Essays on Business Networks in the Multi-level Marketing Industry

by

Eunsoo Kim

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in the University of Michigan
2017

Doctoral Committee:

Professor Puneet Manchanda, Chair
Associate Professor Anocha Aribarg
Associate Professor Yves F. Atchadé
Professor Professor Peter Lenk
Assistant Professor Eric Schwartz

Eunsoo Kim
eunsoo@umich.edu
ORCID iD: 0000-0001-9332-5653

© Eunsoo Kim 2017

Dedication

This dissertation is dedicated to my dear family for all their love and unconditional support; to Lisa and Rosa, whom I shared most of my time with during this journey; and to all of those who cherish relationships with others.

Acknowledgements

First and foremost, I would like to express my utmost gratitude to my advisor, Puneet Manchanda for his years of guidance and sresupport. He not only helped me develop the skills and attitudes needed for research, but also provided me emotional support. His mentorship led me to be a better researcher and a better person. I would also like to thank Anocha Aribarg with whom I've been working from the beginning of mt Ph.D, program to the very end. I am greatly inspired by her positive drive, thoughtfulness and persistence, and I am very thankful for her for standing by my side. Moreover, I truly appreciate my dissertation committee members, Anocha Aribarg, Peter Lenk, Yves F. Atchadé and Eric Schwartz for taking their precious time providing me helpful comments and feedback to move the dissertation forward. Moreover, I would like to thank marketing faculty members and my fellow Ph.D. students in the doctoral program at the Ross School of Business for all of their sincere encouragement.

Last but not least, I would like to thank my parents, sister and brother for all their endless love. Without their support, I would not have reached anywhere near where I am right now.

Contents

List of Tables

List of Figures

Abstract

This dissertation contributes to our knowledge of how social factors (e.g., homophily, the existence of others) influence individual business behaviors. To explore the influence of social factors, the Multi-level Marketing (MLM) Industry is especially suitable because social and business networks are interrelated. MLM is a major type of direct selling wherein Independent Business Owners (IBOs) recruit others, who become IBOs themselves. IBOs in MLM are pivotal because they both serve end-customers and expand and maintain the IBO network. However, as these IBOs are not firm employees, understanding the relationships among them is critical. Two important questions are addressed. How is the IBO network created (Essay One)? How do IBOs influence the activity of other IBOs (Essay Two)?

Essay One, "Social versus Economic Factors in Business Network Formation: An Empirical Analysis of the Multi-level Marketing Industry," explores how IBOs form networks between IBOs and new joiners. Unlike social networks, where relationships are mainly driven by social motives, IBOs and new joiners in MLM also expect economic returns. Thus, I investigate network formation in terms of social and economic factors. Moreover, I investigate how the relationship value between the IBO and the new joiner affects the new joiner's future business activities. By employing a structural Bayesian two-sided matching model, I find that both factors significantly affect network formation. However, economic factors are more important than social factors, playing a significant role in determining the network relationship. Also, social factors are significant in forming the network relationship, but are less influential driving new joiners to engage in business activities in the future. Finally, new joiners who are driven primarily by economic factors at the formation stage show significantly more business activity.

Essay Two, "The Impact of Existence of Others on Inactive Behavior in the Multi-level Marketing Industry," examines one specific aspect of IBOs' turnover behavior. There is a vast literature on turnover, yet its main focus is on traditional hierarchical organizations. MLM has a unique industry environment, which suggests that previous findings may not apply. Specifically, relationships among IBOs are highly emphasized; they are often seen as personal, yet competition among IBOs remains despite the importance of coordinating efforts. Given these unique features, I focus on how IBOs' inactivity can be quantitatively understood in relation to three different dimensions of others (Individual, Network family, and Proximity). One focal aspect is status, which is awarded by MLM firms to successful IBOs to motivate IBOs. Analyzing county-level and individual-level data, I find consistent evidence that being near high status IBOs decreases inactivity — a protective effect. The results hold true at the network family level and for proximity. Also, IBOs with a larger family tended to stay active. However, having more IBOs in proximity increased inactivity, indicating that competition in proximity seems inevitable. Lastly, neural network is applied to predict inactivity, which reveals a similar conclusion about the role of high status IBOs. These findings have implications for the MLM industry, which has an interest in maintaining the IBO sales force.

Chapter 1 Social versus Economic Factors in Network Formation: An Empirical Analysis of the Multi-level Marketing Industry

1.1 Abstract

Multi-level Marketing (MLM) networks are unlike other business networks, where relationships are mainly driven by economic benefits. According to the existing understanding of MLM networks, relationship formation between distributors (IBOs) and new joiners is based mainly on social aspects. I leverage a novel dataset from a large MLM firm to examine the importance of economic returns versus social aspects in determining the formation of networks. I also examine the role of social and economic factors, which contribute to the value of relationships between the IBOs and the new joiners, in explaining the new joiners' future business activity. T consider the disparate perspectives of the current IBOs and the new joiners, I use a two-sided matching model using hierarchical Bayesian estimation. The results suggest that network formation is determined by potential economic returns on both sides. Business-support-related information matters in business relationship formation for new joiners. IBOs, on the other hand, place importance on information that could signal the future potential income from new joiners. As social factors, demographic homophily and geographic distance also had a significant impact on business relationship formation, yet their importance was about one third that of economic factors. Using the recovered social and economic factors, I also share evidence that the new joiners facing business relationships with greater emphasis on economic factors tend to engage actively in the business.

1.2 Introduction

Individuals are part of both personal and professional (or business) networks. Business networks are different from purely social networks in that there is economic and social exchange between individuals in these networks (Grayson 1996). In other words, both the static and dynamic structure of relationships has an impact on economic flows between individuals in these networks. A very large and important industry where such networks determine economic outcomes is the Multi-level Marketing (MLM) industry. In this industry, the main mechanism of distribution is an organized network of independent distributors (typically referred to as Independent Business Owners or IBOs). Firms in this industry do not employ a dedicated sales force, relying instead on the network of IBOs to market and sell their products. Thus they prefer to give up direct control in exchange for lower cost distribution and advertising/retailing costs (Biggart 1989, King and Robinson 2000, Jain et al. 2015). For the IBOs, the economic returns come from two major sources - the margin from selling products to end customers and bonuses generated based on the selling activities of recruited sub-distributors (Grayson 1996, Peterson and Wotruba 1996, Jain et al. 2015). This creates an incentive for IBOs to grow (and maintain) the network by recruiting sub-distributors. The prevalent belief among participants in this industry is that IBOs should leverage their social networks to find sub-distributors (Clothier 1997, Grayson 1996). However, recent industry publications have begun to emphasize that IBOs need to focus less on social connections or intimates, noting that most successful IBOs do not have a significant proportion of family and/or friends in their networks (Lilyquist 2016a). Other researchers have noted that depending on beliefs about the friendship, existing social ties can either hurt or improve business effectiveness (Grayson 2007). However, there is little research documenting the role of economic versus social aspects in the development of these business networks.

In this research, I examine business network formation in the MLM industry, focusing on the relative roles of economic versus social factors. This specific industry provides a good environment to examine the formation of the individual business relationship, when there can be a conflict or co-presence between the social and economic aspects. I leverage a novel dataset from a large MLM company spanning almost two years of network formation among its IBOs. Specifically, I quantify the role of economic returns, which I term the economic factor, and social homophily (demographic and geographic), which I term the social factor, in determining network relationships. Through this, I shed light on how IBOs select individuals to partner with or vice versa, and I investigate whether the prevalent belief about using social connections to build networks has empirical support. Such social-professional relationships differ from other non-individual-based business relationships, which mainly emphasize business potential, such as partners that can generate significant values with better sourcing process abilities (Ni and Srinivasan 2015) or experiences (Sorensen 2007). In addition, using the recovered social and economic factors determining the relationship value between an IBO and a new joiner, I explore whether these factors affect the new joiner's business activity. There has been some work focusing on the MLM industry in general. Some of the topics include the motivation to leave or join the business (Jain et al. 2015, Wotruba and Tyagi 1991), institution setting comparison (Brodie et al. 2002), determinants of business outcomes post relationship formation (Frenzen and Davis 1990), and size and profitability of the network (Coughlan and Grayson 1998). However, business relationship formation has not been a main focus. According to the MLM Executives Industry Survey, the average IBO uses 76% of his or her total work hours in building business networks and managing an existing network of IBOs as opposed to engaging in retail selling (Coughlan and Grayson 1993, 1998). This indicates the importance of examining how IBOs form business relationships with one another. There is some evidence that whom one chooses as a business partner is critical for success (Coughlan and Grayson 1998, Crittenden and Crittenden 2004, Allen 2016), and that IBOs and the new joiners engage in purposive seeking behavior (King and Robinson 2000). Also, in order to grow the distributor network from the MLM company's perspective, it is crucial to understand what type of sponsoring relationship leads the new joiners to actively engage in the business later on. I conduct an exploratory analysis to examine the question. Due to lack of behavioral data, a vast majority of previous works on the MLM industry relied on simulation-based approaches (Legara et al. 2008, 2009), survey data or experiments (Coughlan and Grayson 1998, Jain et al. 2015, Grayson 2007) or interviews (Albaum and Peterson 2011, Grayson 1996).

The MLM industry is controversial, especially as many of its critics confuse it with pyramid schemes or Ponzi schemes (Bonoma 1991). The latter are illegal while the former is legal, as clarified by the United States Federal Trade Commission in 1979 (see also Albaum and Peterson (2011), Keep and Vander Nat (2014), Peter J. Vander Nat (2002) and Bosley and McKeage (2015)). While it is not easy to distinguish one from the other (Taylor 2011), the main distinction comes from the compensation rule vis-a-vis recruiting. Specifically, if performance bonuses are generated from the legitimate sales of goods to consumers, or to oneself, or from the sales generated by people who are recruited, as opposed to the mere act of recruiting other distributors, then this activity is legal (Xardel 1993, Albaum and Peterson 2011, Legara et al. 2008, Taylor 2011).[1] The industry is also seen as unpleasant to work in due to the high pressure of recruiting new people, the difficulty of selling products and the alienation of family and friends (Grant 2012). Though understandable, there are a lot of misconceptions regarding these aspects of the industry, and other observers have tried to set the record straight (Coughlan 2012). Another misconception is that this industry is based on a "flawed" business model and is therefore doomed; however, the industry in the United

[1] Xardel (1993) explains in detail the criteria to distinguish legitimate MLM companies from illegal pyramid schemes - see Appendix 8. Most MLM companies provide specific guidelines that clarify the legality of their business, and some note explicitly that bonuses are a function of a threshold level of business revenues generated by an IBO's network.

States has been operating for 50 years and continues to grow, recording a 4.8% increase in sales in 2015 relative to 2014, reaching $36 billion and accounting for 37% of the retail sales generated globally in the MLM industry (Tortora 2015). A recent estimate from the World Federation of Direct Selling Association (WFDSA 2016) suggests that there were more than 103 million agents all across the globe in 2015 — an increase of more than 20% from just five years ago.

In the MLM industry, IBOs do not receive fixed compensation from the firm. The firm gives IBOs freedom to build their networks (i.e., hire sub-distributors), does not set fixed quotas and allows them to decide on the amount of time they spend on recruiting versus selling. IBOs therefore have direct control over the time and effort they put into the business and also the consequent returns from the business. However, the firm sets clear incentives, rewarding IBOs with performance bonuses based on the product sales resulting from network activity. Thus, network formation is a central issue in this industry (King and Robinson 2000). Having larger networks is critical, as it allows IBOs to grow the business, "multiplying" their efforts by penetrating new markets through advertising, distribution and sales (Peterson and Wotruba 1996, Jain et al. 2015). Also, the challenge of IBOs is to build and maintain one's own network of IBOs and thereby benefit from the work of those recruited (Brodie et al. 2002).

As noted above, both industry insiders and outsiders believe that growth of this network is best achieved via leveraging the IBO's social network because these people are the easiest to reach and convince (Clothier 1997). However, other researchers have also pointed out that leveraging social relationships can create potential conflicts, since business dealings can distort the existing relationship and the partners can have different role expectations (Bloch 1996, Grant 2012, Grayson 2007). Other research notes that selecting the "right" individuals, maintaining their motivation, and developing their skills are crucial to success in this industry (Crittenden and Crittenden 2004). This argues for choosing IBOs less for social reasons than for business reasons. There is also evidence that the network structure contains information and therefore can potentially affect indivudals' actions (Centola 2010).

In order to investigate these issues, I specify a two-sided matching model and take it to the data. Two-sided matching models are often used to model the underlying preferences driving the observed match between two sides (with mutual agreement). Such models have been applied in various settings, e.g., the marriage market (Logan et al. 2008, Hsieh and Lee 2012), the education market (Boyd et al. 2013), the mutual fund market (Park 2008, Chen 2013), the sourcing market (Ni and Srinivasan 2015) and the sports recruiting market (Yang and Goldfarb 2015). The data comprise matches (between existing IBOs and new joiners) in a specific geographic area of the Unites States over a twenty-two month period. I estimate the model using hierarchical Bayesian methods.

In brief, results show that while social factors (measured via demographics, income similarity, and geography) do play a role, both parties are impacted significantly by economic considerations in finalizing matches. New joiners face a higher relationship value when joining IBOs that can provide higher levels of business support. IBOs tend to prefer new joiners who are likely to be more motivated to make economic returns. Interestingly, contrary to expectations, the level of the distributor in the network hierarchy does not influence the preference of new joiners. To quantify the relative importance of the economic and social factors from the realized matches, the social factors had about one-third the importance of the economic factors at the aggregate level. Also, based on my exploratory analysis of whether economic or social factors explain the new joiners' future business activity, I find that new joiners whose relationships have higher economic value engage more actively in the business.

Essay One adds to the literature on business networks, network formation, and the MLM industry on multiple dimensions. First, I am able to model both economic and social factors that go into network formation. While the industry and previous research argues that social factors are the primary force in network formation, I find evidence that economic factors matter substantially as well. The two-sided matching approach isolates the different considerations on both sides of the match. Second, I also examine the role of network structure as a signal enabling network formation. Third, using the recovered social and economic factors, which constitute the relationship value between IBOs and new joiners at the relationship formation stage, I am able to draw an inference on the new joiner's MLM business engagement. Finally, to the best of my knowledge, this is one of the first papers to use actual behavioral data from the MLM industry.

The remainder of the paper is organized as follows. In section 1.3, I describe the institutional setting as well as the conceptual and empirical industry background and conceptual framework used to guide the empirical analysis. I describe the research setting and data in section 1.4. Section 1.5 describes the modeling framework and results. I discuss findings in section 1.6. Section 1.7 concludes with a discussion of the study's limitations and suggestions for future research.

1.3 Institutional Background

1.3.1 The Multi-level Marketing Industry

The Direct Selling Association broadly defines direct selling as the marketing of products or services directly to the end users involving face-to-face sales, where the products and services are distributed, advertised, and sold by the distributors without a distinct retail store (Peterson and Wotruba 1996). In 2015, the total retail sales from direct selling stood at $36 billion, which represents an annual increase of just over 4.8%. In general, the sales of the industry have been growing steadily since 2009. The number of people involved in direct selling as distributors has also been increasing, with 20.2 million people in 2015 representing an 11% annual increase (DSA 2015a). This figure is around 6.2% of the total U.S. population. MLM, also known as network marketing, is one method of organizing and compensating salespeople in a direct selling business. Specifically, in this method, the distributor not only sells the products, but also recruits, trains and manages other recruited distributors, who will replicate such activities. The use of the moniker MLM is due to the characteristics of the commission plan in this setting - the commission is paid on sales made at "multiple levels" or "multiple tiers" in the hierarchically structured tree-like distributor network, which the distributor created through recruiting (King and Robinson 2000, Biggart 1989). Yet in most cases, each distributor has a direct access to the firm to purchase the product and does not have to go through the multiple tiers of distributors. Previous literature has used the terms "MLM" and "direct selling" interchangeably until recently (Brodie et al. 2002), probably because MLM firms dominate the direct selling industry by 95.7% in terms of percent of firms following an MLM-style commission plan, and by 99.8% in terms of the number of distributors working in the industry (DSA 2015b)).[2]

Given the commission structure in MLMs, the IBOs (distributors) focus on developing their own networks, recruiting new distributors, motivating and managing them and also encouraging them to develop their own network in turn, via compensation, special incentives and recognition. Through these activities, an IBO attains not only monetary success but also intrinsic feelings of personal satisfaction by forming a productive and growing network organization (Brodie et al. 2002). This is in contrast to a single-level direct selling setting, where distributors do not build their own network via recruiting and training, but rather focus on achieving compensation based on their own direct sales to the end customer (Brodie et al. 2002). MLM businesses are therefore run by self-employed IBOs, who distribute and/or

[2]The percentage of sales explained by MLM is also high - 97.1% (2015 Growth & Outlook Report: U.S. Direct Selling in 2014, Direct Selling Association).

sell products made or selected by the MLM company. Unlike employees in the traditional business domain, IBOs have a lot of freedom in how they conduct their business. For instance, the IBOs can sell in any territory or region, using their own operations without any direct supervision, receiving the appropriate commission as long as they operate within the policies and procedures defined in the MLM company regulations. Some of the examples of these regulations include product claims that can be made, the type and amount and content of advertising, specific locations where sales can or cannot be made, as well as the specifics of the compensation plan (Biggart 1989, King and Robinson 2000).

New joiners become part of the business through multiple routes. They can meet their future sponsor distributor if they are already end customers, or if they are friends or relatives or business acquaintances. They may be invited to a meeting organized to explain the sales and marketing plan (some IBOs, typically the successful ones, arrange these meetings periodically). Others travel great distances to meet people and explain the MLM setting to them (Xardel 1993). Potential joiners can also leverage the company's website, which reveals distributors' contact information. If a new joiner is interested in the sales and marketing plan of the MLM, s/he agrees to be sponsored by an existing distributor. Figure 1.1 depicts the network formation process. In Figure 1.1(a), each distributor in the business is labeled by a number according to his/her position in the structural hierarchy. In industry parlance, this number is denoted as a "level." When the new joiner joins the business, his/her level is determined by the level of the sponsoring distributor. For example, if in Figure 1, Red's level is 5, all the IBOs who are directly recruited by Red (both Blues) are one level below – represented numerically as level 6. Level does not necessarily reflect how long the distributor has been in the business as it is solely dependent on the level of the sponsor. The Blues are now denoted as Red's "direct downlines" while Red is each Blue's "upline." The new joiner can be connected only to one upline and in most cases cannot change his/her upline. All the IBOs who are directly or indirectly connected to Red from below (e.g., Blues and Oranges) are denoted as Red's "total downlines."

Each distributor has two potential sources of income. The first source is via the sales of the products directly to the end consumer, by selling products at a margin. The second source is via the sales of his or her recruits, the sales of the people they in turn recruit, and so on. For example, in Figure 1.1(a), Red's second source of income arises from the sales of his total downlines (Blues and Oranges) as a form of bonus. As noted earlier, this second source of income makes MLM different from single-level direct selling. This second source also provides the impetus and motivation to each distributor to manage and expand his/her network. In general, the relative contribution of the two sources of income varies by status, hierarchical position within the network and experience in the business. Typically, the higher the status, the higher the hierarchical position and the longer the experience, the higher the contribution of income from the network.[3] I now detail the social and economic factors that could influence new joiners and potential uplines in terms of network formation in MLM settings.

1.3.2 Social Factors

Social factors have generally been considered very important in the MLM industry. The relatinoship between the upline (IBO) and the direct downline (new joiner) has been characterized as the "key" social relationship in the MLM industry (Sparks and Schenk 2001). Biggart (1989) notes that in MLM networks, the business nature of the relationship comes second to the social nature of the relationship. In terms of network formation, many MLM firms suggest that the best way to find new distributors in the initial stage is to "list people you know and begin from there." (Lilyquist 2016a) Also, social connections, who are considered more credible than business connections, often attest to the quality of the product and share their positive experiences with the firm. In the MLM framework, the end customer's willingness to pay is highly dependent on relationship marketing; therefore customers prefer to purchase products from acquaintances, relatives, etc. rather than strangers (Legara et al. 2009, Jain et al. 2015). Business relationships often begin from there (Legara et al. 2009).

While I have information on the business network links among distributors, I do not have direct information about their social networks. I use previous research showing that homophily, i.e., the phenomenon where people prefer those who share observable characteristics, leads to a higher probability of forming connections (Ansari et al. 2011). The observable characteristics I use are demographic characteristics and Zip-code location.

1. <u>Demographics</u>:

   Network formation based on common observables has been documented in both empirical and theoretical papers (Christakis et al. 2010, Currarini et al. 2009). As noted above, distributors are "pushed" to interact with friends and family in the hope that these social contacts will become part of their business networks (King and Robinson 2000). It is also easier to interact with others who are similar to oneself (Harrison and Klein 2007), and such interactions may lead to a future business partner. Thus, I expect that common observable demographic characteristics will increase the potential to build these relationships, especially as previous research has documented that this commonality helps to build trust (Narayan et al. 2011).

[3] For example, in the estimation sample, the high status IBOs get 80% of their income from the network, while the average distributor (across all status levels) gets 31% of his/her income via the network.

2. Geographic Distance: Along with demographics, geographic distance has often been incorporated in the matching settings as a proxy for homophily as similar people are likely to be in close physical proximity in terms of home and/or work (Boyd et al. 2013, Geweke et al. 2003). Also, geographic proximity has been identified as one of the main factors in engendering word of mouth towards product or service adoption (Nam et al. 2010, Bell and Song 2007, Albuquerque et al. 2007, Choi et al. 2010). New joiners in the MLM industry rely heavily on word-of-mouth information before taking the plunge (Legara et al. 2009). Thus, if the new joiners are located close to IBOs, they may have a higher chance of hearing product testimonies. Closer physical proximity also increases the probability that the distributor is known to the new joiner.

1.3.3 Economic Factors

Economic factors are also likely to have an effect on network formation. Contrary to industry beliefs, there is evidence that successful IBOs tend to have very few friends and family involved in the business, and in most cases, friends and family who are in the business join after seeing the success of the business (Lilyquist 2016a). In addition, some studies suggest that in MLM settings, business-related factors (e.g., extra money, product quality, work freedom) drive joining the business, whereas social factors (e.g., building rapport, CSR, fun) do not contribute in motivating people to join the business (Jain et al. 2015), and that economic support from existing IBOs leads to stronger relationships (Bandura 1962). Other evidence suggests that focusing too much on strengthening social ties and building unit cohesion can have adverse effects on individual sales (Sparks and Schenk 2006).

These findings support the claim that network formation is driven by economic or business-related factors as well. Consistent with this, business support from the potential upline has been identified as one of the major factors that influence the satisfaction of IBOs in MLM marketing (Lee et al. 2016). I examine the factors that could signal economic factors for both new joiners and IBOs below.

1.3.3.1 Distributors' Characteristics (New Joiners' Economic Return)

New joiners are often advised not only to look for the right company with the right product (Collamer 2013, Lilyquist 2016a), but also to engage in thorough research to find the right person to work with as a sponsoring upline distributor (Allen 2016). In the MLM industry, the key leadership relationships are those between IBOs and the member who recruited them into the organization (i.e., their uplines). As mentioned earlier, each distributor has a lot of motivation to help the new joiners to succeed in their business, mainly because the distributor receives a commission based on the sales made by his/her recruits and all the recruits the new joiners will eventually sponsor. In addition, in most MLM companies, the distributor Rules of Conduct *require* all uplines to train and supply IBOs they have recruited/registered, as well as offer multiple ways to participate (Xardel 1993, Crittenden and Crittenden 2004). The more a distributor is willing to build and succeed in the business, the more she will focus not only on recruiting, but also on training the new recruits on best business practices (King and Robinson 2000). As a result, new joiners will experience higher relationship value when partnering with IBOs with established reputations and demonstrated business success (Bonabeau 2002). The challenge for joiners is to assess the level of business support prior to choosing which distributor to link with, based on physical contact, word of mouth or online research (Lilyquist 2016a). They typically do this by examining the following characteristics of existing IBOs.

1. <u>Status from Business Performance</u>: As the founder of a leading MLM firm said, "We have two forms of reward in this world, one is recognition, and the other is dollars. I employ them both in the business" (Klebnikov 1991). Consistent with the academic definition of status (Hu and Van den Bulte 2014), status in this industry reflects social stratification based on economic success as well as reputation and recognition. Specifically, status is awarded based on two metrics — one's own financial activity and that of one's connected downlines. All the IBOs who devote enough time and effort to the business are qualified to be ranked in the higher status, reflected by different signifiers.[4] Based on their financial achievement, IBOs not only earn additional financial rewards, but also receive a significant amount of social recognition (Xardel 1993, Amway 2015). Status does not follow from the formal lines of authority as in conventional organizations, but is a function of ability and performance, creating a social hierarchy that is different from network structural hierarchy (Pratt 2000). The recognition manifests itself in being highlighted at company conventions and meeting, being asked to instruct other IBOs who are not as successful, etc. In general, the higher the status, the more the time spent by an distributor in personally teaching new joiners how to build a successful business (Pratt 2000).

   Thus, from a new joiner's perspective, status of the uplines is likely to be a signal about the the uplines' degree of success, and how much the distributor is recognized and respected in the business. The higher the status, therefore, the more desirable the connection.

[4]This signifier varies by firm. One large MLM firm uses job titles to signal status. Another one uses precious gemstones as signifiers.

2. <u>Duration in the Business</u>: The MLM industry is known for a relatively high turnover rate (Wotruba and Tyagi 1991, King and Robinson 2000). There could be multiple reasons why this is the case. First, compared to other industries, there is a high degree of freedom for each distributor and there are no barriers to exit. Second, as more than 90 percent of IBOs are engaged in the business as a part-time job supplementing their main income (DSA 2015b, Dir 2017), it may be unnecessary for them to stay in the business for a long period of time. Third, some IBOs, after joining, question the income potential in the MLM industry and exit. In this situation, long tenure by an existing distributor signals that s/he has probably exceeded the income threshold that is required for him/her to stay with the MLM company. In other words, s/he is financially successful (Wotruba and Tyagi 1991). Thus, the longer a distributor remains in the business, the more likely s/he is to attract new joiners. On the other hand, IBOs who are relatively new to the business may be more enthusiastic about all aspects of the business (Legara et al. 2008), and this could be valued highly by the new joiners. Thus, it is an empirical question to examine how and to what degree long tenure in the business generates higher relationship value in the new joiners' network formation decisions.

3. <u>Relative Level Based on the Network Structure</u>: Level is a measure reflecting one's hierarchical location in the tree-like network structure or network hierarchy (see Figure 1.1). The structure of an MLM network is hierarchical, which may suggest that IBOs located at a higher level would earn more. However, given the compensation system, the profit opportunity is technically the same for each IBO, regardless of where s/he is located in the network hierarchy (Lilyquist 2016b). In addition, as explained earlier, an IBO's level is determined by his or her sponsoring IBO when s/he joins the business, and it does not change over time. Thus, distributors can even possibly earn more than those who bring them into the organization (King and Robinson 2000).

   So while this suggests that network hierarchy should not play a role after controlling other correlated aspects, it may be hard for new joiners to understand this *ex ante*. Prior literature has also suggested that the network structure contains (perceived) information on power, knowledge dissemination and innovation within firms and can affect adoption (Albert and Barabási 2002, Van den Bulte and Wuyts 2007). The predictions for how network structure affects outcomes are mixed, with Legara et al. (2008) showing (via an analytical model) that earnings are independent of location in the network hierarchy, while Daquis et al. (2013) show (via simulation) that higher levels can make more profits. None of these predictions are tested with behavioral data or in terms of network formation, which gives us an opportunity to do so in our approach.

4. <u>Distributor's Focus</u>: I incorporate variables that may capture distributors' differential focus on business. As mentioned earlier, there are two main sources of income: product-sales based commissions generated through networking activity, and selling the products to the end customer. IBOs can freely rely on one income source more than the other by engaging in different business activities, depending on the distributor's focus in the business. Of course, I cannot tell whether an IBO pursued and attained a certain focus, or whether an IBO is in the process of achieving a certain focus. Yet such different styles of business may carry different economic meanings in creating business relationships. Distributors who rely on building networks (and earn commissions) will be reflected by the number of direct downlines and the existence of indirect networks, whereas those who rely on selling the product will be reflected by number of registered retail customers.

   The number of direct downlines can signal what kind of income source the potential upline is capable of. Specifically, the number of direct downlines can be mapped to the "degree" measure in the network literature. Degree has often been examined in various network-related studies reflecting popularity, reputation, etc. (Albert and Barabási 2002). IBOs who are already supervising a larger number of direct downlines may have better chances of attracting new joiners as the new joiner may prefer to be connected with a more popular entity, which results in preferential attachment (Albert and Barabási 2002). In addition, a large set of direct downlines can testify to the distributor's recruiting skill.

   However, having a large number of direct downlines may also have negative effects on the potential uplines' attractiveness. This is because the larger number of direct downlines may imply that the new joiner may receive less attention and support, assuming that the uplines have finite capacity and time to invest in this MLM business. Therefore, the size of the existing direct downlines to the potential upline distributor may have multiple effects on new joiners' perceptions of relationship value, and whether the upline is considered to be attractive or not due to the number of existing direct downlines will depend on the relative magnitude of the above-mentioned effects.

   The existence of indirect networks captures whether the total number of downlines exceeds the number of direct downlines or not. If the upline is focusing more on expanding the whole network, s/he will spend time teaching and motivating connected downlines to engage in building network activity. Potentially, the focal upline's income source will greatly depend on the indirect network that others created. Will the new joiners prefer to be attached to a distributor who has successfully built the business multiple levels below by transmitting the business practices with high management skills, when the downlines could be perceived as possible competitors?[5]

While the number of direct downlines and the existence of indirect networks can signal the IBO's focus and their ability in terms of recruiting and managing/expanding the network, the number of retail customer proxies for the IBO's ability to sell the products to registered retail customers.

1.3.3.2 New joiners' Characteristics (IBOs' Economic Return)

For the success of the business, selecting the right individual matters not only for new joiners, but also for IBOs (Crittenden and Crittenden 2004). The MLM Executives Industry Survey shows that an average IBO uses 44% of his or her total business hours in managing the existing network of IBOs — the most time-consuming activity in recruiting and retailing selling (Coughlan and Grayson 1993, 1998). Managing network involves teaching, training and motivating downline distributors to both sell the product and recruit new IBOs. Preserving and building the business network not only takes time but also requires psychological commitment (King and Robinson 2000, Pratt 2000). Thus, IBOs need to maintain a good balance among selling the product, maintaining the current network, and recruiting new direct downlines (King and Robinson 2000). Because of the time commitment and psychological investment required, the ability to find the right prospective IBOs has been recognized as one of the keys to success in the business (Lilyquist 2016b). Thus, IBOs will try to seek for the most suitable direct downlines to recruit given their circumstances.

Whereas new joiners have abundant information to judge the possible uplines, current uplines do not have much information to judge whether the new joiner will be productive enough to contribute to their income steam in the long term. Individuals join the MLM industry for various reasons, which in turn leads to different levels of time and effort allocation in the business (DSA 2015c). For example, some people enter the industry to make it their main source of income, while others enter it to obtain supplemental income. This variation is one of the reasons the turnover rate is especially high, and it accounts for the relatively small average earnings in the industry (Albaum and Peterson 2011). Given that the IBO's commission is based on the downlines' product

[5] A similar argument holds for the number of total downlines — the number of direct downlines. But unlike the number of direct downlines, which IBOs know, there is uncertainty on information about the entire network below them. The result is consistent with what I report in the results section.

selling and recruiting activities and that they need to put a lot of effort into motivating and training new joiners (Coughlan and Grayson 1998), IBOs will prefer new joiners who are more committed to the business.

1.4 Data

Data come from a leading global MLM firm that has been a pioneer in this industry.[6] This firm has more than 3 million IBOs operating in 100 countries. The main product categories that the firms sells via its IBOs are nutrition, health & beauty, bath & body and home decor. The data cover a twenty-two-month period from January 2006 to October 2007, with activity recorded at the monthly level. The data include detailed information on each distributor's identification number, hierarchy and status, network-related characteristics, demographic variables (age, gender, marriage status, language, ethnicity) and Zip code of the distributor. The time-series data allow me to map out the network formation and growth for distributor over this twenty-two-month period. I first explain how I set up the data for the model presented in §3 and then present descriptive statistics.

1.4.1 Network Formation via Matching

The estimation sample is restricted to IBOs in the Chicago area. As is common in this literature (Sorensen 2007, Yang et al. 2009, Agarwal 2015), I include the set of IBOs who formed relationships during the sample period with complete information. More specifically, the focal new joiners at month t are defined as those who joined the business in the Chicago area at month t and whose connected uplines are located within a 35-mile radius of their location. The goal of examining the relationship within the geographic boundary is to capture the formation based on local interaction. The limitation of this approach is that I do not include relationship formation between distributors who are located farther away, which is also possible via phone or internet. However, this approach allows me to include about 80% of total newly formed relationships in the Chicago area during this period.

Table 1.1 shows the number of uplines and the new joiners in each month in the sample. I have a total of 884 observations for distributors (uplines) and 941 observations for new joiners (downlines). Among those 884 observations, 685 unique distributors are represented. It is also possible for a new joiner to become an upline, if s/he is successful in recruiting the new joiners (within the same month or in a subsequent month). As our data are at the monthly level, the number of new joiners who became uplines within the same month is

[6] I cannot reveal the name of the firm due to the non-disclosure agreement under which the data were obtained.

indicated in the column 'Transitioned IBOs.' Among the 941 new joiners, 7.43% successfully added a new direct downline within the same month.

Note that in any month, an upline can match with one or more downlines. The most common type of match is a one-to-one match, where an upline matches with only one joiner. For example, if Gray would like to match an existing distributor for the network in Figure 1.1-a), she can match with four potential uplines - Red, Blue, Orange 1 and Orange 1. If she chooses Red, then the network looks like Figure 1.1-b). However, it is possible for Red to acquire more than one downline in a month. Table 1.2 shows the number of downlines matched with uplines in each month. While the majority of matches are one-to-one, 6% of our matches are one to more than one. The modeling approach therefore needs to accommodate these types of matches as well.

1.4.2 Matching Related Variables

1.4.2.1 Social Factors: Demographic Homophily & Geographic Distance

I have data on three demographic variables - age, language and ethnicity. Table 1.3 shows the proportion of language users and ethnicity in the estimation sample, including both uplines and new joiners. Three languages (English, Spanish and Korean) and seven ethnic groups (European, Hispanic, Asian, African American, Non-Orient Asian/Polynesian, Jewish and Arab) account for a majority of data. The age distribution across our sample can be seen in Figure 1.3.

In order to see whether networks in this industry showed homophily, I looked at patterns of these three variables across uplines and downlines. There is significant and positive correlation (0.34) between the age of uplines and downlines. The contingency table for language and ethnicity is shown in Table 1.4 and Table 1.5. The cells showing the highest percentage from the downline's perspective are marked with $\star$. These data patterns on language and ethnicity suggest that there is evidence for demographic homophily in the MLM network. To incorporate this into the model, for each distributor-joiner pair, I used binary variables to denote whether the pair shared the same language and ethnicity. For age, I used the absolute difference in age between each distributor-joiner pair. As I focus on the homophily effect based on the clear pattern in the data, I use the absolute difference, but I acknowledge that the specification does not capture possible age asymmetry effects, for instance, whether seniority plays a role in the choice of business partner.

Lastly, I incorporate dyad-level income for following three reasons. First, people may interact at the work site and learn about supplementary income opportunities. According to a survey data from Raymond and Tanner Jr. (1994), although most of the consumers made product purchases at home, the workplace was the second-most-frequent location for the business interaction, accounting for 39.5% of respondents. Second, it helps to capture members of the social inner circle with similar income levels. Third, people with similar income level are highly likely to live nearby in the same neighborhood. I transformed nine income groups shown Table 1.3 into a dyad-level binary variable indicating membership in the same income group or not, to capture income similarity.

Even within the Chicago area, geographic proximity is likely to affect both potential new joiners' and uplines' preferences. I compute the centroid based on distance between the new joiner and the existing uplines based on Zip-code information. The histogram on the right-hand side in Figure 1.3 shows the realized distance distribution in our sample from the new joiners' perspective. The figure shows that most new joiners and uplines who formed a relationship were physically fairly close to each other.[7]

1.4.2.2 Distributor Economic Signifiers for New Joiners

Now let's turn our focus to the variables that signal the attractiveness of distributors to new joiners on economic grounds. I begin with the descriptive statistics of the high status in Table 1.6.

In addition to the level in the network — a structural hierarchy, other measures of hierarchy exists, which is the status. High status signifies success, typically based on earnings and other contributions to the firm. The firm provided data on the most important signifier, denoted here as "High status." This is an extremely difficult signifier to earn as it is reserved for the most accomplished IBOs. For example, among 884 uplines in our sample, only 7.5% of IBOs currently hold (or have held) high status.[8] As a result, achieving this status is highly aspirational for all IBOs. The earnings of high status IBOs were much higher than those of low status IBOs with an average monthly payment of \$2669 versus \$17 (based on 2007-08 data provided to us by the firm). High status IBOs also get a "recognition pin" and are featured in distributor publications, get to participate in the annual growth-incentive program and are invited to an all-expense-paid conference to obtain useful resources and connections to help them grow as leaders.

Duration in the business is measured as the number of months the distributor has been part of the business. I see a wide variation in the distributor focus related variables. The

[7] As noted earlier, I have censored the distance at 35 miles, covering about 80% of network formation in the Chicago area. Without censoring, the distance (in miles) between each pair is distributed as follows: minimum = 0, 1st quartile = 1.9, median = 6.1, mean = 130.7, 3rd quartile = 24.320, max = 7371.

[8] Status is evaluated every fiscal year, and the data provided do not record the year-by-year status of distributor. However, I have data on whether the distributor ever achieved high status and his/her status at the end of the data. In addition, a new joiner cannot have high status. Using these features, I am able to obtain the status of a distributor quite precisely.

number of direct downlines and the number of total downlines reflect the degree to which the distributor has been focusing on building network activity, and the number of retail customers indicates that of product-selling activity (Figure 1.4). Among 884 uplines, 19.5% do not have any prior recruiting experience before the match I observe (having zero direct downlines) and 63.8% of uplines do not have any retail customers.

In contrast to the prevailing beliefs in the MLM industry, Table 1.7 shows that being a high status member is relatively highly correlated with the distributor focus variables (the number of total downlines, direct downlines, and retail customers) but not very correlated with level (-0.05). This implies that location in the network hierarchy is not a predictor of business success. As expected, I see a high correlation between experience and the number of direct downlines. In terms of level, people who are located in the lower part of the tree generally have less experience in the business.

1.4.2.3 New Joiner Economic Signifiers for Uplines

In order to proxy for a new joiner's economic potential, I use data on his or her annual financial earnings filed as part of his or her application. Using this variable, uplines can infer the degree to which a downline relies on this business and "hunger" to succeed in the MLM business. Table 1.3 shows that about a quarter of all applicants have very low annual incomes (less than $15,000), while about 40% earn above $50,000 annually.

1.5 Modeling Approach

1.5.1 Two-sided matching model

As shown in the data section, my modeling framework has two sides - IBOs and new joiners. Each IBO can recruit multiple new joiners (who become direct downlines), but each new joiner can be matched with only one distributor, or has a one-to-many match. Thus, I employ an empirical two-sided matching game called the College Admission Model, wherein the IBOs and new joiners decide with whom to match. This model is an extended version of the Marriage Model that allows only a one-to-one match (Gale and Shapley 1962, Roth and Sotomayor 1992).

The two-sided matching model is chosen because of the two distinct characteristics that differentiate it from discrete choice models. First, IBOs' and new joiners' decisions are not independent (Park 2008). To illustrate, in the typical discrete choice model setting, each retail consumer chooses a product that maximizes his or her utility. In general, such a consumer's decision is not affected by other consumers' choices, or by the availability of the product caused by other consumers' choices, as the exact same product can be offered to the consumer. In the two-sided matching model, however, the business relationship decision of an IBO or a new joiner affects how others make their decisions, because no business partners are exactly the same, and the formation of one relationship limits others' choice sets. More importantly, unlike a shopping setting where consumers choose the product, here, both IBOs and new joiners need to mutually agree in order to form a match. Second, unlike the discrete choice model, where the choice set is usually exogenously pre-specified, in this model, a decision maker (an IBO or a New joiner) cannot take the choice set as given as (potentially) not all decision makers on the other side may be willing to match with him/her. This issue has been termed in various ways as the endogenous choice set issue or the simultaneity issue in the two-sided matching literature (Echenique et al. 2010, Hsieh and Lee 2012). What this implies is that the preferences of the IBOs and the new joiners determine the willing partners for each IBO and new joiner, which constitute a choice set or an opportunity set, as defined in a later section). In our approach, I use Bayesian Markov Chain Monte Carlo methods to simulate these unobserved opportunity sets.

The major components of the model are the IBOs' and the new joiners' latent match utility, which is the relationship value that reflects preferences on both sides. Together with the rules of the matching game, I recover the latent match utility and the associated parameters that determine the equilibrium match, as reflected in the dataset. Using this model, I can understand the determinants of the business relationships in the MLM industry. [9]

1.5.2 Two Sides, Utilities, and Equilibrium

The general rules in the matching model are that any pair of agents on the opposite side of the market (e.g., IBOs - New joiners) may be matched to one another bilaterally, if they mutually agree. Also, the observed match in the data is the outcome I want to explain through the model. Here I lay out the model with three building blocks: two sides, latent match utilities, and equilibrium.

[9] This approach is consistent with previous approaches such as those of Logan et al. (2008) and Hsieh and Lee (2012) in the marriage market and Yang et al. (2009) in the professional team sports market (the match between sports team, and players), where understanding the match itself is the focus. This is in contrast to other works such as Park (2008), Sorensen (2007), Akkus et al. (2014), and Pan (2014), where the objective is to supplement the outcome equation by explicitly outlining the match between the two sides to permit non-random matching while addressing the endogeneity problem and/or inferring a causal relationship between the match and the focal outcome.

1.5.2.1 Two Sides

Two separate sides are involved in the model — the IBOs' side and the new joiners' side - and I model the relationship formation between the two sides. In forming a set of people on each side, I have no information on a potential set of IBOs and new joiners who wanted to form business relationships, but failed to do so. Therefore, the model is based on the observed matches, as is often the case with empirical matching models. This means that the realized relationship identifies the set of people on the IBO and new joiner sides, and nobody on the IBO and new joiner side remains unmatched (Sorensen 2007, Akkus et al. 2014, Yang et al. 2009, Boyd et al. 2013). In turn, I do not address what motivates an individual to become part of the MLM industry or why a certain distributor decides to enter the recruiting market at a specific time, as I condition on the set of actual matches.

Let I_t and J_t denote a set of IBOs (or potential uplines, from the new joiners' perspective) and a set of new joiners (or potential direct downlines from the uplines' perspective) at time t, which constitutes a matching market, where t = 1,2,....,T. The set of new joiners, J_t, consists of all the newly joined business owners who are connected to one of the existing distributors by the end of time t. The set of IBOs, I_t, includes those who formed a relationship with the counterpart new joiners at least once in month t. I_t is composed of two types of IBOs: the first type, the IBOs who entered the MLM business before month t, and the second type, those who entered the business during month t as new joiners and in turn successfully recruited other new joiners within the same month t, or so-called transitioned IBOs. Table 1.1 column 'Number of transitioned IBOs' indicates such IBOs.

Due to the one-to-many match, the unique aspect of this industry, the new joiners can be connected to only one of the existing IBOs by joining the business (exclusivity in recruiting). However, the existing IBOs can recruit several (s_{it}) new joiners in a given market t. The set of all the potential matches is $M_t = I_t \times J_t$, where a matching μ_t is a subset of M_t and the realized match in the data. IBO i's set of recruited new joiners is defined as $\mu_t(i)$, and new joiner j's connected IBO is denoted as $\mu_t(j)$. Following this notation, the match between an IBO i and a new joiner j in market t can be expressed in three equivalent ways: $ij \in \mu_t$, $\mu_t(i) = j$, and $\mu_t(j) = i$. I define the temporal duration of market t as one month to match the granularity of the data. To keep the study tractable, I model the market as independent.

1.5.2.2 Latent Match Utilities (Relationship Value)

Each new joiner and distributor on each side seeks to form a relationship with an individual from the opposite side. Formation of the relationship is governed by the latent utility of the relationship, or the relationship value for the set of all the potential matches M_t, which is represented in the V_t matrix. The number of rows in V_t matches the number of IBOs in I_t (n_{I_t}), and the number of columns matches the number of new joiners in J_t (n_{J_t}). Specifically, the relationship value between distributor i and new joiner j in month t is denoted as V_{ijt}, which constitutes the V_t matrix. This V_t is unobservable, and it is what I am trying to recover from the dataset. For ease of presentation, I suppress the month t notation.

$$V = \begin{pmatrix} V_{11} & \dots & V_{1j} & \dots & V_{1n_J} \\ \vdots & \ddots & \dots & \dots & \vdots \\ V_{i1} & \dots & V_{ij} & \dots & V_{in_J} \\ \vdots & \dots & \dots & \ddots & \vdots \\ V_{n_I1} & \dots & V_{n_Ij} & \dots & V_{n_In_J} \end{pmatrix} \tag{1.1}$$

I assume that the latent utilities, or relationship values, are pair-dependent, and shared between the two sides by the fixed sharing rule. In other words, IBO i's utility of being connected with new joiner j is the same as new joiner j's utility of being connected with IBO i - the variation across these relationship values explains the network formation. Thus, increasing the utility of either the IBO's or the new joiner's side will generate a higher relationship value. Even if the relationship values are defined at the pair level, IBO i will prefer a match that confers relatively higher V_{ik} as compared across the characteristics of new joiners $k \in J$. Similarly, new joiner j will prefer a match that provides relatively higher V_{kj} as compared across the IBOs $k \in I$.

In order to make the model tractable, I assume a complete information game on each side, which is a common practice in the empirical matching model. It seems plausible that new joiners who are somewhat interested in the business engage in gathering information on the potential upline IBOs. Some companies reveal the existing IBOs' information on the website to facilitate the decision making. Also, from the IBOs' perspective, in addition to drawing future direct downlines from existing social relationships, there are different ways to invite the new joiners into the business. One way is through a meeting organized to explain the sales and marketing plan (Xardel 1993). Although I have evidence of purposive searching between new joiners and IBOs (King and Robinson 2000, Allen 2016), as I do not have information about the detailed matching process (e.g., whom a person on one side approached on the other side in order to settle down with a final observed match), I rely on this assumption. However, the opportunity set over which the decision makers maximize their utility will be a subset of the people on the other side, which is explained in the following section.

This pairwise utility is assumed to be shared between the two sides by the fixed sharing rule. This assumption has been used in different settings in the empirical matching literature involving unassortative matching (Sorensen 2007, Park 2008). Even if the relationship value is specified at the pair level, the preference of each new joiner over the potential IBOs will be affected only by the characteristics of the IBOs (including the pair-dependent characteristics) and the error term, as the values related to the joiner's own characteristics are fixed. Similarly, the preferences of potential IBOs regarding the new joiners will be affected only by the characteristics of a set of new joiners and the error term.

Methodologically, this use of pairwise latent utilities with a fixed sharing rule assumption is known to handle the multiple equilibria issue, which is often raised in the two-sided nature of the empirical matching model (Uetake and Watanabe 2012b, Hsieh and Lee 2012). Few studies have carried out empirical estimation, suggesting broadly three different ways aside from this approach. The most prevalent approach is utilizing a transfer between the two sides in forming a relationship (e.g., a wage or price that gets transferred between the two sides in creating the match). This transfer allows us to derive a joint utility, and it can be estimated using non-parametric maximum score estimation, so-called the Fox estimator (Echenique et al. 2013, Fox 2008, Yang et al. 2009, Pan 2014, Yang and Goldfarb 2015, Zamudio et al. 2013). Some rely on *ex ante* knowledge of how the match is formed, or assume a type of equilibrium, or use a partial identification strategy (Uetake and Watanabe 2012b, Boyd et al. 2013, Uetake and Watanabe 2012a). The last approach relies on a vertical preference structure in the market, where everyone agrees on the preference ordering of the other side and vice versa (an assortative matching) (Clark 2006, Chen 2013, Lee 2014, Agarwal and Diamond 2014, Ni and Srinivasan 2015).

However, business relationship formation in the MLM industry does not involve explicit transfer. I do not have information on the business relationship formation. Also, the social variables in our study are pair dependent (e.g., demographic homophily, geographic distance), which prohibits me from imposing vertical preference restrictions on both or one side of the market. I therefore follow utility specifications that model the joint utility of both sides, as if both sides are sharing the common utility generated by the relationship, which is often encountered under matching without transfer (Sorensen 2007, Park 2008). Together with the fixed sharing rule, the utility of two sides align, ensuring unique equilibria without the need to assume transfer, an equilibrium selection rule, or a vertical market structure. Also, the utility of a business relationship is often determined by the synergy of the characteristics of the IBOs and the new joiners.

1.5.2.3 Equilibrium Outcome

In the structural two-sided matching model, the observed match is the optimal equilibrium outcome. The solution concept I adopt is a pairwise stability, which is the common equilibrium notion in the matching literature (Roth and Sotomayor 1992). According to this concept, no IBO and new joiner prefer each other over their observed relationship in the data. Or, a matching μ is stable if it cannot be improved by any IBO-new joiner pair (Roth and Sotomayor 1992). In the college admission model, a stable matching always exists (Roth and Sotomayor 1992), and the fixed sharing ensures that the model has a unique stable solution. When an IBO and a new joiner prefer to abandon their existing relationships with their own business partners, which is observed in the data, and form a new relationship with each other, they form a blocking pair. In order for the match to be stable, such blocking pairs should not exist. Thus, the pairwise stability is defined by a no-blocking-pair condition and can be specified with a set of inequalities.[10]

The set of inequalities that reflects the non-blocking condition is based on intuitive logic. To illustrate, let's assume that an IBO i and a new joiner k are *unmatched*. The reason i and k did not form a business relationship is that at least one side had an alternative business partner generating higher utility, and the alternative partner is what we observe in the data. An analogous example would be a marriage market. The reason a woman and a man did not marry each other is that at least one person had an alternative partner generating a higher utility, which led him or her to marry the alternative partner. Based on this logic, the relationship value between unmatched i and k (V_{ik}) should be lower than that of i and k's formed business partners observed in the data from at least one side, which can then justify the unmatched outcome. Thus the relationship values for all the *unmatched relationships* will be restricted with *upper bounds*.

On the other hand, let's assume that an IBO i and a new joiner j are *matched* with each other, which we observe in the data. Based on i's perspective, retrospectively, i must have built a relationship with j, who offers the highest relationship value among potential new joiners with whom i could have formed relationships but left behind. Now, as this is a two-sided matching model, we also need to consider j's perspective at the same time, as both perspectives must have been satisfied in order to form a business relationship. Similarly, j must have been connected to the IBO i, who offers the highest relationship value (V_{ij}) among those potential IBOs with whom j could have formed a relationship. Thus, following this

[10] In theory, the stability is defined by the following two conditions: individual rationality and the no-blocking-pair condition (Roth and Sotomayor 1992). The meaning of individual rationality is that no upline or new joiners form a match that is less preferable than being unmatched. However, individual rationality does not apply in our set-up, as I assume that agents prefer to create relationships.

logic, the relationship values for all the *matched relationships* will be restricted with *lower bounds.*

As the above illustration shows, a little complication comes from the one-to-many matching characteristic; namely, an IBO can gather more than one new joiner, whereas a new joiner can be connected to only one IBO. To present the non-blocking-pair condition mathematically, it is convenient to define the two concepts: $\overline{V}_{ij}$ and $\underline{V}_{ij}$. I keep the notation consistent with that used in previous literature. First, $\overline{V}_{ij}$ is an upper bound when i and j are *unmatched pairs.* $\overline{V}_{ij}$ is the pairwise opportunity cost of *deviating* from one of their current relationships that i and j have - observed in the data - and hypothetically creating a new match with each other. Mathematically,

$$\overline{V}_{ij} \equiv max \left[min_{j' \in \mu(i)} V_{i,j'}, V_{\mu(j),j}\right] \tag{1.2}$$

To explain, the first specification within the max bracket in equation (1.2) is from IBO i's perspective. If IBO i were to give up one of his/her existing relationships in order to form a new match with new joiner j, it would be the one that provides him/her the lowest relationship value. Again, as IBO i can recruit more than one new joiner, I need to compare the current relationships IBO i has, in order to compute IBO i's cost of deviation (the first part within the max bracket). The second part within the max bracket is from new joiner j's perspective. Whatever s/he needs to give up in order to form a hypothetical relationship with IBO i is the utility from his/her observed relationship, $V_{\mu(j),j}$. After acquiring the abandoning cost from both i and j's perspective, the maximum value of the two is the opportunity cost of deviating from their current match and forming a new match with each other.

Based on this explanation, for the observed *unmatched* pair ij, the relationship value $V_{i,j}$ must be lower than the opportunity cost of 'creating' the new relationship with each other and abandoning their current match, $\overline{V}_{ij}$. If $V_{i,j}$ were greater than $\overline{V}_{ij}$, the value that makes it worthwhile for both i and j to give up the observed relationship, then i and j would have an incentive to give up (one of) their current partners and form a relationship with each other, causing a conflict with the no-blocking-pair condition.

Second, $\underline{V}_{ij}$ is a lower bound when i and j are unmatched pairs. $\underline{V}_{ij}$ is the pairwise opportunity cost of i and j *remaining* together. In order to create the relationship $ij \in \mu$, i and j both must have given up all the other potential business relationships that they otherwise could have formed. Mathematically, the remaining cost is computed as

$$\underline{V}_{ij} \equiv max \left[max_{j' \in O(i)} V_{i,j'}, max_{i' \in O(j)} V_{i',j}\right], \tag{1.3}$$

where O(i) and O(j) are opportunity sets for i and j. In the two-sided matching model, each entity on each side chooses a relationship so as to maximize his or her relationship value over the set of available opportunities, which is defined by opportunity set $O(\cdot)$.[11] For instance, if IBO i could have provided a higher relationship value than new joiner j"s observed business parter based on data, $\mu(j')$, or $V_{i,j'} \leq V_{\mu(j'),j'}$, j' is in i's opportunity set. In equation (1.3), the first portion in the max bracket is the maximum relationship value that IBO i had to give up for the relationship with new joiner j. The second portion is defined from the perspective of new joiner j in a similar way.

As I mentioned before, there are two types of IBOs in I_t: the IBOs who entered the market before month t, and those who entered during month t and successfully recruited new joiner(s) within the same month. The transitioned IBOs appear both in I_t and J_t. Although I do not observe the actual opportunity sets for both sides and thus rely on the simulation, I do know that depending on when the new joiner applied to the business, the maximum pool of IBOs to consider may change - the new joiners who became part of the business after the transitioned IBO's entry might have faced a larger pool of IBOs. Thus, I adjust the opportunity sets accordingly, which is captured by the application order, d.[12] Mathematically, $O(\cdot)$ is defined as below. The opportunity sets of IBO i and new joiner j are:

$$O(i) = \left\{ j' \in J : V_{i,j'} > V_{\mu(j'),j'} \ \& \ (j' \neq i) \right\} \tag{1.4}$$

$$O(j) = \left\{ i' \in I : (V_{i',j} > min_{j' \in \mu(i')} V_{i',j'}) \ \& \ (i' \neq j) \ \& \ (d_{i'} < d_j) \right\} \tag{1.5}$$

In computing $\underline{V}_{ij}$, I need to consider only those who are part of the opportunity set, not the pairwise utilities generated from all the agents on the other side. This is because each IBO and new joiner have the opportunity to be paired only with anybody from the other side who wishes to be connected with him/her, or prefers him/her to his/her current match, as it is a mutual process.

Now using $\overline{V}_{ij}$ and $\underline{V}_{ij}$, the non-blocking pair condition is characterized by a set of inequalities following the two rules below (Sorensen 2007).

1. (Unmatched i and j) μ is stable if and only if $V_{ij} < \overline{V}_{ij}$ for $\forall ij \notin \mu$.

2. (Matched i and j) μ is stable if and only if $V_{ij} > \underline{V}_{ij}$ for $\forall ij \in \mu$.

If all the pairwise utilities in V satisfy the two preceding equality conditions, no IBOs

[11] Some researchers term it a feasible set or a choice set.

[12] I acknowledge that the slight modification that I incorporate to include the IBOs who entered during the month may introduce variations into the uniqueness condition. But instead of removing those IBOs who entered during the month, I included them with the minor restriction to reduce the bias and reflect reality. Observing more than twenty markets, along with using hierarchical Bayes estimation to pool information from other markets in the inference can alleviate this potential problem.

or new joiners believe that they can improve their matches by dissolving their current relationship and forming a new one with each other, and the resulting match is stable. In other words, these conditions require that each IBO match with the best new joiner among those willing to match with the IBO, and that each new joiner match with the best IBO among those willing to create the relationship.

1.5.3 Empirical model

As one of the objectives is to understand the determinants of the network formation between IBOs and new joiners, the relationship value is modeled based on previously mentioned characteristics. The pairwise relationship value of each pair $ij \in M_t$ is modeled as:

$$V_{ijt} = X'_{ij}\alpha_t + \epsilon_{ij}, \tag{1.6}$$

where $X_{ijt} \in R^k$ is a vector of characteristics applied for IBO i and new joiner j to form a paired utility, including $[Duration_i, DistributorFocus_i, Status_i, Level_i, EarningPotential_j, DemographicHomophily_{ij}, Distance_{ij}, IncomeSimilarity_{ij}]$, reflecting both social and economic variables. The list of variables included in the relationship value function is specified in Table 1.8. The parameters that govern relationship values could vary by the composition of IBOs and new joiners at a specific month. Thus, a vector of preference parameters α_t is defined to be month t specific. As the parameters are defined at the distributor-new joiner pair level, α_t cannot be individual specific. Also, α_t cannot be defined at the pair level because there is only one observation from each pair at a given month.

In order to estimate a vector of preference parameters, $\alpha_t \in R^k$, ϵ_{ij} is assumed to follow iid normal error with no loss of generality and to ease the computation complexity. The ϵ implies the pairwise unobserved deviations from the mean utilities across the agents and months.

1.5.3.1 Identification

The matching model can be seen as a variant of the discrete choice model; thus, in order to identify the model, the standard assumptions need to be applied in the empirical setting (McCulloch et al. 2000, Imai and van Dyk 2005, McCulloch and Rossi 1994). As is the case in the discrete choice model, only differences of characteristics across IBOs or new joiners have identifiable effects, as one's relative utility (or preference) will not be affected by both scale and level. Thus, in order to fix the scale, I standardized the scale of normal error, ϵ, to be 1.

In terms of identifying the IBOs and new joiners who did get their best match and who did not, the variation in the pair-dependent variables allows us to identify those two groups of people. If the new joiners care only about the IBO's economic factors, then everybody will have the same preference across the IBOs, and the preference over the potential IBOs would be the same. The same holds for the potential IBOs, the assortative matching case (Ni and Srinivasan 2015). What makes the preference different across each individual are the pair dependent variables, such as geographic distance and demographic homophily, and that identifies the IBOs who did get their best match and those who did not. Also, the slots in the IBO's side help us incorporate the situation where not everybody can possibly be matched with their most preferred counterpart.

1.5.3.2 Bayesian inference using Gibbs sampling

I employ Bayesian inference using a Gibbs sampling algorithm to estimate the two-sided matching model. Bayesian estimation is highly beneficial when facing integration of high dimensional latent variables (Sorensen 2007, Logan et al. 2008, Ni and Srinivasan 2015). This is especially useful in a two-sided matching set-up, which involves high dimensional integration (Sorensen 2007, Park 2008).

In the estimation process, I use data augmentation to recover the latent relationship value V_t, which can nicely handle the difficulty of integrating high dimensional latent relationship value variables (Geweke 1999, Sorensen 2007). Data augmentation can ease the complexity of drawing V_t, which satisfies the stability condition of the college admission model. I impose a conjugate normal prior for tractability in sampling the latent relationship value. Thus α_t follows normal distribution, $\alpha_t \sim \mathcal{N}(\mu_\alpha, \Sigma_\alpha)$, $\mu_\alpha \sim \mathcal{N}(\overline{\mu}_\alpha, A^{-1})$, and $\Sigma_\alpha \sim \mathcal{IW}(\nu, W)$. Especially with the normal error assumption in the utility and the normal conjugate prior on α_t, the marginal conditional augmented posterior distributions for V in the Gibbs sampling are truncated normal.

Also, in drawing latent pairwise utility V, it is necessary to simulate the simulating opportunity set (Uetake and Watanabe 2012b). However, not every IBO or new joiner has the same opportunities; the set of available opportunity varies for each person and depends on the relationship value of the other person in the two-sided matching market.[13] Thus, unlike a general one-sided discrete choice model where the choice set is fixed and agents are assumed to maximize the utility over a given choice set, a direct inference based on the observed choice is not possible, as I do not observe the relevant choice set for each individual,

[13] For instance, based on equation (1.4), new joiner j's opportunity set is all the IBOs who could gain higher utility if they were matched with new joiner j as opposed to their current match, $\mu(i)$. If new joiner j proposes to form a relationship to one of the IBOs in O(j), the IBO will agree to do so as s/he can gain higher utility compared to that of their current match. Thus, IBO i's choice set is the opportunity set and is dependent on the relationship value of the other side.

nor do I have supplemental data. Hsieh and Lee (2012) term this endogeneity in choice set. The ML-based estimation approach proposed by Uetake and Watanabe (2012b) involves approximating the opportunity set for each agent on one side based on all the preferences of the other side through simulation, which determines the willing partners. However, Bayesian estimation with Gibbs sampling can nicely handle such a complicated region of integration and can embed such interdependency of relationship values when forming opportunity sets, as I simulate the opportunity set conditioning on the previously updated α_t and the values from other pairwise V elements for each MCMC interation.

1.6 Results

1.6.1 Model Estimates

I now discuss parameter estimates in Table 1.9. The model estimates in the utilty specification indicate how the relationship is formed between new joiners and potential upline IBOs. Business relationship formation can be explained through the preference parameters in the pair-utility specification. After obtaining the estimates of the preference parameters, I can quantify the marginal effect of each variable holding other variables constant.

1.6.1.1 Social Factors

I first focus on the role of social factors, represented by demographic homophily and geographic proximity. Except the age difference and geographic distance, all variables are coded as binary dummy variables, either sharing the same characteristics (= 1) or not (= 0). Our sample represents three types of language speakers, seven types of ethnicities, and nine income groups, previously shown in Table 1.3. These variables capture pre-existing social networks and homophily aspects. Based on the results, the coefficients for social factors, represented by language, age difference, ethnicities, income groups and geographic distance, are all significant.

In more detail, the coefficients for the same language and the same ethnicities are strongly positively significant, and the same income group is marginally positively significant. This indicates that IBO - new joiner pairs with similar backgrounds generated higher pairwise utility, or relationship value, leading to a business relationship. The coefficients for age difference and geographic distance are strongly negatively significant. IBO-new joiner pairs exhibit a higher utility with a smaller age gap, implying that the distributor network is often created among existing friends or individuals with similar backgrounds, who could enter into a friend relationship, or who are already friends. When it comes to the geographic distance at the Zip-code level, IBO-new joiner pairs who live close by tend to generate higher utility.

Although the estimated coefficients show how significant each variable is in explaining the pairwise utility of the relationship value, it is unclear how to interpret the estimated coefficient, which is an issue previously raised in the matching model. Therefore, following Sorensen (2007) and Park (2008), I computed the marginal probability of the coefficient to observe the marginal effect of each variable, holding other variables constant.[14] This shed lights on the impact of each variable on a choice probability based on a simple choice scenario. To illustrate, let's assume that there are two possible choice alternatives (i.e., two new joiners). If the two new joiners have identical observed characteristics, the IBO's choice will depend entirely on the unobserved factors, and the probability that the IBO prefers one new joiner to the other will be a random 50% (the same is true for a new joiner establishing a relationship with one of the IBOs). Now, from the perspective of a new joiner, comparing two otherwise identical IBOs, the probability that the new joiner prefers the IBO having the same language as a business partner is 79%, 29% higher from the random 50%. Similarly, the magnitude of the approximate marginal probability increase for those of the same ethnicity, while other characteristics remain unchanged, is 28.8%. The marginal difference of a variable in the choice scenario belonging to the social factors is reported in the right-most column in Table 1.9. Based on this marginal probability, distance showed the highest absolute marginal effects, followed by the same language.

The empirical results on social factors support the idea that unlike people's general belief about how the business relationship should be formed, those with similar backgrounds and living close by tend to associate with each other as business partners. This result is consistent with MLM industry practices whereby distributors tend to initiate business opportunities from a list of acquaintances (Biggart 1989, King and Robinson 2000, Lilyquist 2016a), which is a unique feature in this MLM industry. The result appears to be consistent with findings from Frenzen and Davis (1990), who find that having social ties can increase the likelihood of selling/purchasing products in this type of industry. However, unlike Frenzen and Davis (1990), I offer empirical evidence that these social aspects play a role in business relationship formation as well, which has not been examined before, although I do not observe actual social ties. This reasonably follows from previous findings that the new joiners are often exposed to business opportunities through purchasing products (Legara et al. 2009). My empirical results indicate that both new joiners and distributors show a tendency to start a business with someone who is probably familiar to them, possibly sharing a similar lifestyle,

[14]Given new joiner j, the probability that matching with i (ij) is preferred to matching with i' (i'j) is $Pr(X'_{ij}\alpha + \epsilon_{ij} > X'_{i'j}\alpha + \epsilon_{i'j})$. From this, I can derive the marginal effect, dP/dX, using the chain rule. For dummy variables, marked with $\star$ in the results in Table 1.9, I approximate the marginal effect by calculating the difference, rather than using a derivative.

having a similar socio-economic background, or having potential commonalities.

1.6.1.2 Economic Factors

New joiner's economic returns (IBO characteristics)

I now turn to the economic aspects of the IBOs' characteristics. I examine the following four variables: distributor's duration in the business, distributor's focus, status, and network structural hierarchy (level). For the distributor's duration in the business, I had competing hypotheses: Do potential upline distributors who have been in the business longer have a higher valuation or not? I find that the potential uplines who have been in the business less time are relatively highly regarded in terms of business relationships on average, partly because the younger ones tend to be more enthusiastic and strongly motivated to expand the business (Legara et al. 2008).

For the distributor's focus, which represents the IBO's dependence on building networks or selling products, the IBOs who already recruited a large number of direct downlines appear to be less desirable, because the relationship value decreases with the number of direct downlines (-2.96), after I control for other related variables. Also, whether the IBOs built a multi-level business network below (reflected by the indirect network) is insignificant. In other words, new joiners faced a higher relationship value when the IBOs had a small number of direct downlines, other things held constant. This finding is somewhat consistent with guidance to MLM distributors. Given the effort and time required for salesforce management, some MLM experts advise IBOs to recruit and develop a small number of excellent direct downlines, by sponsoring their business rather than trying to sign up everybody — 'Sponsor, Don't Recruit' (Failla and Hardwick 1995, Lilyquist 2016a). Whereas the previous advice was from the perspective of increasing management efficiency, my findings support this claim from a slightly different perspective: Having more direct downlines can have a negative impact on relationship values and thus on relationship formation. This also reflects the situation where the new joiners have to face competition with other direct downlines who share the same sponsoring upline, and may receive less attention from the upline. If a new joiner has not gained attention and support from an upline, s/he will not have the confidence nor the training to build her own downline network. Lastly, when IBOs have a relatively large number of retail customers, it helps marginally to increase the relationship value.

Relationships with potential upline distributors who have attained higher status in the business (currently or in the past) generate higher relationship value. Depending on the interpretation of status, this may be because IBOs with high status have acquired better skills, or because status signals higher inherent business ability. In general, however, attaining high status can directly signal business performance-related ability, increasing the relationship value of high status distributors as their sponsoring upline as compared to low status distributors. Using the estimates, I find that the probability that a new joiner becomes a business partner with a high status IBO (all else being equal) is twice as high as when the new joiner partners with a low status IBO (Table 1.9).

The results on the network structural hierarchy (level) suggest that, after controlling for other possible confounding variables, such as the duration in the business, distributor's focus, status and other social factors, the level does not have a significant impact on network formation. Even if the distributor's compensation scheme does not depend on the network location, previous literature shows mixed results on the impact of level on the distributor's own performance. At least for distributor network formation, the results seem to indicate that there is no bias toward a specific network location.[15]

IBO's economic returns (new joiner characteristics)

From the upline IBOs' perspective, IBOs will seek new joiners who have future earning potential. The annual income before joining the business can serve as a proxy for the new joiner's financial motivation to join the business or how easy it is for an IBO to persuade the person to join the business. Our results show that new joiners whose annual income is more than \$50,000 had the lowest relationship value from the potential upline IBOs' perspective, as compared to the low-income (less than \$15,000) and mid-income (between \$15,000 and \$50,000) groups. On average, the high-income group is least preferred, followed by the mid-income group, although the coefficient is not significantly different from that of the lowest income group. Based on these results, IBOs prefer the new joiners who are more in need of additional income sources and may stay longer in the business; these new joiners generate relatively higher relationship value. This is consistent with previous survey findings in Goodman and Jolson (1973) on consumers' resistance toward a direct-selling effort, which show that resistance increases with household income.[16]

1.6.2 Social versus Economic Factors in Network Formation

In order to examine which factors play a more significant role in individual business network formation, especially in an industry where the social factors are vaguely understood as a main

[15] I acknowledge that a better test of this would be to vary the topological features of the network exogenously via a controlled experiment and then measure outcomes (Centola 2010). However, it is hard to imagine an MLM field setting where this would be feasible.

[16] Of course, I need to be cautious in interpreting the income effects as I cannot extend these results to infinity, especially for the lowest income group. Some might argue that, in general, income is correlated with one's ability to perform. However, I am not trying to overgeneralize the income effect to negative infinity. What I am claiming is that conditional on some boundary conditions that make a new joiner attractive, financial income could go the other way around.

factor driving the relationship, I compute the aggregate value of each factor. Reordering the equation 1.6 based on the social and economic factors gives us $V_{ijt} = Xsocial'_{ij}\alpha_{1t} + Xecon'_{ij}\alpha_{2t} + \epsilon_{ij}$. Using the draws of α_{1t} and α_{2t}, along with the associated X at each market t, social and economic factors from the realized business relationships are computed. From the equation, $Xsocial'_{ij}\alpha_{1t}$ is termed the social factor and $Xecon'_{ij}\alpha_{2t}$ is defined as the economic factor. Next, I average the two factors across the markets to examine the aggregate relative importance between the two.

The estimates suggest that the average social factor is 0.50, whereas the average economic factor is 1.56 — social factors have approximately one-third the importance of economic factors.[17] So although social factors are important to a certain degree, economic factors convey a significant portion of information in determining the relationship value, which explains the business-relationship formation.

1.6.3 Social versus Economic Factors and Future Business Behavior

MLM firms offer distributors financial incentives to provide guidance to their new joiners, and Sparks and Schenk (2001) characterize these relationships as the key social relationships in MLMs. Given the business relationship, some distributor-new joiner pairs may be formed by relying more on the social factors, whereas other pairs may benefit more from the economic factors. Now that I have recovered the value of social and economic factors, it would be useful to examine further what type of relationship at the relationship-formation stage leads to higher business activity in the future. This is an important question because the firm wants distributors to expand the network and sell the products.

Figure 1.5 shows the histogram of the new joiners' individual-level social and economic factors, along with the relationship value. As a reminder, heterogeneity in the model is incorporated by each month; thus α_t parameter is month specific. Yet in computing the factors, individual differences within each month are reflected through the IBO-new joiner characteristics. Based on the histogram, the economic factor is more dispersed than the social factor. To examine relationship types, the scatter plot in Figure 1.6 shows all 941 new joiners' social and economic factors recovered during the relationship value estimation. For expository purposes, Table 1.10 shows the categorization of each new joiner based on the sign of social and economic factors. Of the 941 new joiners, 32% show that the relationship value with the IBO stems from the social factor (positive social, negative economic factor),

[17] I have also computed the aggregate social and economic factors using both realized and potential matches. The estimates suggest that the average social factor is 0.82, whereas the average economic factor is 1.78 - still maintaining the consistency.

whereas 35% show the relationship value stems from the economic factor (positive economic, negative social). The correlation between the two factors is -0.26. These percentages, along with smaller percentages in both positive or both negative conditions, show that IBOs and new joiners are creating business relationships that rarely satisfy both factors (in a relative sense), and the business relationship is often formed emphasizing one factor or the other.

With the recovered social and economic factors, which explain the relationship value with IBOs, I examine whether the two factors influence the new joiner's future business activity. I specifically focus on the new joiner's number of direct downlines, registered customers, and total downlines 6 months after joining the business, as these activities reflect the expansion of the MLM network. I tracked these measures for the 582 new joiners who entered the MLM business during the first 15 months within the sample. Table 1.11 shows the expansion of the business after 6 months. Almost 50% of the new joiners did not engage in business expansion activities, perhaps utilizing the MLM business only for their own consumption or staying inactive.

The question is, what type of relationship at the relationship-formation stage leads to higher business activity in the future? I regress the recovered social and economic factors on the number of direct downlines, total downlines and registered customers. The results are presented in Table 1.12. I find that the economic factor that explained part of the relationship value in the formation stage influenced the new joiner's engagement in MLM business activities, in terms of engaging in recruiting direct downlines and registering customers. Interestingly, social factors were significant only in forming the relationship; they did not explain whether the new joiners actively engaged in MLM business activities.[18] This finding implies that emphasizing social aspects to create business relationships can hinder the potential overall business growth.

1.7 Conclusion and Limitations

While most research attention has been placed on networks where the main driver of the the network formation is the social purpose, I have a unique set-up where economic incentives potentially influence the in network formation as well, with possible trade-offs. In fact, the MLM industry provides us a perfect set-up to examine the interplay between the social and economic aspects in the distributor-network relationship. With explicit measures that

[18] I also find similar results on the financial outcomes, which are commissions received and retail margin. However, due to missing data, I do not formally report the results. For the non-missing commissions data, which includes 76% of the 582 new joiners, economic factors were significant(Estimate = 0.427 (0.123)***). For the non-missing retail margin data, which includes 31% of the 582 new joiners, only economic factors were significant (Estimate = 0.916 (0.354)**).

could reflect the economic aspects stemming from the two key participants, new joiners and IBOs, I am able to study the role of both social and economic factors in network formation. Furthermore, I provide empirical evidence on whether people involved in business are behaving irrationally by creating business relationships based solely on social factors for the business purposes), or rationally by considering future economic aspects.

The research highlights one of the important behavioral aspects, distributor network formation in the MLM industry, which has been understudied in the marketing domain. In the MLM industry, expanding the distributors (IBOs) is an important issue, because IBOs are the ones who advertise, distribute, and sell the product directly to the end user without going through all the intermediaries. Unlike conventional firms that rely on a sales force, the new joiners are recruited by the IBOs, not hired. So from a firm's perspective, expanding the network through the IBOs' efforts is important, and both new joiners and IBOs have good reasons to 'select' the right business partner.

Taken together, the results of the main analysis show strong evidence of social factors playing a role, namely, demographic and geographic homophily. In addition, economic returns on each side (new joiners and IBOs) affect the network formation. I examine under what circumstances the relationship value (pairwise utility) becomes the highest, which reflects the nature of the distributor network formation. Also, by computing the marginal effect of social factors, and economic factors on each side, I quantify the attractiveness lift in terms of forming the relationship. This research is especially relevant in the currently changing recruiting/distributor signing environment, where there is more purposive searching and exploratory courtship between the new joiners and the IBOs (King and Robinson 2000).

To highlight the findings so far, I find that on the new joiner's side, IBOs having a high status and being relatively new to the business, increase the relationship value between IBOs and new joiners. Also, having a small set of direct downline and constantly engaging in retail product selling increase the value of the relationship. All these IBO characteristics are driven by the new joiners' economic concerns. From the IBOs' perspective, new joiners with high income generated the lowest relationship value. Unlike previous mixed findings, my results show that the network structure (e.g., where the distributor is located in the hierarchy) does not add extra information to network formation after controlling for some of the possible confounding variables, including experience, status, and the distributor's focus.

Consistent with the descriptive analysis on the demographic homophily, I find that sharing the same language, ethnicity and similar age increase the relationship value significantly, while income similarity is marginally significant. Geographic proximity increases the likelihood of being matched by generating higher relationship value.

Using the recovered social and economic factors, the follow-up analysis shows that in explaining the network in MLM, although IBOs and new joiners place a high relationship value on people like themselves (i.e., on social factors), economic factors played a larger role in forming relationships - approximately three times larger. Finally, understanding the type of relationship value between IBOs and new joiners that can lead to active business participation is crucial in expanding the distributorship. In the follow-up analysis, I provide evidence that new joiners with a high economic factor, which is a part of relationship value, tend to show more active business activities.

Managerially, this research can help the MLM industry and IBOs in three ways. First, at the industry level, this research can change people's perceptions about how the business works. The results suggest that people involved in this industry are motivated more by rational (business) reasons than by social factors. Second, this research can help both current IBOs and new joiners to understand how to better design and manage their business by increasing relationship value. Specifically, it can help distributors to understand how strongly they are positioned in terms of attracting new people, and what they can do about it if they are not strongly positioned. Also, the new joiners can benefit by understanding what type of IBOs may be suitable for their needs and for their future success. Both sides can now understand how to increase the relationship value. Third, it can help the firm to recommend the right business partners, especially for those who do not have a specific partner in mind yet. The aim is not only to create the relationships with a high relationship value, but also to generate new joiners' active participation, which can expand the business.

Methodologically, in common with previous work on empirical matching models, my approach has the following limitations. First, I assume away search costs as in Sorensen (2007). The results could be impacted if search costs exist and interact with the social and economic factors that I consider. Second, I assume that each distributor is knowledgeable about all agents (and their associated information) on the other side who are willing to form a match as in Boyd et al. (2013). Therefore, during the estimation stage, the stability condition is imposed on the all the agents on the other side, but only those who are interested in forming the match are used in drawing V. Third, as is typical with such models, my estimation uses observed outcomes. The data make it hard to identify *ex ante* the potential set of IBOs the new joiner has approached and vice versa, as well the actual social network of all agents. I acknowledge that the use of demographic variables and geographic distance is an imperfect proxy for the network. Finally, utility specification is at the level of joint utility. Recent studies note that in the empirical matching model, it is still hard to recover the true preference of each side separately (Hsieh and Lee 2012), unless there is a restrictive matching setting in the market. Thus I am unable to distinguish which side - the IBO or the new joiner - is receiving higher utility from the match. Additionally, the variables that are pair dependent (e.g., demographic homophily or geographic proximity) could have an asymmetric impact on each side, which I do not distinguish in this paper. I hope that future work can relax these assumptions.

1.8 Tables

Table 1.1: Number of IBOs or New Joiners in Each Month

	Allowing one-to-many match		
Month	# IBOs	# New Joiners	# Transitioned IBOs
1	23	23	1
2	19	19	0
3	12	12	0
4	18	19	2
5	16	16	4
6	21	21	2
7	53	54	2
8	45	47	4
9	61	65	3
10	46	51	5
11	33	37	1
12	46	51	1
13	55	58	2
14	49	51	2
15	54	58	4
16	47	52	4
17	60	62	4
18	63	65	9
19	67	77	11
20	43	45	3
21	53	58	6
Total	884	941	70

Table 1.2: Types of Match Observed in the Data

Month	1 to 1 match	1 to 2 match	1 to 3 match	1 to 4 match
1	23	0	0	0
2	19	0	0	0
3	12	0	0	0
4	17	1	0	0
5	16	0	0	0
6	21	0	0	0
7	52	1	0	0
8	43	2	0	0
9	58	2	1	0
10	43	2	0	1
11	29	4	0	0
12	41	5	0	0
13	52	3	0	0
14	47	2	0	0
15	50	4	0	0
16	43	3	1	0
17	58	2	0	0
18	61	2	0	0
19	58	8	1	0
20	41	2	0	0
21	49	3	1	0
Total	833	46	4	1
Percentage	94.23%	5.20%	0.45%	0.11%

Table 1.3: Language, Ethnicity and Income Group

Language	1) English	68.32%
	2) Spanish	31.07%
	3) Korean	0.60%
Ethnicity	1) European	22.62%
	2) Hispanic	44.36%
	3) Asian	7.79%
	4) African American	14.23%
	5) Non-Orient Asian/Polynesian	7.99%
	6) Jewish	0.87%
	7) Arab	2.15%
Income	1) Income less than $15,000	23.96%
	2) Income more than $15,000 and less than $50,000	35.68%
	3) Income more than $50,000	40.36%

Table 1.4: Contingency Table on Language (in Percentage)

	Upline Language				
Downline Language	English	Spanish	Mandarin	Korean	Total
English	71.42*	2.01	0.24	1.25	74.93
Spanish	2.08	22.09*	0.01	0.02	24.19
Mandarin	0.04*	0	0.02	0	0.07
Korean	0.29	0	0	0.52*	0.81
Total	73.83	24.11	0.27	1.79	100

Table 1.5: Contingency Table on Ethnicity (in Percentage)

	Upline Ethnicity								
Downline Ethnicity	Missing	European	Hispanic	Asian	African American	Non-Orient Asian /Polynesian	Jewish	Arab	Total
Missing	1.79	4.23	1.92	0.63	0.4	0.61	0.11	0.14	9.82
European	4.09	31.92*	2.86	1.02	1.55	0.68	0.61	0.24	42.96
Hispanic	1.87	3.47	23.11*	0.37	0.25	0.68	0.1	0.08	29.94
Asian	0.51	1	0.21	5.98*	0.07	0.22	0.02	0.07	8.08
African American	0.47	1.93*	0.24	0.11	1.32	0.14	0.05	0.07	4.33
Non-Orient Asian/Polynesian	0.41	0.41	0.51	0.17	0.05	1.46*	0.01	0.06	3.08
Jewish	0.1	0.57*	0.07	0.02	0.03	0.02	0.07	0.01	0.88
Arab	0.13	0.22*	0.06	0.07	0.06	0.15	0.01	0.2	0.9
Total	9.36	43.76	28.97	8.36	3.73	3.97	0.98	0.87	100

Table 1.6: Summary Statistics on Business Support and Network Structure

	High status	Duration in the business (Month)	# Direct Downlines	# Total Downlines	# Customers	Level
Min.	0.00	0.00	0.00	0.00	0.00	15.00
1st.Q	0.00	1.00	1.00	1.00	0.00	28.00
Median	0.00	5.00	1.00	2.00	0.00	33.00
Mean	0.02	24.75	2.79	20.18	2.77	33.07
3rd.Q	0.00	26.00	3.00	8.00	2.00	38.00
Max.	1.00	344.00	25.00	1375.00	143.00	97.00

Table 1.7: Correlation Table on Business Support and Network Structure

	High status	Duration in the business	# Direct Downlines	# Total Downlines	# Customers	Level
High status	1.00	0.28	0.41	0.54	0.32	-0.05
Duration	0.28	1.00	0.64	0.39	0.32	-0.22
# Direct Downlines	0.41	0.64	1.00	0.44	0.46	-0.17
# Total Downlines	0.54	0.39	0.44	1.00	0.14	-0.12
# Customers	0.32	0.32	0.46	0.14	1.00	-0.09
Level	-0.05	-0.22	-0.17	-0.12	-0.09	1.00

Table 1.8: A List of Variables in the Matching Model

1. Social Factors

Demographic homophily

a. Age difference

b. Language
(Same = 1, else = 0)

c. Ethnicity
(Same = 1, else = 0)

d. Income group
(Same = 1, else = 0)

Geographic homophily

a. Zip code distance

---

2. Economic Factors

Distributor characteristics (New joiners' Economic Return)

a. Duration in the business
(# months)

b. Distributor's Focus
Network (# Direct downline, Existence of indirect network)
Product Selling (# Customers)

c. Status from Business Performance
(High status =1, Low status = 0)

d. Level based on the Network Structure

New joiner characteristics (Distributor's Economic Return)

a. Annual Income before joining the business
Low: Income <$15,000
Medium: $15,000 ≤ Income <$50,000
High: $50,000 ≤ Income

Table 1.9: Estimates of the Matching Model

	Variable	Mean		Marginal probability
Social Factors	Language⋆	1.16	**	0.29
	Absolute age difference (1/10)	-0.33	**	-0.09
	Ethnicity⋆	1.13	**	0.28
	Distance (miles/10)	-1.65	**	-0.46
	Income group similarity⋆	0.31	*	0.09
Economic Factors (Distributor characteristics)	Duration(months/10)	-0.68	**	-0.19
	Direct downline (1/10)	-2.96	**	-0.84
	Indirect network⋆	-0.14		-0.04
	# Customers (1/10)	0.99	*	0.26
	Status⋆	3.91	**	0.50
	Relative level (*10)	-0.09		-0.03
Economic Factors (New joiner characteristics)	Medium income⋆	-0.21		-0.06
	High income⋆	-0.34	*	-0.09

Notes. 1. * significant at 0.10,** significant at 0.05. 2. I report the Bayesian equivalence of a p-value to indicate statistical significance (Rossi et al. 1996). Using the MCMC estimation procedure, I empirically evaluate the posterior distributions of the parameters or their functions (e.g., choice probabilities) to conduct a hypothesis test. 4. Generated 700,000 draws and kept every 10th draw of the last 300,000 draws to compute the posterior means of the parameters. 5.⋆ indicates the binary variable. Thus, in computing the marginal probability, the variables with ⋆ is evaluated for the discrete change.

Table 1.10: Relationship Types Based on Estimated Social and Economic Factors (%)

		Economic Factor		
		Positive	Negative	Total
Social Factor	Positive	19.66	31.99	51.65
	Negative	34.64	13.71	48.35
	Total	54.30	45.7	100

Table 1.11: New Joiner's Business Activities After 6 Months of Entry

	# Direct Downline	# Total Downline	# Registered Customer
Min.	0	0	0
1st	0	0	0
Median	0	0	0
Mean	0.4536	1.656	0.4691
3rd	1	1	0
Max.	10	53	17

Table 1.12: Estimates of Social and Economic Factors on Business Outcomes

	# Direct Downline		# Total Downline		# Registered Customer	
	Estimate	SD	Estimate	SD	Estimate	SD
(Intercept)	0.452	(0.035)***	1.696	(0.232)***	0.467	(0.068)***
Social Factor	0.003	(0.009)	-0.040	(0.062)	0.004	(0.018)
Economic Factor	0.012	(0.003)***	0.029	(0.022)	0.012	(0.006)*
Adjusted R square	0.020		0.001		0.002	

1.9 Figures

Figure 1.1: Multi-level Marketing Network Structure

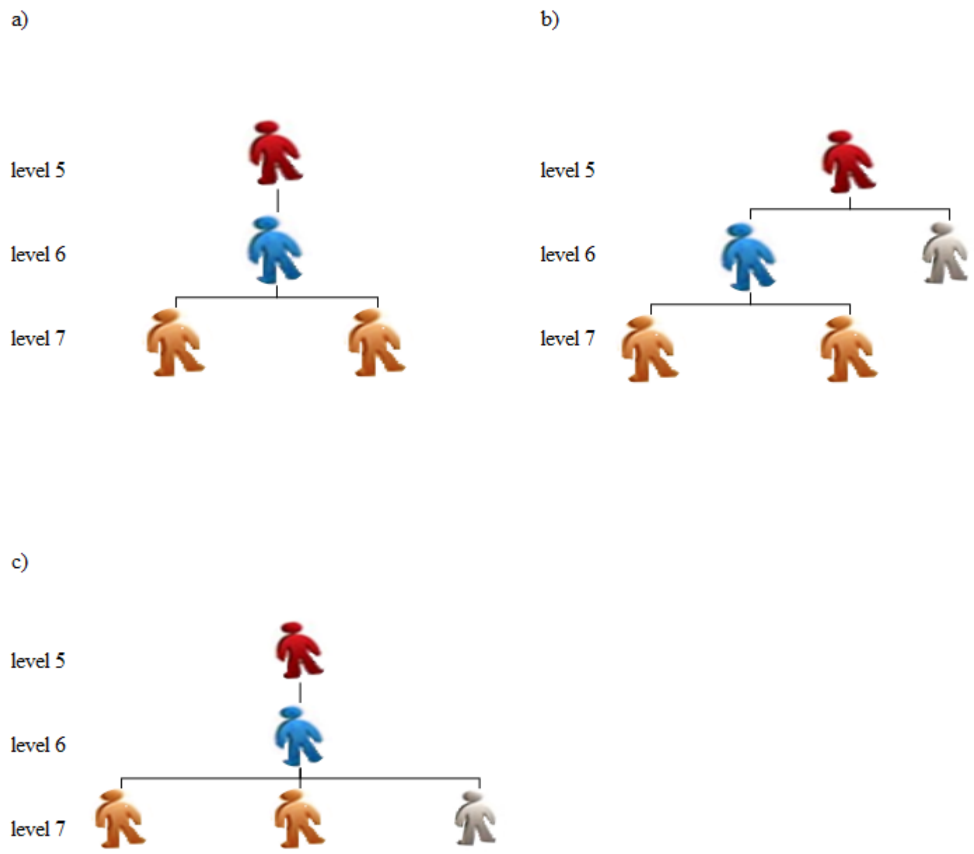


Figure 1.2: Approximate 35 miles Radius from Chicago

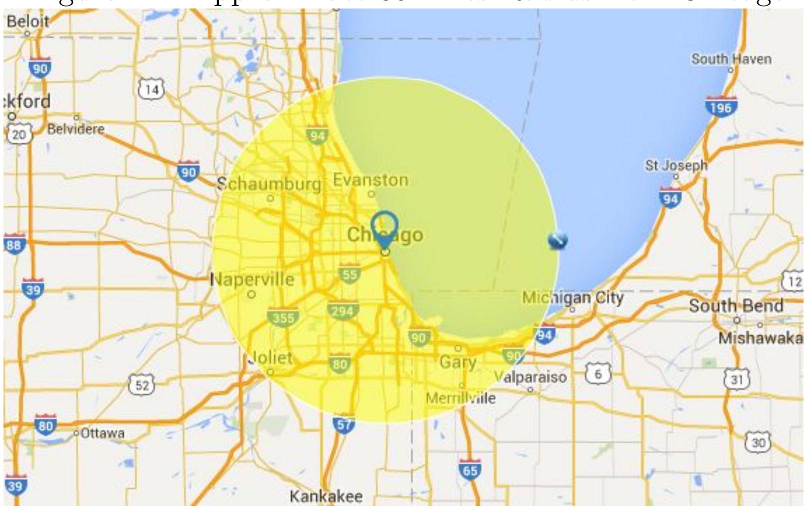


Figure 1.3: Age and Distance Distribution

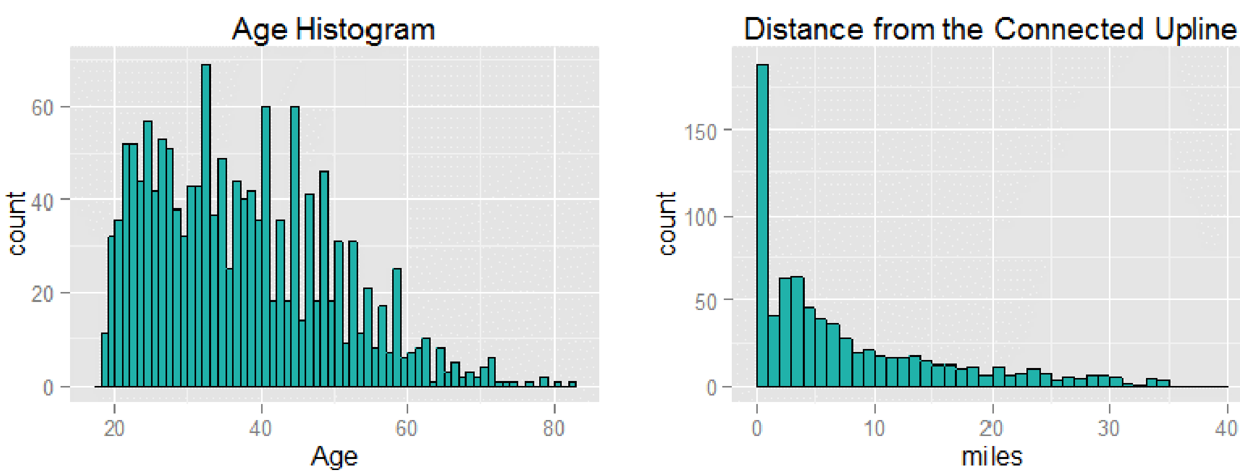


Figure 1.4: Number of Direct Downlines, Total Downlines and Customers

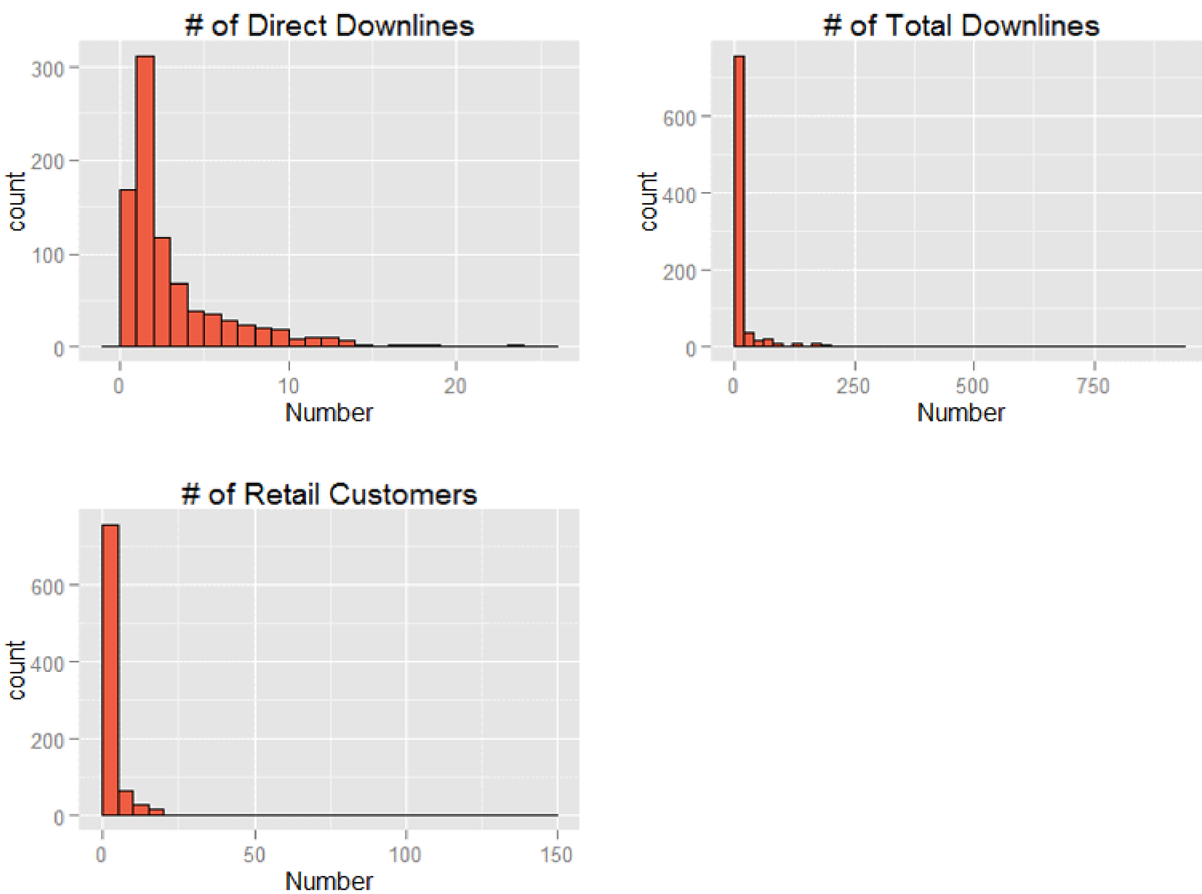


Figure 1.5: New Joiner's Estimated Social and Economic Factors in Relationship Formation

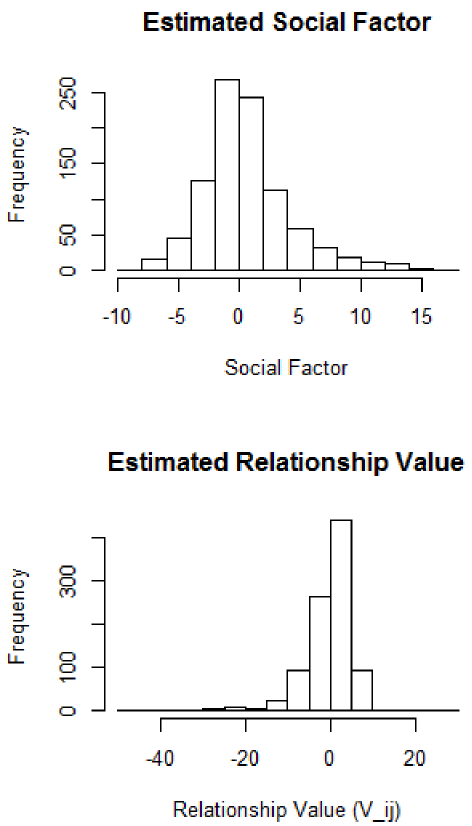
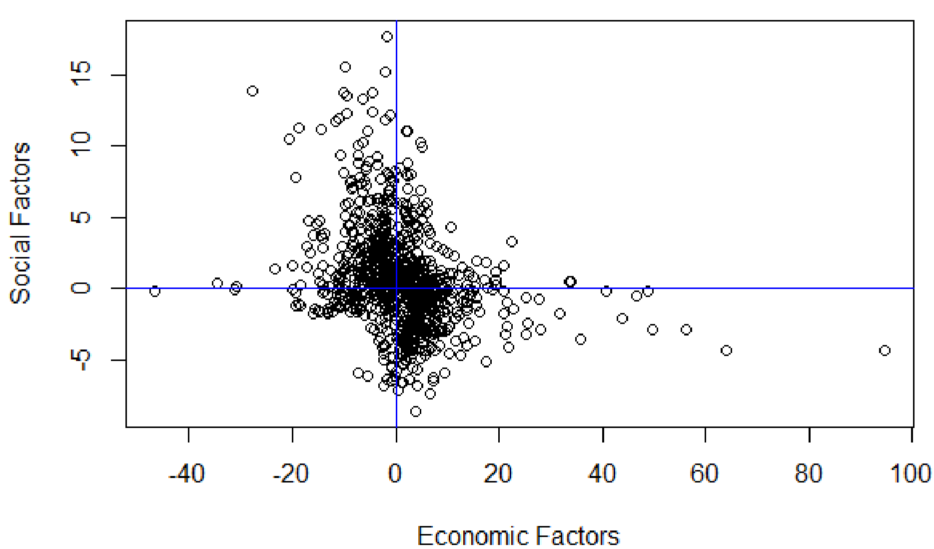


Figure 1.6: New Joiner's Relationship Types Based on Estimated Social and Economic Factors



Chapter 2 The Impact of Existence of Others on Inactive Behavior in the Multi-level Marketing Industry

2.1 Abstract

Turnover behavior has long been a topic of interest in organizational studies. Understanding the behavior in the context of multi-level marketing (MLM) poses a unique challenge, however, due to the unique characteristics of the industry, which differs significantly from traditional hierarchical organizations. MLM firms work closely with a network of distributors (Independent business owners, or IBOs), who either sell products to the end customers or recruit others into the network. Although IBOs in the MLM industry are 'independent' business owners, their business activity can be subject to a significant influence from others. This stems, in part, from the cooperative, social nature of business relationships, where there is a high degree of exposure to the IBOs. To understand this influence, this paper examines external social factors, namely, how the existence of different others affects an IBO's inactivity. In particular, I examine three different dimensions of others: Individuals, Network family, and Proximity. Using both county-level analysis and individual-level survival analysis, I find evidence that the existence of others in proximity increased an IBO's inactive behavior. In contrast, when an IBO is associated with a larger sized network family, IBOs remained active. Another factor influencing the inactivity is the existence of 'successful others' (with high status) in proximity and network family dimensions, which have a protective effect from inactive behavior. Machine learning analysis has also been applied to predict future inactivity behavior. The model is found to be useful in distinguishing those IBOs who will not experience the inactivity versus those who will, and also in detecting an inactive period occurring within a quarter. Consistent with results from statistical analysis, the relative importance of the variables (See Machine Learning section) obtained from machine learning reveals that the percentage of high status IBOs within a family still plays an important role, even in predicting a future inactivity period.

2.2 Introduction

"A few companies are beginning to realize that sales turnover is a sleeping giant, inconspicuously swallowing a significant portion of their productivity and profits" (Futrell and Parasuraman 1984). Despite the fact that the quote is from the early eighties, it is still true today. Salesperson turnover has been a topic of interest in the organizational literature since the late seventies for two main reasons: cost and the phenomenon itself (DeConinck and Johnson 2009, Holtom et al. 2008). Cost is of interest because losing each employee entails losses for the firm, which are estimated to vary from a few thousand dollars to more than twice the employee's salary, depending on the industry, the content of the job, the availability of replacements and other factors (Hinkin and Tracey 2000). This includes the cost involved with hiring and training a new employee (Griffeth and Hom 2001).

Apart from the issue of cost, interest in turnover in sales organizations has been ongoing, making it a notable because the estimated turnover rate among salespeople is double that for other jobs (Richardson 1999). Although some aspects of turnover may be positive, such as the elimination of ineffective performers (Jolson et al. 1987), in general, researchers and industry managers are eager to understand how to reduce it because of its impact on the size and continuity of revenue generation (Jolson et al. 1987, Wotruba et al. 2005, Darmon 2008, Holtom et al. 2008). Thus, substantial research attention has been placed on identifying correlates and predictors of employee turnover in traditional hierarchical organization settings (Jolson et al. 1987).

Within the sales industry, one specific type of sales organization suffers from a turnover rate that is even higher than that of the sales force of the U.S. manufacturers, the traditional organizations (Coleman 1989): Multi-level marketing (MLM or Network marketing) organizations (Raymond and Tanner Jr. 1994, Wotruba et al. 2005). MLM is a type of direct selling (the other type is Single-level marketing) (Biggart 1989, Brodie et al. 2002), which accounts for more than 95 percent in terms of the number of firms, number of sales people, and retail sales within the direct selling industry (DSA 2015b).[1] Some well-known companies that employ the MLM business model are Amway, Mary Kay and Herballife.

MLM organizations differ from traditional hierarchical organizations along several dimensions. In MLM, distributors (Independent Business Owners or IBOs) are not viewed as employees, but rather as an independent contractors or entrepreneurs, with physical and psychological independence from the firm (Jolson et al. 1987). As such, they are not subject to the same kind of monitoring and control found in traditional sales departments in hierarchical organizations, although they are required to follow the legal and ethical guidelines

[1] Please refer to institutional background in essay one for detailed industry facts.

informally set forth by the company organization (Pratt 2000). IBOs not only sell products but also recruit and supervise other IBOs who become part of the IBO network. This industry setting is well reflected in the compensation plan, where IBOs do not earn a fixed income as an employee in a hierarchical firm; rather, their earnings are based on commissions and a personally set retail margin. Commissions are based on the product sales carried out by the IBO and the connected network of IBOs, which is expanded though recruiting. This structure provides IBOs a flexible schedule with a better work-life balance, while also providing motivation to expand the network (Biggart 1989). In order for the MLM firm to motivate IBOs from the flexible working environment, monetary and non-monetary recognition are utilized, which include the systematic commission plan and status.

Given the different industry settings, the turnover costs of the MLM is unique, even differing from other sales organizations. In particular, the burden of the turnover costs, discussed in more detail in the literature review, falls on the shoulder of both the MLM firms and the IBOs. The risk to the firms is mainly due to the IBOs being the main retail outlet to the end customers, while IBOs being 'independent.' In the 2016 10-K, Herballife disclosed risks related to their business and their very first point was to acknowledge that their distributors may voluntarily terminate their distributor agreements at any time: "Our failure to establish and maintain Member and sales leader relationships for any reason could negatively impact sales of our products and harm our financial condition and operating results" (Herballife 2016). The risk from the IBOs' side is due to their role in recruiting and training efforts that themselves, rather than the firm, undertake and the unique mentor relationship between recruiter-recruited IBOs. Unlike in a traditional firm, where employees must work before they decide to quit (apart from a leave of absence), in an MLM, one can remain inactive without permission from the manager or the company, since there is no formal quitting procedure from the distributor and no official employment termination notice from the company (Wotruba 1990b, Wotruba et al. 2005).

This concept is consistent with being one's own boss as a self-employed independent contractor, with the individual having a choice in terms of how much to engage with the business. Thus, IBOs who are registered under the business, but at a certain point became less interested without engaging in any activities, can be an indicator of turnover. Research has shown that many MLM firms specify some length of inactivity as a potential indicator of turnover (Wotruba 1990b, Wotruba et al. 2005). Despite the uniqueness of the MLM and IBO structure, little academic literature has examined turnover behavior in the MLM industry. The few existing studies on direct selling turnover rely on subjective survey data to explore this phenomenon, which provides a room for quantitative data-driven research.

Utilizing a unique dataset from a leading MLM firm, I examine turnover, more precisely, inactive behavior in the MLM industry. Previous literature on MLM turnover suggests a strong relationship with job performance (Jolson et al. 1987, Wotruba 1990b), which can be attributed to the business characteristics of an individual. Although the business characteristics of an individual provide high explanatory power, other factors may play a role in turnover. On the one hand, an MLM is social, encouraging cooperative relationships with others (Sparks and Schenk 2006), which has even been described as cult-like behavior (Bhattacharya and Mehta 2000). On the other hand, selling the same product from the firm, IBOs may inevitably be influenced by other IBOs surrounding them. Jolson et al. (1987) have attempted to examine the degree of perceived competition using a one item survey but did not find evidence for relationships between competition and turnover. Aside from this study, to the best of my knowledge, no previous literature documents how individual turnover behavior can be affected by the presence of others in the MLM industry, either in a cooperative or a competitive manner.

I examine the existence of others along three different dimensions based on MLM industry characteristics: Individuals, Network family and Proximity. For the first dimension, I examine two types of key individuals unique to the MLM industry—upline and high status IBOs. An upline is an individual with whom a prospective IBO creates a business relationship. When a prospective IBO joins the MLM, s/he is connected with only one upline, which creates a mentor-type relationship. A high status IBO is a successful IBO approved by the MLM firm who takes on the role of leader of a sub-group of IBOs. Unlike in a traditional organization, having the high status does not imply having more authoritative power: there is no subordinate relationship between a high status IBO and IBOs without that status (Biggart 1989). The second dimension, the network family, is a motivational network organization usually set up by accomplished IBOs in the MLM business to help other IBOs become successful. Lastly, the third dimension, geographical proximity, indicates those who are located nearby. Across these three dimensions, I examine the influence of high status IBOs on an IBO's inactive behavior.

To answer the question, I rely on a county-level regression analysis and an individual-level survival analysis. The two analyses show similar findings for the effects of others explaining the focal IBO's inactive behavior. Consistent with industry concerns (e.g., Msweli and Sargeant (2001)), I find competition among close IBOs to be inevitable within a particular area, as reported in the franchise literature (Pancras et al. 2012). However, in both the descriptive county-level regression and individual-level survival analysis, I find the interesting role of status related to inactive behavior, which is a feature unique to the industry indicating of success. Exposure to high status IBOs decreased inactive behavior in general (based on county-level analysis). Specifically, the percentage of high status IBOs along the network family dimension and the proximity dimension decreased inactive behavior (based on individual-level analysis). In other words, the more high status IBOs within the same network family and within proximity, the more likely that IBOs would remain active in the business—a protective effect. However, along the individuals dimension, the distance from the structurally nearest high status IBO did not have a protective effect, whereas the distance from an upline sponsoring IBO was found to be positively related to active behavior. In addition, IBOs associated with a larger network family tended to stay active in the business, potentially by fulfilling the need for social connection (DSA 2015c), or by fulfilling a desire to have access to potential network resources (Jolson et al. 1987). Moreover, I find that when the market has a high penetration rate—relatively more people are involved in this business—IBOs are less likely to face the inactivity hazard, potentially due to awareness about the company or its products, which makes it easier for IBOs to do business activities (Jolson et al. 1987).

Turning my attention from understanding the inactivity phenomenon to predicting a period of IBOs' future inactivity, I conducted a set of machine learning analyses, focusing on neural networks. Based on one of the analyses, I detect and predict a duration for encountering an inactive period in the future by quarter. The model is designed to allow us to pinpoint the upcoming quarter when the IBO is going to be inactive. The model is sufficient for distinguishing the IBOs in a specific month that do not follow an inactivity period, and the IBOs that are currently under the inactivity period, attaining an overall accuracy of 81.44%. Based on one approach I take, the model is also sufficient for predicting the inactive period occurring within a quarter. However, there exists some room for improvement for a long term prediction—predicting IBOs in a specific month, experiencing the inactive period after a quarter. Based on the results, the relative importance across variables shows that one of the top five most important variables in the neural network is the percentage of high status members within a network family. Variables related to 'others' also were ranked relatively high, reinforcing my main claim that exposure to other IBOs has a role in explaining the focal IBO's inactive behavior. This results may be due to the nature of the industry, where social interaction and status is highlighted, and industry features that help motivate IBOs to be successful in the business.

This study contributes to the turnover literature, in particular, the MLM industry. Interestingly, some of the industry specific features that do not appear or do not exactly match those of traditional hierarchical firms, such as status or network family, were found to have a protective effect in preventing IBOs' inactivity. MLM firms can potentially utilize the findings in this research to reduce inactivity behavior. Methodologically, this study is the first empirical study based on hard data to examine turnover in MLM; previous studies relied on in-person surveys. This unique aspect allowed me to explore the environmental factors that have not been thoroughly studied. Lastly, I examine the applicability of machine learning techniques to the MLM turnover literature. I was able to draw similar conclusions using both traditional methods and machine learning, showing that how the focal IBO is situated among other IBOs, in particular, among successful others, plays a role in explaining and predicting inactivity.

The rest of the essay is organized as follows. Section 2.3 provides a brief review of MLM turnover cost, prior literature on turnover, and the three different dimensions, namely, individuals, network family, and proximity. Section 2.4 presents the data. Section 2.5 introduces the county-level and individual-level analyses. I then move on to Section 2.6, which focuses on inactivity prediction using machine learning. Section 2.7 concludes the second essay.

2.3 Literature Review

2.3.1 Turnover and Costs in MLM

Given the distinct institutional setting, the cost of turnover in MLM should be different from that in traditional organizations. What is unique to MLM is that IBO turnover imposes costs on both the MLM firm and on connected IBOs. From the firm's perspective, first, the firm loses the direct distribution channel to the end customer, which is the main role of IBOs. The other roles that IBOs play, such as recruiting other IBOs, training, motivating, advertising, managing customer loyalty, maintaining good image and reputation as a representative of a firm and its products, are also pivotal in an MLM business. This situation differs from that of traditional hierarchical firms, where these functions are carried out by the administrative staff in Sales, Human Resources, and Marketing departments (Msweli and Sargeant 2001). Thus, MLM firms see the managing distributor relationship as the most critical aspect of their business, as firms rely heavily on IBOs to deliver their product and service.

The second reason MLM firms are concerned about turnover is that they lose existing customers. Based on the industry fact sheet from the Direct Selling Association (2015), the person-to-person sales strategy accounted for more than seventy percent among all other sales strategies employed in the direct selling, the most prevalent sales strategy in the direct selling the industry, followed by the party plan (DSA 2015a). Therefore, because of the lack of other retail channels, the IBO is very important intermediary for generating customer loyalty and repeated sales.

The problem with this person-to-person sales strategy is that, unless the customers are registered online, they are personally taken care of by their own personal IBO, which reveals that firms are unable to track each end customer the IBO is serving. Raymond and Tanner Jr. (1994) highlight the need to follow up with the customers of IBOs who no longer work for the firm. Even if the customers appreciate the products the firm is offering, they may have been poorly serviced as a result of thier IBO's inactivity or leave. Then the firm not only loses the current sales by losing the contact, but also risks future sales, if no channel is provided for continuing the relationship. Another higher order issue that naturally follows from dissatisfied customers is the unfavorable perceptions arising from the frustration experienced when products are no longer easily available (Raymond and Tanner Jr. 1994).

From the IBO's perspective, when a recruited downline remains inactive or leaves, considerable investment must be made once more in the recruitment and training of new IBOs, as opposed to focusing on selling the product. Unlike the traditional hierarchical firms, this recruiting and training cost burden falls on the IBOs rather than on the firms, resulting in an inefficient distribution outlet. In contrast, when a recruiter IBO stays inactive or leaves, the recruited IBOs will be exposed to a bad example and fail to receive productive guidance, which in turn leads to a potentially less successful distribution outlet.

2.3.2 Prior Literature on Turnover

Due to the high cost of turnover in organizations, much previous research has developed frameworks to understand some of the reasons behind sales force turnover and ways to tackle this issue. Theories have been developed to shed light on the antecedents and consequences of turnover in the organization and the processes that result in turnover. More specifically, theories have revealed both immediate and distal causes and consequences. Within this research, a range of constructs have been examined, from individual attitudes or business characteristics to a higher level of interest such as variables at the organization level (Griffeth et al. 2000, Holtom et al. 2008). Mobley (1992) proposed four categories of factors that are correlated with turnover: organizational work environment, individual work related, individual non-work related, and external environmental variables (Price 1983). Major determinants include job satisfaction (Cotton and Tuttle 1986, Lee et al. 1999, Chen et al. 2011), performance (Futrell and Parasuraman 1984, Ladik et al. 2000), or organization commitment (Griffeth et al. 2000). Conceptually, Zoltners et al. (2013) identified three segments of sales people experiencing the turnover, namely, low performers with low potential, low performers with significant future potential, and turnover among high performers, and provided different ways to cater to each of the segments. In essence, though, most of the research examining the turnover issue focuses on traditional hierarchical firms (Holtom et al. 2008).

Among the lines of research examining traditional organizations, some evidence emphasizes the role of others in turnover behavior. For instance, in a meta-analysis of turnover, work environment factors appeared, such as leadership (leader member exchange, supervisory satisfaction) and co-worker related variables (work group cohesion, and co-worker satisfaction) (Griffeth et al. 2000). These factors indicate that one's decision to stay in the business is affected by other individuals with whom one highly interacts—supervisor or co-workers. In addition, Holtom et al. (2008) pointed out from their meta-analysis that one of the trends explaining the turnover behavior in organization studies is the emphasis on interpersonal relationships, such as leader-member exchange or interpersonal citizenship behaviors. Based on this general background, in the following section, I introduce additional literature on traditional organizations that can be mapped onto the suggested three dimensions of others in the MLM industry.

While focus has mostly been on turnover in traditional firms, including some evidence for the role of others, some prior research has touched upon direct (or MLM) salespeople turnover. Looking at various types of sales organizations[2], including salespeople serving in-home consumers, Jolson et al. (1987) carried out an exploratory investigation survey of eighty possible correlates of sales force tenure in general, looking at four different categories: individual non-work related, external economic factors, individual work related, and organizational variables. Of the eighty, only four were shown to be significant: age (+), earnings (+), popularity/acceptance of product/company (+) and sales force size (+). In addition to this study, which comprehensively examines related factors, other factors were also highlighted in other studies. In terms of *job performance*, the relationship between job performance and inactivity-proneness is found to be negative only for the low performers (the poorer, not the better, performers are more likely to become inactive) (Wotruba 1990b). The findings on the relationship between job performance and turnover in direct selling is consistent with general findings in the professional selling literature in traditional hierarchical organizations (Futrell and Parasuraman 1984, Ladik et al. 2000, Jackofsky et al. 1986). *Unmet expectations* has been examined both cross sectionally and longitudinally, along with *job satisfaction*, which is a well-established relationship in general sales force settings (Wotruba and Tyagi 1991, Koroth 2014). *Involvement in the business* has also been identified as an influencer of turnover in MLMs. In particular, the work compared part-timers with full-timers. The finding is that part-timers had greater job satisfaction, performed better measured by earnings per hour worked, and were less likely to quit (Wotruba 1990a). *How one perceives how others view their job or selling behavior* was also related with inactivity-proneness in direct selling (Wotruba 1990b).

[2]Examined four different sales force types depending on the main customer base: (1) resellers, including wholesalers and retailers, household consumers, (2) household consumers contacted at their residences—mainly direct sellers, (3) manufacturers, (4) other organizational end users (business, goverment, etc.).

These constructs are based mostly on an IBO's individual characteristics and were examined using a qualitative survey approach. This research differs from the previous work in that I examine the IBO's connection to three different dimensions of others in explaining their inactive behavior: Important individuals, Network family and Proximity. Although Jolson et al. (1987) had a one-item survey question related to 'perceived degree of competition', it proved to be insignificant and was dropped from further discussion.

In a slightly different context, turnover behavior can be linked to the customer churn management literature because the goal of both turnover and customer churn management is to make people keep connected with the company; whether that being employees, IBOs or customers. In particular, in the MLM turnover, IBOs can be considered as consumers as well as distributors, so previous findings on continuing relationship with existing consumers with exposure to other consumers can contribute to understanding the MLM turnover behavior. Haenlein (2013) examined social interactions in customer churn behavior in mobile phone provider and found that when a focal actor has an out-going call relationship with other individuals who have recently defected, the focal actor is significantly more likely to defect from the provider. While much attention has been placed on social interactions in the customer acquisition process (e.g., Iyengar et al. (2011)), based on the network data, research shows the role of social interaction in churning behavior. Taking a similar stance, in my second essay, I examine the role of the existence of three different dimensions of others on inactive behavior in the MLM industry, based on empirical data.

In Section 2.3.3, I will explain the three different dimensions of others that can possibly affect IBO's turnover behavior in MLM: Important individuals, Network family, and Proximity.

2.3.3 Three Different Dimensions of Others

2.3.3.1 Individual

At the individual dimension, I examine two important individuals in the MLM industry who potentially have an influence on the focal IBO's inactivity behavior: an upline and a high status IBO.

Geographic Distance from an Upline

In MLMs, IBOs are situated in a unique network structure, where IBOs are nodes and are connected as in an inverse tree structured network. In accordance with this network structure, commissions are calculated based on products sold by the IBOs located below. A sponsoring IBO is located one level higher compared to a joiner. Sponsors (or recruiters) are called uplines; joiners are referred to as direct downlines. Important here is that one new joiner can be connected with only one upline.

The relationship between the upline and the recruited IBO is quite special. It is similar to that of mentor-mentee, where each is an independent entrepreneur. Once the new joiner becomes a part of the network under a certain IBO – a sponsoring upline, it is the incumbent on the upline to help the new joiner initiate the business, maintain motivation, and provide a long- lasting support (Crittenden and Crittenden 2004). Because of this intricate relationship, new joiners are often advised to find the 'right' sponsoring upline to be successful in the business (Allen 2016). I thus hypothesize that IBOs located near their own connected upline IBOs show a tendency to stay active in the business.

Geographic Distance from a High Status IBO

IBOs usually receive a supplementary bonus for sales generated by recruiting on top of retail margin, which motivates them to build networks and sell products – a key characteristic of MLMs that distinguishes them from single-level direct selling (Wotruba et al. 2005). But beyond the supplementary bonus, firms utilize additional motivational stimuli in the form of high status as a designation of success. As the MLM firms do not have direct control over the IBOs, firms award multiple tiers of status based on IBOs' business activities that entail not only financial benefits but also high social recognition (King and Robinson 2000). More specifically, the status provides attention and respect from other IBOs and the MLM firm. This status generates social recognition, which explains one of the motivations for being part of the MLM business (Wotruba and Tyagi 1992, DSA 2015c). At the same time, these high status IBOs are bound by additional responsibilities along with selling the products and expanding the network. They play a pivotal role in teaching and motivating other IBOs at a group level through leadership responsibilities and setting a good example of success in the MLM by being an 'admirable entity' (Amway 2015, Biggart 1989). Based on the associated characteristics of high status, I hypothesize that IBOs who are geographically close to a high status IBO will tend to stay active in the business.

2.3.3.2 Network Family

Network family is commonly known as the Business Support System or motivational organization, and is generally an independent organization initiated and funded by successful IBOs to offer support, training, and professional development programs to other IBOs within the same family. Structurally, imagine multiple mutually exclusive network trees. Due to its independent nature, the MLM firm provides an accredited program. An IBO's family and an upline's family may not necessarily be the same; however, it is rare for someone of lower status to have a family different from that of his/her upline. In organizational studies, organization structure (e.g., size, flat/tall hierarchy) is believed to affect the behavior of members, such as turnover, absenteeism, even performance. However, the findings about organization size are mixed (Dalton et al. 1980). Among different network family characteristics that may affect one's business activity, I examine two characteristics, namely, size of the family and percentage of high status IBOs within the family.

Network Family size

Well-documented evidence exists for the friendly, personal, and cooperative nature of relationships among IBOs (Biggart 1989, Sparks and Schenk 2006, Grayson 1996), and the success of the MLM business model is largely dependent on IBOs' forming cooperative relationships (Sparks and Schenk 2006). IBOs may feel connected by becoming part of a big group, providing a relatively higher family-based social satisfaction, which is an additional utility IBOs receive by becoming part of an MLM business (Bhattacharya and Mehta 2000).

Previous literature on traditional organizational studies have supporting claims on the role of socialization on turnover. Research on socialization tactics (i.e., seeking performance feedback and information seeking, building relationships and networks) shows that the tactics help the organization to embed employees into the organization. Utilization of such tactics increases new joiners' perceived level of accommodation and helps them to adjust to the new work environment. Also, it increases the level of organizational commitment, which is strongly related to job satisfaction, job performance and turnover. In particular, tactics related to collectivism were positively related to job embeddedness, which mediates the relationship between the socialization tactics and turnover (Allen 2006, Menguc et al. 2007).

Job embeddedness is found to be negatively correlated with turnover (Mitchell et al. 2001). The concept indicates the degree of attachment and inertia with the organization, and is composed of three concepts: link (formal or informal connection within an organization), sacrifice (perceived loss of material or psychological benefits by quitting the job) and fit (compatibility with corporate and external environment) (Mitchell et al. 2001). Similar to the job embeddedness construct, McPherson et al. (1992) specifically examine the social network within the organization and find that the more social ties one has within the organization, the less likely one is to leave the organization.

Hom et al. (2009) corroborate the prevailing views of the importance of interrelationship in turnover that social exchange (e.g., shared investment, mutual trust, enduring relationship among employee-organization relationships) and job embeddedness in employee-organization relationships increase workforce commitment and loyalty, which reduces the intention to quit.

Although the institutional setting is different, based on the previous literature from organizational studies, the connection can be made: IBOs with a larger family will have more opportunities to be exposed to socialization tactics, to form social ties, and in general, be connected to other IBOs within the same family. Thus, IBOs with a larger family tend to stay active in the business.

In a previous survey on tenure in organizations, Jolson et al. (1987) find evidence from sales people serving in-home customers showing that sales force size is positively related to an IBO's tenure. Their argument is that the larger the sales force, the more improved the quality of the sales managers and the training programs, and the more effective the sales support activities (Jolson et al. 1987). As mentioned earlier, unlike in traditional firms where workers meet one another on a daily basis, as IBOs are not hired by the firm, they have relatively limited contact with the firm or with other IBOs. Instead, IBOs (especially high status IBOs) often organize informal collaborative work groups, usually through the same family, to exchange information and selling tips, hold sales workshops to create shared opportunities, and assist one another (Sparks and Schenk 2001). Of course, the finding was based on a single-item survey, and uncertainty exists in terms of how the survey respondents interpreted sales force size. To come to a more robust understanding of turnover, I examine the number of others both within the family and in proximity, which I will explain in Section 2.3.3.3.

Percentage of high status IBOs within the family

In general, the larger the family size, the greater the likelihood of having successful people in the family (Jolson et al. 1987) showing a good example. Specifically, when those successful people exist within the same family, their status will become more prominent to the other IBOs. Identifying successful IBOs is done by the MLM firm.

As mentioned earlier, leadership-related variables such as supervisory satisfaction or leader-member exchange were a component explaining the turnover behavior in traditional organizations (Griffeth and Hom 2001). This institutional setting, along with the findings in organizational studies, lead me to expect that the higher the percentage of high status within one family, the less likely one is to become inactive because of the potential exposure to direct and indirect supervision and possible interaction with successful members.

2.3.3.3 Proximity

Even if an IBO is associated with a network family and receives helpful support and motivation, his/her business activity will also greatly depend on the proximity of each individual. Family connection is a global network aspect, while proximity is a local business environment. In an MLM, there are no constraints as to which market they will serve, unlike in a traditional sales force with assigned territories. Nevertheless, IBOs located close to others will have a higher chance of serving a similar customer base, with the possibility of an overlapping social network. In addition, since the compensation plan is highly dependent on how much an IBO and his/her recruitee sell the product, it is likely that IBOs located close to others will face competition. Previous research on sales force turnover attempted to examine the negative effect of perceived degree of competition on turnover, but the results turned out to be insignificant (Jolson et al. 1987). However, Msweli and Sargeant (2001) and Sparks and Schenk (2006) acknowledge the competition prevailing between IBOs, and the importance of coordinating between IBOs in the MLM organization (Msweli and Sargeant 2001), despite the need to be cooperative in the MLM industry (Sparks and Schenk 2006), a fundamental trait of IBO relationships (Biggart 1989, Grayson 1996). This concern of competition indicates that empirically supported competition may exist among IBOs within a proximate distance. I would thus expect to see local competition within a proximate area.

On the other hand, IBOs may stay active longer when high status IBOs from the same network family are located in proximity, both because of the role of high status explained earlier and the fact that the IBOs have a better chance of learning about their success stories compared to high status IBOs not sharing the same family.

2.4 Data

Data is available from one of the leading MLM firms in the U.S. for 318,363 IBOs who joined the MLM business from January 2006 until October 2007. Unlike previous churn literature where a clear cut definition of churn exists, as shown in Table 2.1, the data itself does not indicate whether an IBO dropped out of the business completely or not. Thus, I examine the concomitants of inactive behavior, that is, those who did not engage in the business for six consecutive months (i.e., two quarters).

More specifically, if six consecutive inactive data points are observed at the end of a particular period, from month (t+1) to (t+6), the IBO is categorized as inactive from month t. Some IBOs may not have engaged in the business for six consecutive months in the middle of the sequence but returned later on. These returned IBOs were then considered to be active

Table 2.1: A Sample of Churn Related Literature

	Industry	Churner
Coussement and Van den Poel (2009)	Newspaper Subscription Services	when his/her subscription is not renewed within four weeks after the expiry date (Company doesn't allow ending the subscription prior to the maturity date)
Lemmens and Croux (2006)	U.S. wireless telecommunications company	when a mature subscriber (who has been with the company for at least six months) churns during the period of 31- 60 days after the sampling date
Braun and Schweidel (2011)	Telecom provider	observe the month in which the customer was acquired and defined as churner if the customer churned while under observation.
Haenlein (2013)	Mobile phone industry	when churn happened during the data period
Jamal and Bucklin (2006)	Direct-to-home satellite TV provider	when churn happened during the data period (drawn from among those customers who had subscribed to the firm's service during the past 12 months)
Neslin et al. (2006)	Wireless telecom industry	Churn was measured in the fifth month

from the beginning of the data.[3] Engagement in the business is defined according to the following two criteria : (i) showing no point accumulation either through personal sales or group level sales, or (ii) no evidence of earning bonus income from the firm. Based on a six-month screening, 89,176 IBOs became inactive in the business, which is about 26 percent of the joiners. Figure 2.1 shows the tenure of the IBOs in the business who eventually became inactive.

Both a county-level and an individual-level analyses are presented in Section 2.5. In this data section, the three dimensions for the individual-level analysis is discussed, except for the variables that appear in both analyses. Data relevant to the county-level analysis is presented under the county-level analysis section (Section 2.5.1).

2.4.1 Individuals

Two important individuals in an MLM have a potential influence on the focal IBO's inactive behavior: an upline and a high status IBO. For each IBO, an identification code indicates the connected upline IBO and a relevant high status IBO (a sub-group leader). Based on the very last data point, October 2007, among 606,551 IBOs, 3,901 — approximately 0.6%,

[3]Active was defined as (personal sales point > 0 & non-missing value) OR (group sales point > 0 & non-missing value) OR (earned bonus > 0 & non-missing value). One needs to be in inactive state for 6 consecutive months toward the end of the data.

a very small number were awarded high status .

To capture the influence of these two individuals, geographic distance between them is computed based on zip-code information. The descriptive statistics for the upline distance and the high status distance in miles is provided in Table 2.4. As can be seen, in general, the high status IBO is located farther away from the focal IBO compared to the upline (mean (High status distance) = 346.12 versus mean (Upline distance) = 149.891), perhaps because MLM business relationships are often created with someone located nearby. Based on a survey from Raymond and Tanner Jr. (1994), most consumers made product purchases at home, followed by the workplace as the second most frequent location for business interactions, which can lead to MLM business relationships.

Another to note is the correlation between the distance from a high status IBO and the percentage of high status IBOs within the family or the percentage of high status IBOs within the proximity. Tables 2.2 and 2.3 show the correlations. The two correlations are directionally negative (2.2corr(distance from a high status IBO, percentage of high status IBO within the family) = -0.005), Table 2.3 corr(distance from a high status IBO, percentage of high status IBOs within proximity) = -0.131). However, the distance from a high status IBO is not necessarily highly correlated with the percentage of high status IBO within the family nor the percentage of high status IBOs within proximity, meaning that the distance from the high status IBO is capturing a different aspect from others in different dimensions.

2.4.2 Network Family

Network family is a motivational organization, usually set up by successful IBOs. Each IBO is associated with a network family, which can be identified with a label. The variables related to a family (family size, high status ratio within the family) are computed using the associated family label for each individual. In the data, there are 21 families during the period of 2006 – 2007. Figures 2.2 and 2.3 show the family size and the percentage of high status members within a family based on the very last observation, October 2007. As can be seen, a great deal of variation exists across the families. In terms of the size of the network, the largest family is 184 times larger than the smallest network: the largest family includes 101,400 IBOs; the smallest has only 550 IBOs. Examining the family network size variation across time shows that the growth rate across the family from January 2006 to October 2007 also varies considerably, ranging from 7.4% to 257%, and is not necessarily correlated with the family size. Given the variation in growth rates, it seems reasonable to conclude that an IBO's involvement in recruiting activity varies depending on the IBO's network family.

Based on full data from 2006 until October 2007, a family tend to have few high status members, if any, with a maximum percentage being around 2.07% at most. As shown in

Tables 2.2 and 2.3, the correlation between the size of family and the number of high status IBOs is 0.907; I thus examine the percentage of high status IBOs to capture the relative degree of exposure to them. Figure 2.3 shows the percentage of high status IBOs in each family at the very last data point. The small percentage implies that attaining high status is an indicative of significant success, which involves a rather difficult process. Even within the small percentage, however, there is variation across families. As an example, one of the families has no high status members. Another notable characteristic is that the family size and percentage of high status IBOs do not necessarily have a linear correspondence. For example, family 37 has about half the number of IBOs compared to family 17, the largest family; yet, the percentage of high status IBOs in family 37 is slightly greater.

Not only do the size and the composition of status differ in families, but also how the IBOs within a family are distributed varies as well. Figure 2.4 shows the geographic distribution of two exemplar families in October 2007 within the U.S. As can be seen, depending on where an IBO is located, the exposure to own versus other family members varies considerably. Also these families are dispersed across the U.S. and that the density or the highly populated area may be different depending on the family,. Table 2.4 shows the summary statistics of family-related variables using the time series data, using from January 2006 to October 2007.

2.4.3 Proximity

In order to determine the existence of other IBOs within a small boundary, proximity is defined as a 30-mile radius from an IBO's zipcode. Figure 2.5 shows the approximate 30-mile radius from a zip-code. Proximity (own family ratio and the number of total IBOs within 30 miles) was computed using the individual-level family label and zip-code information. The 30-mile radius was drawn at the zip-code level, allowing us to possibly examine micro-level competition among IBOs.

Table 2.4 shows the descriptive statistics of IBO's 30-mile proximity at the individual level. IBO's zipcodes are used to calculate the number of IBOs within the same family, the number of total IBOs regardless of the family, and the percentage of own family. Tables 2.2 and 2.3 show the correlation among the variables representing the two dimensions of others: Network family and proximity. Since the high correlation between the number of an IBO's own family and the number of total IBOs within a 30-mile radius is highly correlated (corr= 0.77), the percentage of own family is used instead of the raw number for the individual level analysis. Included is the percentage and the number of high status IBOs from the same family for future usage.

The number of total IBOs varies widely. For instance, an area with the zip code 86502 in Arizona has only 5 total IBOs within a 30-mile radius (Figure 2.6 Left. Red boundary.), whereas an area with the zip code 90631 in California has 44,791 IBOs (Figure 2.6 Right. Red boundary.). Similarly, wide variation exists in the percentage of own family and the number of total IBOs. For some IBOs, for example, less than 1 % of surrounded IBOs are from the same family, while for other IBOs, 90 % of surrounded IBOs are from the same family.

The row Market in Table 2.4 indicates the market penetration rate by the IBOs. It is constructed from the Total number of IBOs/the number of labor force within a county. This figure captures how popular and widely accepted the MLM firm and its product are. I have supplemented the yearly county level labor force data from the Bureau of Labor Statistics. Thus, IBOs within the same county will have the same figure. As the county level is the granularity of data I have for the labor force, the total number of IBOs were also computed at the county level.

2.5 Existence of Others

MLM Turnover is primarily voluntary and, as mentioned earlier, often not a formally documented event: it is simply the consequence of permanent inactivity (Wotruba 1990b). To examine turnover, previous studies of turnover among traditional salespeople studies have used either a dichotomous measure (quit, not quit), a single scale likelihood to quit in the next x months (Msweli-Mbanga 2004), an estimate of the expected duration of future employment (Jolson et al. 1987), an intentions-to-quit (or intentions-to-stay) scale (Wotruba and Tyagi 1992), scales measuring thinking of quitting or attitude toward quitting, or an interval scale of inactivity-proneness (I have never thought about quitting – I am no longer active in the job)(Wotruba 1990b). These prior studies typically take the form of self-reported surveys, thus relying on scales asking direct or indirect questions related to the turnover or future tenure. As explained earlier, inactive behavior was defined as those IBOs who did not engage in the MLM business for six consecutive months. There are two reasons for using this measure. First, there is no indication of explicit drop-out in the dataset. Second, the measure is managerially meaningful to examine because, from a firm or a sponsoring IBO's perspective, a six-month span enables the possibility of intervening before an IBO becomes inactive, and thus increase the chance of remaining active.

2.5.1 County-level Analysis

The overall pattern of inactivity is examined using an aggregate-level regression analysis at the county level. The number of counties represented by the IBOs who joined after

December 2005 is 2,877 among from 3,220 counties in the U.S.. The dependent variable is the percentage of IBOs *active* during a given month for each county. The effect of family or proximity cannot be determined here as the data is available only at the individual-level (see discussion in the follow-up analysis). Instead, an initial relationship is shown between the existence of other IBOs and the overall activity within a county.

For each county, the number of IBOs (regardless of family) and the percentage of high status IBOs are computed to examine the role of others in inactivity behavior. The reason for using the percentage of high status IBOs instead of the raw number is due to a high correlation between the number of high status within a county and the number of IBOs (Table 2.6).

To capture market characteristics, the unemployment rate of the county and the penetration rate within the county are included. The penetration rate (Total number of IBOs/number of Labor force) reflects the popularity/awareness of the firm or products (demand), as explained in the last part of the data section.

In addition, as mentioned in the literature review section, there is wide consensus on the relationship between performance and turnover, both in the sales force tenure literature and the MLM tenure literature. To control for this, the average estimated earnings from products selling in a given a county were calculated. While each IBO can set his or her own retail margin, absent the individual level mark-up information, it is assumed to be fixed at a 20 percent, which corresponds to the minimum markup the survey respondents responded in Coughlan and Grayson (1998).[4] The reason for using the earnings from personal product selling is because of the high correlation between the percentage of high status IBOs and bonus earnings, and the percentage of high status IBOs and total earnings, as shown in Table 2.6 (cor(% high status, bonus) = 0.731, cor(% high status, earning) = 0.583, cor(% high status, personal selling) = 0.159).[5]

In addition to the influence of other IBOs, there could be influence from others who are not necessarily associated with the same MLM firm; for example, competitors in a similar business. To partially capture the degree of competitors working for other MLM firms, I used competitors' headquarters information, under the assumption that there would be many IBOs working with the MLM company if the headquarters were located nearby. Using the 2016 Direct Selling News North America 50 List, which is created based on the North America 50 ranking for the 2016 DSN Global 100 (DSN 2016),[6] I identified possible competing MLM firms for 2006 - 2007. Among the 50 firms in the 2016 list, 39 competing firms existed

[4]Based on Coughlan and Grayson (1998), the firm suggested markup is reported to be 40 - 50 percent, and the average respondents' markup was within the range.

[5]The same analysis was also run using bonus instead, and the results were consistent.

[6]ranked based on the 2015 revenues

during those years, which matched my data span. Next, I found the location of the headquarters and matched it with the county. From 39 competing firms' locations, 27 counties were identified as having other competing firms. To reflect the competition, two metrics were used. The first is the shortest distance from one of these competing firms' headquarters. The distance is calculated based on the center of the county using fips information. Figure 2.7 shows the distribution of the shortest distance from the competitors' headquarters in miles (min = 0, Median = 170.300, Mean = 295.400, max = 725.300). The second is the number of competing firms' headquarters within a certain boundary, a similar approach to that of Seim (2006). The results with both metrics are presented in Tables 2.7 and 2.8. Lastly, year, month and county-level fixed effects were included. The descriptive statistics on variables in the county-level analysis are reported in Table 2.5.

Results The results based on the two competitor related metrics are presented in Tables 2.7 and 2.8. The analysis shows that the counties with a smaller number of IBOs in the previous month resulted in a higher percentage of IBOs staying active in the business. This implies that even though IBOs are not assigned/constrained to a fixed geographic region in expanding the network or serving end customers, IBOs are inevitably competing with one another within a similar geographic region by selling the same product purchased from the firm, as in other franchise business. Also, the percentage of active IBOs were found to be higher when there are relatively more high status IBOs within the county, which may be the case for two potential reasons. First, having a stratified status system based on an IBO's achievement can encourage other surrounding IBOs to stay in the business — showing a good example of success. To put it differently, one can naturally think, "s/he achieved it, why can't I?." Second, high status does not only mean receiving more financial benefit, but also having a greater responsibility to guide others and become a role model. Receiving such guidance and tips to run the business may encourage IBOs to stay active in the business.

In general, the higher penetration rate (number of total IBOs adjusted for the labor force size of a county), the less likely to show inactivity. Thus, when a county has a high demand either for the brand or the product, the percentage of IBOs staying active is higher. Also, counties with higher average previous IBOs' earnings from own product selling at (t-1) showed a higher percentage of active IBOs at (t), consistent with the previous literature.

In terms of competitors from other MLM firms, the estimate for the shortest distance from the competing firm's headquarters is positively significant implying that the farther away from the competing headquarters, the higher the active IBO percentage. Based on the descriptive statistics of minimum distance from the competing firms' headquarters (median = 178.6, mean = 247.2), I also tried using the number of competing firm's headquarters within 200 miles for the robustness check, shown in Table 2.8. The overall pattern remains the same. The number of competitor headquarters is negatively statistically significant as expected (coef. = -0.069, t-value = -3.583), indicating that when a county is highly exposed to competing firms' business activity, the percentage of IBOs staying active is lower.

Although initial relationship can be seen between the inactivity behavior and the business environment, in particular the existence of other IBOs, there are some limitations to this approach. This analysis does not allow us to examine the different dimensions of others, namely, network family and geographic proximity, at the individual level. Also, as it is an aggregate level analysis, individual level heterogeneity cannot be controlled for, which leads us to the individual-level analysis.

2.5.2 Individual-level Analysis

In the following subsections, statistical survival analysis is used to examine how the existence of other IBOs influences one's inactivity in the MLM business. As mentioned in the previous section, three others of interest are to be focused on: important individuals, network family and proximity. To carry out this analysis, each IBO's inactive time information will be used as the key dependent variable.

To control for the individual specific business related characteristics, an individual IBO's own status, earnings, network structure (level) and business style (# frontline, # downline, # customer) are used. Drawing from previous research on the relationship between job performance and tenure, these individual characteristics are controlled for. For the market related variables, county-level penetration rate and unemployment rate are incorporated — the same variables as in the county-level analysis.

To account for the influence of other competitors from different MLM firms, similar to the county-level analysis, the shortest distance from the competing firm's headquarters is used, where the distance is calculated based on the individual zip-code rather than the center of the county. Individual-level socio-demographic information, including age, five groups of language (English, Spanish, Mandarin, Spanish, Korean), three groups of marital status (Married, Single, Unknown), three groups of gender (Male, Female, Unknown), eight groups of ethnicity (European, Hispanic, Asian, African American, Non-orient Asian/Polynesian, Jewish, Arab, Unknown), and ten groups of income level (min = less than \$15,000, max = more than \$125,000) were also controlled for.[7] Last, other control variables were included such as time fixed effects, state fixed effects, family fixed effects, and entry fixed effects

[7] To circumvent dropping observations due to missing socio-demographic information, an additional category for each type of socio-demographic variables were used to indicate the observations with a missing value.

(cohort effects).

2.5.2.1 Time-independent hazard model

I first employed a time-independent Cox hazard model by extracting summary statistics from the individual level time-series panel data to explain the duration of IBOs staying active in the business. Unlike other parametric survival analysis models, the Cox hazard model is known for not relying on distributional assumptions in modeling hazard (Cox 1992). Three types of models were employed, with the mean, initial, and the last value of the covariates for each IBO. If the results from these three models were consistent, then there would not be not much need to incorporate the time varying nature of covariates, as the results could be loosely generalized. Right censored observations were also incorporated in the model. Table 2.9 shows the results.

The results show that the sign and significance across the models are consistent for some variables; for example, own family ratio within a proximity, distance from two important individuals, penetration rate, and the shortest distance from the competitors' headquarters. In general, however, using the mean, initial and last value of the covariates for each individual produced inconsistent estimates, with some even flipping the sign, which led to the need to investigate a model that incorporates time-varying covariates.

2.5.2.2 Time-varying hazard model

Here, I employed a modified version of the Cox hazard model with time-dependent covariates, also referred to as an extended Cox-model (Therneau et al. 2017). Specifically, for my monthly data, a discrete-time hazard rate model was applied that incorporates the impact of time-varying covariates (e.g., density of other IBOs, own time-varying characteristics) on the likelihood of becoming inactive. It is similar to Gupta (1991), Nikolaeva et al. (2009) in the sense that the nature of time dependence is incorporated; it differs, however, as it modifies the Cox hazard model, which does not assume a specific distribution to the hazard rate. A Cox model with time-independent covariates would compare the survival distributions between those who do not become inactive (ever) to those who become inactive for each month. An IBO's status at the end of the period of data would determine which category they were in for the entire duration. However, in the time-varying covariate case, the risk of an event between becoming inactive versus not are compared at each event month, and would re-evaluate which risk group each IBO belongs to based on whether IBOs became inactive by that month.

Conditional on an IBO's time-dependent covariate $X_i(t)$, the Cox model for hazard is:

$$h(t; X_i(t)) = \lambda(t) \times \phi(X_i(t)), \tag{2.1}$$

where $\phi(X_i(t)$ is a time-varying covariate function and $\lambda(t)$ is the base hazard rate. Each IBO's hazard rate will depend on its characteristics (e.g., one's own business activity) and proximity market factors (e.g., competitive/supportive IBO density), some of which may vary over time. Thus, the impact of covariates in IBO i's hazard rate in period t is captured in function ϕ:

$$\phi(X_i(t)) = exp[\beta' X_i(t)] \tag{2.2}$$

where $X_i(t)$ are time-dependent covariates, β are the response coefficients. The inactivity hazard at time t depends only on the value of the covariates at that time ($X_i(t)$); β is constant over time. A positive $\hat{\beta}$ implies that an increase in the value of the covariate by one unit will increase the conditional probability of the IBO's inactivity by $exp(\hat{\beta}) \times 100\%$, whereas a negative $\hat{\beta}$ implies that there will be a decrease in the probability of inactivity by $exp(\hat{\beta}) \times 100\%$ compared to the baseline hazard.

Suppose there are K distinct inactivity occasions and let $(t_1, t_2, ..., t_K)$ indicate K ordered inactivity starting months from the entry. The risk set $R(t_k)$ denotes the set of IBOs who are at risk for inactivity (still active) after t_k months of entry.[8] For a particular inactivity at month t_k, conditional on the risk set $R(t_k)$, the partial likelihood can be specified as

$$L(\beta) = \prod_{i=1}^{N} \left[\frac{exp[\beta' X_i(t_k)]}{\Sigma_{j \in R(t_k)} exp[\beta' X_j(t_k)]} \right]^{\delta_i}$$

where δ_i is the inactivity starting point indicator. Thus, at each event month, the model compares the current covariate values of the IBOs who had became inactive at month t_k to the current values of all other IBOs who were at risk at that t_k, or who had not yet become inactive (Therneau et al. 2017, Xu 2016).

2.5.2.3 Results and Discussion

Table 2.10 shows the individual level survival analysis results incorporating the time varying covariates. Each column from (1) to (7) indicates a model by gradually adding different types of variables, where Model (7) is the full model. Model (1) is the baseline model including only the time, state, family fixed effects and cohort effect (entry time). Model (2) adds individual-level dummy-coded socio-demographic variables, and Model (3) adds individual business-related characteristics. Model (4) includes market and competitor information.

[8] Entry time of each IBO may be different.

Model (5) adds distance information from specific types of connected IBOs, namely upline and high status. Last, models (6) and (7) include proximity and network family variables, respectively. Across the models, the signs and the magnitude remain quite consistent.

Table 2.11 is based on Model (7), but reports more detailed results. I begin by discussing the most comprehensive model (7). After discussing the model, I will compare Models (1) – (7). Table 2.11, exp(coeff) reports the change in hazard rate from the baseline if there is a change in a variable by one unit. Let me begin by examining how the existence of others at the individual level affect an IBO's inactive behavior. It is followed by network family, proximity, and other variables that are worth examining.

Individual

As mentioned earlier, two individual IBOs are of interest based on the institutional setting: Upline and high status IBO. An upline is the person an IBO created the business relationship with. When a prospective IBO joins the MLM, s/he is connected with only one IBO, namely an upline. A high status IBO is a successful IBO approved by the MLM firm who takes on a role as the leader of a sub-group. The zip code distance between these two IBOs and a focal IBO are measured, respectively. The results in Table 2.11 reveal that, consistent with the hypothesis, the zip code distance from the sponsoring upline IBO slightly decreased the inactivity hazard compared to the baseline. The percentage change in the hazard rate from the baseline hazard is quite small, probably because the unit of the distance in this analysis is one mile. On the other hand, when the IBO is closer to the high status IBO, the inactivity hazard slightly increased, unexpectedly. This does not support the hypothesis raised in the literature review, that IBOs who are geographically close to a high status IBO will tend to stay active in the business. Further investigation of this issue will be needed to understand the underlying mechanism. Generally, the absolute effect size is larger for the distance to the upline distance rather than to the high status IBO, indicating that IBOs' activity is more largely influenced by the direct business sponsor.

Network Family

Based on the results in Model (7), I find that the larger the network family size, the lower the inactive hazard. Compared to the baseline hazard, increasing the family size by one unit results in a slight decrease in inactive hazard, 0.962 times less inactive compared to the baseline. This is consistent with previous survey results on organization size and turnover of salespeople serving in-home customers (Jolson et al. 1987), and literature based on general organization turnover, in which workers tend to stay in the organization when there are socialization tactics, high job-embeddedness, and high social ties within the organization.

Being associated with a larger network family seems to provide such socialization opportunities with a larger group of other IBOs, in turn serving as an inspirational and motivational group — having a protective effect.

The percentage of high status IBOs within the same network family is estimated to be negative, which indicates that an increase in the percentage of high status IBOs within the family by one unit will significantly decrease the hazard, 0.078 times less hazard compared to the baseline. Based on the results, it seems that the high status IBOs are not regarded as potential competitors but rather aspirational entities who motivate IBOs to be like them. High status IBOs within a network family also show the protective effect.

Proximity

Having more IBOs within the 30-mile radius from the same family seems to increase the inactivity hazard. This result is somewhat consistent with county-level results, where the percentage of active IBOs and the number of total IBOs within the same county were found to be negatively correlated. However, in the individual-level analysis, the total number of IBOs within proximity did not significantly affect an IBO's behavior. In other words, the more IBOs 'whom you have a high chance to know' are around you, makes it worse. Although the cooperative relationship is advised (Sparks and Schenk 2006, Msweli and Sargeant 2001), IBOs may perceive other IBOs nearby as potential competitors.

In order to examine the role of high status IBOs within the 30 mile radius, I included the percentage of high status IBOs in the full model.[9] Adding this variable did not change the sign and significance of other variables. Interestingly, the percentage of high status IBOs who share the same family within the 30- mile radius reduced the IBOs' risk of inactivity. Not only having relatively more high status IBOs within the family as a whole network but also having more of them physically near by seems to help motivate IBOs to engage in the business.

Other variables

Other variables are also worth highlighting in Table 2.11. First, the penetration rate, which captures the relative density of IBOs among the labor force within a county, was found to be negative, moderately decreasing the hazard 0.812 times. This finding is consistent with the results from the county-level analysis, where the counties with a high penetration rate showed a higher percentage of active IBOs. Again, having more IBOs relative to the total number of labor force may imply a stronger presence of a brand or firm in a market with

[9]High status % = (# of high status IBOs from own family within 30mi/ Total # of IBOs within 30mi)

a higher awareness and familiarity either to end customers or future potential IBOs. This creates an accommodating environment for initiating sales to end-customers or expanding business networks (Msweli and Sargeant 2001). This finding also aligns with previous survey findings that the popularity or acceptance of a product or company is positively related to sales force tenure (Jolson et al. 1987).

In terms of competitors, the minimum distance to the closest competing MLM firm's headquarters is negatively significant, which is similar to the findings in the county-level analysis. Touching upon the individual business related variables, I see that IBOs who are successful in the business — either through financial earnings or business activities (i.e., building a network (# of frontline and/or # of downline), selling the product (# of retail customer))- tended to stay active in the business. Although the high status variable was insignificant, probably due to the correlation among other variables signifying success, the result is consistent with literature that sheds light on the relationship between job performance and turnover in MLMs (Jolson et al. 1987, Wotruba 1990b).

Table 2.12 shows the gradual increase in the AIC, as more information was added to the model. Due to the different number of samples, the difference between AIC of two consecutive columns are computed to approximate the overall value of information type added in explaining the IBO's inactivity behavior. Based on the Δ AIC, individual business-related information (e.g., earnings, business style reflected by the number of direct downlines, total downlines and registered customers) had the highest contribution in explaining inactivity (Δ AIC = 17,556). Based on the solid literature on job performance and turnover in organization studies, this finding is expected.

This is followed by the two important individuals related variables, namely the distance from upline and high status IBOs (Δ AIC = 7,643). The improvement is significantly large considering the number of variables added. The improvement may be because these two individuals are either the most closely connected to an IBO, or the most closely observable success story.

Next, the third largest increase is from the network family information, such as the size of family and the percentage of high status IBOs within the family. Interestingly enough, the gain is even slightly larger than that of the socio-demographic information, which has a considerably larger set of variables. This improvement implies that, aside from individual business-related characteristics, the network related information — network size, and network composition — has a positive influence on an IBO's activity.

Lastly, and unexpectedly, proximity related information, (e.g., the total number of IBOs, the percentage of own family, and the percentage of high status IBOs from own family within 30 miles) had the least explanatory power reflecting inactivity behavior. Based on these results, aside from the two individual IBOs, the existence of others at the network family level appears to dominate the others nearby.

2.6 Prediction Using Machine Learning

This section applies a machine learning technique to predict IBO's active duration in the business. The advantage of using machine learning is that, with the availability of big data, it can model real world phenomena without being limited to a pre-defined functional form. Strict assumptions from traditional statistics, such as normaility (Hua et al. 2007), are also unnecessary. Although some concerns have arisen about the relative paucity of econometric theory and the difficulty of interpretation (Bajari et al. 2015b), interest in utilizing machine learning in economics and business has been on the rise. Bajari et al. (2015a,b) surveyed some of the popular methods in machine learning, which include random forest and support vector machines, and its applicability to demand estimation. They showed that utilizing machine learning produced the smallest out-of-sample prediction error.

My review of the cost of turnover in MLMs highlighted the importance of preventing MLM inactivity, a precursor of turnover. Learning from IBOs' past behavior and predicting the probability of staying inactive can potentially mitigate inactivity. Thus, in this section, I predict an IBO's future inactivity using three different definitions of inactivity. The first approach is binary incidence prediction, determining whether an IBO will go through an inactive period or not in the following month. The results are compared using a binary discrete choice model. The second and third approaches are based on duration prediction; in particular, determining when IBOs will become inactive in the future, in quarters. A more detailed explanation of the difference between the second and the third approach will be provided in the analysis section.

Note that the second and third approaches predict the exact quarter when the IBO is going to be inactive, similar to Chaplot et al. (2015) in the Massive Open Online Courses (MOOCs) literature, rather than whether an IBO is going to stay in the business or not (Sinha et al. 2014). Predicting the exact stopping period (student attrition) has been studied in MOOCs (Chaplot et al. 2015). However, the problem setting is quite distinct. In the MOOCs' context, the drop-out can be clearly defined, as each student is facing the same online lecture and homework, and the end point is clearly defined. In MLM, the issue becomes more complicated, with no specific steps to follow due to the amount of freedom given to the IBOs and no clear end point. Thinking about different ways to predict behavior will allow the MLM firm and sponsoring uplines to take the necessary steps to prevent or reduce IBO attrition while they are in the business.

Whether an IBO stays in the business or becomes inactive, the precursor to turnover, cannot be determined by a small set of variables. Thus, in this section, for prediction purposes, more detailed information from the data, such as recency or frequency of business activities, is included. In addition, because inactivity is most likely the outcome of the interactions between multiple factors such as past business activity, network structure and the market environment, the prediction of whether an IBO stays active in the business or not is a function of non-linear relationships between multiple predictor variables. Machine learning methods have thus been applied to our analysis, as they have proven to be appropriate for such models (Bajari et al. 2015a). Specifically, I focus on one methodology, neural networks.

2.6.1 Neural Networks

Artificial neural network (or simply neural network) is one type of machine learning. How a neural network works is similar to how our brain works. In the brain, neural pathways become strengthened if they are useful. In neural networks, pathways also become strengthened if they are useful; however, it is the weight between neurons that changes (increases) to reinforce a successful neural pathway, and such adjustment (training) in weight is done through a back-propagation algorithm. In other works, a neural network seeks a set of weights to derive the best possible fit through observations of the training data set (Wu et al. 2006). From a statistical point of view, a neural network is closely related to generalized linear models; however, a neural network is much more flexible in capturing non-linear relationships among variables, and uses a different estimation procedure (feed forward and back prop- agation) (West et al. 1997).[10]

This methodology has been applied to numerous fields, such as pattern recognition, medical diagnosis, forecasting for tourism demand, and stock market returns (Deng et al. 2008). Its application to the business-related topics have a relative short history.[11] Some examples include the prediction of corporate financial distress (Hua et al. 2007, Chen and Du 2009), consumer choice prediction (West et al. 1997), and evaluation in bank branch efficiency (Wu et al. 2006). Due to the empirical success of artificial neural networks at predicting the outcome of complex nonlinear processes, the methodology has become recognized for its superior forecasting ability (West et al. 1997, Deng et al. 2008, Maciel and Ballini 2008, Chaplot et al. 2015).

A neural network is composed of three different types of layers — a set of neurons — and an activation function. First is the input layer, which can be considered as the independent

[10] West et al. (1997) introduces a definitive description and provides pros and cons of using this methodology relative to statistical procedures.

[11] Vellido et al. (1999) has a good amount of survey on the applications in business context.

variable. Second is the output layer, which is the dependent variable or the desired output. Last is the hidden layer(s), which connects the input layer to the output layer. This layer determines the non-linear mapping relationship (Chen and Du 2009). Figure 2.8 shows an example of a neural network with one input layer, one hidden layer and one output layer. Multiple hidden layers can be incorporated to capture more complex data patterns. Previous studies indicate that a neural network with even one hidden layer with a sufficient number of neurons can approximate any continuous function (Kaastra and Boyd 1996). The number of neurons in the hidden layer should fall somewhere between the input layer size and the output layer size, but there is no rule of thumb. Each layer is connected with weights, which I recover through algorithms. Figure 2.8 shows a partial set of weights, labeled $w_{10}, ..., w_{13}$, which connects the hidden layer and the output layer. Lastly, the activation function allows us to generate actual network output, which is used to match with the desired value. Given Figure 2.8, the activation function plays a role in neurons in the hidden layer and the output layer, colored in orange. Different types of functions, such as sigmoid, linear, logit, hyperbolic tangent (West et al. 1997), can be used as the activation function.

2.6.2 Data for Machine Learning

Table 2.13 describes independent variables in the machine learning analysis. A total of 225 explanatory variables are included. These variables number about twice as many compared to the previous survival analysis, with 114 variables. These variables are the neurons in the input layer (Figure 2.8).

The independent variables — either time variant or invariant — are divided into nine categories and these variables. As mentioned in the previous data section, the three dimensions of others are also included — the categories of individuals, network family, and proximity. I also have market-level variables and variables that capture competition with other competing firms, measured based on either the zip code distance from the headquarters or the number of headquarters within a certain boundary. Individual business-related variables are also incorporated, but compared to the previous analysis, significantly more variables that capture IBOs' business activity trajectories are included. To control for individual characteristics not necessarily related to the business, socio-demographic dummy variables are included.

Additional new category has been added, labeled 'Recency, Frequency, and Monetary value'. In the literature on customer churning, Fader et al. (2005) shows that customer's past behavior is an important predictor for one's future behavior. It is common practice to summarize customers' past behavior in terms of Recency, Frequency and Monetary value(RFM) characteristics (Fader et al. 2005, Coussement and Van den Poel 2009). Thus, in order to capture the past business-related interaction, RFM related variables are incorporated.

Recency (e.g., the elapsed time since last purchase or recruiting) and Frequency (e.g., the number of months with prior purchases or with recruiting) are the new category added, whereas Monetary value (Total amount of selling, Total number of people recruited) appears under the Individual business-related category.

Lastly, other variables include time, month, network family and state fixed effects. The same set of variables shown in Table 2.13 is used across the three different analyses presented.

2.6.3 Analysis and Results

In general, four important statistics are examined to the total number of predictions: the percentage of true negatives (TN), true positives (TP), false negatives (FN), and false positives (FP). In the inactive behavior setting, positives indicate the inactive observations — incidents we are aiming to predict — whereas negatives indicate the active observations. For the turnover case, in general, false negatives — the percentage of predictions erroneously predicted to be active — are substantially more costly than false positives — percentage of predictions erroneously predicted to be inactive. Thus, minimizing the false negative rate is important.[12] While minimizing the false negative rate, it is also necessary to maintain overall accuracy so as to not produce too many false positives for the organization to handle. The metrics below will be used to report the model fit. Precision indicates among the observations that were predicted as inactive, the percentage of correctly identified observations. Recall indicates among the inactive observations, the percentage of correctly identified inactive observations.

Accuracy = (True positive + True negative)/ Total Observation = # of correct prediction / Total Observation
Precision = True positive / Positive Prediction
Recall = True positive/ Positive data

2.6.3.1 Approach One – Incidence Prediction

This section compares neural network with other methods. Here, the purpose of the analysis is to provide a way to pin-point the IBOs who are *likely to become inactive in the following month.* The output of the algorithm is 1 indicating the IBO will become inactive following month for exact six consecutive months, and 0 otherwise. Figure 2.9 shows the visual representation of the dependent variable. The orange box indicates the period without engagement for less than six months, thus treated as active. The purple boxes indicate

[12] A false negative rate of 0.x means that I correctly identify(100-(100*0.x))%

the period without engagement six months or more. The Y_{next} is coded as 1 if the month is followed by six consecutive no-engagement months. If an IBO later came back to the market, just like the observation between the two purple boxes in Figure 2.9, the observation is again treated as active until it becomes inactive the following month for exactly six consecutive months.

In order to provide a comparison between the binary logistic model and the neural network model, prediction using the traditional method was also implemented. The same set of independent variables is used for the two models. The logistic model is widely used for binary predictions. However, unless one pre-specifies interaction terms among variables, the method is restricted in being able to capture the non-linear relationships among all the variables. A neural network can naturally incorporate such non-linear relationships among the variables through non-linear activation functions that connect neurons in different layers (orange highlighted circles in Figure 2.8) and also by having multiple hidden layers.

For all the neural network models, I used H2O, which is a fast, scalable open source machine learning tool (Candel et al. 2016). H2O can be used in R, Python, and other platforms, and it is known to provide faster speed compared to other neural network packages in R (e.g., nnet, neuralnet). For training, the data was randomly split into a training set (70%), a validation set (15%), and a test set (15%). The validation set was used to overcome the over-fitting problem in the model. Two hidden layers with 130 and 90 neurons respectively, were used.

Table 2.14 shows a direct comparison of the prediction on the test data between the neural network and binary logistic regression. Across all metrics, the neural network outperformed the logistic regression results, particularly on the precision metric. Precision indicates among the predicted inactivities how precisely the prediction reflects the true data. Thus, the neural network outperformed in making a precise prediction. Table 2.15 shows the contingency table based on the neural network.

2.6.3.2 Approach Two – Duration Prediction

The second approach is consistent with what was used in the previous sections. With no evidence of engaging in the business for a consecutive 6 months (t+1) – (t+6), and *having no further indication of returning to the business*, the observation right before the consecutive 6 months, (t) is identified as an inactive point. Following this approach, the dependent variable is constructed based on the very last inactivity behavior for each individual. Figure 2.10 shows an example. The purple boxes indicate inactive period, where there is no engagement for six months or longer. The orange box indicates no engagement, yet the duration is less than six months. For each individual and for each month, the dependent variable is coded as the number of quarters left until start of the inactivity (Y_q1 in Figure 2.10).

The reason for using the quarter-based dependent variable (Y_q1) instead of predicting the exact month (Y_{m1}) is to the reduce the unbalanced data problem. For example, 91% IBOs were categorized as active. With so few observations facing the inactive period, categorizing the duration by month, would leave only a small number of observations for each month category to learn from. Lumping the remaining duration specified in month into a quarter results in relatively more observations for each quarter category to learn from.

Combining the months into quarters seemed to be the best approach for now. These quarters indicate when an IBO will face the inactive period, such as within one quarter, two quarters, up to five quarters. They are treated as categorical variables, similar to solving the classification problem. The output of the algorithm is a quarter-based prediction: how many quarters left until the inactive month. An additional category, 'active' is also defined, in order to keep the observations that are not followed by the inactive period.

For training, the data was randomly split into a training set (70%), a validation set (15%), and a test set (15%,), consistent with the previous approach. The validation set is used to overcome the over-fitting problem in the model. Ten-cross validation was used and two hidden layers was specified with 100 and 30 neurons respectively. For the activation function, rectifier was used. One issue that arise is that of the unbalanced sample,as shown in column 'Data %' in Table 2.17. Such data imbalance leads to a bias of the classifier towards the majority class. To handle this discrepancy, the training data has been balanced so that the minority classes are over-sampled and the majority classes are under-sampled.

The out of sample prediction is shown in Table 2.17. Based on the second approach, the accuracy of 79.9% was quite high. In general, the method did well in distinguishing those observations without future inactive behavior from those with upcoming behavior (Active vs the rest). In addition, the model was sufficient for predicting the inactivity period beginning within a quarter: among the data with an inactive period starting within a quarter, 77.9% were correctly identified. However, it did not perform well in distinguishing the future quarters in which an IBO might become inactive. Lack of precision for the longer term prediction could be due to the definition of inactivity, which leaves smaller number of observations to learn the IBOs' behavior for the longer term prediction.

Table 2.18 shows the top twenty-five relative importance of the input variables. The relative importance can be approximated using the estimated weights. Even though the impression is that often the neural network is a black box, there have been attempts to recover the relative importance utilizing the estimated weights (Olden et al. 2004, West et al. 1997). The method reported here relies on Gedeon's method (Gedeon 1997), implemented by H2O. Based on the method, the importance of input i on output k is computed as the product of connecting weights going through r, where r is all possible hidden neurons. The relative importance is obtained by normalizing Q_{ik}s so that the max is one.[13]

For illustration purpose, when there are m input neurons and one hidden layer with n hidden neurons, measure for the contribution of an input i to a hidden neuron j is defined as P_{ij} where

$$P_{ij} = \frac{|w_{ij}|}{\sum_{p=1}^{m} |w_{pj}|} \tag{2.3}$$

Similarly, the contribution of a hidden neuron j to an output neuron k can be defined as P_{jk} where

$$P_{jk} = \frac{|w_{jk}|}{\sum_{r=1}^{n} |w_{rk}|} \tag{2.4}$$

The contribution of an input neuron i to an output neuron k is then,

$$Q_{ik} = \sum_{r=1}^{n} (P_{ir} \times P_{rk}). \tag{2.5}$$

Although the model did not precisely predict when future inactivity would take place, it is still meaningful to examine which variables determined the prediction the most. Interestingly, along with the individual business characterstics (e.g., Activation days, # frontline) and socio-demographic variables (e.g., age, income groups, which appear more often than other socio-demographic variables), variables that are related to network family (high status percentage within a family, network family fixed effects) and important individuals (distance from high status and upline) are ranked relatively high. The variables related to others are boldfaced. Although the statistical significance cannot be tested using machine learning methods, and figuring out signs for variables are on-going question to answer, given the results, how IBOs are exposed/connected to other IBOs seems to affect inactive behavior based on the relative importance of the variables.

2.6.3.3 Approach Three – Duration Prediction

The third approach is to detect every inactive period equal to or longer than six consecutive months. IBOs can go through multiple inactive periods. This approach differs from the second one in how I treat such return to activity. In the second approach, if a return is observed after the first inactive period, the first inactive period is categorized as active, until the very last inactive period is observed, showing no evidence of a return to the business.

[13] Based on the H2O Java code, the relative importance is computed based on the weights from the first two hidden layers possibly for the sake of simplicity and speed. Thus, if the model contains more than two hidden layers, it is recommended to compute the relative importance manually rather than relying on H2O.

However, in the third approach, each inactive period is treated as a separate occasion. Thus, for each IBO, there may be multiple starting points of inactive periods. Similar to the second approach, the dependent variable is coded as the number of quarters left until the start of inactivity (Y_{q2} in Figure 2.10).

Unlike approach two, one additional category is added on top of quarters and active category, which indicates the time when an IBO is going through the inactive period. In Figure 2.10, such category is identified as zero.

Similar to approach two, the data was again randomly divided into a training set (70%), a validation set (15%), and a test set (15%). In order to keep the consistency, the random seed was remained unchanged. Ten-cross validation is applied with two hidden layers with 100 and 30 neurons, respectively. For the activation function, hyperbolic tangent is used. The classification results are shown in Table 2.19. Prediction from the third approach is 81.4%, slightly higher than that in the second approach (79.9%). Similar to the second approach, the model is effective in selecting the observations (IBO at a specific month) within the inactive period, or the observations that are not followed by inactive period (active); however, the duration prediction is often confused with the active category.

A similar conclusion can be drawn from the relative variable importance. Based on Table 2.20, the role of high status is more prominent: the percentage of high status within a family, and in proximity both appeared to have relatively high importance. In addition, the percentage of high status IBOs within the same family appeared in both approach two and three, with a fairly high rank. Although a argument for causality may be hard to make, there seems to be consistency in what was found in thee survival analysis, namely, that the IBOs' inactivity behavior is related to the surrounding others.

2.7 Conclusion and Limitations

Through the county-level and individual-level descriptive analysis, I examined how different dimensions of others are related to IBOs' inactive behavior. The three dimensions are specific individuals (who are important in the MLM business), network family (a motivational organization an IBO belongs to), and proximity (a geographic boundary). For each dimension, I focused on the role of high status IBOs in explaining inactivity. In particular, network family and status are unique aspects of the MLM industry, where both are supposed to invigorate IBOs' passion in the business. A network family is usually set up by successful IBOs who meet the qualification criteria to create and sell business support materials, hold meetings and educational seminars, to support educate and motivate the IBOs (Amway 2015). Because MLM firms have no control over these IBOs, the IBO's status is utilized to infuse additional stimuli to keep IBOs engaged in the business. High status not only comes with a financial benefit but also social recognition. High status IBOs sell products and recruit others, just as no-status IBOs. Thus, they differ from the high level managers of traditional organization, who have different roles, with stronger vertical authority. The high status IBOs, however, face additional responsibility to gather a subgroup of IBOs.

Using both descriptive county-level analysis regression and individual-level survival analysis, I found consistent evidence that being near high status IBOs decreased the likelihood of inactive behavior — a protective effect. The results held at the network family level (percentage of high status within a family) and within a given proximity (percentage of high status from the same family within 30 miles), although the distance from a high status IBO did not necessarily decrease inactivity behavior. As for individuals, staying near a sponsoring upline showed a decrease in inactivity hazard. In addition, IBOs associated with a larger network family tended to stay active in the business, potentially by fulfilling the need for social connection (Bhattacharya and Mehta 2000, DSA 2015c). Or, possibly a large network family may have a larger number of successful IBOs, easily developing improved training programs and more effective sales support activities (Jolson et al. 1987). Yet geographic competition was inevitable within a certain boundary. One difference in the competition effect in MLM is that only IBOs associated with the same network family increased the inactive hazard (percentage of own family within 30 miles). Moreover, I found that when the market has a high penetration rate, IBOs are less likely to face the inactivity hazard. In terms of the relative importance across the different dimensions in explaining inactivity, aside from individual business characteristics, which are known to highly predict turnover, individual level others had the strongest explanatory power, followed by network family, and then proximity.

Following the county and individual level analysis, I redirected the problem from an understanding of the inactive phenomenon to the prediction of inactivity (predicting inactive period incident and duration). I applied one machine learning technique, a neural network, to examine its applicability to this type of problem. The model did a decent job in terms of prediction, resulting in roughly 81% accuracy; in particular, given a time point, it distinguished well which IBOs would stay active (approach two and three), or would fall during inactive period (approach three). Based on approach one and two, the model was also sufficient for predicting the inactive period occurring within a quarter. However, predictions of when IBOs would experience an inactivity period in the distant future were inconclusive, in particular, for inactive periods occurring after a quarter (approach three). The relative importance across more than two-hundred variables showed that the variables related to the existence of others ranked quite high, even in the neural network models. In particular, the percentage of high status within a family consistently appeared as one of the top five variables explaining future inactivity behavior.

This study contributes to the turnover literature of the MLM industry. Some of the industry-specific characteristics of MLM do not exist or do not align with traditional hierarchical firms. Findings such as status or network family having a protective effect against IBOs' inactivity can perhaps be used as a building block for the MLM firm's strategy and can shed light on how the MLM industry could preserve their workforce. Even if some research touches upon MLM turnover (e.g.,Jolson et al. (1987), Wotruba et al. (2005)), to the best of my knowledge, this paper is the first attempt to understand an IBO's behavior through the lense of intricate relationships with other IBOs, such as distance from important individuals, connection in a network family, or exposure to other IBOs near by. In addition, methodologically, this study fills a gap in the MLM turnover literature. Previous studies have examined turnover behavior through in-person surveys, relying on a qualitative approach. The quantitative analysis that has been provided in this study, however, allows us to further exploit some of the factors mentioned above, such as important individuals, network family, and proximity, which cannot be directly addressed through surveys. Based on the empirical data, our study can provide data-driven suggestions for better understanding turnover in MLMs. For example, despite the associated cost, MLM firms could consider relaxing the high status standard to increase the number of high status IBOs, which would managerially motivate existing IBOs to remain active. Of course, the firm may need to consider the decrease in the value of scarcity and symbolic meaning of high status caused by the increase of high status IBOs. Alternatively, MLM firms could take an initiative to manage the network families — providing ways for small families to collaborate with one another to share manpower, and give IBOs a sense of belonging to a big network family. In addition, although the explanatory power was low, potential competition among IBOs sharing the same family within a geographical region should be carefully scrutinized in order to prevent inactivity.

Some limitations in this research can be addressed in future research. First, the model used to run the individual analysis is a variant of the Cox model with time-varying covariates, which perhaps is not as robust compared to other survival models relying on distributional assumptions. Despite the drawbacks of using this model, I find the interesting evidence for influence of others. Moving further, expanding the model using a parametric hazard model (Gupta 1991, Nikolaeva et al. 2009) can potentially enrich the research. Extending the survival model with a parametric baseline hazard, may allow me to directly compare the time to inactivity from both methods. Second, the results are based on how the inactivity was defined. Perhaps defining it differently might result in different implications. Trying different ways to see if the results are robust will be helpful for generalizing the results. Third, in defining proximity, 30-mile radius was used as the boundary. However, other potential boundaries should be tested for future robustness check. Potentially, referring to how other industries define the sales territory could help set the boundary for proximity. Lastly, moving forward, finding a better way to predict the exact time of upcoming inactivity behavior will be a question that remains to be answered.

2.8 Tables

Table 2.2: Individual-level Correlation Table a.

		Upline dist.	High status dist.	Family size	# High status	% High status
Individual	Upline distance	1.000	0.348	-0.031	-0.031	-0.016
	High status distance	0.348	1.000	0.019	0.009	-0.005
Family	Family size	-0.031	0.019	1.000	0.907	-0.089
	# High status	-0.031	0.009	0.907	1.000	0.280
	% High status	-0.016	-0.005	-0.089	0.280	1.000
Proximity	% Own family	-0.148	-0.187	0.207	0.207	0.064
	# Total IBO	-0.040	-0.035	0.210	0.131	-0.159
	% High status	-0.055	-0.131	0.034	0.091	0.154
	# High status	-0.070	-0.123	0.251	0.262	0.052
Market	Penetration	-0.046	-0.023	0.087	-0.080	-0.367

Table 2.3: Individual-level Correlation Table b.

		% Own family	# Total IBO	% High status	# High status	Penetration
Individual	Upline	-0.148	-0.040	-0.055	-0.070	-0.046
	High status	-0.187	-0.035	-0.131	-0.123	-0.023
Family	Family size	0.207	0.210	0.034	0.251	0.087
	# High status	0.207	0.131	0.091	0.262	-0.080
	% High status	0.064	-0.159	0.154	0.052	-0.367
Proximity	% Own family	1.000	-0.046	0.320	0.268	0.228
	# Total IBO	-0.046	1.000	-0.004	0.564	0.249
	% High status	0.320	-0.004	1.000	0.634	0.131
	# High status	0.268	0.564	0.634	1.000	0.213
Market	Penetration	0.228	0.249	0.131	0.213	1.000

Table 2.4: Individual-level Descriptive Statistics

		Min.	1st Q.	Median	Mean	3rd Q.	Max.
Individual	Upline distance	0	4.142	13.638	149.891	54.745	5747.027
	High status distance	0	16.09	80.06	346.12	393.02	5091.19
Family	Family size	248	20312	37616	42850	61199	100724
	# High status	0	0	3	12.95	11	190
	% High status	0%	0.691%	0.799%	0.865%	0.987%	2.058%
Proximity	# Own family	1	122	405	1342	1310	15450
	% Own family	0.008%	8.783%	19.451%	23.136%	33.884%	89.189%
	# Total IBO	5	854	2770	6091	7273	44791
	% High status	0%	0%	0.094%	0.215%	0.272%	13.864%
	# High status	0	0	3	12.95	11	190
Market	Penetration	0.000	0.003	0.004	0.004	0.005	0.024

Table 2.5: County-level Descriptive Statistics

County level	Min.	1st Q	Median	Mean	3rd Q	Max.
Total # of IBOs	1	349	1297	4380	4177	34580
High status IBOs ratio	0	0.003	0.006	0.006	0.008	0.250
# High status IBOs	0	2	9	33.7	36	340
# of Own Family IBOs	1	67	259	1130	1002	9974
# of Other Family IBOs	0	324	1126	3968	3877	34578
# of Own Family IBOs/ Total #	0	0.095	0.220	0.251	0.358	1.000
# of Other Family IBOs/ Total #	0	0.642	0.780	0.749	0.905	1.000
Penetration rate	0	0.003	0.005	0.005	0.007	0.024
LaborForce	300	104535	363266	826438	902047	4864160
Closest distance from competitor's HQ	6.5	101.7	178.6	247.2	315.6	2946.0
Unemployment rate	1.5	3.8	4.5	4.665	5.2	18.1

Table 2.6: County-level Correlation Table

	% Active IBO	# High status	% High status	# Total IBO	Unemployment	Penetation	Mean Personal	Mean Bonus	Mean earning
% Active IBO	1.000	-0.005	0.058	-0.039	0.000	-0.009	0.011	0.097	0.071
# High status	-0.005	1.000	0.184	0.690	-0.029	0.111	0.069	0.169	0.149
% High status	0.058	0.184	1.000	0.060	-0.033	0.042	0.159	0.731	0.584
# Total IBO	-0.039	0.690	0.060	1.000	-0.034	0.136	0.059	0.074	0.075
Unemployment	0.000	-0.029	-0.033	-0.034	1.000	-0.150	-0.089	-0.067	-0.084
Penetation	-0.009	0.111	0.042	0.136	-0.150	1.000	0.088	0.109	0.105
Mean Personal	0.011	0.069	0.159	0.059	-0.089	0.088	1.000	0.453	0.363
Mean Bonus	0.097	0.169	0.731	0.074	-0.067	0.109	0.453	1.000	0.749
Mean earning	0.071	0.149	0.584	0.075	-0.084	0.105	0.363	0.749	1.000

Table 2.7: County-level Results (a)

	Estimate	Std. Error	t-value	P-value
(Intercept)	0.868	0.025	34.439	0.000
lag.Total IBO	-0.00001	0.000	-5.405	0.000
lag.High status %	0.166	0.071	2.344	0.019
lag.Penetration rate	4.890	0.488	10.017	0.000
lag.UnempRate	0.001	0.001	0.972	0.331
lag.meanPersonal$	0.000003	0.000	2.104	0.035
Competitor shortest distance	0.00027	0.000	3.583	0.000
Year dummy		Y		
Month dummy		Y		
County dummy		Y		
Adjusted R-squared		0.773		

Table 2.8: County-level Results (b)

	Estimate	Std. Error	t-value	P-value
(Intercept)	0.986	0.014	70.277	0.000
lag.Total IBO	-0.00001	0.000	-5.405	0.000
lag.High status %	0.166	0.071	2.344	0.019
lag.Penetration rate	4.890	0.488	10.017	0.000
lag.UnempRate	0.001	0.001	0.972	0.331
lag.meanPersonal$	0.000003	0.000	2.104	0.035
# of Competitor	-0.069	0.019	-3.583	0.000
Year dummy		Y		
Month dummy		Y		
County dummy		Y		
Adjusted R-squared		0.773		

Table 2.9: Individual-level Analysis Results with Time Independent Covariates

		mean X		first X		last X	
Variable type	X	coef	se(coef)	coef	se(coef)	coef	se(coef)
Family	Family size/1000	-0.156***	0.001	0.006***	0.002	-0.137***	0.001
	High status %	8.226***	0.072	-0.087	0.073	5.016***	0.062
Proximity	Own family ratio	0.413***	0.040	0.236***	0.039	0.473***	0.040
	Total IBO/1000	-0.007***	0.001	0.002	0.001	-0.005***	0.001
	High status %	0.027	0.017	-0.019	0.015	-0.002	0.019
Individuals	Upline_distance	-0.0004***	0.000	-0.0005***	0.000	-0.0004***	0.000
	High status_distance	0.0002***	0.000	0.0004***	0.000	0.0003***	0.000
Market	TotalPenetCounty %	-2.386***	0.049	-0.259***	0.048	-2.648***	0.045
	UnempRate	-0.027***	0.005	0.011**	0.005	-0.025***	0.005
Competitor	shortestdistcomp	-0.001***	0.000	-0.0002***	0.000	-0.001***	0.000
Individual business	High status	-0.007	0.225	-0.017	2.552	-0.002	0.115
	Earning	0.0002***	0.000	-0.0003***	0.000	0.0002***	0.000
	relative level (by LOA)	0.002	0.041	-0.673***	0.037	-0.220***	0.042
	# Frontline	-0.088***	0.014	-0.317***	0.018	-0.084***	0.011
	# Total downline	-0.325***	0.006	-0.317***	0.009	-0.257***	0.005
	# Retail customer	-0.131***	0.005	-0.270***	0.006	-0.065***	0.004
Socio-demographic	Age	Y		Y		Y	
	Language (4 groups)	Y		Y		Y	
	Marital status (3 groups)	Y		Y		Y	
	Gender (3 groups)	Y		Y		Y	
	Ethnicity (8 groups)	Y		Y		Y	
	Income level (10 groups)	Y		Y		Y	
Other controls	Entry dummy	Y		Y		Y	
	State dummy	Y		Y		Y	
	LOA dummy	Y		Y		Y	
	Log likelihood	-557163.8		-624098.8		-544212	

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 2.10: Individual-level Survival Analysis Results

Variable type	Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Network Family	Family size/1000							-0.039***
	High status %							-2.540***
Proximity	Own family ratio						0.315***	0.262***
	Total IBO/1000						-0.0008	0.001
	High status %						-0.0511***	-0.0442***
Individual	Upline_distance					-0.00006***	-0.00005***	-0.00005***
	High status_distance					0.00002**	0.00003**	0.00003***
Market	TotalPenetCounty %				-0.238***	-0.236***	-0.269***	-0.202***
	UnempRate				0.014***	0.014***	0.011**	0.012***
Competitor	shortestdistcomp				-0.0002**	-0.0002**	-0.0003***	-0.0002**
Individual business characteristics	High status			-0.015	-0.015	-0.015	-0.015	-0.016
	Earning			-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	relative level (by LOA)			-0.453***	-0.447***	-0.442***	-0.463***	-0.512***
	# Frontline			-0.262***	-0.261***	-0.261***	-0.261***	-0.257***
	# Total downline			-0.235***	-0.235***	-0.235***	-0.235***	-0.238***
	# Retail customer			-0.118***	-0.118***	-0.118***	-0.119***	-0.122***
Socio-demographic characteristics	Age		Y	Y	Y	Y	Y	Y
	Language (4 groups)		Y	Y	Y	Y	Y	Y
	Marital status (3 groups)		Y	Y	Y	Y	Y	Y
	Gender (3 groups)		Y	Y	Y	Y	Y	Y
	Ethnicity (8 groups)		Y	Y	Y	Y	Y	Y
	Income level (10 groups)		Y	Y	Y	Y	Y	Y
Other controls	Entry dummy	Y	Y	Y	Y	Y	Y	Y
	Year dummy	Y	Y	Y	Y	Y	Y	Y
	Month dummy	Y	Y	Y	Y	Y	Y	Y
	State dummy	Y	Y	Y	Y	Y	Y	Y
	LOA dummy	Y	Y	Y	Y	Y	Y	Y
Log-Likelihood		-625211.1	-624571.4	-615207.3	-615375.7	-611551.9	-611516.4	-610849.7
N Obs.		1996546	1996546	1996453	1995957	1981810	1981810	1981810

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 2.11: Individual-level Survival Analysis Results - Model (7)

Variable types	Variables	coef	exp(coef)	se(coef)	z
Network family	Family size/1000	-0.039***	0.962	0.001	-27.413
	High status %	-2.540***	0.079	0.080	-31.674
Proximity (30 mi)	Own family ratio	0.262***	1.299	0.038	6.805
	Total IBO/1000	0.001	1.001	0.001	1.159
	High status %	-0.0442***	0.957	0.017	-2.539
Individual	Upline distance	-0.00005***	1.000	0.000	-4.362
	High status distance	0.00003***	1.000	0.000	3.632
Market	County penetration rate %	-0.202***	0.817	0.043	-4.699
	UnempRate	0.012***	1.012	0.005	2.654
Competitor	shortestdistcomp	-0.0002**	1.000	0.000	-2.554
Individual business	High status	-0.016	0.985	1.467	-0.011
	Earning	-0.0001***	1.000	0.000	-3.385
	relative level (by LOA)	-0.512***	0.599	0.040	-12.748
	# Frontline	-0.257***	0.773	0.012	-22.237
	# Total downline	-0.238***	0.788	0.005	-50.506
	# Retail customer	-0.122***	0.885	0.004	-31.6

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 2.12: Information Gain by Variable Types

Information types	Δ AIC
Socio-Demographic characteristics ((2) - (1))	-1231
Individual business characteristics ((3) - (2))	-17556
Market and competitor ((4)-(3))	-818
Individual ((5)-(4))	-7643
Proximity ((6)-(5))	-65
Network Family ((7)-(6))	-1330

Table 2.13: Variables for Machine Learning

IV	
Network Family	Family size, lag family size, network growth from (t-1) High status %, # Min, max level, Depth of a family (min-max) Own family %, #
Proximity	Total IBO High status from own family %, # Total %, # of high status Network Family count % of own high status among # of high status in proximity
Individual	Distance from a high status IBO & Upline
Market (County)	TotalPenetCounty %, Labor force, UnempRate, # employed Diversity across the families withint a county (2) TotalIBOcounty
Competitor	The shortest distance from competing firms' headquarters # of competitors' headquarters by each band (100mi - 500mi) # of competitors' headquarters within 300mi, 500mi
Individual business	High status duration from entry (running variable) financial report (t): Total Earnings, Product sales(Point, $), Bonus(Point, $) lag financial report (t-1) Cumulative average of financial report until (t-1) Missed value in finanical (t-1) Activation of MLM business (30 day, 90day, 120 day) relative level (by LOA), absolute level # Frontline, # Total downline, # Retail customer Addition/removal of # Frontline, # Total downline, # Retail customer from (t-1) Cumulative addition/removal since the beginning Approximated bonus bracket at (t) Bonus bracket attainment (first 3%, 9%, 18%, 25%) Month since the bonus bracket attainment
Socio-demographic	Age Language Marital status Gender Ethnicity Income level
Recency, Frequency	Total no engagement months until t Total engagement months until t Continuous engagement # until t # of engagement after the first inactive duration, if applicable Cumulative average of no engagement maximum rest until t after the first inactive duration (dummy, count), if applicable
Other controls	Entry dummy (cohort) Year Month State Family

Table 2.14: Model Comparison between NN and Logistic Model

	Neural Network	Logistic
Accuracy	0.859	0.839
Recall	0.844	0.843
Precision	0.738	0.699

Table 2.15: Contingency Table for Approach One (Out of Sample Prediction)

		Prediction				
		0	1	Row sum	% Data	Row correct %
Data	0	119736 (59.7%)	18610 (23.02%)	138346	69%	87%
	1	9717 (4.8%)	52510 (26.2%)	62227	31%	84%
Colum sum		129453	71120	200573		

Table 2.16: Approach Two Data Description (Full Data)

Total	Stay	Q1	Q2	Q3	Q4	Q5
2482116	2264231	134059	49774	21912	9242	2898
	0.91	0.05	0.020	0.009	0.004	0.001

Table 2.17: Contingency Table for Approach Two (Out of Sample Prediction)

		Prediction								
		Active	Q1	Q2	Q3	Q4	Q5	Row sum	Data %	Row correct %
Data	Active	**108818**	16013	215	1068	358	37	126509	62.66%	86.0%
	Q1	14217	**51853**	93	302	67	8	66540	32.96%	77.9%
	Q2	3549	2015	**89**	126	10	4	5793	2.87%	1.5%
	Q3	1409	390	14	**363**	27	0	2203	1.09%	16.5%
	Q4	440	63	2	53	**164**	4	726	0.36%	22.6%
	Q5	70	7	2	1	14	**32**	126	0.06%	25.4%
	Column sum	128503	70341	415	1913	640	85	201897	100%	

Table 2.18: Variable Importance Based on Approach Two

	Variable	Relative importance
1	Year 2007	1.00
2	**High status % within a family**	0.98
3	Activation of MLM (30 days)	0.95
4	Entry cohort effect (200603)	0.94
5	Age	0.92
6	Income group 6	0.88
7	# Frontline	0.88
8	Entry cohort effect (200604)	0.86
9	Income group 1 (the Lowest)	0.86
10	Income group 9 (the Highest)	0.85
11	**# of family within proximity**	0.84
12	Income group 8	0.83
13	**Network family dummy (37)**	0.83
14	Entry cohort effect (200601)	0.82
15	**Distance from a high status IBO**	0.82
16	Gender missing	0.82
17	Entry cohort effect (200602)	0.81
18	Income group 7	0.81
19	Level	0.81
20	Cumulative average of no engagement	0.80
21	# of competitors' 200mi	0.79
22	**Distance from an Upline**	0.78
23	State fixed effect (Georgia)	0.77
24	Income group 4	0.76
25	Ethnicity missing	0.76

Table 2.19: Contingency Table for Approach Three (Out of Sample Prediction)

		Prediction									
		Inactive	Q1	Q2	Q3	Q4	Q5	Active	Row sum	Data %	Row correct %
	Inactive	**52931**	0	0	0	0	0	7890	60821	30.12%	87.03%
	Q1	1	**1383**	59	16	2	0	11040	12501	6.19%	11.06%
	Q2	0	80	**52**	14	1	0	1426	1573	0.78%	3.31%
Data	Q3	0	9	9	**34**	1	0	333	386	0.19%	8.81%
	Q4	0	6	0	13	**6**	0	70	95	0.05%	6.32%
	Q5	0	1	0	1	3	**1**	6	12	0.01%	8.33%
	Active	15001	1104	215	146	24	0	**110019**	126509	62.66%	86.97%
	Column sum	67933	2583	335	224	37	1	130784	201897	100%	

Table 2.20: Variable Importance Based on Approach Three

		Relative importance
1	Engagement	1.00
2	Bonus bracket	0.54
3	Age	0.50
4	# Frontline	0.50
5	**High status % within a family**	0.48
6	State fixed effect (FL)	0.48
7	Income group 4	0.47
8	Entry cohort effect (200607)	0.46
9	Diff (#Downline IBO)	0.46
10	Ethnicity missing	0.45
11	Entry cohort effect (200604)	0.45
12	**Network family dummy (2)**	0.45
13	Income group 6	0.45
14	Activation of MLM (30 days)	0.45
15	Income group 3	0.45
16	State fixed effect (IL)	0.44
17	Entry cohort effect (200603)	0.44
18	Level	0.44
19	**Network family dummy (37)**	0.44
20	# of competitors' 200mi	0.43
21	Entry cohort effect (200601)	0.43
22	**High status % within a family in proximity**	0.43
23	**Network family dummy (21)**	0.43
24	Engagement after inactivity	0.42
25	State fixed effect (NC)	0.42

2.9 Figures

Figure 2.1: Number of Months in the Business Conditional on Being Inactive

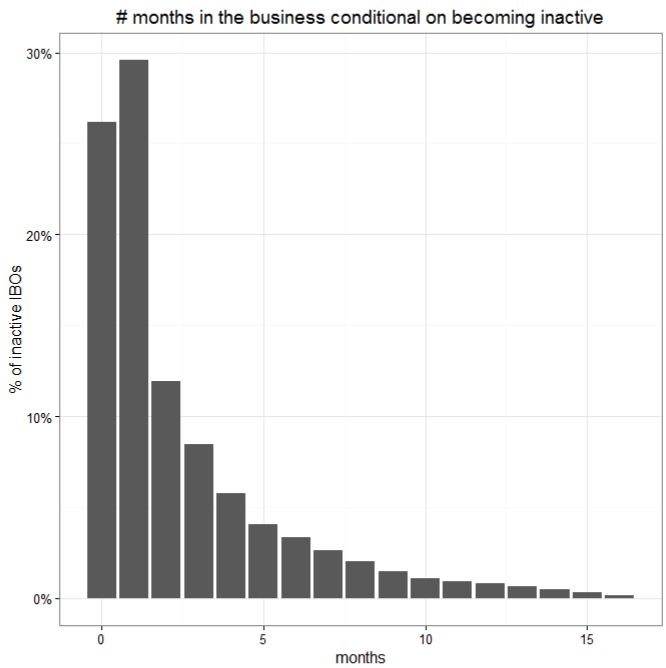


Figure 2.2: Network Family Size - Oct. 2007

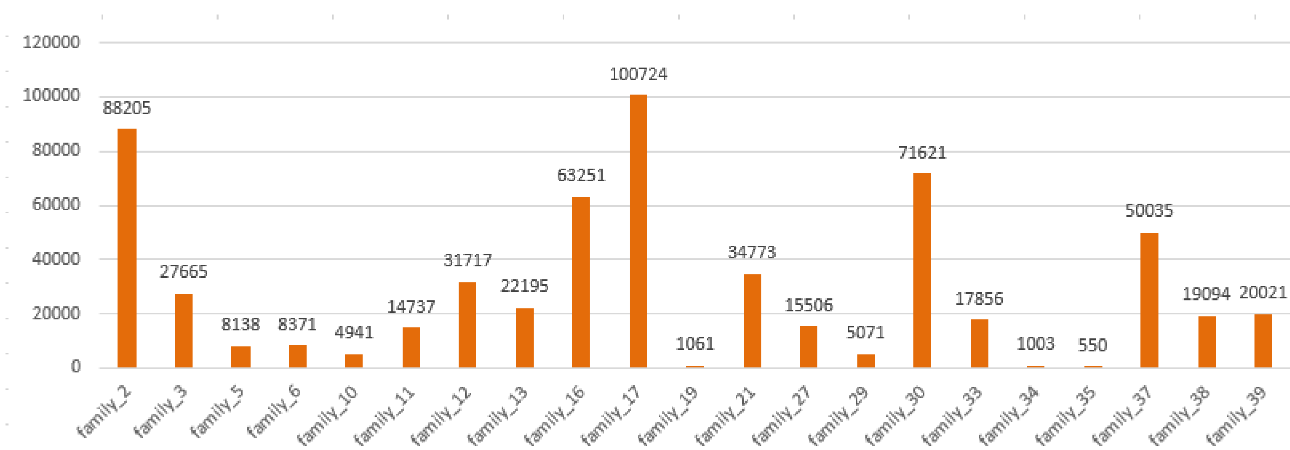


Figure 2.3: Percent of High Status Members Given a Network Family - Oct. 2007

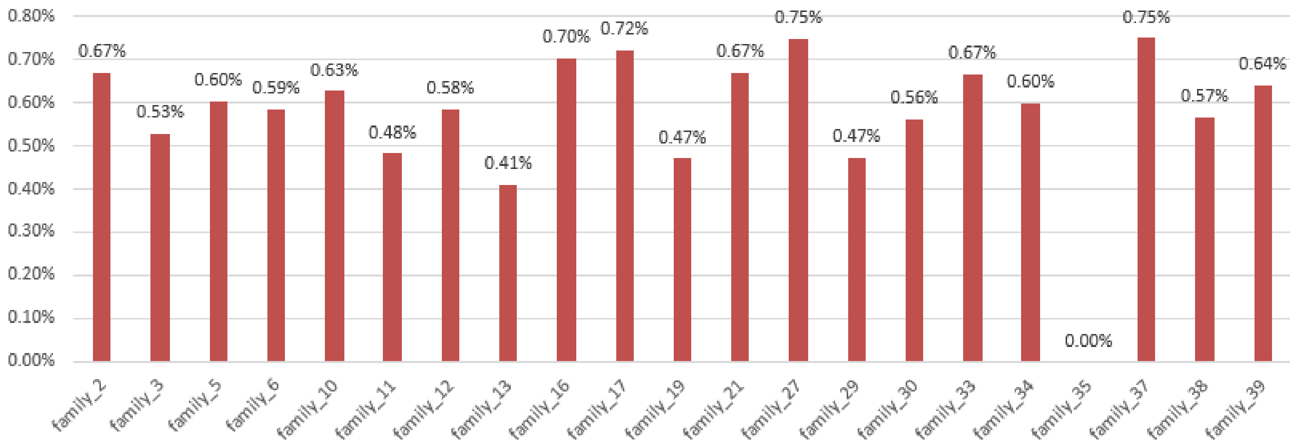


Figure 2.4: Family Distribution Examples

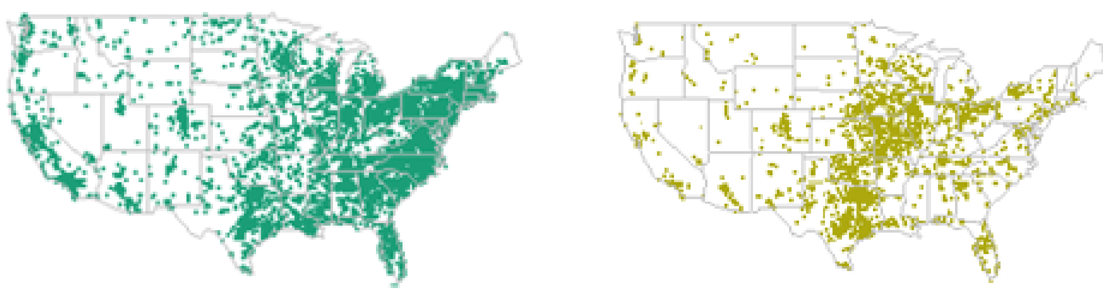


Figure 2.5: Example of 30-mile Radius

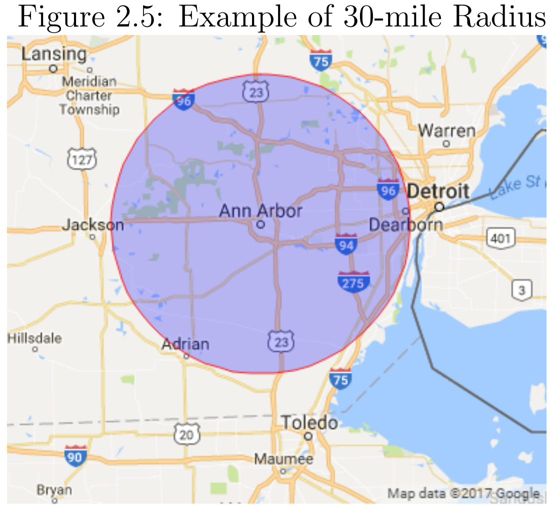


Figure 2.6: Example of the Lowest and the Highest Total IBOs in Proximity

Figure 2.7: Histogram of Minimum Distance from Competitors' Headquarters

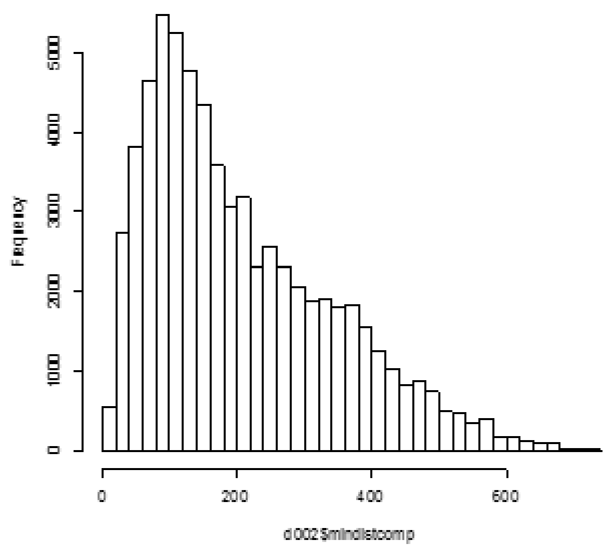


Figure 2.8: Example of Neural Network with Three Layers

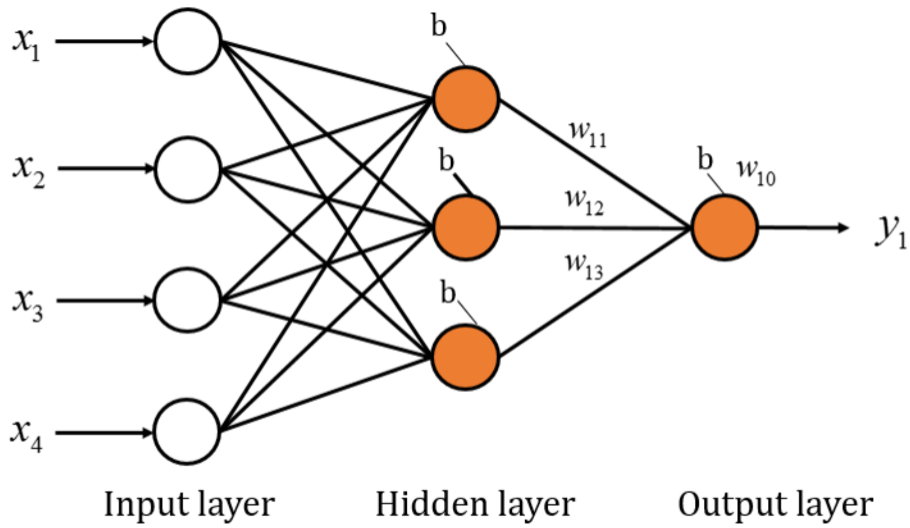


Figure 2.9: Example of Approach One

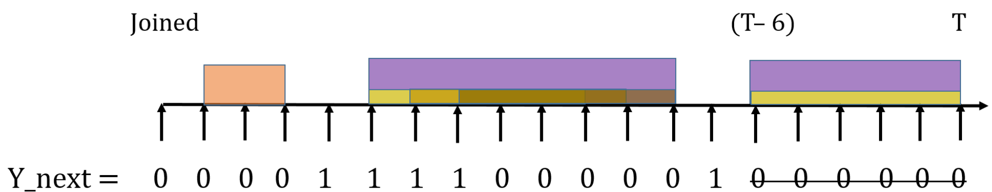
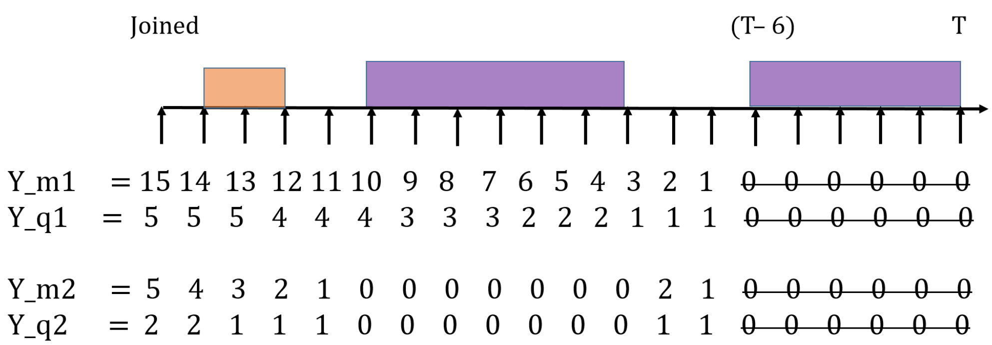


Figure 2.10: Example of Approach Two and Three



Bibliography

2017. Direct sales industry statistics. URL `https://www.directsalesaid.com/articles/industry-statistics`.

Agarwal, Nikhil. 2015. An empirical model of the medical match. *The American Economic Review* **105**(7) 1939–1978.

Agarwal, Nikhil, William Diamond. 2014. Identification and estimation in two-sided matching markets. .

Akkus, Oktay, J Anthony Cookson, Ali Hortacsu. 2014. Endogenous matching, underwriter reputation, and the underpricing of initial public offerings. *Underwriter Reputation, and the Underpricing of Initial Public Offerings (September 18, 2014)* .

Albaum, Gerald, Robert A Peterson. 2011. Multilevel (network) marketing: An objective view. *The Marketing Review* **11**(4) 347–361.

Albert, Réka, Albert-László Barabási. 2002. Statistical mechanics of complex networks. *Reviews of modern physics* **74**(1) 47.

Albuquerque, Paulo, Bart J Bronnenberg, Charles J Corbett. 2007. A spatiotemporal analysis of the global diffusion of iso 9000 and iso 14000 certification. *Management science* **53**(3) 451–468.

Allen, David G. 2006. Do organizational socialization tactics influence newcomer embeddedness and turnover? *Journal of management* **32**(2) 237–256.

Allen, Scott. 2016. Too good to be true? - 6 questions to check out an mlm/network marketing opportunity. URL `http://entrepreneurs.about.com/cs/multilevelmktg/a/toogoodtobetrue.htm`. [about.com; Updated 22-March-2016].

Amway. 2015. Amway business reference guide 2015. URL `https://www.amway.com/en/ResourceCenterDocuments/Visitor/ops-amw-gde-v-en--BusinessReferenceGuide.pdf`.

Ansari, Asim, Oded Koenigsberg, Florian Stahl. 2011. Modeling multiple relationships in social networks. *Journal of Marketing Research* **48**(4) 713–728.

Bajari, Patrick, Denis Nekipelov, Stephen P Ryan, Miaoyu Yang. 2015a. Demand estimation with machine learning and model combination. Tech. rep., National Bureau of Economic Research.

Bajari, Patrick, Denis Nekipelov, Stephen P Ryan, Miaoyu Yang. 2015b. Machine learning methods for demand estimation. *The American Economic Review* **105**(5) 481–485.

Bandura, Albert. 1962. Social learning through imitation. .

Bell, David R, Sangyoung Song. 2007. Neighborhood effects and trial on the internet: Evidence from online grocery retailing. *Quantitative Marketing and Economics* **5**(4) 361–400.

Bhattacharya, Patralekha, Krishna Kumar Mehta. 2000. Socialization in network marketing organizations: is it cult behavior? *The Journal of Socio-Economics* **29**(4) 361–374.

Biggart, Nicole Woolsey. 1989. *Charismatic capitalism: Direct selling organizations in America*. University of Chicago Press.

Bloch, Brian. 1996. Multilevel marketing: what's the catch? *Journal of Consumer Marketing* **13**(4) 18–26.

Bonabeau, Eric. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences* **99**(suppl 3) 7280–7287.

Bonoma, Thomas. 1991. This snake rises in bad times. *Marketing News* **25**(4) 16.

Bosley, Stacie, Kim K McKeage. 2015. Multilevel marketing diffusion and the risk of pyramid scheme activity: The case of fortune hi-tech marketing in montana. *Journal of Public Policy & Marketing* **34**(1) 84–102.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, James Wyckoff. 2013. Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics* **31**(1) 83–117.

Braun, Michael, David A. Schweidel. 2011. Modeling customer lifetimes with multiple causes of churn. *Marketing Science* **30**(5) 881–902. URL `http://www.jstor.org/stable/23011984`.

Brodie, Stewart, John Stanworth, Thomas R Wotruba. 2002. Comparisons of salespeople in multilevel vs. single level direct selling organizations. *The Journal of Personal Selling and Sales Management* 67–75.

Candel, Arno, Viraj Parmar, Erin LeDell, Anisha Arora. 2016. Deep learning with h2o.

Centola, Damon. 2010. The spread of behavior in an online social network experiment. *Science* **329**(5996) 1194–1197.

Chaplot, Devendra Singh, Eunhee Rhim, Jihie Kim. 2015. Predicting student attrition in moocs using sentiment analysis and neural networks. *AIED Workshops*.

Chen, Gilad, Robert E Ployhart, Helena Cooper Thomas, Neil Anderson, Paul D Bliese. 2011. The power of momentum: A new model of dynamic relationships between job satisfaction change and turnover intentions. *Academy of Management Journal* **54**(1) 159–181.

Chen, Jiawei. 2013. Estimation of the loan spread equation with endogenous bank-firm matching. *Structural Econometrics Models (E. Choo and M. Shum, eds.)* **31** 251–290.

Chen, Wei-Sen, Yin-Kuan Du. 2009. Using neural networks and data mining techniques for the financial distress prediction model. *Expert systems with applications* **36**(2) 4075–4086.

Choi, Jeonghye, Sam K Hui, David R Bell. 2010. Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer. *Journal of Marketing Research* **47**(1) 75–89.

Christakis, Nicholas A, James H Fowler, Guido W Imbens, Karthik Kalyanaraman. 2010. An empirical model for strategic network formation. Tech. rep., National Bureau of Economic Research.

Clark, Simon. 2006. The uniqueness of stable matchings. *Contributions in Theoretical Economics* **6**(1) 1–28.

Clothier, Peter J. 1997. *Multi-level Marketing: A Practical Guide to Successful Network Selling*. 3rd ed. Kogan page.

Coleman, Lynn G. 1989. Sales force turnover has managers wondering why. *Marketing News* **23**(25) 6–7.

Collamer, Nancy. 2013. Can you really make money in direct sales? URL `http://www.forbes.com/sites/nextavenue/2013/04/01/can-you-really-make-money-in-direct-sales/2/18246f154cf1`. [forbes.com; Updated 1-April-2013].

Cotton, John L, Jeffrey M Tuttle. 1986. Employee turnover: A meta-analysis and review with implications for research. *Academy of management Review* **11**(1) 55–70.

Coughlan, Anne T. 2012. 7 things you should know about direct selling. *Wordpress, http://distribunomics.wordpress.com/2012/11/21/7-things-you-should-know-about-direct-selling* .

Coughlan, Anne T, Kent Grayson. 1993. Multi-level marketing executives industry survey summary report. *Unpublished report, Northwestern University, Evanston* .

Coughlan, Anne T, Kent Grayson. 1998. Network marketing organizations: Compensation plans, retail network growth, and profitability. *International Journal of Research in Marketing* **15**(5) 401–426.

Coussement, Kristof, Dirk Van den Poel. 2009. Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications* **36**(3) 6127–6134.

Cox, David R. 1992. Regression models and life-tables. *Breakthroughs in statistics*. Springer, 527–541.

Crittenden, Victoria L, William F Crittenden. 2004. Developing the sales force, growing the business: The direct selling experience. *Business Horizons* **47**(5) 39–44.

Currarini, Sergio, Matthew O Jackson, Paolo Pin. 2009. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* **77**(4) 1003–1045.

Dalton, Dan R, William D Todor, Michael J Spendolini, Gordon J Fielding, Lyman W Porter. 1980. Organization structure and performance: A critical review. *Academy of management review* **5**(1) 49–64.

Daquis, John Carlo P, Angelique O Castañeda, Nelson D Sy, Ranier Joseph V Abgona. 2013. Profitability and growth topology analysis of unilevel-type of network marketing structures. *The Philippine Statistician* **62**(2) 13–29.

Darmon, René Y. 2008. The concept of salesperson replacement value: A sales force turnover management tool. *Journal of Personal Selling & Sales Management* **28**(3) 211–232.

DeConinck, James B, Julie T Johnson. 2009. The effects of perceived supervisor support, perceived organizational support, and organizational justice on turnover among salespeople. *Journal of Personal Selling & Sales Management* **29**(4) 333–350.

Deng, Wei-Jaw, Wen-Chin Chen, Wen Pei. 2008. Back-propagation neural network based importance–performance analysis for determining critical service attributes. *Expert Systems with Applications* **34**(2) 1115–1125.

DSA. 2015a. Direct selling in 2015: An overview. URL `http://www.dsa.org/docs/default-source/research/dsa_2015factsheetfinal.pdf?sfvrsn=8`.

DSA. 2015b. Dsa 2015 growth & outlook report: U.s. direct selling in 2014. URL `http://www.dsa.org/docs/default-source/research/researchfactsheet2007-2014.pdf?sfvrsn=0`.

DSA. 2015c. Dsa 2015 motivation - in direct selling, success is different for different people. URL `http://www.dsa.org/docs/default-source/advocacy/dsa-successisdifferentfactsheetv4.pdf?sfvrsn=2`.

DSN. 2016. 2016 dsn north america 50 list. URL `http://directsellingnews.com/index.php/view/2016_dsn_north_america_50_list1`.

Echenique, Federico, SangMok Lee, Matthew Shum. 2010. Aggregate matchings .

Echenique, Federico, SangMok Lee, Matthew Shum. 2013. Partial identification in two-sided matching models .

Fader, Peter S, Bruce GS Hardie, Ka Lok Lee. 2005. Rfm and clv: Using iso-value curves for customer base analysis. *Journal of Marketing Research* **42**(4) 415–430.

Failla, Don, Joe Hardwick. 1995. *How to Build a Large Successful Multi-level Marketing Organization*. Multi-Level Marketing International, Incorporated.

Fox, Jeremy T. 2008. Estimating matching games with transfers. Tech. rep., National Bureau of Economic Research.

Frenzen, Jonathan K, Harry L Davis. 1990. Purchasing behavior in embedded markets. *Journal of Consumer Research* **17**(1) 1–12.

Futrell, Charles M, A Parasuraman. 1984. The relationship of satisfaction and performance to salesforce turnover. *The Journal of Marketing* 33–40.

Gale, David, Lloyd S Shapley. 1962. College admissions and the stability of marriage. *American mathematical monthly* 9–15.

Gedeon, Tamás D. 1997. Data mining of inputs: analysing magnitude and functional measures. *International Journal of Neural Systems* **8**(02) 209–218.

Geweke, John. 1999. Using simulation methods for bayesian econometric models: inference, development, and communication. *Econometric Reviews* **18**(1) 1–73.

Geweke, John, Gautam Gowrisankaran, Robert J Town. 2003. Bayesian inference for hospital quality in a selection model. *Econometrica* **71**(4) 1215–1238.

Goodman, Charles S, Marvin A Jolson. 1973. Overdue attention to a neglected step-child of marketing.

Grant, Kelli B. 2012. 10 things direct sales marketers won't say. *Wall Street Journal Market Watch* .

Grayson, Kent. 1996. Examining the embedded markets of network marketing organizations.

Grayson, Kent. 2007. Friendship versus business in marketing relationships. *Journal of Marketing* **71**(4) 121–139.

Griffeth, Rodger W, Peter W Hom. 2001. *Retaining valued employees*. Sage Publications.

Griffeth, Rodger W, Peter W Hom, Stefan Gaertner. 2000. A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of management* **26**(3) 463–488.

Gupta, Sunil. 1991. Stochastic models of interpurchase time with time-dependent covariates. *Journal of Marketing Research* 1–15.

Haenlein, Michael. 2013. Social interactions in customer churn decisions: The impact of relationship directionality. *International Journal of Research in Marketing* **30**(3) 236–248.

Harrison, David A, Katherine J Klein. 2007. What's the difference? diversity constructs as separation, variety, or disparity in organizations. *Academy of management review* **32**(4) 1199–1228.

Herballife. 2016. Form 10-k, herbalife ltd. URL `http://ir.herbalife.com/annuals.cfm`.

Hinkin, Timothy R, J Bruce Tracey. 2000. The cost of turnover: Putting a price on the learning curve. *The Cornell Hotel and Restaurant Administration Quarterly* **41**(3) 144–21.

Holtom, Brooks C, Terence R Mitchell, Thomas W Lee, Marion B Eberly. 2008. 5 turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *Academy of Management annals* **2**(1) 231–274.

Hom, Peter W, Anne S Tsui, Joshua B Wu, Thomas W Lee, Ann Yan Zhang, Ping Ping Fu, Lan Li. 2009. Explaining employment relationships with social exchange and job embeddedness. *Journal of Applied Psychology* **94**(2) 277.

Hsieh, Chih-Sheng, Lung-Fei Lee. 2012. A structural modeling approach for network formation and social interactions with applications to students' friendship choices and selectivity on activities. *Job market paper, Department of Economics, Ohio State University* .

Hu, Yansong, Christophe Van den Bulte. 2014. Nonmonotonic status effects in new product adoption. *Marketing Science* **33**(4) 509–533.

Hua, Zhongsheng, Yu Wang, Xiaoyan Xu, Bin Zhang, Liang Liang. 2007. Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications* **33**(2) 434–440.

Imai, Kosuke, David A van Dyk. 2005. A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of econometrics* **124**(2) 311–334.

Iyengar, Raghuram, Christophe Van den Bulte, Thomas W Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2) 195–212.

Jackofsky, Ellen F, Kenneth R Ferris, Betty G Breckenridge. 1986. Evidence for a curvilinear relationship between job performance and turnover. *Journal of Management* **12**(1) 105–111.

Jain, S, B Singla, S Shashi. 2015. Motivational factors in multilevel marketing business: A confirmatory approach. *Management Science Letters* **5**(10) 903–914.

Jamal, Zainab, Randolph E Bucklin. 2006. Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *Journal of Interactive Marketing* **20**(3-4) 16–29.

Jolson, Marvin A, Alan J Dubinsky, Rolph E Anderson. 1987. Correlates and determinants of sales force tenure: an exploratory study. *Journal of Personal Selling & Sales Management* **7**(3) 9–28.

Kaastra, Iebeling, Milton Boyd. 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* **10**(3) 215–236.

Keep, William W., Peter J. Vander Nat. 2014. Multilevel marketing and pyramid schemes in the united states: An historical analysis. *Journal of Historical Research in Marketing* **6**(2) 188–210.

King, C.W., J.W. Robinson. 2000. *The New Professionals: The Rise of Network Marketing as the Next Major Profession*. Prima Soho. URL `https://books.google.com/books?id=O6fWk-_FbrwC`.

Klebnikov, Paul. 1991. The power of positive inspiration. *Forbes* **9** 244–251.

Koroth, Abdul Assis. 2014. Antecedents of distributors turnover in multilevel marketing. *Indian Journal of Commerce and Management Studies* **5**(1) 62.

Ladik, Daniel M, Greg W Marshall, Felicia G Lassk, William C Moncrief. 2000. The relationship of satisfaction and performance to salesforce turnover: A replication and extension. *American Marketing Association. Conference Proceedings*, vol. 11. American Marketing Association, 232.

Lee, Kwee-Fah, Teck-Chai Lau, Kai-Yin Loi, et al. 2016. Driving distributors' satisfaction in multilevel marketing (mlm) companies. *International Journal of Academic Research in Business and Social Sciences* **6**(2) 105–122.

Lee, SangMok. 2014. Incentive compatibility of large centralized matching markets .

Lee, Thomas W, Terence R Mitchell, Brooks C Holtom, Linda S McDaneil, John W Hill. 1999. The unfolding model of voluntary turnover: A replication and extension. *Academy of Management journal* **42**(4) 450–462.

Legara, Erika Fille, Anthony Longjas, Rene Batac. 2009. Competition in a social structure. *International Journal of Modern Physics C* **20**(01) 1–7.

Legara, Erika Fille, Christopher Monterola, Dranreb Earl Juanico, Marisciel Litong-Palima, Caesar Saloma. 2008. Earning potential in multilevel marketing enterprises. *Physica A: Statistical Mechanics and its Applications* **387**(19) 4889–4895.

Lemmens, Aurélie, Christophe Croux. 2006. Bagging and boosting classification trees to predict churn. *Journal of Marketing Research* **43**(2) 276–286.

Lilyquist, Mindy. 2016a. 10 MLM success tips for home business success. URL `http://homebusiness.about.com/od/homebusinessprofiles/a/10-Mlm-Success-Tips.htm`. About.com; Updated 05-April-2016.

Lilyquist, Mindy. 2016b. MLM success starter guide - tips, tools and resources to help you build your mlm business. URL `http://homebusiness.about.com/od/homebusinessprofiles/tp/Mlm-Success-Starter-Guide.htm`. About.com; Updated 16-May-2016.

Logan, John Allen, Peter D Hoff, Michael A Newton. 2008. Two-sided estimation of mate preferences for similarities in age, education, and religion. *Journal of the American Statistical Association* **103**(482) 559–569.

Maciel, Leandro Santos, Rosângela Ballini. 2008. Design a neural network for time series financial forecasting: Accuracy and robustness analysis. *Instituto de Economía, Universidade Estadual de Campinas, Sao Paulo-Brasil* .

McCulloch, Robert, Peter E Rossi. 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* **64**(1) 207–240.

McCulloch, Robert E, Nicholas G Polson, Peter E Rossi. 2000. A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics* **99**(1) 173–193.

McPherson, J Miller, Pamela A Popielarz, Sonja Drobnic. 1992. Social networks and organizational dynamics. *American sociological review* 153–170.

Menguc, Bulent, Sang-Lin Han, Seigyoung Auh. 2007. A test of a model of new salespeople's socialization and adjustment in a collectivist culture. *Journal of Personal Selling & Sales Management* **27**(2) 149–167.

Mitchell, Terence R, Brooks C Holtom, Thomas W Lee, Chris J Sablynski, Miriam Erez. 2001. Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of management journal* **44**(6) 1102–1121.

Mobley, William H. 1992. Employee turnover: Causes, consequences, and control .

Msweli, Pumela, Adrian Sargeant. 2001. Modelling distributor retention in network marketing organisations. *Marketing Intelligence & Planning* **19**(7) 507–514.

Msweli-Mbanga, P. 2004. Predicting turnover behaviour of direct salespeople. *Southern African Business Review* **8**(3) 14–25.

Nam, Sungjoon, Puneet Manchanda, Pradeep K Chintagunta. 2010. The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service. *Marketing Science* **29**(4) 690–700.

Narayan, Vishal, Vithala R Rao, Carolyne Saunders. 2011. How peer influence affects attribute preferences: A bayesian updating mechanism. *Marketing Science* **30**(2) 368–384.

Neslin, Scott A, Sunil Gupta, Wagner Kamakura, Junxiang Lu, Charlotte H Mason. 2006. Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of marketing research* **43**(2) 204–211.

Ni, Jian, Kannan Srinivasan. 2015. Matching in the sourcing market: A structural analysis of the upstream channel. *Marketing Science* **34**(5) 722–738.

Nikolaeva, Ralitza, Manohar U Kalwani, William T Robinson, S Sriram, et al. 2009. Survival determinants for online retailers. *Review of Marketing Science* **7**(1) 360.

Olden, Julian D, Michael K Joy, Russell G Death. 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* **178**(3) 389–397.

Pan, Yihui. 2014. The determinants and impact of executive-firm matches. *Available at SSRN 1571892* .

Pancras, Joseph, Srinivasaraghavan Sriram, Vineet Kumar. 2012. Empirical investigation of retail expansion and cannibalization in a dynamic environment. *Management Science* **58**(11) 2001–2018.

Park, Minjung. 2008. An empirical two-sided matching model of acquisitions: Understanding merger incentives and outcomes in the mutual fund industry. *Unpublished manuscript available at http://www. econ. umn. edu/˜ mpark/MergerPaper. pdf* .

Peter J. Vander Nat, William W. Keep. 2002. Marketing fraud: An approach for differentiating multilevel marketing from pyramid schemes. *Journal of Public Policy & Marketing* **21**(1) 139–151. URL http://www.jstor.org/stable/30000715.

Peterson, Robert A, Thomas R Wotruba. 1996. What is direct selling? definition, perspective, and research agenda. *Journal of Personal Selling & Sales Management* **16**(4) 1–16.

Pratt, Michael G. 2000. The good, the bad, and the ambivalent: Managing identification among amway distributors. *Administrative Science Quarterly* **45**(3) 456–493.

Price, James L. 1983. Reviewed work(s): Employee turnover: Causes, consequences, and control. *Journal of Marketing Channels* 506–507.

Raymond, Mary Anne, John F Tanner Jr. 1994. Selling and sales management in action: Maintaining customer relationships in direct sales: Stimulating repeat purchase behavior. *Journal of Personal Selling & Sales Management* **14**(4) 67–76.

Richardson, Robert. 1999. Measuring the impact of turnover on sales. *The Journal of Personal Selling and Sales Management* 53–66.

Rossi, Peter E, Robert E McCulloch, Greg M Allenby. 1996. The value of purchase history data in target marketing. *Marketing Science* **15**(4) 321–340.

Roth, Alvin E, Marilda Sotomayor. 1992. Chapter 16 twosided matching. *Handbook of Game Theory with Economic Applications, Elsevier, vol* **1** 485–541.

Seim, Katja. 2006. An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics* **37**(3) 619–640.

Sinha, Tanmay, Patrick Jermann, Nan Li, Pierre Dillenbourg. 2014. Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. *arXiv preprint arXiv:1407.7131* .

Sorensen, Morten. 2007. How smart is smart money? a two-sided matching model of venture capital. *The Journal of Finance* **62**(6) 2725–2762.

Sparks, John R, Joseph A Schenk. 2001. Explaining the effects of transformational leadership: an investigation of the effects of higher-order motives in multilevel marketing organizations. *Journal of Organizational Behavior* **22**(8) 849–869.

Sparks, John R, Joseph A Schenk. 2006. Socialization communication, organizational citizenship behaviors, and sales in a multilevel marketing organization. *Journal of Personal Selling & Sales Management* **26**(2) 161–180.

Taylor, Jon M. 2011. The case (for and) against multi-level marketing: The complete guide to understanding the flaws – and proving and countering the effects – of endless chain "opportunity" recruitment, or product-based pyramid schemes. URL `http://www.mlmwatch.org/01General/taylor.pdf`. Consumer Awareness Institute.

Therneau, Terry, Cynthia Crowson, Elizabeth Atkinson. 2017. Using time dependent covariates and time dependent coefficients in the cox model .

Tortora, Andrea. 2015. Direct selling's strength in the world's billion dollar markets. *Direct Selling News* .

Uetake, Kosuke, Yasutora Watanabe. 2012a. Entry by merger: Estimates from a two-sided matching model with externalities. *Available at SSRN 2188581* .

Uetake, Kosuke, Yasutora Watanabe. 2012b. A note on estimation of two-sided matching models. *Economics Letters* **116**(3) 535–537.

Van den Bulte, Christophe, Stefan Hendrik Katrin Wuyts. 2007. Social networks in marketing. *MSI Relevant Knowledge Series* .

Vellido, Alfredo, Paulo JG Lisboa, J Vaughan. 1999. Neural networks in business: a survey of applications (1992–1998). *Expert systems with applications* **17**(1) 51–70.

West, Patricia M, Patrick L Brockett, Linda L Golden. 1997. A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science* **16**(4) 370–391.

WFDSA. 2016. Global direct selling - 2015 retail sales URL `http://www.wfdsa.org/files/pdf/global-stats/SalesReport2015v25-31-2016.pdf`. World Federation of Direct Selling Associations.

Wotruba, Thomas, Stewart Brodie, John Stanworth. 2005. Differences in turnover predictors between multilevel and single level direct selling organizations. *The International Review of Retail, Distribution and Consumer Research* **15**(1) 91–110.

Wotruba, Thomas R. 1990a. Full-time vs. part-time salespeople: A comparison on job satisfaction, performance, and turnover in direct selling. *International Journal of Research in Marketing* **7**(2-3) 97–108.

Wotruba, Thomas R. 1990b. The relationship of job image, performance, and job satisfaction to inactivity-proneness of direct salespeople. *Journal of the Academy of Marketing Science* **18**(2) 113–121.

Wotruba, Thomas R, Pradeep K Tyagi. 1991. Met expectations and turnover in direct selling. *The Journal of Marketing* 24–35.

Wotruba, Thomas R, Pradeep K Tyagi. 1992. Motivation to become a direct salesperson and its relationship with work outcomes. *Journal of Marketing Channels* **2**(2) 41–56.

Wu, Desheng Dash, Zijiang Yang, Liang Liang. 2006. Using dea-neural network approach to evaluate branch efficiency of a large canadian bank. *Expert systems with applications* **31**(1) 108–115.

Xardel, D. 1993. *The Direct Selling Revolution*. Business: Blackwell, Blackwell Business. URL `https://books.google.com/books?id=-1pkQgAACAAJ`.

Xu, Ronghui (Lily). 2016. Lecture 7 time-dependent covariates in cox regression URL `http://www.math.ucsd.edu/~rxu/math284/slect7.pdf`.

Yang, Yupin, Avi Goldfarb. 2015. Banning controversial sponsors: Understanding equilibrium outcomes when sports sponsorships are viewd as two-sided matches. *Journal of Marketing Research* **28**(6).

Yang, Yupin, Mengze Shi, Avi Goldfarb. 2009. Estimating the value of brand alliances in professional team sports. *Marketing Science* **28**(6) 1095–1111.

Zamudio, César, Yu Wang, Ernan Haruvy. 2013. Human brands and mutual choices: an investigation of the marketing assistant professor job market. *Journal of the Academy of Marketing Science* **41**(6) 722–736.

Zoltners, Andris A., PK Sinha, Sally E. Lorimer. 2013. How to make sense of sales force turnover. *Harvard Business Review* ¡a href="https://hbr.org/2013/06/how-to-make-sense-of-sales-for"¿.