticketing carriers. I use 'ticket_carrier' shown in non-stop flight tickets as an indicator of an incumbent of a city pair. With this criterion, strictly speaking, entry in a market means that an airline sells a ticket (rather than operating a direct flight) for the market. I use ticketing carrier since there are a few markets in which a subsidized airlines of a major airline carrier operates, for example American Eagle Airlines. Moreover, some such small airlines are operating different flights as multiple major airline carriers. For example, ExpressJet Airlines operates as AA, DL and UA. By using the ticketing carrier criterion, I can easily verify the major airline for which those flights of subsidized airlines operate.

- Population: This variable comes from the Annual Estimates of the Population of Metropolitan and Micropolitan Statistical Areas: April 1, 2010 to July 1, 2011 (CBSA-EST2011-01), I take average population of two endpoint areas, and use it as a demographic variable.
- Income per capita and income change: These variables comes from the census data.
- Non-stop Distance: The non-stop distance between two airports can be found in DB1B data. When there are two or more airports in a statistical area, I use the average distance over all possible combinations. For example, from New York to Dallas, I take average distance over all 8 airports combinations: JFK-DAL, JFK-DFW, EWR-DAL, DWR-DFW, LGA-DAL, LGA-DFW, ISP-DAL and ISP-DFW. For some markets for which DB1B data does not contain their distances, I obtain the missing distance from http://www.webflyer.com/travel/mileage_calculator/.

4.4.2 Network Measures

I include the following network variables as a measure of each airline's competitiveness in a market. Table 4.2 provides descriptive statistics for the variables explained in this section. First, airport presence or hub-size, $hubsize_{n,ij}$ is measured by the average number of airports which airline n flies from or to the endpoint airports of market ij. In a mathematical expression, it is written as

$$Hubsize_{n,ij} = \frac{1}{2} \sum_{k \neq i,j} (y_{n,ik} + y_{n,jk}).$$

Since DB1B provides a ticket data, the number of direct destinations of firm i from an airport can be collected. I use the average of hub size of firm i at origin and destination airports as one of the measures for airport presence of firm n.

The variable $City2$ is a dummy variable which takes value one if an airline operates both areas, and zero otherwise. This is another measure for airport presence of firm n.

$$City2_{n,ij} = 1[\sum_{k \neq i} y_{n,ik} \geq 0] \times 1[\sum_{k \neq j} y_{n,jk} \geq 0].$$

The number of one-stop flights between area i and j is obtained by squaring each firm's adjacency matrix Y_n. Note that this is the maximum possible number of one-stop flights of airline n in a market. In a mathematical expression, the ith row and jth column element of Y_n^2 represents all possible one-stop flights between city i and j. That is,

$$Onestop_{n,ij} = Y^2_{n,ij} = \sum_k y_{n,ik} y_{n,jk}.$$

Note that this is not necessarily a practical number of one-stop flights due to the following reasons. First, some markets are too close to have one-stop flights. For example, there are many possible one-stop routes of major airlines for LA-San Diego. In an extreme case, one could fly from LA to New York and then New

	AA	DL	UA	US	WN
Average # of one-stop flights	1.0843	2.3812	2.5932	8.0138	1.652
Average hub-size	11.69	10.69	25.57	33.48	19.33
Sum of "City2"	2701	3321	3403	3403	2145

Table 4.2: Descriptive statistics: one-stop flights, hub-size, City2

York to San Diego. However, the non-stop distance of the market is short, so this type of one-stop routes may not be attractive. On the other hand, it is still possible that one would take a flight from LA to Phoenix and then next flight from Phoenix to San Diego if the price is reasonable.[7] Hence, I cannot rule out all possible one-stop flights. Rather, I try to reduce this problem by controlling for the distance between two areas.

Second, airlines may not sell one-stop flight tickets for a certain markets. Or, even if airlines sell a ticket, no sales are realized in the period. In this case, the econometrician cannot observe those one-stop flights from the data. For these reasons, the number of one-stop flights derived from airline networks is not a perfect measure, but I consider it as a good proxy after controlling for the distance.

Finally, I include the eigenvector centrality (also called Bonacich centrality) as a firm's competitiveness measure in the profit function. The eigenvector centrality of an area i in airline n's network can be obtained from the following equation.

$$Y_n v_n = \lambda_n v_n,$$

where λ_n is the largest eigenvalue of Y_n and v_n is the corresponding right eigenvector to λ. The ith element of v_n, i.e. $v_{n,i}$, is the eigenvector centrality of area i in

[7]In addition, there are a small portion of people who want to collect mileage with a minimum amount of spending. If a one-stop flight gives much higher mileage relative to its price, they would not mind flying through the one-stop flights.

airline n's network. This eigenvector centrality measures how influential each area is in the network. I conjecture that if two areas are more influential in an airline's network, the airline is more likely to operate a direct flight to connect those two areas.[8]

4.5 Estimation

The estimation procedure in this chapter follows Berry (1992). Although Berry (1992) explains the estimation method, it is still worth reviewing the procedure and explaining potential computation problems for practice. The estimation relies on the method of simulated moments (MSM) introduced by McFadden (1989) and Pakes and Pollard (1989).

First, draw R set of unobserved variables $\varepsilon_{ij}^{(r)} = (\varepsilon_{ij,0}^{(r)}, \varepsilon_{ij,1}^{(r)}, \cdots, \varepsilon_{ij,N}^{(r)})'$, $r = 1, \cdots, R$ for all ij, from the independent and identically distributed standard normal distribution. In practice, I use $R = 20, 100$. It is important that this R set of unobservables remain fixed throughout the estimation process. For each draw of unobservables, compute $\pi_{n,ij}^{(r)}$ for each firm and each market, where the superscript (r) indicates the rth simulation error. Note that minimization algorithm starts with a pre-specified starting values of parameters. It is also important to try many starting values to obtain a global minimum. Next, compute the equilibrium number of firms $\hat{n}(W_{ij}, \theta, \varepsilon_{ij}^{(r)})$ corresponding to each simulation error, and the estimated equilibrium number of firms $\hat{N}_{ij} = \frac{1}{R}\sum_r \hat{n}(W_{ij}, \theta, \varepsilon_{ij}^{(r)})$. The estimated market error is then computed as $\hat{\varepsilon}_{0,ij}(\theta) = \sum_q y_{q,ij} - \hat{N}_{ij}$.

The next step is to find an order of entry for each market. Let $q(n, W_{ij}, \theta, \varepsilon_{ij}^{(r)})$

[8] Estimation with eigenvector centrality will be studied later.

be the ranking of firm n in market ij corresponding to the rth simulation. Then, the estimated probability that firm n is an incumbent in market ij is $\hat{P}_{n,ij} = \frac{1}{R}\sum_r \hat{p}_n(W_{ij}, \theta, \varepsilon_{ij}^{(r)})$, where $\hat{p}_n(W_{ij}, \theta, \varepsilon_{ij}^{(r)}) = 1[\hat{n}(W_{ij}, \theta, \varepsilon_{ij}^{(r)}) \geq q(n, W_{ij}, \theta, \varepsilon_{ij}^{(r)})]$. Finally, the estimated firm- and market- specific error $\hat{\varepsilon}_{n,ij}$ is obtained by $\hat{\varepsilon}_{n,ij}(\theta) = y_{l,ij} - \hat{P}_{n,ij}$.

Now, I compute the moments. Let $g_{ij}(\theta)$ be the ijth sample moment such that

$$g_{ij}(\theta) = \left[\hat{\varepsilon}_{0,ij}(\theta)H_0(W_{ij})', \hat{\varepsilon}_{1,ij}(\theta)H_1(W_{ij})', \cdots, \hat{\varepsilon}_{N,ij}(\theta)H_N(W_{ij})'\right]',$$

where $H(W_{ij})$ is a vector of functions of instruments, or exogenous variables. For the market-specific component $\hat{\varepsilon}_{0,ij}(\theta)$, I use four market-specific variables, the Herfindahl indices in terms of the number of one-stop flights and hub-size, and the sum of dummy variables for operating both cities across all airline carriers. For each firm's unobservable profit shifter, $\hat{\varepsilon}_{n,ij}(\theta)$, I use the observed characteristics of that firm corresponding to the profit function specifications in addition to the variables mentioned above. Hence, the number of total moments is $7 + (K + 7) \times N$, where K is the number of parameters for firm characteristics. This number is bigger than the dimension of parameter vector to assure identification.

The objective function is written as

$$Q(\theta) = G(\theta)'\Omega G(\theta),$$

where $G(\theta) = [\varepsilon_0(\theta)H_0(W)', \varepsilon_1(\theta)H_1(W)', \cdots, \varepsilon_N(\theta)H_N(W)']$. Let $G_M(\theta) =$

$\frac{1}{M}\sum_{j\neq i}\sum_i g_{ij}(\theta)'$. Then, the sample objective function is written as

$$Q_M(\theta) = G_M(\theta)'\Omega G_M(\theta), \tag{4.4}$$

where V is an weighting matrix. In order to estimate $\hat{\theta}_{MSM}$, I first use an identity matrix as Ω and find $\tilde{\theta}$ that minimize the above objective function. Then, compute the estimated weighting matrix

$$\hat{\Omega} = \frac{1}{M}\sum_{j\neq i}\sum_i \left(g_{ij}(\tilde{\theta}) - G_M(\tilde{\theta})\right)\left(g_{ij}(\tilde{\theta}) - G_M(\tilde{\theta})\right)'.$$

Next, plug $\tilde{V}$ into the equation (4.4), and obtain $\hat{\theta}_{MSM}$ that minimize the objective function.

To obtain the asymptotic variance of $\hat{\theta}_{MSM}$, I use the formula in Pakes and Pollard (1989). They show that

$$\sqrt{M}(\hat{\theta}_{MSM} - \theta) \Rightarrow N(0,\ (\Gamma'\Gamma)^{-1}\Gamma' V\Gamma(\Gamma'\Gamma)^{-1}),$$

where Γ is the derivative matrix of $G(\theta)$, V is the variance-covariance matrix of $G(\theta)$. Following Pakes and Pollard (1989), I use

$$\Gamma_{k,M}(\theta) = \sqrt{M}\left\{G_M(\hat{\theta} + \sqrt{M}e_k) - G_M(\hat{\theta})\right\}$$

as an estimator for the kth column of Γ, and

$$\hat{V}_M = (1 + \frac{1}{R})\frac{1}{M}\sum_{j\neq i}\sum_i \left(g_{ij}(\hat{\theta}) - G_M(\hat{\theta})\right)\left(g_{ij}(\hat{\theta}) - G_M(\hat{\theta})\right)'$$

Variable	Estimate
Population	-0.3085
	(0.4134)
Income per capita	2.9411***
	(1.1439)
Growth rate	-2.8889**
	(1.3692)
Distance	0.0015
	(0.0026)
Hub-size	0.0081
	(0.0080)
City2	9.2581***
	(2.0570)
δ	-5.1108***
	(1.5191)
ρ	0.6291
	(1.0613)
Constant	0.2202
	(4.0961)
# of markets	3486

Table 4.3: Results when entry is defined as in Berry (1992)

10%, ** 5%, *** 1% level of significance. Standard errors in parenthesis

as an estimator for V. Note that $(1+\frac{1}{R})$ is multiplied to adjust estimation precision error due to simulation.

4.6 Results

As a benchmark, I use the entry definition of Berry (1992) and estimate the model. Since the definition of entry include one-stop flights, I did not include the number of one-stop flights in the post-entry profit function. Although I have tried to use similar specification to one in Berry (1992), some variables differ due to data limitation.

Table 4.3 shows the estimation results. Note that I use the DB1B data for the second quarter of 2013, and Berry (1992) use the data for the second quarter of 1980. Since more than three decades have passed, it is reasonable to expect several changes in estimate. Especially, compared to Berry (1992), the signs of coefficients on population and constant have been changed, but they are not significant. The changes may also be originated from including more market specific variables such as income per capita and the income growth rate. More interestingly, the effect of hub-size turns out to be small and insignificant. This result provides another reason for me to include other network variables to separate positive and negative network externalities. The number of firms in a market has large negative effects and significant as expected. The coefficients on the variable *City*2 and the number of firms in a market are large in magnitude and significant.

Now, I run the empirical model with the definition of entry provided in this chapter. I use two different specifications, one of which is the same as the previous one. As we have seen in Table 1, there are huge gap in the number of markets served by airlines between two different entry definitions. Hence, it is reasonable to expect obtaining different estimation results from those in Table 4.2.

Table 4.4 shows the results. In the second column of the table, the coefficients on the market-specific variables have been changed substantially. The signs of those coefficients are all different from Table 4.3 and significant. The network variables, hub-size and City2 give positive profits and significant. The coefficients δ on the number of firms in a market becomes positive. This result seems counter-intuitive. I conjecture that it may be because a large market provides large profits.

When I include one-stop flights in the profit function (the third column), the

Variable	Without One-stop	With One-stop
Population	0.2327***	0.6113
	(0.0563)	(1.9971)
Income per capita	-1.7500***	2.3449***
	(0.5124)	(0.2565)
Growth rate	1.3522*	-1.3088
	(0.7076)	(1.4957)
Distance	0.0002	-0.0052***
	(0.0012)	(0.0001)
One-stop		-0.0493
		(0.0879)
Hub-size	0.0894***	0.2057***
	(0.0157)	(0.0333)
City2	0.8058**	0.3952
	(0.3911)	(0.5235)
δ	0.8257***	0.9813
	(0.2066)	(0.8210)
ρ	0.3156***	-0.4058
	(0.0813)	(1.2239)
Constant	-1.7488**	0.6113
	(0.8064)	(1.9971)
# of markets	3486	

Table 4.4: Results when entry is defined as operating a direct flight

10%, ** 5%, *** 1% level of significance. Standard errors in parenthesis

estimation results are much different from the model specification (the second column) without one-stop flights. It may be because firm-specific profit shifters take some amount of explanation power of market-specific variables. The most interesting result is that the magnitude of the coefficient on hub-size becomes more than two times larger, and that of City2 variable becomes about a half and insignificant. On the other hand, the coefficient on one-stop flights is negative although it is insignificant. As I mentioned earlier in Section 4.2, when the model includes the variable hub-size alone, it could contain both positive and negative effects. By adding the variable, one-stop, it seems that the negative effect of having one-stop flights is decomposed from hub-size. In terms of magnitude of those two coefficients, having approximately four one-stop flights between two areas cancels out the positive effects of two flights from or to one of two endpoints airports, or cancels out the positive effects of one flight from both endpoints airports. Again, the number of firms in a market has a positive effect in this specification, but insignificant.

4.7 Conclusion

I estimate network externalities in the U.S. airline industry with an entry game and the weak notion of a stable network. The stability notion allows the econometrician to include various network measures as a firm's competitiveness into the post-entry profit function. I focus on the decomposition of the negative effects and the positive effects of having large airport presence. I find that the effect of hub-size is positive and lager in magnitude when I include one-stop flights compared to when the one-stop variable is not included. This result suggests that negative externalities is very likely to exist. For more precise estimation results, an empirical model with heterogeneous competitive effects is recommended as a further study.

APPENDIX A

APPENDIX FOR CHAPTER 2

A.1 Proofs

A.1.1 Proof of Proposition 2.3.3

Before proving the proposition, I introduce a couple of definitions from Hellmann (2012). Let A^N be the complete network (the network with all possible links), and $\mathcal{A}$ be the set of all possible network configurations. Let $L_i(A) := \{ij \in A | j \in \mathcal{N}\}$ be the set of player i's links in A. I use $L_{-i}(A) := A \backslash L_i(A)$ to denote the set of links in A except individual i's links, and $mu_i(A + l, l)$ to denote i's marginal utility of adding a set of new link l to A. It is either a set of links or a single link. The ordinal convexity (concavity) of U in own links is defined as below.

Definition (Definition 5 in Hellmann (2012)) A utility function U_i of player i is ordinal convex (concave) in own links if $\forall A \in \mathcal{A}$, $\forall l_i \subseteq L_i(A^N - A)$, and $\forall ij \notin A + l_i$,

(i) $mu_i(A + ij, ij) \geq 0 \implies (\Longleftarrow)\, mu_i(A + l_i + ij, ij) \geq 0$,

(ii) $mu_i(A + ij, ij) > 0 \implies (\Longleftarrow)\, mu_i(A + l_i + ij, ij) > 0$.

Another definition from Hellmann (2012) is the ordinal strategic complements (substitutes) property.

Definition (Definition 6 in in Hellmann (2012)) A utility function U_i of player i satisfies the ordinal strategic complements (substitutes) property if for all $A \in \mathcal{A}$ and any set of links $l_{-i} \subseteq L_{-i}(A^N - A)$ it holds that

(i) $mu_i(A + ij, ij) \geq 0 \implies (\Longleftarrow) mu_i(A + l_{-i} + ij, ij) \geq 0$,

(ii) $mu_i(A + ij, ij) > 0 \implies (\Longleftarrow) mu_i(A + l_{-i} + ij, ij) > 0$.

Lemma A.1.1. *(Theorem 1 in Hellmann (2012)) Suppose a profile of utility functions $U = (U_1, \cdots, U_N)$ satisfies ordinal convexity in own links and the ordinal strategic complements property. Then:*

(i) There does not exist a closed improving cycle.

(ii) There exists a PS network.

According to the above lemma, if $U = (U_1, \cdots, U_N)$ satisfies ordinal convexity in own links and the ordinal strategic complements property, then there exists a pairwise stable network. Thus, to show the existence of a pairwise stable network, it is enough to show that the utility function satisfies those two properties.

First consider the ordinal convexity in own links. From (2.2), I get $mu_i(A + ij, ij) = u_{ij} + \varepsilon_{ij} + \Delta \sum_{k \notin \mathbf{N}_i(A)} a_{ik}a_{jk}$. Likewise, $mu_i(A + l_i + ij, ij) = u_{ij} + \varepsilon_{ij} + \sum_{k \notin \mathbf{N}_i(A+l_i)} \Delta_{ik}a_{ik}a_{jk}$. Let $\tilde{A} = A + l_i$ and $\tilde{a}_{ij}$ be the i, jth element of $\tilde{A}$. Then, I know that $a_{jk} = \tilde{a}_{jk}$ for all $j, k \neq i$, since l_i is a set of links which i must involve in. Also, I can easily see that $\tilde{a}_{ij} \geq a_{ij}$ for all i and k, since $A + l_i$ is simply a network which adds extra links of i to A. Hence, $\sum_{k \notin \mathbf{N}_i(A+l_i)} a_{ik}a_{jk} \geq \sum_{k \notin \mathbf{N}_i(A)} a_{ik}a_{jk}$. I know that $\Delta_{ik} \geq 0$ for all i and k, and its magnitude does not depend on the underlying network structure. Thus, $\sum_{k \notin \mathbf{N}_i(A+l_i)} \Delta_{ik}a_{ik}a_{jk} \geq \sum_{k \notin \mathbf{N}_i(A)} \Delta_{ik}a_{ik}a_{jk}$, and $mu_i(A + l_i + ij, ij) \geq mu_i(A + ij, ij)$. If $mu_i(A + ij, ij) \geq 0$ then $mu_i(A + l_i + ij, ij) \geq 0$. The strict inequality relation also holds. Therefore, U_i satisfies ordinal convexity.

Similarly, by letting $\tilde{A} = A + l_{-i}$ I can easily verify that $\sum_{k \notin \mathbf{N}_i(A+l_{-i})} a_{ik}a_{jk} = \sum_{k \notin \mathbf{N}_i(\tilde{A})} \tilde{a}_{ik}\tilde{a}_{jk} \geq \sum_{k \notin \mathbf{N}_i(A)} a_{ik}a_{jk}$. Hence, if $mu_i(A+ij, ij) \geq 0$ then $mu_i(A+l_{-i}+ij, ij) \geq 0$.

It is also true for the strict inequality relation. Therefore, U_i satisfies ordinal strategic complements property. By Lemma 3, there exists a pairwise stable network, and the existence does not depend on the realization of ε. Q.E.D.

A.1.2 Proof of Proposition 2.4.1

Proof. For the ease of exposition, define

$$\begin{aligned} x_1 &= (1,\, x_{1j},\, \cdots,\, x_{lj}) \\ x_2 &= (1,\, x_{1i},\, \cdots,\, x_{li}) \\ z &= ((x_{1i} - x_{1j})^2,\, \cdots,\, (x_{li} - x_{lj})^2). \end{aligned}$$

Let $w_1 = (x_1, z)$, $w_2 = (x_2, z)$, and $\theta = (\alpha, \beta)$, where β includes Δ to make notation simpler. $\tilde{X}_\theta$ can be rewritten as

$$\begin{aligned} \tilde{X}_\theta = \{(w_1, w_2) \in \mathbf{X} \times \mathbf{X} :\ & x_1\alpha + z\beta < 0 \leq x_1\alpha_0 + z\beta_0, \\ \text{and } & x_2\alpha + z\beta < 0 \leq x_2\alpha_0 + z\beta_0\} \end{aligned}$$

Let x_k and x_m be explanatory variables with full support. I use x_{-k} and α_{-k} to denote regressors except x_k and its corresponding coefficients, respectively. Define $x_{-k,m}$ and $\alpha_{-k,m}$ analogously. Without loss of generality, let $\alpha_{0,k} > 0$. The other case is symmetric. I have three cases to consider.

(i) Case $\alpha_k < 0$:

$$x_1\alpha + z\beta < 0 \leq x_1\alpha_0 + z\beta_0 \quad \text{and} \quad x_2\alpha + z\beta < 0 \leq x_2\alpha_0 + z\beta_0 \tag{A.1}$$

$$\begin{aligned} x_{1,-k}\alpha_{-k} + z\beta + x_{1k}\alpha_k < 0 \quad &, \quad 0 \leq x_{1,-k}\alpha_{0,-k} + z\beta_0 + x_{1k}\alpha_{0k}, \\ x_{2,-k}\alpha_{-k} + z\beta + x_{2k}\alpha_k < 0 \quad &\text{and} \quad 0 \leq x_{2,-k}\alpha_{0,-k} + z\beta_0 + x_{2k}\alpha_{0k}. \end{aligned}$$

Equivalently,

$$\begin{aligned} x_{1k} > -\frac{x_{1,-k}\alpha_{-k} + z\beta}{\alpha_k} \quad &, \quad x_{1k} \geq -\frac{x_{1,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}, \\ x_{2k} > -\frac{x_{2,-k}\alpha_{-k} + z\beta}{\alpha_k} \quad &\text{and} \quad x_{2k} \geq -\frac{x_{2,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}. \end{aligned}$$

From the full support condition of x_k, probability that the above inequalities hold is strictly positive. Thus, $\Pr[(x_{ij}, x_{ji}) \in \tilde{X}_\theta] > 0$, and θ_0 is point identified.

(ii) Case $\alpha_k = 0$:

Solving (A.1) now gives

$$\begin{aligned} x_{1,-k}\alpha_{-k} + z\beta < 0 \quad &, \quad x_{1k} \geq -\frac{x_{1,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}, \\ x_{2,-k}\alpha_{-k} + z\beta < 0 \quad &\text{and} \quad x_{2k} \geq -\frac{x_{2,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}. \end{aligned}$$

$$\begin{aligned} x_{1,-k,m}\alpha_{-k,m} + z\beta + x_{1m}\alpha_m < 0 \quad &, \quad x_{1k} \geq -\frac{x_{1,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}, \\ x_{2,-k,m}\alpha_{-k,m} + z\beta + x_{2m}\alpha_m < 0 \quad &\text{and} \quad x_{2k} \geq -\frac{x_{2,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}. \end{aligned}$$

$$x_{1,-k,m}\alpha_{-k,m} + z\beta + x_{1m}\alpha_m < 0 \quad , \quad x_{1k} \geq -\frac{x_{1,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}},$$

$$x_{2,-k,m}\alpha_{-k,m} + z\beta + x_{2m}\alpha_m < 0 \quad \text{and} \quad x_{2k} \geq -\frac{x_{2,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}. \tag{A.2}$$

Solving the above inequalities for x_{1m} and x_{2m} respectively gives conditions for x_{1m} and x_{2m}. Since $f_{x_m|x_k=x}$ has full support for all x, probability that the four inequalities in (A.2) hold simultaneously is strictly positive. Thus, $\Pr[(x_{ij}, x_{ji}) \in \tilde{X}_\theta] > 0$, and θ_0 is point identified.

(iii) Case $\alpha_k > 0$:

Solving (A.1) now gives

$$x_{1k} < -\frac{x_{1,-k}\alpha_{-k} + z\beta}{\alpha_k} \quad , \quad x_{1k} \geq -\frac{x_{1,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}, \tag{A.3}$$

$$x_{2k} < -\frac{x_{2,-k}\alpha_{-k} + z\beta}{\alpha_k} \quad \text{and} \quad x_{2k} \geq -\frac{x_{2,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}}. \tag{A.4}$$

Since θ is not a scalar multiple of θ_0, two bounds in (A.3) and (A.4) are not the same. Hence, the above inequalities provide an interval only if the following conditions hold.

$$-\frac{x_{1,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}} < -\frac{x_{1,-k}\alpha_{-k} + z\beta}{\alpha_k} \quad \text{and} \quad -\frac{x_{2,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}} < -\frac{x_{2,-k}\alpha_{-k} + z\beta}{\alpha_k}$$

I rewrite the first inequality as

$$\frac{x_{1,-k,m}\alpha_{0,-k,m} + z\beta_0 + x_{1,m}\alpha_{0m}}{\alpha_{0k}} > \frac{x_{1,-k}\alpha_{-k} + z\beta + x_{1m}\alpha_m}{\alpha_k}.$$

Solving this for $x_{1,m}$ gives

$$x_{1m} > (\frac{x_{1,-k}\alpha_{-k} + z\beta}{\alpha_k} - \frac{x_{1,-k,m}\alpha_{0,-k,m} + z\beta_0}{\alpha_{0k}})/(\frac{\alpha_{0m}}{\alpha_{0k}} - \frac{\alpha_m}{\alpha_k}),$$

or

$$x_{1m} < (\frac{x_{1,-k}\alpha_{-k} + z\beta}{\alpha_k} - \frac{x_{1,-k,m}\alpha_{0,-k,m} + z\beta_0}{\alpha_{0k}})/(\frac{\alpha_{0m}}{\alpha_{0k}} - \frac{\alpha_m}{\alpha_k}),$$

I can solve the second inequality A.4 for x_{2m} and get similar results. Since $f_{x_m|x_k=x}$ has full support for all x, there always exists x_k, x_m and $x_{-k,m}$ such that

$$-\frac{x_{1,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}} < x_{1k} < -\frac{x_{1,-k}\alpha_{-k} + z\beta}{\alpha_k},$$

and

$$-\frac{x_{2,-k}\alpha_{0,-k} + z\beta_0}{\alpha_{0k}} < x_{2k} < -\frac{x_{2,-k}\alpha_{-k} + z\beta}{\alpha_k}.$$

Hence, $\Pr[(x_{ij},\ x_{ji}) \in \tilde{X}_\theta] > 0$, and θ_0 is point identified.

A.2 Distance Based Utility

A.2.1 Model with Distance-Based Utility

In this appendix, I discuss the model specification under distance-based utility. Individual's utility of a network has a close framework to the one in the connections model of Jackson and Wolinsky (1996). Let x_i be an $L \times 1$ vector of observed characteristics of i. X is an $L \times N$ matrix of observed characteristics of all individuals. The utility function of agent i by forming a network A is written as;

$$U_i(A|X, \varepsilon; \theta) = \sum_{j=1}^{N} a_{ij}(u_{ij} - c + \varepsilon_{ij}) + \sum_j a_{ij} \sum_{k \notin \mathcal{N}_i(A)} a_{jk} \delta v_{ik}, \tag{A.5}$$

where $\mathcal{N}_i(A)$ denotes the set of individuals in which individuals have a direct link with i under the current network configuration A. One can also call this set as the neighborhood of i. i.e.

$$\mathcal{N}_i(A) = \{k \in \mathcal{N} | \, a_{ik} = 1\}. \tag{A.6}$$

The base utility u_{ij} has the same form as (2.3), and c is the cost of maintaining link. I specify the linear cost function for simplicity. The second term exhibits utility from indirect friends, or friends of friends. The decay parameter $\delta \in (0, 1]$ represents the decreasing rate of benefits with respect to the distance between i and k.

I simplify the intrinsic value v_{ik} of indirect friend k to i as $v_{ik} = u_{ik}/n_{ik}$, where n_{ik} is the number of common friends of i and k. By doing so, I equally distribute the benefits from indirect friends equally to all of common friends of i and j. For example, if Paul gives $u_{Paul,Tom}$ amount of utility to Tom as being a friend of Tom's friend, it will be divided by $n_{Paul,Tom}$, or the number of common friends of

Paul and Tom. Individual i's marginal utility of j is

$$U_i(A + ij|X, \varepsilon; \theta) - U_i(A - ij|X, \varepsilon; \theta) = u_{ij} - c + \varepsilon_{ij} + \delta \sum_{k \notin \mathcal{N}_i(A)} a_{jk} \frac{u_{ik}}{n_{ik}}. \quad \text{(A.7)}$$

Note that I cut off one's utility from indirect friends after friends of friends. The reason is mostly computational. Computation of this marginal utility requires counting all possible paths of length two from i to j's neighbors. This computation is burdensome once I include indirect friends with length more than two. Intuitively, an individual knows friends of her friends, but he or she may neither know nor care about friends of friends of her friends and so on. Thus, I do not consider friends of friends of friends in i's utility.

A.2.2 Existence of a Pairwise Stable Network

There are a few differences in distance-based utility. First, the utility function (A.5) allows i's utility of an indirect friend k, or v_{ik} to depend on both i and j's characteristics. Second, the utility function has a form of distance-based utility, so that I can estimate the decay parameter δ and the cost c of having a friend. In spite of these advantages, the utility specification is limited since it does not guarantee the existence of pairwise stable networks. If I restrict $v_{ik} = v \geq 0$ in (A.5), then I can show that there exists at least one pairwise stable network.

Proposition A.2.1. *Consider the utility function (2.3). If $v_{ik} = v \geq 0$ for all i and k, then for all ε there exists at least one pairwise stable network .*

Proof. Proving the proposition is similar to the proof of Proposition 2.3.3. According to Lemma A.1.1 above, if $U = (U_1, \cdots, U_N)$ satisfies ordinal convexity in own links and the ordinal strategic complements property, then there exists a pairwise stable network. Thus, it is enough to show that the utility function satisfies those two properties.

First consider the ordinal convexity in own links. From (2.8) and (2.18), I get $mu_i(A + ij, ij) = u_{ij} - c + \varepsilon_{ij} + \delta\nu \sum_{k \notin \mathcal{N}_i(A)} a_{jk}$. Likewise, $mu_i(A + l_i + ij, ij) = u_{ij} - c + \varepsilon_{ij} + \delta\nu \sum_{k \notin \mathcal{N}_i(A+l_i)} a_{jk}$. Since $a_{jk} \in A + l_i$ greater than equal to $a_{jk} \in A$ for all j and k, $mu_i(A + ij, ij) \leq mu_i(A + l_i + ij, ij)$. Thus, the utility profile U is convex in own links. Second, I consider the strategic complements property. Suppose $mu_i(A+ij, ij) = u_{ij} - c + \varepsilon_{ij} + \delta\nu \sum_{k \notin \mathcal{N}_i(A)} a_{jk} \geq 0$. The marginal utility of adding a link ij to $A + l_{-i}$ can be computed as $mu_i(A + l_{-i} + ij, ij) = u_{ij} - c + \varepsilon_{ij} + \delta\nu \sum_{k \notin \mathcal{N}_i(A+l_{-i})} a_{jk}$. Since $A + l_{-i}$ has at least as many links as A, $\sum_{k \notin \mathcal{N}_i(A+l_{-i})} a_{jk} \geq \sum_{k \notin \mathcal{N}_i(A)} a_{jk}$. Thus, $mu_i(A + l_{-i} + ij, ij) \geq 0$. The strong inequality part holds similarly. Hence, the utility profile U satisfies the strategic complements property, and therefore a pairwise stable network exists.

Although the restriction $\nu_{ik} = \nu \geq 0$ guarantees the existence of a pairwise stable network, I do not impose it. Instead, I use the following proposition to guarantee the existence of a pairwise stable network.

Proposition A.2.2. *Fix X and θ. For any network A, there always exists a set $\mathcal{E}_A$ of $\varepsilon = (\varepsilon_{12}, \varepsilon_{13}, \cdots, \varepsilon_{1N}, \varepsilon_{23}, \cdots, \varepsilon_{N-1,N})$ with $\Pr(\varepsilon \in \mathcal{E}_A | X; \theta) > 0$ such that A is pairwise stable.*

Proof. Pick an arbitrary network A. Since X and θ are fixed, the deterministic parts of marginal utilities for all pairs are fixed. That is, the first three terms on the RHS of the below equation are fixed:

$$U_i(A+ij|X,\varepsilon;\theta) - U_i(A-ij|X,\varepsilon;\theta) = u_{ij} - c + \delta \sum_{k\notin \mathcal{N}_i(A)} a_{jk}\frac{u_{ik}}{n_{ik}} + \varepsilon_{ij}. \tag{A.8}$$

Let $mu_i^d(j|A,X;\theta) = u_{ij} - c + \delta\sum_{k\notin\mathcal{N}_i(A)} a_{jk}\frac{u_{ik}}{n_{ik}}$. I know this term is bounded. If $a_{ij} = 1$, then for $\varepsilon_{ij} \in (-mu_i^d(j|A,X;\theta),\infty) \neq \emptyset$, i's marginal utility of j given X and A_{-ij} is positive. Likewise for $\varepsilon_{ji} \in (-mu_j^d(i|A,X;\theta),\infty) \neq \emptyset$, j's marginal utility of i given X and A_{-ij} is positive. So, $a_{ij} = 1$ is pairwise stable for $(\varepsilon_{ij},\varepsilon_{ji}) \in \mathcal{E}_{ij}^1 = (-mu_i^d(j|A,X;\theta),\infty) \times (-mu_j^d(i|A,X;\theta),\infty)$.

If $a_{ij} = 0$, then for $\varepsilon_{ij} \in (-\infty, -mu_i^d(j|A,X;\theta)) \neq \emptyset$ and $\varepsilon_{ji} \in (-\infty,-mu_j^d(i|A,X;\theta)) \neq \emptyset$ $i's$ marginal utility and j's marginal utility are non-positive, respectively. Thus, for $(\varepsilon_{ij},\varepsilon_{ji}) \in \mathcal{E}_{ij}^0 = \left\{(-\infty,-mu_i^d(j|A,X;\theta)) \times \mathbb{R}\right\} \cup \left\{\mathbb{R} \times (-\infty,-mu_j^d(i|A,X;\theta))\right\}$, $a_{ij} = 0$ is pairwise stable. Let $\mathcal{E}_A = \times_{ij}\mathcal{E}_{ij}^{a_{ij}}$. It is trivial that $\mathcal{E}_A$ is nonempty. Therefore, for all $\varepsilon = (\varepsilon_{12},\varepsilon_{13},\cdots,\varepsilon_{1N},\varepsilon_{23},\cdots,\varepsilon_{N-1,N}) \in \mathcal{E}_A$, A is pairwise stable.

Proposition A.2.2 states that for any observed network there exists a region of unobserved variables in which the network is pairwise stable. In other words, instead of restricting the utility function, I may restrict the stochastic structure of the model to guarantee the existence of pairwise stable networks. Point identification can be obtained similarly to the results of Proposition 2.4.1. I omit the proof of identification for this reason.

APPENDIX B

APPENDIX FOR CHAPTER 3

B.1 Proofs

B.1.1 Proof of Proposition 3.3.1

In order to prove the proposition, 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 (2002b) to the multigraph framework.

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

(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 $Y = Y_1, \cdots, Y_K = Y'$ such that for any $k \in 1, \cdots, K-1$, $Y_{k+1} = Y_k - Y_{k,ij} + Y'_{k,ij}$ 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 3.3.1) 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.

B.1.2 Proof of Proposition 3.3.2

Proof. I use the contra-positive to prove the proposition. Suppose that there is a cycle, $\{Y = Y_0, Y_1, \cdots, Y_K = Y\}$. 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.

B.1.3 Proof of Proposition 3.3.3

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. Let $\tilde{u}^{(s)}_{ij} := u^{(s)}(W_{ij}, V_{ij}, X_{ij}, \varepsilon_{ij})$. Define

$$\begin{aligned}
\omega(Y) \quad = \quad & \sum_{i}^{n}\sum_{j\neq i} a^{(1)}_{ij}\tilde{u}^{(1)}_{ij} + \sum_{i}\sum_{j} a^{(1)}_{ij}\left(\frac{1}{2}\gamma^{(1,1)}\sum_{k\neq i,j} a^{(1)}_{ik}a^{(1)}_{jk} + \gamma^{(2,1)}\sum_{k\neq i,j} a^{(2)}_{ik}a^{(2)}_{jk}\right) \\
& + \sum_{i}^{n}\sum_{j\neq i} a^{(2)}_{ij}\tilde{u}^{(2)}_{ij} + \sum_{i}\sum_{j} a^{(2)}_{ij}\left(\gamma^{(1,2)}\sum_{k\neq i,j} a^{(1)}_{ik}a^{(1)}_{jk} + \frac{1}{2}\gamma^{(2,2)}\sum_{k\neq i,j} a^{(2)}_{ik}a^{(2)}_{jk}\right) \\
& + \sum_{i=1}^{n} a^{(1)}_{ij}a^{(2)}_{ij}\delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij}, \varepsilon_{ij}).
\end{aligned}$$

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

Individual i's marginal utility of forming the first type relationship with j is

$$U_i(Y_{ij} = (1,0)) - U_i(Y_{ij} = (0,0)) \quad = \quad \tilde{u}^{(1)}_{ij} + \gamma^{(1,1)}\sum_{k\neq i,j} a^{(1)}_{ik}a^{(1)}_{jk} + \gamma^{(2,1)}\sum_{k\neq i,j} a^{(2)}_{ik}a^{(2)}_{jk}$$

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

$$mu_{ij} \quad = \quad \tilde{u}^{(1)}_{ij} + \tilde{u}^{(1)}_{ji} + 2\gamma^{(1,1)}\sum_{k\neq i,j} a^{(1)}_{ik}a^{(1)}_{jk} + 2\gamma^{(2,1)}\sum_{k\neq i,j} a^{(2)}_{ik}a^{(2)}_{jk}$$

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

$$\begin{aligned}
\omega(Y') - \omega(Y) \quad = \quad & \tilde{u}^{(1)}_{ij} + \tilde{u}^{(1)}_{jj} + 2\times\frac{1}{2}\gamma^{(1,1)}\sum_{k\neq i,j} a^{(1)}_{ik}a^{(1)}_{jk} + 2\gamma^{(2,1)}\sum_{k\neq i,j} a^{(2)}_{ik}a^{(2)}_{jk} \\
& +2\times\frac{1}{2}\gamma^{(1,1)}\sum_{k\neq i,j} a^{(1)}_{ik}a^{(1)}_{jk}.
\end{aligned}$$

Note that the first two lines are the sum of the marginal utilities of i and j except multiplying $\frac{1}{2}$ to $\gamma_1^{(s)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)}$. 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,

$$\begin{aligned} U_i(Y_{ij} = (1,0)) - U_i(Y_{ij} = (0,1)) \quad &= \quad \tilde{u}_{ij}^{(1)} - \tilde{u}_{ij}^{(2)} + \gamma^{(1,1)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} + \gamma^{(2,1)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} \\ &\quad -\gamma^{(1,2)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} - \gamma^{(2,2)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)}. \end{aligned}$$

$$\begin{aligned} mu_{ij} \quad &= \quad \tilde{u}_{ij}^{(1)} - \tilde{u}_{ij}^{(2)} + \tilde{u}_{ji}^{(1)} - \tilde{u}_{ji}^{(2)} + 2\gamma^{(1,1)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} + 2\gamma^{(2,1)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} \\ &\quad -2\gamma^{(1,2)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} - 2\gamma^{(2,2)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)}. \end{aligned}$$

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

$$\begin{aligned} \omega(Y') - \omega(Y) \quad &= \quad \tilde{u}_{ij}^{(1)} - \tilde{u}_{ij}^{(2)} + \tilde{u}_{ji}^{(1)} - \tilde{u}_{ji}^{(2)} + 2 \times \frac{1}{2}\gamma^{(1,1)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} + 2\gamma^{(2,1)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} \\ &\quad -2\gamma^{(1,2)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} - 2 \times \frac{1}{2}\gamma^{(2,2)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)} \\ &\quad +2 \times \frac{1}{2}\gamma^{(1,1)} \sum_{k \neq i,j} a_{ik}^{(1)} a_{jk}^{(1)} - 2 \times \frac{1}{2}\gamma^{(2,2)} \sum_{k \neq i,j} a_{ik}^{(2)} a_{jk}^{(2)}. \end{aligned}$$

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 $\sum_{i=1}^{n} a_{ij}^{(1)} a_{ij}^{(2)} \delta^{(1,2)}(W_{ij}, V_{ij}, X_{ij}, \varepsilon_{ij})$ 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

B.1.4 Definitions in Random Set Theory

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 $Q : \Omega \mapsto C$, from the probability space to the collection of closed sets $\{F \in C : F \cap D \neq \emptyset\}$ for all $D \in \mathcal{K}$.

Definition A map Q from a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ to C is called a *random closed set* if its inverse function $Q^{-}(D)$ satisfies,

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

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

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

Definition (i) A functional $T_X(D) : \mathcal{K} \to [0,1]$ given by

$$T_X(D) = P(X \cap D \neq \emptyset), \ D \in \mathcal{K},$$

is called the capacity functional of X.

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

$$C_X(F) = P(X \subset F),\ F \in \mathcal{C},$$

is called the containment functional of X.

B.1.5 Proof of Proposition 3.5.1

Proof. Since all admissible duples (θ, P_ε) are in $\mathrm{H}[(\theta, P_\varepsilon)]$, I only need to prove that a duple $(\theta', P'_\varepsilon)$ which is not admissible by the model is not in $\mathrm{H}[(\theta, P_\varepsilon)]$. 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

$$\begin{aligned} \mathrm{H}[(\theta, P_\varepsilon)] &= \Big\{(\theta, P_\varepsilon) \in \Theta \times \mathcal{P} | \, P_\varepsilon(F) \leq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \cap F \neq \emptyset;\ F^0], \\ &\quad \forall F \in \mathcal{C}(\mathbb{R}^{2^p-1})\, a.e.\, z \in \mathcal{Z}\Big\}. \end{aligned}$$

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

B.1.6 Proof of Proposition 3.5.2

Proof. Recall the sharp identification region (3.20)

$$\begin{aligned} \mathrm{H}[(\theta,\ P_\varepsilon)] &= \Big\{(\theta, P_\varepsilon) \in \Theta \times \mathcal{P} | \, P_\varepsilon(F) \leq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq F;\ F^0], \\ &\quad \forall F \in C(\mathbb{R}^{2p-1})\, a.e.\, z \in \mathcal{Z}\Big\}. \end{aligned}$$

First, I want to rule out sets that are not the union of sets in $C_1(\theta) = \{\mathcal{E}_\theta(y,\ w,\ z) : y \in \mathcal{Y}, w \in \mathcal{W}, \text{and } z \in \mathcal{Z}\}$. Let T_1 be an arbitrary set which is not a union of sets in $C_1(\theta)$. I can find the largest possible union of sets in $C_1(\theta)$, which is a subset of T_1. Denote the largest possible union F_1. From Assumption 3.2.(ii), the distribution of ε is continuous, and ε has everywhere positive density. Hence, $P_\varepsilon(F_1) \leq P_\varepsilon(T_1)$. Now consider $\Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq T_1;\ F^0]$ and $\Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq F_1;\ F^0]$. If $\mathcal{E}_\theta(y,\ w,\ z)$ is a subset of T_1, then it is also a subset of F_1 since F_1 is the largest possible union. So, I have

$$P_\varepsilon(T_1) \geq P_\varepsilon(F_1) \geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq F_1;\ F^0] = \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq T_1;\ F^0],$$

Thus, $P_\varepsilon(F_1) \geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq F_1;\ F^0]$ is not informative. Therefore I can rule out sets that are not the unions of sets in $C_1(\theta)$.

Next, I want to rule out non-connected sets among the unions of sets in $C_1(\theta)$. Pick an arbitrary non-connected union T_2. Without loss of generality, suppose E_1 and E_2 are non-connected, and $T_2 = E_1 \cup E_2$. For E_1 and E_2, I know

$$\begin{aligned} P_\varepsilon(E_1) &\geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq E_1;\ F^0], \\ P_\varepsilon(E_2) &\geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq E_2;\ F^0]. \end{aligned}$$

Also, I have

$$\begin{aligned} P_\varepsilon(T_2) &\geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq T_2;\ F^0] \\ &= \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq E_1;\ F^0] + \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq E_2;\ F^0], \end{aligned}$$

by construction of the containment functional, i.e.

$$\begin{aligned} \Pr[\mathcal{E}_\theta(Y,\ w,\ z) \subseteq F; F^0] &= \sum_{w \in \mathcal{W}} \left\{ \sum_{y \in \mathcal{Y}} (1[\mathcal{E}_\theta(y,\ w,\ z) \subseteq F] \right. \\ &\qquad \left. \times \Pr[Y = y | W = w,\ Z = z; F^0] \Big) \Pr(W = w | Z = z;\ F^0) \right\}. \end{aligned}$$

Since E_1 and E_2 are disjoint, $P_\varepsilon(T_2) = P_\varepsilon(E_1) + P_\varepsilon(E_2)$. Hence, I have

$$\begin{aligned} P_\varepsilon(T_2) &\geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq T_2;\ F^0] \\ &\Leftrightarrow \\ P_\varepsilon(E_1) + P_\varepsilon(E_2) &\geq \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq E_1;\ F^0] + \Pr[\mathcal{E}_\theta(Y,\ W,\ z) \subseteq E_2;\ F^0]. \end{aligned}$$

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.

B.2 Pairwise Stability of a Multigraph with Non-transferable Utility

Definition *(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 with non-transferable utility (PSMnt) 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 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 $Y'_{ij} \neq (0, \cdots, 0)'$.

B.3 Core Determining Class

In this appendix, I explain how to obtain a core determining class in practice. With the utility function (3.24), $C_2(\theta)$, or the power set of $C_1(\theta) = \{\mathcal{E}_\theta(y, w, z) : y \in \mathcal{Y}, w \in \mathcal{W}, \text{and } z \in \mathcal{Z}\}$ has cardinality equal to $2^8 = 256$. The cardinality of $\mathcal{Z}$ 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 $\mathcal{E}_\theta(y, w, z)$'s as $\{E_1, \cdots, E_8\}$. I construct the adjacency matrix G_θ among $\{E_1, \cdots, E_8\}$ based on the connectedness of each pair (E_i, E_j). That is, $G_{\theta,ij} = 1$ if E_i and E_j are connected, and zero otherwise. Then, among those 2^8 sets, I delete non-connected sets and duplicated sets based on G_θ. 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 θ, or more specifically $\beta_3^{(1)}$ and $\beta_3^{(2)}$.

The core determining class differs by the value of the parameter vector θ. 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 $E_1, \cdots, E_8$ as follows. $E_1 = \mathcal{E}_\theta((0, 0), 0, z)$, $E_2 = \mathcal{E}_\theta((0, 0), 1, z)$, $E_3 = \mathcal{E}_\theta((1, 0), 0, z)$, $E_4 = \mathcal{E}_\theta((1, 0), 1, z)$, $E_5 = \mathcal{E}_\theta((0, 1), 0, z)$, $E_6 = \mathcal{E}_\theta((0, 1), 1, z)$, $E_7 = \mathcal{E}_\theta((1, 1), 0, z)$, and $E_8 = \mathcal{E}_\theta((1, 1), 1, z)$.

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

  Figure B.1 shows how the area of unobservables is divided into 8 different regions. Table B.4 shows the adjacency matrix G_θ for Case 1, which characterizes the connectedness of E_i's. The core determining class $\mathcal{M}(\theta)$ is described in Table B.2.

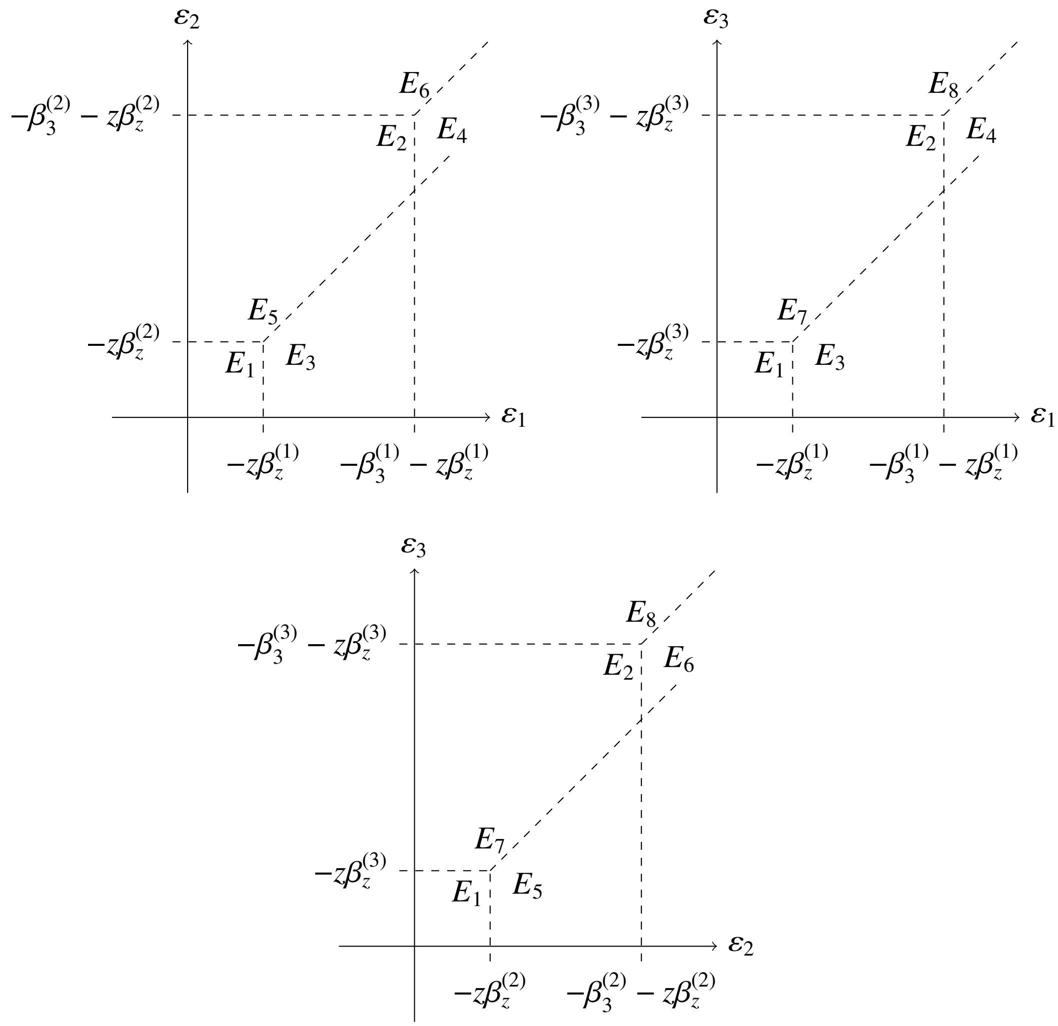

Figure B.1: Regions of Unobservables Corresponding to $\mathcal{E}_\theta(y, w, z)$ given z when $\beta_3^{(2)} < \beta_3^{(1)} < 0$.

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 $\beta_3^{(1)} + \beta_3^{(2)} > 0$,

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8
E_1	0	1	0	0	0	0	0	0
E_2	1	0	1	0	1	0	1	0
E_3	0	1	0	1	0	0	0	0
E_4	0	0	1	0	1	0	1	0
E_5	0	1	0	1	0	1	0	0
E_6	0	0	0	0	1	0	1	0
E_7	0	1	0	1	0	1	0	1
E_8	0	0	0	0	0	0	1	0

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

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8		E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8
F_1	■	■	■	■			■	■	F_{19}	■	■	■	■				
F_2	■	■	■			■	■	■	F_{20}	■	■	■		■	■		
F_3	■	■	■				■	■	F_{21}	■	■	■		■			
F_4	■	■			■	■	■	■	F_{22}	■	■	■					
F_5	■	■				■	■	■	F_{23}	■	■			■	■		
F_6	■	■					■	■	F_{24}	■	■			■			
F_7			■	■	■	■	■	■	F_{25}	■	■						
F_8			■	■		■	■	■	F_{26}	■							
F_9			■	■			■	■	F_{27}			■	■	■	■		
F_{10}				■	■	■	■	■	F_{28}			■	■	■			
F_{11}				■		■	■	■	F_{29}			■	■				
F_{12}				■			■	■	F_{30}			■					
F_{13}					■	■	■	■	F_{31}				■	■	■		
F_{14}						■	■	■	F_{32}				■	■			
F_{15}							■	■	F_{33}				■				
F_{16}								■	F_{34}					■	■		
F_{17}	■	■	■	■	■	■			F_{35}					■			
F_{18}	■	■	■	■	■				F_{36}						■		

Table B.2: The core determining class $\mathcal{M}(\theta)$ when $\beta_3^{(2)} < \beta_3^{(1)} < 0$.

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 $\beta_3^{(1)} + \beta_3^{(2)} < 0$.

B.4 Detailed Estimation Procedures

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

B.4.1 Drawing Parameters

For the simulated annealing procedure, I draw more than 500 initial values of θ from $Unif(-1,1)$ in order to make sure that all regions in Θ 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 Λ, or the Cholesky decomposition of Σ_ε, i.e. $\Sigma_\varepsilon = \Lambda\Lambda'$. Note that I have only five parameters from Σ_ε, since scale normalization is done by setting $var(\varepsilon_1) = 1$.

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.3.

B.4.2 Simulation of Unobservables

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

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

$$m_{ij}^{(r)}(\theta) = (m_{ij,1,1}^{(r)}(\theta), m_{ij,1,2}^{(r)}(\theta), \cdots, m_{ij,1,36}^{(r)}(\theta), \cdots, m_{ij,k,l}^{(r)}(\theta), \cdots, m_{ij,36,36}^{(r)}(\theta))',$$

where

$$m_{ij,k,l}^{(r)}(\theta) \quad = \quad 1[\varepsilon_{ij}^{(r)} \in F_l] - \sum_{w \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \{1[\mathcal{E}_\theta(y, w, z) \subseteq F_l]$$
$$\times 1[Y_{ij} = y | W_{ij} = w, Z_{ij} = z_k] 1[W_{ij} = w | Z_{ij} = z_k]\Big\}.$$

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

$$m_{ij,k,l}(\theta) \quad = \quad \frac{1}{R} \sum_{r=1}^{R} \{1[\varepsilon_{ij}^{(r)} \in F_l] - \sum_{w \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \{1[\mathcal{E}_\theta(y, w, z) \subseteq F_l]$$
$$\times 1[Y_{ij} = y | W_{ij} = w, Z_{ij} = z_k] 1[W_{ij} = w | Z_{ij} = z_k]\Big\}.$$

I still do not use a notation $\bar{m}$ at this stage, since I have not taken the sample average over pairs.

B.4.3 Computing Sample Moments and Test Statistic

I take the sample average over pairs as

$$\bar{m}_{\tilde{N}}(\theta) = (\bar{m}_{\tilde{N},1,1}(\theta), \bar{m}_{\tilde{N},1,2}(\theta), \cdots, \bar{m}_{\tilde{N},1,36}(\theta), \cdots, \bar{m}_{\tilde{N},k,l}(\theta), \cdots, \bar{m}_{\tilde{N},36,36}(\theta))',$$

$$
\begin{aligned}
\bar{m}_{\tilde{N},k,l}(\theta) \quad = \quad & \tilde{N}^{-1} \sum_{m=1}^{M} \sum_{j>i}^{n_m} \sum_{i=1}^{n_m-1} \frac{1}{R} \sum_{r=1}^{R} \{1[\varepsilon_{ij}^{(r)} \in F_l] \\
& - \sum_{w \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \{1[\mathcal{E}_\theta(y,\, w,\, z) \subseteq F_l] \\
& \times 1[Y_{ij} = y | W_{ij} = w,\, Z_{ij} = z_k] 1[W_{ij} = w | Z_i = z_k]\Big\},
\end{aligned}
$$

where $\tilde{N} = \sum_{m=1}^{M} \binom{n_m}{2}$. I also compute the variance-covariance matrix $\hat{\Sigma}_{\tilde{N}}(\theta)$ of $m_{ij}(\theta)$. In practice, I only need to compute the variance of each $m_{ij,k,l}(\theta)$.

$$
\hat{\Sigma}_{\tilde{N}}(\theta) = \tilde{N}^{-1} \sum_{m=1}^{M} \sum_{j>i}^{n_m} \sum_{i=1}^{n_m-1} (m_{ij}(\theta) - \bar{m}_{\tilde{N}}(\theta))(m_{ij}(\theta) - \bar{m}_{\tilde{N}}(\theta))'.
$$

From AS, I have several choices of functions for computing the test statistic. For computational efficiency, I choose S_3 in AS, which is

$$
S_3(m,\, \Sigma)) = \sum_{j=1}^{J} [m_j / \sigma_j]_{-}^{2},
$$

where σ_j is the square root of jth diagonal element of Σ, and

$$
[x]_{-} = \begin{cases} x & \text{, if } x \leq 0 \\ 0 & \text{, otherwise.} \end{cases}
$$

Then, the test statistic $T_{\tilde{N}}(\theta_0)$ is defined as

$$
\begin{aligned}
T_{\tilde{N}}(\theta_0) \quad &= \quad S_3(\sqrt{\tilde{N}} \bar{m}_{\tilde{N}}(\theta),\, \hat{\Sigma}_{\tilde{N}}(\theta)) \\
&= \quad \sum_{k=1}^{K} \sum_{l=1}^{L} [\sqrt{\tilde{N}} \bar{m}_{\tilde{N},k,l}(\theta) / \hat{\sigma}_{k,l}]_{-}^{2},
\end{aligned}
$$

where $K = 36$ and $L = 36$ in practice.

B.4.4 Bootstrap and Moment Selection

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

First, I simulate $B = 100$ (nonparametric i.i.d) bootstrap samples. $\{Y^*_{ij,b}, W^*_{ij,b}, Z^*_{ij,b}, \varepsilon^*_{ij,b}, \forall ij \leq \tilde{N}\}$, where ε^*_{ij} is the R set of unobservables drawn at the first stage.

Compute

$$
\begin{aligned}
M^*_{\tilde{N},b}(\theta_0) &= \sum_{k=1}^{K}\sum_{l=1}^{L}[\sqrt{\tilde{N}}\bar{m}^*_{\tilde{N},k,l}(\theta)/\hat{\sigma}_{k,l}]^2_-, \\
\hat{\Omega}^*_{\tilde{N},b}(\theta_0) &= \hat{D}^{*-1/2}_{\tilde{N}}(\theta_0)\hat{\Sigma}^*_{\tilde{N}}(\theta_0)\hat{D}^{*-1/2}_{\tilde{N}}(\theta_0),
\end{aligned}
$$

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

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

Finally, I take the critical value $\hat{c}_{\tilde{N}}(\theta_0, 1-\alpha)$ to be the $1-\alpha$ sample quantile of $\{S\left(M^{**}_{\tilde{N},b}(\theta_0), \hat{\Omega}^{**}_{\tilde{N},b}(\theta_0)\right) : b = 1, \cdots, B\}$. If $T_n(\theta_0) \leq \hat{c}_{\tilde{N}}(\theta_0, 1-\alpha)$, then θ_0 should be included in the $1-\alpha$ confidence set.

I repeat the above procedures for all values of parameters drawn in Appendix B.4.1., when using SVM. For simulated annealing, these procedures are repeated until the optimization process ends.

B.5 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,

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

and

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

where $c_0 = -\infty$ and $c_{\kappa_w} = \infty$. The unobservable ε 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 $z = \{-1,\ 1\}$. For the three parameters, I construct a grid of approximately 930,000 values and plot the sharp identification region. When β approaches to zero, i.e. the model has no endogenous explanatory variable, the sharp identification

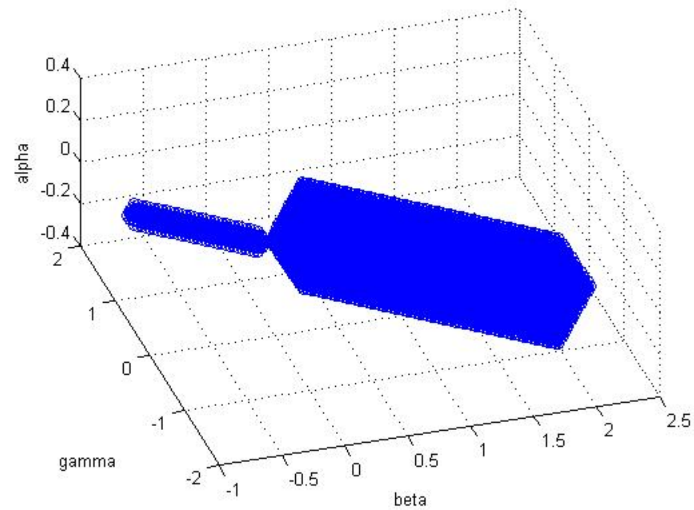

Figure B.2: The sharp identification region of θ when $\kappa_w = 2$

region of γ 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 $z = \{-1, 1\}$. $c = (-\infty, -1/2, 0, 1/2, \infty)$. Again, for the three parameters, I construct a grid of approximately 1.2 million values and plot the sharp identification region. As β approaches to zero, i.e. the model has no endogenous explanatory variable, the sharp identification region of γ shrinks to the empty set. Figure B.3 shows the sharp identification region.

As Figures B.2 and B.3 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.

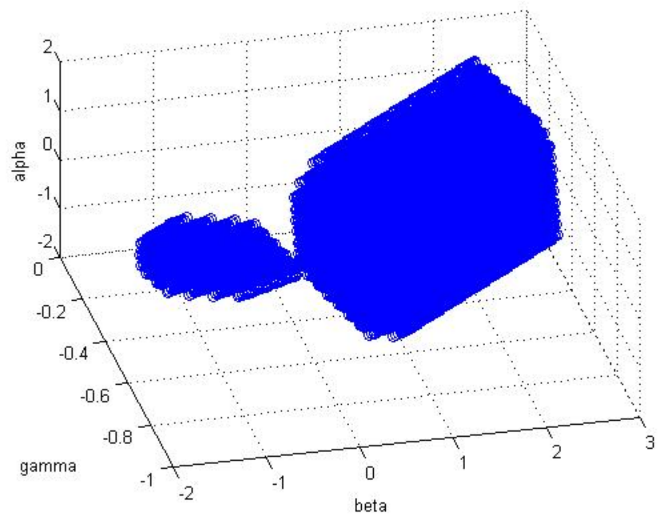

Figure B.3: The sharp identification region of θ when $\kappa_w = 4$

APPENDIX C

APPENDIX FOR CHAPTER 4

C.1 The Lists of Markets and Airports

Metropolitan Statistical Areas	Population	Airports
New York-Newark-Jersey City, NY-NJ-PA	19,567,410	JFK, EWR, LGA, ISP
Los Angeles-Long Beach-Anaheim, CA	12,828,837	LAX, BUR, LGB, SNA
Chicago-Naperville-Elgin, IL-IN-WI	9,461,105	ORD, MDW
Dallas-Fort Worth-Arlington, TX	6,426,214	DAL, DFW
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	5,965,343	PHL
Houston-The Woodlands-Sugar Land, TX	5,920,416	HOU, IAH
Washington-Arlington-Alexandria, DC-VA-MD-WV	5,636,232	DCA, IAD, BWI
Miami-Fort Lauderdale-West Palm Beach, FL	5,564,635	MIA, FLL
Atlanta-Sandy Springs-Rosewell, GA	5,286,728	ATL
Boston-Cambridge-Newton, MA-NH	4,552,402	BOS, MHT
San Francisco-Oakland-Hayward, CA	4,335,391	SFO, SJC, OAK
Detroit-Warren-Dearborn, MI	4,296,250	DTW
Phoenix-Mesa-Scottsdale, AZ	4,192,887	PHX, IWA
Seattle-Tacoma-Bellevue, WA	3,439,809	SEA
Minneapolis-St. Paul-Bloomington, MN-WI	3,348,859	MSP, STC
San Diego-Carlsbad, CA	3,095,313	SAN
St. Louis, MO-IL	2,787,701	STL
Tampa-St. Petersburg-Clearwater, FL	2,783,243	TPA, PIE
Denver-Aurora-Lakewood, CO	2,543,482	DEN
Pittsburgh, PA	2,356,285	PIT
Portland-Vancouver-Hillsboro, OR-WA	2,226,009	PDX
Charlotte-Concord-Gastonia, NC-SC	2,217,012	CLT
Sacramento-Roseville-Arden-Arcade, CA	2,149,127	SMF
San Antonio-New Braunfels, TX	2,142,508	SAT
Orlando-Kissimmee-Sanford, FL	2,134,411	MCO, SFB, DAB

Table C.1: 84 Metropolitan Statistical Areas and Corresponding Airports, 2011

Metropolitan Statistical Areas	Population	Airports
Cincinnati, OH-KY-IN	2,114,580	CVG
Cleveland-Elyria, OH	2,077,240	CLE
Kansas City, MO-KS	2,009,342	MCI
Las Vegas-Henderson-Paradise, NV	1,951,269	LAS
Columbus, OH	1,901,974	CMH
Indianapolis-Carmel-Anderson, IN	1,887,877	IND
Austin-Round Rock, TX	1,716,289	AUS
Virginia Beach-Norfolk-Newport News, VA-NC	1,676,822	ORF
Nashville-Davidson–Murfreesboro–Franklin, TN	1,670,890	BNA
Providence-Warwick, RI-MA	1,600,852	PVD
Milwaukee-Waukesha-West Allis, WI	1,555,908	MKE
Jacksonville, FL	1,345,596	JAX
Memphis, TN-MS-AR	1,324,829	MEM
Oklahoma City, OK	1,252,987	OKC
Louisville/Jefferson County, KY-IN	1,235,708	SDF
Hartford-West Hartford-East Hartford, CT	1,212,381	BDL
Richmond, VA	1,208,101	RIC
New Orleans-Metairie, LA	1,189,866	MSY
Buffalo-Cheektowaga-Niagara Falls, NY	1,135,509	BUF
Raleigh, NC	1,130,490	RDU
Birmingham-Hoover, AL	1,128,047	BHM
Salt Lake City, UT	1,087,873	SLC
Rochester, NY	1,079,671	ROC
Grand Rapids-Wyoming, MI	988,938	GRR
Tucson, AZ	980,263	TUS
Tulsa, OK	937,478	TUL
Fresno, CA	930,450	FAT
Bridgeport-Stamford-Norwalk, CT	916,829	HPN
Albuquerque, NM	887,077	ABQ
Albany-Schenectady-Troy, NY	870,716	ALB
Omaha-Council Bluffs, NE-IA	865,350	OMA
Bakersfield, CA	839,631	BFL
Knoxville, TN	837,571	TYS
Greenville-Anderson-Mauldin, SC	824,112	GSP
Allentown-Bethlehem-Easton, PA-NJ	821,173	ABE

Table C.2: 84 Metropolitan Statistical Areas and Corresponding Airports, 2011 (Continued)

Metropolitan Statistical Areas	Population	Airports
El Paso, TX	804,123	ELP
Baton Rouge, LA	802,484	BTR
Dayton, OH	799,232	DAY
McAllen-Edinburg-Mission, TX	774,769	MFE, HRL
Columbia, SC	767,598	CAE
Greensboro-High Point, NC	723,801	GSO
Akron, OH	703,200	CAK
North Port-Sarasota-Bradenton, FL	702,281	SRQ
Little Rock-North Little Rock-Conway, AR	699,757	LIT
Charleston-North Charleston-Summerville, SC	664,607	CHS
Syracuse, NY	662,577	SYR
Colorado Springs, CO	645,613	COS
Wichita, KS	630,919	ICT
Cape Coral-Fort Myers, FL	618,754	RSW
Boise City-Nampa, ID	616,561	BOI
Madison, WI	605,435	MSN
Des Moines-West Des Moines, IA	569,633	DSM
Jackson, MS	567,122	JAN
Augusta-Richmond County, GA-SC	564,873	AGS
Scranton–Wilkes-Barre–Hazleton, PA	563,631	AVP
Harrisburg-Carlisle, PA	549,475	MDT
Palm Bay-Melbourne-Titusville, FL	543,376	MLB
Chattanooga, TN-GA	528,143	CHA
Spokane-Spokane Valley, WA	527,753	GEG

Table C.3: 84 Metropolitan Statistical Areas and Corresponding Airports, 2011 (Continued)

BIBLIOGRAPHY

Aguirregabiria, V. and C. Ho (2010). A dynamic game of airline network competition: Hub-and-spoke networks and entry deterrence. *International Journal of Industrial Organization 28*(4), 377–382.

Aguirregabiria, V. and C.-Y. Ho (2012). A dynamic oligopoly game of the us airline industry: Estimation and policy experiments. *Journal of Econometrics 168*(1), 156–173.

Aguirregabiria, V. and P. Mira (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics 156*(1), 38–67.

Andrews, D. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Andrews, D. W. and P. J. Barwick (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica 80*(6), 2805–2826.

Artstein, Z. (1983). Distributions of random sets and random selections. *Israel Journal of Mathematics 46*(4), 313–324.

Bajari, P., H. Hong, and D. Nekipelov (2010). Game theory and econometrics: A survey of some recent research. In *Econometric Society World Congress in Shanghai*.

Bajari, P., H. Hong, and S. Ryan (2010). Identification and estimation of a discrete game of complete information. *Econometrica 78*(5), 1529–1568.

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica 68*(5), 1181–1229.

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2012). The diffusion of microfinance. Technical report, National Bureau of Economic Research.

Bar, H. and F. Molinari (2013). Computational methods for partially identified models via data augmentation and support vector machines.

Beresteanu, A., I. Molchanov, and F. Molinari (2011). Sharp identification regions in models with convex moment predictions. *Econometrica 79*(6), 1785–1821.

Beresteanu, A., I. Molchanov, and F. Molinari (2012). Partial identification using random set theory. *Journal of Econometrics 166*(1), 17–32.

Beresteanu, A. and F. Molinari (2008). Asymptotic properties for a class of partially identified models. *Econometrica 76*(4), 763–814.

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica 60*(4), 889–917.

Bjorn, P. and Q. Vuong (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. Technical report.

Bloch, F. and M. Jackson (2006). Definitions of equilibrium in network formation games. *International Journal of Game Theory 34*(3), 305–318.

Blume, L. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior 5*(3), 387–424.

Borenstein, S. (1989). Hubs and high fares: Dominance and market power in the us airline industry. *RAND Journal of Economics 20*(3), 344–365.

Boucher, V. and I. Mourifié (2012). My friend far far away: Asymptotic properties of pairwise stable networks. *Available at SSRN 2170803*.

Bresnahan, T. and P. Reiss (1990). Entry in monopoly market. *The Review of Economic Studies 57*(4), 531.

Bresnahan, T. and P. Reiss (1991a). Empirical models of discrete games. *Journal of Econometrics 48*(1-2), 57–81.

Bresnahan, T. and P. Reiss (1991b). Entry and competition in concentrated markets. *The Journal of Political Economy 99*(5), 977–1009.

Calvó-Armengol, A. and İlkılıç (2009). Pairwise-stability and nash equilibria in network formation. *International Journal of Game Theory 38*(1), 51–79.

Chandrasekhar, A. and M. Jackson (2012). Tractable and consistent random graph models. *Available at SSRN 2150428*.

Chartrand, G., L. Lesniak, and P. Zhang (2011). *Graphs and Digraphs*. CRC Press.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models1. *Econometrica 75*(5), 1243–1284.

Chesher, A. and A. M. Rosen (2012). Simultaneous equations models for discrete outcomes: coherence, completeness, and identification.

Chesher, A. and A. M. Rosen (2013). What do instrumental variable models deliver with discrete dependent variables? *The American Economic Review 103*(3), 557–62.

Chesher, A., A. M. Rosen, and K. Smolinski (2011). An instrumental variable model of multiple discrete choice.

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman (2010). An empirical model for strategic network formation. Technical report, National Bureau of Economic Research.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

Currarini, S., M. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Currarini, S., M. Jackson, and P. Pin (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences 107*(11), 4857.

De Weerdt, J. (2002). Risk-sharing and endogenous network formation. Technical report, WIDER Discussion Papers//World Institute for Development Economics (UNU-WIDER).

De Weerdt, J. and S. Dercon (2006). Risk-sharing networks and insurance against illness. *Journal of Development Economics 81*(2), 337–356.

Ellickson, P. B., S. Houghton, and C. Timmins (2013). Estimating network economies in retail chains: a revealed preference approach. *The RAND Journal of Economics 44*(2), 169–193.

Fafchamps, M. (1992). Solidarity networks in preindustrial societies: Rational peasants with a moral economy. *Economic Development and Cultural Change 41*(1), 147–174.

Fafchamps, M. and F. Gubert (2007). The formation of risk sharing networks. *Journal of Development Economics 83*(2), 326–350.

Fafchamps, M. and S. Lund (2003). Risk-sharing networks in rural philippines. *Journal of Development Economics 71*(2), 261–287.

Fox, J. (2008). Estimating matching games with transfers. Technical report, National Bureau of Economic Research.

Fox, J. and D. Santiago (2008). A toolkit for matching maximum score estimation and point and set identified subsampling inference. Technical report, working paper.

Galichon, A. and M. Henry (2011). Set identification in models with multiple equilibria. *The Review of Economic Studies 78*(4), 1264–1298.

Grieco, P. (2010). Discrete games with flexible information structures: An application to local grocery markets. Technical report, Working papers.

Hellmann, T. (2009). Convexity and complementarity in network formation: Implications for the structure of pairwise stable networks. *Working Papers*.

Hellmann, T. (2012). On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, 1–27.

Hendricks, K., M. Piccione, and G. Tan (1995). The economics of hubs: The case of monopoly. *The Review of Economic Studies 62*(1), 83–99.

Hendricks, K., M. Piccione, and G. Tan (1997). Entry and exit in hub-spoke networks. *The RAND Journal of Economics*, 291–303.

Hendricks, K., M. Piccione, and G. Tan (1999). Equilibria in networks. *Econometrica 67*(6), 1407–1434.

Imbens, G. and C. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*(6), 1845–1857.

Jackson, M., T. Rodriguez-Barraquer, and X. Tan (2012). Social capital and social quilts: Network patterns of favor exchange. *The American Economic Review 102*(5), 1857–1897.

Jackson, M. and A. Watts (2001). The existence of pairwise stable networks. *Seoul Journal of Economics*.

Jackson, M. and A. Watts (2002a). On the formation of interaction networks in social coordination games. *Games and Economic Behavior 41*(2), 265–291.

Jackson, M. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*(1), 44–74.

Jackson, M. O. and A. Watts (2002b). The evolution of social and economic networks. *Journal of Economic Theory 106*(2), 265–295.

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

Khan, S. (2010). Distribution free estimation of heteroskedastic binary response models using probit/logit criterion functions. Technical report, mimeo, Duke University.

Kinnan, C. and R. Townsend (2012). Kinship and financial networks, formal financial access, and risk reduction. *The American Economic Review 102*(3), 289–293.

Kline, B. (2012). Identification of complete information games.

Koenig, M. (2012). The formation of networks with local spillovers and limited observability. Technical report.

Komarova, T. (2012). Binary choice models with discrete regressors: Identification and misspecification. Technical report, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Lazarsfeld, P. and R. Merton (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society 18*, 18–66.

Leung, M. (2013). Two-step estimation of network-formation models with incomplete information. *Available at SSRN 2254145*.

Manski, C. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*(3), 205–228.

Manski, C. (1985). Semiparametric analysis of discrete response:: Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*(3), 313–333.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531.

Manski, C. and E. Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica 70*(2), 519–546.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 995–1026.

Mele, A. (2010). A structural model of segregation in social networks.

Miyauchi, Y. (2012). Structural estimation of a pairwise stable network formation with nonnegative externality. *Available at SSRN 2165884*.

Molchanov, I. (2005). *Theory of Random Sets*. Springer.

Molchanov, I. and F. Molinari (2013). Applications of random set theory in econometrics. *Annual Review of Economics*. forthcoming.

Myerson, R. (1991). *Game theory: analysis of conflict*. Harvard University Press.

Norberg, T. (1992). On the existence of ordered couplings of random sets with applications. *Israel Journal of Mathematics 77*(3), 241–264.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Politis, D. and J. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics 22*(4), 2031–2050.

Politis, D., J. Romano, and M. Wolf (1999). *Subsampling*. Springer Verlag.

Romano, J. and A. Shaikh (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference 138*(9), 2786–2807.

Rosen, A. M. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics 146*(1), 107–117.

Sheng, S. (2012). Identification and estimation of network formation games. *Unpublished Manuscript*.

Storn, R. and K. Price (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization 11*(4), 341–359.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica 77*(4), 1299–1315.

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

Townsend, R. M. (1994). Risk and insurance in village india. *Econometrica*, 539–591.

Watts, A. (2001). A dynamic model of network formation. *Games and Economic Behavior 34*(2), 331–341.