IMT School for Advanced Studies, Lucca

Lucca, Italy

Essays on Inter-firm Ownership Network and Social Media Marketing

PhD PROGRAM IN INSTITUTIONS, MARKETS AND TECHNOLOGIES

Curriculum in Management Science

XXX Cycle

By

Xu Wang

2017

The dissertation of Xu Wang is approved.

Program Coordinator: Prof. Massimo Riccaboni, IMT School for Advanced Studies, Lucca

Supervisor: Prof. Massimo Riccaboni, IMT School for Advanced Studies, Lucca

Supervisor: Prof. Bart Baesens, KU Leuven

Tutor: Dr. Zhen Zhu, IMT School for Advanced Studies, Lucca

The dissertation of Xu Wang has been reviewed by:

Prof. Stefano Schiavo, University of Trento

Prof. Raja Kali, University of Arkansas

Prof. Anders C. Johansson, Stockholm School of Economics

IMT School for Advanced Studies, Lucca

2017

Contents

List of Figures

List of Tables

Acknowledgements

Chapter 2 *"The Effect of Firm Centrality and Business Group Size on Firm Performance: Evidence from Italy"* is co-authored with Massimo Riccaboni and Zhen Zhu.
Chapter 4 is the reproduction of the paper *"On the Optimal Marketing Aggressiveness Level of C2C Sellers in Social Media: Evidence from China*", coauthored with Bart Baesens and Zhen Zhu. The paper has already been submitted to the journal *Omega* and it is now under review.

First, I would like to express my deepest appreciation to my supervisor Prof. Massimo Riccaboni, for his patient guidance and continuous support during my Ph.D. The thesis would not have been possible without his insightful ideas and advice. He not only teaches me how to address research questions, but also deeply impresses me by his great tolerance, good sense of humour and positive attitude whenever we face difficulties.

I would also like to deeply thank Prof. Bart Baesens for all his instructions and supports during my visiting period in KU Leuven. Under his supervision, I quickly acquired a good knowledge of machine learning techniques which helped me accomplish the fourth chapter.

I am particularly indebted to Dr. Zhen Zhu. In the last three years, he spent a lot of personal time discussing with me on my study, always provided constructive suggestions and shared many helpful experiences in doing research. His encouragement and support also helped me go through a difficult time.

I would also like to express my gratitude to the three referees, Prof. Stefano Schiavo, Prof. Raja Kali and Prof. Anders C. Johansson, for the time they spent on reviewing my thesis and the insightful comments they provided.

I am especially grateful to my colleagues Anita Bunea, Yuan Gao and Olena Kulynych, for downloading together the data that are used in the chapters 2 and 3 of the thesis. I would also like to thank Prof. Armando Rungi, Dr. Davide Del Prete, Dr. Valentina Tortolini and the other scholars who provided useful comments during and after my presentations regarding the three chapters.

Many sincere thanks go to Dr. Michelangelo Puliga for spending a lot of time in teaching me the programming skills. I would also like to thank my colleagues Luca Verginer for helping me with programming so many times, and Eugen Stripling for his detailed and patient explanations of the machine learning techniques.

I am also grateful to all the IMT administrative staff for always being so helpful, and to all my friends in the IMT community, for the wonderful pingpong games, for the amazing experiences in the choir, and for all the memorable moments we spent together.

A special thanks goes to my parents and all the other family members for their love and support during all the stages of my life.

Finally, I would like to express my deepest gratitude to my *Shifu* Ven. Master Xuecheng and all the other Buddhist masters and friends from the Longquan Monastery. Their spiritual instructions not only encouraged me to persist through the hardest time in writing the thesis, but also grants me great power to overcome all the difficulties in my future life.

Vita

July 26, 1989	Born, Harbin, China
09/2008-06/2012	B.S. in Statistics Final grade: 3.74/4.00 Renmin University, Beijing, China
09/2012-07/2014	M.S. in Statistics, Actuarial and Financial Sciences Final grade: 110/110 cum laude Università degli Studi di Firenze, Florence, Italy
11/2014-present	Ph.D. in Management Science IMT School for Advanced Studies Lucca, Lucca, Italy
04/2017-06/2017	Visiting Researcher Leuven Institute for Research on Information Systems (LIRIS), KU Leuven, Leuven, Belgium

Presentations

1. X. Wang, B. Baesens, Z. Zhu (2017), “On the Optimal Marketing Aggressiveness Level of C2C Sellers in Social Media: Evidence from China”, presented at KU Leuven, Leuven, Belgium.
2. M. Riccaboni, X. Wang, Z. Zhu (2017), “Ownership Network and Firm Performance”, presented at IMT Lucca, Lucca, Italy.
3. M. Puliga, X. Wang, Z. Zhu (2016), “Web Scraping Using Python: An Example of Taobao”, presented at IMT Lucca, Lucca, Italy.
4. M. Riccaboni, X. Wang, Z. Zhu (2016), “Ownership Network and Firm Performance”, presented at the conference ITSG in IMT Lucca, Lucca, Italy.
5. M. Riccaboni, X. Wang, Z. Zhu (2016), “Ownership Network and Firm Performance”, presented at University of Trento, Trento, Italy.

Abstract

There has been a growing debate on big data and analytics in recent years. Applying analytics to big data creates many opportunities for managers and policy makers to gain greater insight into their business so that they can improve their decision-making. This thesis employs novel approaches of data science to study economic and managerial topics. In particular, we combine the traditional econometric models with novel network measures and machine learning algorithms in exploring the big and high dimensional data of global inter-firm ownership network and Chinese C2C sellers' microblogs in social media, so as to provide managerial strategies for both firm-level and individual business.

The first two studies investigate inter-firm ownership network and firm performance. By analysing the data of Italian firms in the period of debt crisis, the first study provides a deep insight into the relationship between firm performance and the interaction of firm-level centrality and business group size. The findings, together with the novel centrality measure we provide, contribute to the literature on inter-firm ownership network. The second study explores foreign ownership and firm performance from various perspectives. The results reveal that the foreign-owned Italian firms are on average more productive than the ones in domestic-owned MNEs. In addition, we find that the Italian subsidiary with shorter organizational and geographical distance from their foreign owners are on average more productive.

The third study focuses on business in social media. By exploiting the fact that Sina Weibo collaborates with Taobao (Chinas largest C2C e-commerce platform) to provide their sellers an easier way to promote their products using microblogs, we are able to examine the relationship between marketing aggressiveness and marketing popularity. Interestingly, we identify an optimal level of how aggressive Taobao sellers should be when promoting their products over Sina Weibo. This finding contributes to the existing literature on social media marketing, especially in the field of C2C business.

Chapter 1

Introduction

The thesis employs the econometric models, network measures and machine learning algorithms in exploring the big and high dimensional data of global inter-firm ownership network and Chinese online sellers' microblogs in social media. We attempt to provide managerial strategies for both firm-level and individual business for the purpose of improving their performance.

The first two chapters focus on inter-firm ownership network of business groups. Through ownership links, firms can exchange financial capital, superior knowledge and managerial skills with other firms in the same group (Arnold and Javorcik, 2005; Markusen, 1995). However, some studies argue that a business group with diversified subsidiaries may have problems due to weak disclosure requirements, ineffective governance mechanisms, and a poorly developed market for corporate control (Khanna and Palepu, 2000; La Porta et al., 1997, 1998). In the literature, although the role of individual firm position in the inter-firm network and the group-level features on firm performance has been widely discussed, the effect of their interaction has drawn less attention.

In Chapter 2, we attempt to fill this gap by combining the firm-level centrality and business group size and exploring the relationship between their interaction and firm performance. Furthermore, we provide a novel centrality measure based on the harmonic centrality (Newman, 2003; Rochat, 2009; Schilling and Phelps, 2007). Our measure solves the problem of high correlation between centrality and component size using a normalization approach and make the centrality measure more comparable across components, which in our case are business groups composed of different numbers of subsidiaries.

Through an empirical investigation of the performance of 483,835 Italian firms during the debt crisis from 2011 to 2014, we find a positive relationship between firm centrality and performance in small business groups, but not always significantly positive in large groups. What's more, we find that there exists a positive relationship between firm performance and group size for the peripheral subsidiaries, and the group size "premium" for them is larger than for the central ones.

In Chapter 3 we discuss the foreign ownership and firm performance using the same data as in Chapter 2. Based on the location of the ultimate owner, firms involved in a MNE in a certain country can be generally divided into two types: foreign-owned subsidiaries and firms in a domestic-owned MNE. There has been a fruitful discussion on the advantages of MNEs, such as the spillover effect (Bernstein and Mohnen, 1998; Blomström and Sjöholm, 1999), economies of scale (Dunning, 1989; Lovelock and Yip, 1996), and tax reasons (Desai et al., 2004, 2006; UNCTAD, 2015), and consequently, MNEs are proved to have better performance than the domestic groups in many empirical studies. Then we focus on the different types of foreign-owned subsidiaries by considering the location of their direct owners and ultimate owners. UNCTAD (2016) point out that 41% of the foreign affiliates all over the world are owned by direct owners and ultimate owners in different countries.

By analysing the balance sheet data of 564,770 Italian firms, we empirically find that the firms making up part of a multinational group outperform the purely domestic groups. Moreover, among the multinational firms, the foreign-owned Italian firms are on average more productive than the ones in domestic-owned MNEs. Interestingly, we find that the Italian subsidiaries with shorter organizational or geographical distance from their foreign owners are on average more productive. In addition, subsidiaries with multiple cross-border links in the upstream ownership chain are found to be more productive.

Chapter 4 studies the individual online business and their marketing behaviour on social media. Social media has become a widely used marketing tool for reaching potential customers. Because of its low cost, social media marketing is especially appealing to the customer-to-customer (C2C) sellers. Customers can also benefit from social media marketing by learning about products and by interacting with sellers in real time. However, if a seller is too aggressive in promoting her products, customers may get annoyed. Previous literature has mainly focused on the gratification brought by social media (Kaplan and Haenlein, 2010; Mangold and Faulds, 2009). Although some works have pointed out people's negative attitude towards aggressive social media (and online) marketing by making surveys (Akar and Topçu, 2011; Grant, 2005), to the best of our knowledge, no one has empirically provided an optimal marketing aggressiveness level in social media. This paper attempts to fill the gap by analysing the data of Taobao (China's largest C2C e-commerce platform, similar to eBay and Amazon) sellers on Sina Weibo (China's largest microblogging platform). Moreover, our research contributes to the literature on C2C sellers' behaviour in social media by solving the technical problems of identifying the individual sellers and collecting their data.

We identify the 52,187 Taobao sellers on Sina Weibo and collect their microblogs in November 2014. For the 12,744 sellers who add the links of their Taobao shop, we further track their microblogs from July to October in 2016. We define the marketing aggressiveness level as the proportion of a seller's marketing-related microblogs and define the marketing popularity as the average number of likes a seller receives per marketing-related microblog. To classify the microblogs into marketing-related and non-marketing ones, we train different machine learning classifiers, such as decision trees, logistic regression, multinomial naive bayes, and random forest with a manually labelled sample of 5,000 microblogs. The multinomial naive bayes algorithm has the best performance with a ROC score of 0.96 and is used to classify all the microblogs. In modelling the relationship between marketing aggressiveness level and marketing popularity, the linear regression using Yeo-Johnson transformation for the number of followers outperforms the other models such as random forest and neural network. After multiple statistical tests, we empirically confirm that there is an inverted U-shape relationship between the marketing aggressiveness level and the marketing popularity. Specifically, the optimal proportion of marketing microblogs is around 0.3. Moreover, we find a saturation effect of the number of followers on marketing popularity after it reaches around 100,000.

All in all we investigate how the role, position and strategy of firms in business networks impact on their performance. The dissertation is made of two parts. In Chapters 2 and 3 we explore inter-firm ownership network and firm performance. In Chapter 4 we focus on business in social media.

Chapter 2

The Effect of Firm Centrality and Business Group Size on Firm Performance: Evidence from Italy

2.1 Introduction

In an interorganizational network, units can acquire a variety of resources and information by connecting with other units. The ability to access the resources and information highly depends on their network positions. A unit occupying a central position in the network is likely to have more opportunities to access resources and information and benefit from the knowledge spillover effects (Powell et al., 1999; Tsai, 2001; Tsai and Ghoshal, 1998; Zaheer and Bell, 2005). However, units at central positions incur more costs to maintain the relational ties. Furthermore, a high level of connectivity is sometimes considered to derail economic performance by making firms vulnerable to exogenous shocks (Uzzi, 1996, 1997).

The work discusses the ownership network of business groups, which are defined as confederations of legally independent firms linked by multiplex ties (Almeida and Wolfenzon, 2006; Belenzon and Berkovitz, 2010; Belenzon et al., 2013) and belonging to the same owner(s) (Belenzon et al., 2017; Cainelli and Iacobucci, 2011; Cainelli et al., 2006). Firms can exchange financial capital, superior knowledge and managerial skills with other firms in the same group through ownership links (Blomström and Sjöholm, 1999; Markusen, 1995). But some studies argue that a business group with diversified affiliates may have problems due to weak disclosure requirements, ineffective governance mechanisms, and a poorly developed market for corporate control (Khanna and Palepu, 2000; La Porta et al., 1997, 1998).

In literature, the role of individual positions in the inter-firm network and the group-level features has been widely discussed, but the relationship between their interaction and firm performance is underexplored. Though a central firm in large groups can gain more access to information, it is possibly faced with higher costs of acquiring and communicating knowledge (Caliendo and Rossi-Hansberg, 2012) and governance frictions which can reduce the effectiveness of central control (Bethel and Liebeskind, 1998; Patacconi, 2009). The possible trade-off effect may make central firms in a large business group benefit less than expected.

Through an empirical analysis, our research contributes to the existing studies by combining the firm-level centrality and business group size and exploring the relationship between their interaction and firm performance. Furthermore, we provide a novel centrality measure to evaluate a firm's position in the ownership network. The measure is based on the harmonic centrality (Newman, 2003; Rochat, 2009; Schilling and Phelps, 2007), which can avoid the infinite path length problem in a network composed of disconnected components. However, the harmonic centrality value of a certain node is highly correlated with the component size where it is located. Given this drawback, our measure adopts a normalization method to reduce the effect of the component size. In this way, we make the centrality measure more comparable across components.

Our analysis is focused on Italian firms. Italy and some other European countries have experienced a severe debt crisis since the end of 2009. The real GDP growth rate of Italy[1] had been ranked in the last 5 of the EU countries for 3 consecutive years since 2012. From Figure 1 we can find that the economy of Italy was shocked by the crisis since 2011, and the GDP growth rate turned positive again in 2014. During the crisis, the skepticism about the availability of equity capital, as one of the main determinants, results in the decrease in the volume of mergers and acquisitions at the global level (Kostić, 2013), and the ownership structure of most business groups remains stable. It would be prominent to study the network features of the subsidiaries that are more resilient to the exogenous financial shock through an analysis of their performance in the crisis.

Another reason to study Italian firms is that business group, especially controlled by a person or a group of family members, is a common case in Italy (Corbetta and Tomaselli, 1996; Cucculelli and Micucci, 2008). In 2014, the number of the Italian-owned business groups is ranked in the top 10 countries all over the world. What's more, the available databases provide a better coverage for firms in the developed countries and the data quality is also higher than that in the developing world. In summary, we believe that Italy is a representative country to conduct our research.

We manually download the shareholders' data of 17.8 million global firms in 2014 from ORBIS database. Based on their direct inter-firm shareholding relationships, we construct the global ownership network. Then we restrict our sample to the Italian firms and identify 483,835 Italian firms from the global ownership network and collect their financial accounts data from 2011 to 2014 in AIDA (Analisi informatizzata delle aziende italiane), which is the Italian subset of ORBIS. Then we empirically analyse firm centrality, business group size and the relationship between their interaction and firm performance, which is measured as the 3-year average sales growth rate. The results show a positive relation-

[1] Data source: world bank `https://data.worldbank.org/`.

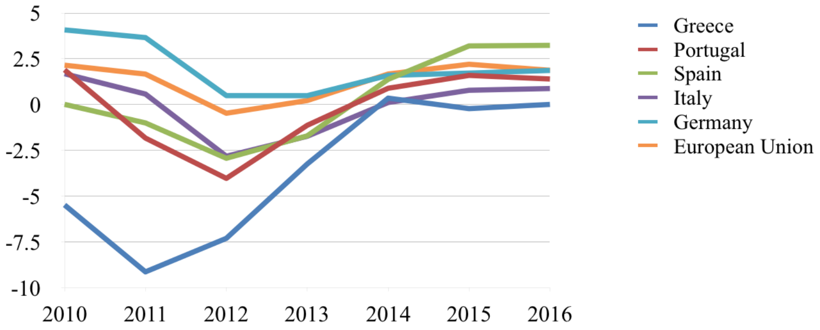

Figure 1: Real GDP Growth Rate of Some European Countries

ship between firm centrality and performance in small business groups, but not always in larger business groups. Moreover, we find that for the peripheral Italian subsidiaries, their performance is positively correlated with the size of the group they belong to. Interestingly, the group size ″premium″ for the peripheral ones is larger than for the central ones. A possible explanation for the findings is that though firms at central positions in larger business groups can facilitate the access to funds and information, they are also faced with higher coordination costs than in small groups. But for firms at a peripheral position in larger groups, the benefit of the reputation effect and the availability to diverse resources can be much larger than that in small groups. Especially during the crisis, being a member of a large business group makes them more resilient to the exogenous financial shocks than in small groups.

The remainder of the Chapter is structured as follows. In Section 2.2 we present the theoretical framework of our research. Then Section 2.3 introduces the sample, centrality and business group size measure and the econometric specifications. In section 2.4 we present some descriptive statistics of the data and then provide the estimation results and analysis. Finally, Section 2.5 concludes the work.

2.2 Theoretical Framework

In the organizational network theory, a vast literature has investigated the effect of unit position and network complexity on unit performance, especially in the context of inter-firm network (Schilling and Phelps, 2007; Zaheer and Bell, 2005). In an inter-firm network, firms are connected through a variety of relationships such as collaboration (Ahuja, 2000; Gulati and Gargiulo, 1999; Gulati et al., 2000), interlocking directorates (Uzzi, 1997; Zaheer and Bell, 2005), ownership (Almeida and Wolfenzon, 2006; Kali and Sarkar, 2011) and credit (Peterson and Rajan, 1994).

Our work is related to the literature on business groups' ownership network and firm performance. Business groups exhibit approximately a pyramidal structure, in which one or more layers of firms are controlled by the same ultimate owner, either directly or through a holding company (Cainelli and Iacobucci, 2011; Goto, 1982). Due to the availability of worldwide shareholding data in recent years, a growing body of research has explored the global ownership network (Glattfelder, 2010; Vitali and Battiston, 2011; Vitali et al., 2011). The existing studies discuss the inter-firm network generally from two perspectives: the individual firm level and the network level (Provan et al., 2007). Though some works have considered the factors of both levels, the relationship between their interplay and firm performance remains unclear. Our research attempts to fill the gap by combining both the individual level and network level factors in the analysis.

2.2.1 Firm-level Centrality

A large number of studies have discussed the benefits of network centrality for a firm. Since network links facilitate the sharing of financial capital (Belenzon and Berkovitz, 2010), superior knowledge and managerial skills (Bernstein and Mohnen, 1998; Markusen, 1995) among firms in the business group, a firm occupying a central position in the network has more opportunities to access diversified knowledge and resources (Powell et al., 1996; Tsai and Ghoshal, 1998) and generate more innovations than the peripheral firms (Owen-Smith and Powell, 2004; Soh, 2003). However, despite the benefits of network ties, a high level of connectivity is sometimes considered to constrain the adaptability of firms and make them vulnerable to exogenous shocks (Uzzi, 1996, 1997).

In literature, diverse measures are used to identify a firm's network position. Gulati and Gargiulo (1999) and Powell et al. (1996) have used closeness centrality, which measures how central a firm is relative to other firms, including both direct and indirect partners. It also measures the firm's reachability to every other firm with the fewest number of intermediate firms. Schilling and Phelps (2007) employ the betweenness centrality measure as a control variable but its effect on subsequent firm patenting fails to achieve statistical significance in any of the estimated models. Mani and Moody (2014) measure a firm's position using mesolevel network structure indicators: disconnected periphery, isolated cluster, small world and nested world. They find that firms residing in the nested core have more multiplex ties and larger transaction volumes compared with firms in the small world or the disconnected periphery. A recent work by Kwon et al. (2016) also adopts the closeness centrality to measure how closely connected a firm is to the rest of the organizations in the inter-firm network, but they haven't found evidence to support that the influence of national trust on alliance governance will decrease as a firm increases its centrality in the international alliance network.

Our research contributes to the literature by providing a novel centrality measure, which is based on the harmonic closeness centrality (Newman, 2003; Rochat, 2009). One advantage of the harmonic centrality is that it can avoid the infinite path length problem compared with the traditional closeness centrality. However, the harmonic centrality value of a node is highly correlated with the size of the component where it is located. Given this drawback, our measure adopts a normalization method to reduce the effect of the component size. By doing so, we make the centrality measure more comparable across components of different sizes. In the case of ownership network, we can better assess how central a firm is in a business group of any size.

2.2.2 Group-level Measures

A strand of literature focuses on the network-level features and tries to understand the impact of the structures and behaviours on individual organizations (Ahuja, 2000; Powell et al., 1996; Provan et al., 2007). Some recent works have developed novel measures to characterize the complexity of a business group's hierarchical structure. Altomonte and Rungi (2013) provide a specific entropy-like measure of organizational complexity of hierarchical chains and find a non-linear relationship between hierarchical complexity and productivity. Belenzon et al. (2013) develop a pyramidal index that measures the distribution of firms by ownership layers. They find that Anglo groups have the most vertical hierarchical structure while the Swiss groups have the most horizontal structures. Mahmood et al. (2017) adopt the intergroup network centralization measure introduced in Freeman (1978) and demonstrate that centralization of equity ties enhances subsidiary performance, but such effects weaken when the environment becomes turbulent.

The existing studies have emphasized the role of individual positions in the network and the group-level features on firm performance, but the relationship between their interaction and firm performance remains unclear. Though a central firm in a large business group can gain more access to information, it has to face possibly higher costs of acquiring and communicating knowledge (Altomonte and Rungi, 2013; Caliendo and Rossi-Hansberg, 2012) and governance frictions which can reduce the effectiveness of central control (Bethel and Liebeskind, 1998; Pataconi, 2009). The possible trade-off effect may make a central firm in a large business group benefit less than expected. On the contrary, peripheral subsidiaries in larger groups may not only enjoy the reputation effect (Chang and Hong, 2000) but also be more resilient to the exogenous financial shocks compared with in small groups. We conjecture that the advantages of being involved in a large business group for a peripheral subsidiary are bigger than for a central one.

2.3 Method

2.3.1 Data and Sample

We derive the ownership data from Bureau van Dijk's (BvD) ORBIS database. ORBIS is the largest and most widely used database, which captures information on companies, especially the private ones, from a wide range of sources. ORBIS database covers information on 17,842,618 global firms' equity ownership structure in 2014.[2] For each firm, there is at least one observation of its direct shareholders, and we collect each of its owners' ID [3] and the ownership shares (see the details of the data form in Table 1). Firms can have multiple direct shareholders, such as firm E and H in Table 1. In total, there are initially 25,635,140 observations and each firm has on average 1.44 direct shareholders.

Table 1: Example of the Data

mark	firm	direct owner	ownership shares
1	B	A	100%
2	C	B	100%
3	D	A	97%
4	E	A	79.35%
5	E	K	20.65%
6	F	E	95%
7	G	F	100%
8	H	F	90%
9	H	L	10%
10	I	H	100%
11	J	A	100%

A strand of studies build the inter-firm ownership network based on the majority rule (Altomonte and Rungi, 2013; Belenzon and Berkovitz, 2010; UNCTAD, 2016). In ORBIS, 81.2% of the firms are owned by ma-

[2]We download the data of all the firms provided with ownership information by ORBIS in January 2015.

[3]BvD identifies each company by a unique ID. The BvD ID number incorporates either the national ID number or the ID provided by their information providers (IP). According to BvD, the ID numbers may change when the national ID numbers change in the official data sources or the BvD IPs decide to switch their ID numbers. In Italy, the BvD ID may change if the company changes address.

jority shareholders. Their ownership links (shares $> 50\%$), representing 56.53% of all the links (see Table 2), are used to build the ownership network in our study.[4] For example, the 5th and 9th row in Table 2 are dropped. For each observation, we construct a directed link from the shareholder to the firm. As illustrated in Figure 2, a directed link starting from shareholder A to firm B is built for the first observation, and then another link from shareholder B to firm C is built for the second observation, and so on. Some firms in the same business group may share the same direct owner. For example, firms B, D, E, J are all owned by firm A in the business group. All firms that belong to the same business group are thus connected through ownership links. As shown in Figure 2, firm A is the parent firm of the business group and controls all the other subsidiaries, and the ownership structure of each business group can be depicted as a pyramid.[5] Since there is no ownership link among firms in different business groups, the global network can be divided into disconnected components.[6]

According to the approach we use in building the ownership network, each firm has 1 shareholder except that the ultimate owner has no shareholder. Thus, the number of entities in each business group is equal to the number of links plus 1, and the number of nodes in the global ownership network is equal to the number of links plus the number of components. For example, the business group in Figure 2 contains 10 firms and 9 ownership links. The global ownership network consists of

[4]We also use other two methods to build the ownership network in which the minority control is considered. If a firm has no majority owner, we rank all its shareholders by their shares $r1, r2, r3, \cdots$. In one method, we retain the link from the top ranking shareholder if its share is larger than the sum of the shares of the second and the third ranking shareholders, that is, $r1 > r2 + r3$. Similarly, in the other method, we retain the link from the top ranking shareholder if its share is larger than the sum of the shares of the second, the third and the fourth shareholders, that is, $r1 > r2 + r3 + r4$.

[5]Due to the restriction of the data, we cannot distinguish the ownership by entities, states, individuals or families. Therefore, the one locates at the top of each corporate ownership structure can be either a parent company whose shareholders are not provided by ORBIS or all have minority ownership, or the individual who is the ultimate beneficial owner of the business group. See more details in 2.6.6.

[6]Since the individual majority shareholders cannot be identified in the data, some groups of size 2 mentioned in the main text of this chapter are actually composed of a standalone firm and its individual shareholder. We perform a new algorithm to filter out these possible links. See more details in 2.6.6.

Table 2: Distribution of Direct Ownership Share

ownership share	frequency	percentage
$(0, 25\%]$	3,997,584	15.59%
$(25\%, 50\%]$	7,144,627	27.87%
$(50\%, 75\%]$	3,615,425	14.10%
$(75\%, 100\%]$	10,877,504	42.43%
Total	25,635,140	100.00%

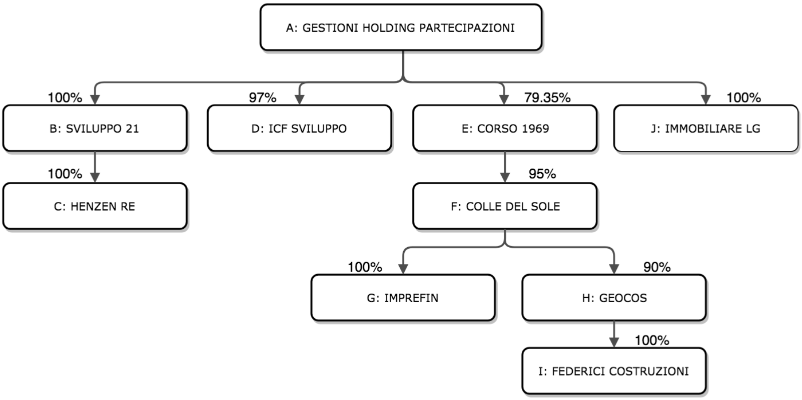

Figure 2: An Italian Business Group's Ownership Structure

25,681,483 nodes, 14,492,929 links and 11,188,554 components.[7] 96.32% of the components contain less than 4 nodes while only 0.05% of them contain more than 50 nodes. The largest component is composed of 8,419 nodes.

As aforementioned, we then restrict our sample to the Italian firms. In our data, Italian-owned business groups represent 3.46% of all the business groups, ranking 8th among all countries.[8] We collect the non-

[7]The network built by using the other two methods is composed of respectively 26,266,255 nodes, 14,873,254 links for the definition $r1 > r2 + r3$, and 26,168,532 nodes, 14,814,218 links for the definition $r1 > r2 + r3 + r4$.

[8]In our data, the 10 most frequent countries where the ultimate owners of all the business groups are located include: United States (12.34%), Russia (9.43%), Norway (4.96%), Poland (4.90%), Australia (4.88%), Germany (4.71%), Bulgaria (4.25%), Italy (3.46%), Roma-

consolidated financial balance sheets data of Italian firms from AIDA, which is also a product of Bureau van Dijk. It contains information on all Italian companies obliged to deposit the balance sheet. We extract the data of 1,164,871 active Italian firms in all sectors and regions in 2014.[9] Among all these Italian firms provided by AIDA in 2014, we identify 483,835 of them in the global ownership network, which is built using the data of ORBIS[10]. These Italian firms belong to 372,109 components. Each of them contains at least one Italian firm and some of them also contain foreign firms. 25 of these groups are composed of more than 1,000 firms (see an example in Figure 3) and the largest one contains 3,260 firms.

2.3.2 Variables

Dependent Variable

Sales growth is a widely used indicator of firm performance (Brush et al., 2000; Collins and Clark, 2003), which reflects how well an organization relates to the environment by successfully expanding their market scope (Ansoff, 1965; Dess and Robinson, 1984; Hofer and Schendel, 1978). We use the 3-year average sales growth rate as the dependent variable. As a robustness check, we also use the logarithm forms of labour productivity (see the definition in Table 12 in Appendix 2.6) and sales per employee as the dependent variable.

Independent Variables

Firm Centrality

---

nia (2.50%) and Spain (1.51%).

[9] In February 2016, we download the Italian firms' balance sheet data of the year 2014. In June 2017, we further download their balance sheet data of the years 2011-2013 but some firms are no longer found in AIDA. See the descriptive statistics in Table 3.

[10] We first match the Italian firms in AIDA with those in the original ORBIS ownership databases by their BvD ID and identify 711,393 firms. Considering that some of them may change the BvD ID number, we further match the rest firms by their names and identify 7,751 firms. In total, we find 719,144 Italian firms with ownership information in AIDA. However, 235,309 of them have no majority owners. According to our definition, we cannot identify which business group these firms belong to. Finally, we retain 483,835 Italian firms for further analysis.

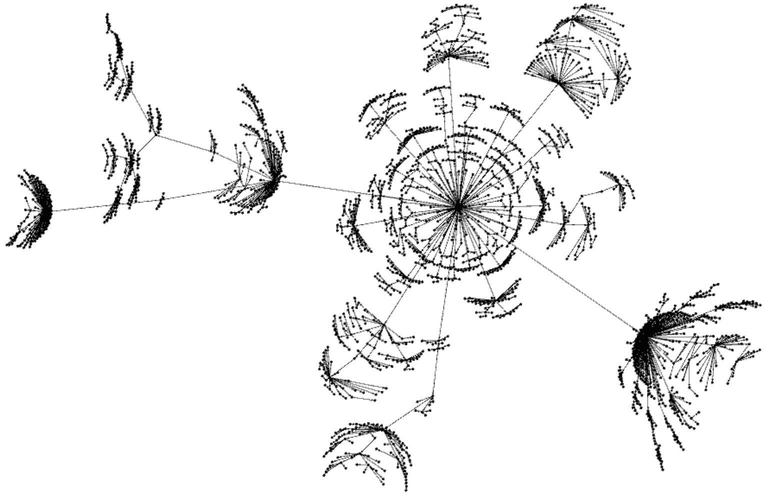

Figure 3: The Ownership Structure of a Large Business Group

Given that the ownership network is directed, the firms that control no subsidiaries in a business group have no connecting path to other firms. Newman (2003) and Rochat (2009) have introduced the harmonic centrality which can avoid this problem by considering the reciprocal of the distance. The definition of the harmonic centrality of a node i is

$$c_i = \sum_j \frac{1}{d_{ij}}$$

where d_{ij} is the number of links in a shortest path connecting node i to node j. If node i has no path to node j, their distance is considered as infinity and the reciprocal is thus zero. Therefore, a node which has no path to any other node in the network has a centrality value of zero.

We notice that nodes in a large component are prone to have larger harmonic centrality values. To reduce the effect of component size on centrality value, we provide a novel measure which normalizes the harmonic centrality to the interval [0, 1]. By doing so, we make it more comparable when assessing how central a firm is across business groups of different sizes.

We define the normalized harmonic centrality as follows:

$$nc_i = \frac{c_i - \min_{j \in g(n_i)}\{c_j\}}{\max_{j \in g(n_i)}\{c_j\} - \min_{j \in g(n_i)}\{c_j\}} \tag{2.1}$$

where n_i is component size, that is, the number of nodes in the component that node i belongs to. $g(n_i)$ is any component of size n_i. Given a component of size n_i, the largest possible value of centrality is reached when a node is directly connected to all the other $n_i - 1$ nodes. In a directed graph [11], the minimal centrality is obviously 0, and formula 2.1 can be rewritten as

$$nc_i = \frac{c_i}{n_i - 1}$$

In our case, n_i is the number of firms in the business group that firm i belongs to. Given a business group of size n_i, the largest possible value of centrality is reached when the ultimate owner directly controls all the $n_i - 1$ subsidiaries. The comparison of harmonic and the normalized centrality values of three examples is illustrated in Figure 4 and 5.

It should be noticed that although our sample is restricted to the Italian firms, 9.9% of them are in multinational groups. Since our centrality measure reflects a firm's position in the business group, the number of steps in the network between an Italian firm and a foreign firm in the same business group is also taken into account when computing the centrality.

Business group size

We adopt the number of firms in a business group to measure group

[11] We build the ownership network based on the control relationship among firms, thus the direction of the ownership link is taken into account. If we ignore the direction and treat the network as an undirected one, the minimal possible centrality is reached when the component is a chain. The node at either end of the chain has the minimal centrality value, which is equal to $\sum_{k=1}^{n_i-1} \frac{1}{k}$. In this case, formula 2.1 can be rewritten as

$$nc_i = \frac{c_i - \sum_{k=1}^{n_i-1} \frac{1}{k}}{n_i - 1 - \sum_{k=1}^{n_i-1} \frac{1}{k}}$$

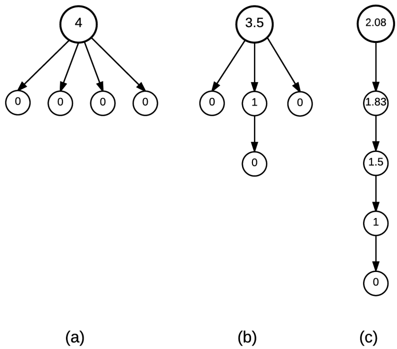

Figure 4: Original Harmonic Centrality

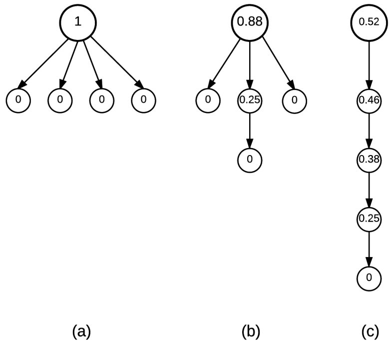

Figure 5: Normalized Harmonic Centrality

size[12] (Belenzon et al., 2017; Del Prete and Rungi, 2015). Since this measure has a long-tail distribution, we perform a log transformation to reduce the variance. Another measure of business group size is the diameter. Diameter is defined as the longest of all the shortest paths in a network, which in our case is a business group. For example, among all the shortest paths, the longest distance of the three prototypes in Figure 5 are respectively 2, 3 and 4. Watts and Strogatz (1998) and Barabási and Albert (1999) argue that the diameter of a network increases logarithmically with the addition of new nodes. Hence, it can be viewed as a proxy of the logarithm of the number of firms in a business group.

Control Variables

The covariates include the firm-level factors in 2014, such as number of employees, capital intensity, firm age, whether the business group is multinational, dummies of sectors and regions. When we use the 3-year sales growth rate as the dependent variable, we also control the logarithm form of sales in 2011.

[12]Since the ownership network of each business group identified by our algorithm may include the individual ultimate owner, the actual number of firms in a group can be 1 less than the number of nodes in its network. We make some further efforts to identify the possible individual ultimate owners. See the details in Appendix 2.6.6

In addition, we use the pyramidal index (PI) introduced by Belenzon et al. (2013) as a control variable. The index reflects the extent to which the organization of subsidiaries is hierarchical by measuring the distribution of subsidiaries in different ownership levels. They define the pyramidal index as $\frac{2(\sum_{i=1}^{N} i \times share_i - 1)}{\#Affiliates - 1}$, where N is the largest number of steps to the ultimate owner, $share_i$ is the ratio of the number of subsidiaries that are located at level i to the total number of subsidiaries in the group, and $\#Affiliates$ is the number of subsidiaries in the group. The value of PI varies from 0 to 1. The higher value indicates that the group structure is more hierarchical. Aghion and Tirole (1997) argue that the hierarchical structures in which decision-making responsibility over non-routine tasks is delegated to local managers result in better decisions.

2.3.3 Empirical Models

We first explore the overall relationship between firm performance and respectively firm centrality and group size through the following linear regression models.

$$Y_i = \alpha + \beta_1\ centrality_i + \beta_2\ \log BG\ size_j + \boldsymbol{\gamma}' \boldsymbol{Z_i} + \epsilon_i \tag{2.2}$$

where Y_i is the performance of firm i, and j is the business group that firm i belongs to. $\mathbf{Z}_i$ are the covariates aforementioned. Since the performance of the firm within the same business group can be correlated, we relax the assumption of the independence of residuals. Instead, we use the cluster-robust standard errors in the estimation by controlling the component they belong to.

Second, we examine the relation between firm centrality and performance given group size. To guarantee that each subsample has enough observations, we merge some business groups of different sizes into few intervals. We run the regression model 2.2 separately over the subsamples, dropping the term $\log BG\ size_j$.

Next we compare the relation between business group size and firm performance given the centrality level. According to the empirical distribution of centrality, we generally select the central firms and the peripheral subsidiaries as two subsamples. Then we separately run regressions, dropping the centrality term from 2.2.

2.4 Results and Discussion

2.4.1 Descriptive Statistics

We report the summary statistics of the variables in Table 3 and their correlation matrix in Table 4.

Table 3: Summary Statistics

Variables	N	mean	sd	p10	p50	p90
3-year sales growth	247,092	0.646	15.85	-0.184	0.010	0.519
centrality	483,835	0.0628	0.215	0	0	0
No. firms in BG	483,835	14.53	109.0	2	2	6
diameter	483,835	1.674	1.592	1	1	3
No. employees	395,395	14.15	167.1	1	3	18
sales 2014	396,447	4,265	111,392	0	255	3,629
sales per labour 2014	395,112	229.6	2,305	0	76	380
sales 2013	308,425	5,273	131,897	27	384	4,603
sales 2012	290,611	5,560	138,663	24	402	4,869
sales 2011	270,594	5,883	131,279	22	432	5,326
value added	396,730	912.9	21,688	-7	67	897
labour productivity	395,395	45.26	700.0	-5.069	22.20	80
fixed assets	396,357	1,641	79,799	0	33	1,554
capital intensity	395,025	370.7	5,675	0	7.250	438
age	483,751	14.27	12.39	3	10	31
mne	483,835	0.0987	0.298	0	0	0
PI	159,883	0.2276	0.3412	0	0.0105	1
ownership level	483,831	1.098	0.535	1	1	2

2.4.2 Main Results

We then perform linear regression to explore the relationship between firm performance and the interaction of centrality and business group size. Table 5 reports the regression results.[13] The coefficient of centrality

[13] All regression results reported in this chapter are based on the majority control network. The results of the other two definitions of network are consistent and available upon request. In the regression model, the observations with the values of their dependent vari-

Table 4: Correlation Matrix

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
1. 3-year sales growth	1.000												
2. (log) labour productivity	0.015	1.000											
3. (log) sales per employee	0.021	0.646	1.000										
4. centrality	0.000	0.149	0.088	1.000									
5. (log) No. firms in BG	0.017	0.227	0.157	0.036	1.000								
6. diameter	0.019	0.222	0.153	0.043	0.944	1.000							
7. (log) No. employees	-0.006	0.209	0.186	0.089	0.101	0.104	1.000						
8. (log) capital intensity	0.006	0.366	0.151	0.155	0.132	0.131	-0.234	1.000					
9. sales 2011	-0.002	0.061	0.080	0.031	0.098	0.100	0.126	0.010	1.000				
10. age	-0.016	0.213	0.135	0.154	0.077	0.083	0.194	0.264	0.042	1.000			
11. mne	0.014	0.212	0.159	0.093	0.606	0.564	0.139	0.086	0.080	0.062	1.000		
12. PI	-0.001	0.022	0.007	0.342	-0.210	-0.145	0.007	0.018	-0.017	0.022	-0.055	1.000	
13. ownership level	0.021	0.118	0.095	-0.367	0.581	0.611	0.058	0.008	0.032	-0.019	0.370	0.185	1.000

is significant and positive in all the specifications, controlling number of firms in the business group or diameter to measure group size, and with or without the PI[14] measure. The results show that both firm centrality and group size are in general positively correlated with sales growth during the crisis. Table 13 and Table 16 in Appendix 2.6 present the results using labour productivity and sales per employee as performance measure and the results are consistent. We also notice that the coefficients of the two specifications are very close. This is mainly due to the high correlation between the two measures of business group size. As shown in Table 4, the correlation between the diameter and the logarithm form of number of firms in the business group is as high as 0.946, which is in line with the theory in Watts and Strogatz (1998) and Barabási and Albert (1999).

We further explore the relation between firm centrality and performance given the group size[15]. To guarantee that each subsmaple has enough observations, we divide the firms according to the following intervals of the group size: [3, 5], [6, 7], [8, 13], [14, 50], [50, $+\infty$). The results are shown in Table 6. We notice that centrality is significant and positive when the component size is smaller than 50. However, the centrality is not significant under the significance level of 0.05 in the subsample of business groups of size between 8 and 13. Table 14 and 17 in Appendix 2.6 show the results using the other two performance measures. We find that when the business group size is larger than 7, the centrality is not always significant and sometimes even negative. The results imply that the centrality of Italian firms in the ownership network and their performance during the crisis are positively correlated if the business group size is smaller than 8.

able lying in the 1% tail of the distribution are excluded. We also make a robustness check by dropping the observations in the 5% tail of the distribution, and the results are similar and hence not reported here.

[14] The results in the rest of this chapter are reduced to the sample of firms making part of groups of size larger than 2. On one hand, the PI measure requires that the business groups have at least three companies. On the other hand, due to the data restriction, some groups of size 2 in our algorithm can be composed of an individual ultimate owner and a company. See Appendix 2.6.6 for more details.

[15] Since the log number of firms is highly correlated with diameter, we use only the former

Table 5: Regression Results of Centrality and Business Group Size

Dependent variable: 3-years average sales growth rate				
VARIABLES	(1)	(2)	(3)	(4)
centrality	0.155***	0.152***	0.0813***	0.0787***
	(0.00667)	(0.00666)	(0.0105)	(0.0105)
(log) No. firms	0.0471***		0.0268***	
	(0.00722)		(0.00692)	
diameter		0.0236***		0.0112***
		(0.00369)		(0.00320)
PI			0.0325***	0.0248***
			(0.00804)	(0.00788)
(log) No. employees	0.213***	0.213***	0.181***	0.181***
	(0.00264)	(0.00264)	(0.00412)	(0.00412)
(log) capital intensity	0.0239***	0.0240***	0.0182***	0.0184***
	(0.000832)	(0.000836)	(0.00145)	(0.00145)
(log) sales 2011	-0.230***	-0.230***	-0.200***	-0.199***
	(0.00252)	(0.00251)	(0.00388)	(0.00387)
mne	0.0872***	0.101***	0.0617***	0.0765***
	(0.00762)	(0.00707)	(0.00861)	(0.00775)
age	-0.00406***	-0.00407***	-0.00381***	-0.00383***
	(0.000135)	(0.000136)	(0.000247)	(0.000247)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	1.248***	1.256***	1.165***	1.176***
	(0.0127)	(0.0128)	(0.0223)	(0.0222)
Observations	244,542	244,542	79,744	79,744
R-squared	0.152	0.152	0.131	0.131

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6: Regression Results of Subsamples by Business Group Sizes

Dependent variable: 3-years average sales growth rate					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Size 3-5	Size 6-7	Size 8-13	Size 14-50	Size 50+
centrality	0.0842***	0.160***	0.0579*	0.101**	0.0859
	(0.0117)	(0.0380)	(0.0335)	(0.0457)	(0.128)
PI	0.0404***	0.0314	0.0718	0.130	1.389***
	(0.0082)	(0.0463)	(0.0659)	(0.117)	(0.487)
(log) No. employees	0.193***	0.152***	0.181***	0.140***	0.165***
	(0.00537)	(0.0136)	(0.0145)	(0.0109)	(0.0117)
(log) capital intensity	0.01797***	0.00470	0.0195***	0.0197***	0.0222***
	(0.00537)	(0.00528)	(0.00486)	(0.00439)	(0.00510)
(log) sales 2011	-0.213***	-0.187***	-0.212***	-0.171***	-0.171***
	(0.0048)	(0.0121)	(0.0140)	(0.0127)	(0.0106)
age	-0.00380***	-0.00450***	-0.00316***	-0.00340***	-0.00414***
	(0.000224)	(0.000659)	(0.00105)	(0.000570)	(0.000571)
mne	0.0506***	0.0158	0.0737***	0.0160	0.0838
	(0.0094)	(0.0204)	(0.0211)	(0.0305)	(0.0724)
sectors	Yes	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes	Yes
Constant	1.265***	1.264***	1.346***	1.121***	0.925***
	(0.0279)	(0.0737)	(0.0882)	(0.0901)	(0.0960)
Observations	54,472	6,505	6,363	6,433	5,971
R-squared	0.137	0.132	0.147	0.138	0.113

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Then we investigate the relationship between firm performance and the size of the group it belongs to, given a certain level of centrality. According to the centrality distribution, we consider the firms with centrality smaller than 0.1 as peripheral and no smaller than 0.5 as central. The first two columns in Table 7 report the regression results by using the subsamples of central and peripheral firms. We notice that the business group size (in $\log$) is significant and takes a positive sign in both specifications, implying a positive relationship between firm's sales growth rate and group size for both the central and peripheral ones.

We further explore their relation by using dummies of group size intervals. The results are shown in the third and fourth column of Table 7. Using the business group size of 3 as benchmark, we find that the dummy variables of group size intervals are statistically significant and take a positive sign. What's more, we notice that the coefficients of the size intervals dummies in the subsample of peripheral firms increase with group size. In other words, as group size increases, the sales growth difference between an subsidiary in a group of size larger than 3 firms and that in a group of 3 firms becomes larger. We check this by using the other two performance measures and find that the coefficients also increase with group size, though it decreases once the group size exceeds 50 (see details in Table 15 and Table 18 in Appendix 2.6). More interestingly, when the group size is no smaller than 8, the coefficients in the subsample of peripheral firms are even larger than those in the subsample of central firms. This still holds when we use the other two measures as dependent variables. The results suggest that there is a larger performance gap between Italian firms located in the periphery of the networks of large (size ≥ 8) and of small business groups (size $= 3$), compared with that of the firms located in central positions in the networks of large and small groups. This is probably due to that though firms at a central position in larger business groups can facilitate the access to funds and information, they face higher coordination costs than in small groups. However, for firms at a peripheral position in larger groups, the benefit of the reputation effect and the availability to diverse

to measure business group size here and below.

resources can be much larger than that in small groups. In the context of financial crisis, being a member of a large group also makes the peripheral firms more resilient to the exogenous financial shocks and faster to recover compared with in small groups.

More interestingly, we notice that the PI measure is significant in all the specifications in Table 7. It takes a positive sign using the subsamples of peripheral firms while a negative sign using those of the central firms. The results imply that the peripheral firms have better performance in groups with more hierarchical structure while the central firms have better performance in groups with flatter structure.

We also conduct another approach to compare the performance of firms in large and small groups given similar position in the network. We first select all the global business groups of size larger than 3 including at least 1 Italian firm. Then we randomly remove 10% of the links from these business groups' ownership structure, and obtain some "faked" business groups of size 3. Based on their position in the ownership structure, we divide them into three sets, which are respectively headquarters, intermediate-level firms and bottom-level subsidiaries. These treated firms are then matched with firms at the corresponding position in the business groups of original size 3 by region, sector, age and sales. We use the Coarsened Exact Matching (CEM) method provided by Iacus et al. (2008), which allows using a regression model to deal with the remained imbalance after matching. Controlling for the number of employees, capital intensity and sales (in log) in 2011, we estimate the treatment effect on sales growth. The aforementioned steps started from the random split are repeated for 30 times and we derive 30 estimation results for each of the three sets. The results of the treatment effect are shown in Figure 6. We can notice that in the bottom-level subsidiaries set, there is a significantly positive effect in almost all samples. While the set of headquarters has the lowest proportion of the significantly positive effect. These findings further support the previous results, that is, the performance gap in large and small groups for the peripheral subsidiaries is bigger than for the central ones.

We further compare the peripheral firms in business groups with the

Table 7: Regression Results of Subsamples by Centrality Levels

	Dependent variable: 3-years average sales growth rate			
	(1)	(2)	(3)	(4)
VARIABLES	Periphery	Central	Periphery	Central
(log) BG size	0.0213***	0.0648***		
	(0.00692)	(0.0121)		
BG size 4-5			0.0380***	0.0586***
			(0.00772)	(0.0153)
BG size 6-7			0.0779***	0.113***
			(0.0125)	(0.0243)
BG size 8-13			0.105***	0.0995***
			(0.0138)	(0.0260)
BG size 14-50			0.121***	0.117***
			(0.0153)	(0.0305)
BG size 50+			0.139***	0.103*
			(0.0217)	(0.0577)
PI	0.0772***	-0.0576***	0.0735***	-0.0402**
	(0.0114)	(0.0156)	(0.0115)	(0.0162)
(log) No. employees	0.186***	0.161***	0.185***	0.162***
	(0.00482)	(0.00842)	(0.00480)	(0.00842)
(log) capital intensity	0.0189***	0.0161***	0.0182***	0.0153***
	(0.00168)	(0.00312)	(0.00167)	(0.00311)
(log) sales 2011	-0.205***	-0.185***	-0.207***	-0.185***
	(0.00450)	(0.00817)	(0.00448)	(0.00817)
age	-0.00425***	-0.00236***	-0.00422***	-0.00229***
	(0.000338)	(0.000388)	(0.000331)	(0.000386)
mne	0.0806***	0.00212	0.0475***	0.000601
	(0.0110)	(0.0124)	(0.0108)	(0.0124)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	1.202***	1.154***	1.222***	1.201***
	(0.0255)	(0.0562)	(0.0264)	(0.0562)
Observations	61,074	13,491	61,074	13,491
R-squared	0.136	0.123	0.138	0.124

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

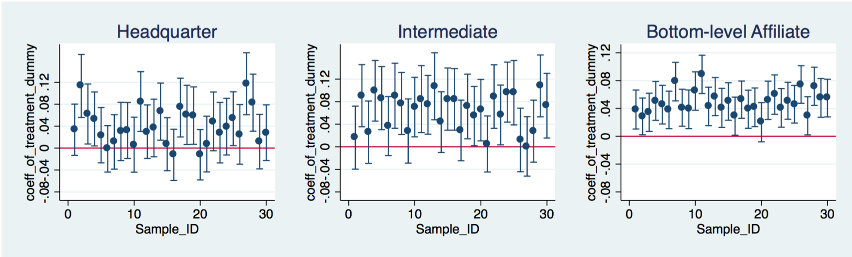

Figure 6: Treatment Effect of Firms Split from Business Groups of Size > 3

standalone ones using a matching approach. 112,792 Italian standalone firms with full financial data in AIDA in 2014 are used as the control group. The standalone ones are matched with the treatment group, that is, the peripheral firms (with centrality value less than 0.1) in business groups by region, sector, age and sales. We also use the CEM method here. The regression results are presented in Table 8. We notice that the coefficient of the treatment group is significant and takes a positive sign in all the specifications. Moreover, the coefficient grows with the business group size, though it decreases a bit once the group size exceeds 50. We make another robustness check by restricting the sample to the firms with no subsidiary, that is, with centrality value of 0. The regression results reported in Table 19 in Appendix 2.6.4 are consistent. The results show that the Italian subsidiaries located in the periphery of large groups' ownership network have a higher sales growth rate than the standalone ones during the crisis. The findings above suggest that it is more beneficial for the peripheral subsidiaries in small business groups and the standalone firms to be merged into larger business groups where they can gain more advantages and are more resilient to financial shocks in the crisis.

Table 8: Regression of Peripheral Firms' Subsamples by BG Sizes after CEM

	Dependent variable: 3-years average sales growth rate					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	BG size 2-3	BG size 4-5	BG size 6-7	BG size 8-13	BG size 14-50	BG size 50+
treatment	0.0203***	0.105***	0.144***	0.161***	0.173***	0.201***
	(0.00272)	(0.00618)	(0.0102)	(0.00987)	(0.00890)	(0.00881)
(log) No. employees	0.219***	0.183***	0.181***	0.184***	0.176***	0.175***
	(0.00160)	(0.00237)	(0.00241)	(0.00231)	(0.00224)	(0.00233)
(log) capital intensity	0.0139***	0.0151***	0.0162***	0.0166***	0.0174***	0.0142***
	(0.000591)	(0.000869)	(0.000903)	(0.000880)	(0.000866)	(0.000870)
(log) sales 2011	-0.229***	-0.190***	-0.189***	-0.184***	-0.183***	-0.181***
	(0.00105)	(0.00154)	(0.00158)	(0.00152)	(0.00152)	(0.00156)
Constant	1.169***	0.966***	0.959***	0.930***	0.936***	0.926***
	(0.00561)	(0.00791)	(0.00807)	(0.00773)	(0.00778)	(0.00794)
Observations	275,678	104,177	94,401	96,236	96,296	95,961
R-squared	0.149	0.129	0.133	0.134	0.133	0.126

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

2.4.3 Comparison between Centrality Measure and Ownership Level

In this part we compare our centrality measure with the one in Belenzon et al. (2017). By analysing the data on the structure of corporate groups in Western Europe, they find that the focal subsidiaries with greater organizational distance from parent companies have lower sales growth rate and their performance is more similar to that of the matched standalones in response to changing industry conditions. They adopt the ownership level to measure the organizational distance in the hierarchical structure of the group, which is defined as the number of intermediate subsidiaries separating a focal subsidiary from parent company plus 1.

We also explore the relation between ownership level and firm growth using the data of Italian firms. The summary statistics of growth rate by ownership level is present in Table 9.[16] We can notice that on average, the Italian subsidiaries with a longer organizational distance from the ultimate owner in the ownership network have a larger sales growth rate during the crisis.

Table 9: 3-years Average Sales Growth by Ownership Level

	Sample: group size ≥ 2			Sample: group size ≥ 3		
Ownership level	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
1	0.133	0.717	10,480	0.126	0.676	3,682
2	0.159	0.716	211,640	0.162	0.736	53,595
3	0.183	0.766	16,521	0.183	0.766	16,521
4	0.182	0.775	4,012	0.182	0.775	4,012
5	0.196	0.822	1,251	0.196	0.822	1,251
6	0.208	0.858	714	0.208	0.858	714
Total	0.160	0.722	244,618	0.167	0.744	79,775

We use the ownership level in place of our centrality measure in the previous models. As Table 10 shows, its coefficient is significant and takes a positive sign in most of the specifications. We also make a robustness check by matching each firm in the sample with a standalone and use the difference of sales growth rate between them as the dependent

[16]Following Belenzon et al. (2017)'s strategy, we replace the values of ownership level larger than 6 with 6.

variable. The results reported in Table 21 in Appendix 2.6.5 also show that the coefficient of ownership level is significantly positive. The results seems to contradict those of Belenzon et al. (2017). This might be due to that their measure is highly correlated with group size, given that their correlation coefficient in our data is as high as 0.581 (see Table 4). Another possible explanation is that we focus on Italian business groups in the crisis period while they consider European groups before the crisis.

Table 10: Regression Results of Ownership level

Dependent variable: 3-years average sales growth rate				
VARIABLES	(1)	(2)	(3)	(4)
ownership level	0.0233***	-0.00530	0.0219***	0.0115*
	(0.00356)	(0.00665)	(0.00414)	(0.00622)
(log) BG size		0.0420***		0.0168**
		(0.00891)		(0.00748)
PI			0.0313***	0.0461***
			(0.00765)	(0.00909)
(log) No. employees	0.218***	0.217***	0.187***	0.186***
	(0.00264)	(0.00267)	(0.00413)	(0.00414)
(log) capital intensity	0.0257***	0.0247***	0.0179***	0.0175***
	(0.000823)	(0.000830)	(0.00143)	(0.00144)
(log) sales 2011	-0.226***	-0.228***	-0.198***	-0.199***
	(0.00241)	(0.00250)	(0.00378)	(0.00383)
industry growth	0.103***	0.0952***	0.108***	0.105***
	(0.0117)	(0.0121)	(0.0148)	(0.0149)
mne	0.154***	0.101***	0.0881***	0.0673***
	(0.00641)	(0.00740)	(0.00763)	(0.00811)
age	-0.00365***	-0.00371***	-0.00339***	-0.00341***
	(0.000130)	(0.000131)	(0.000231)	(0.000231)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	1.171***	1.185***	1.100***	1.093***
	(0.0140)	(0.0150)	(0.0231)	(0.0228)
Observations	244,538	244,538	79,744	79,744
R-squared	0.149	0.150	0.132	0.132

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To separate the effect of group size on the measure, we split the data into subsamples based on the group size. As Table 11 shows, the coefficient of ownership level, though not significant, is negative if group size is small and becomes positive when it exceeds 14. If we use the sales growth rate difference between a focal firm and a matched standalone one as the dependent variable, the results are consistent (see Table 21). By considering the effect of group size, we find that the positive effect of ownership level is mainly in large groups. In other words, the Italian subsidiaries at more bottom level in the ownership network of larger business group (size ≥ 14) have a relatively better performance in the crisis. While in small business groups, their performance is worse than those closer to the headquarters in the network, which is in line with our previous findings.

Compared with ownership level, our centrality measure resolves the problem of its high collinearity with group size. It is also more informative since it takes into account the number of steps of a focal firm to all its direct and indirect controlling subsidiaries, which reflects to a certain extent its coordinating power in the group's ownership network.

2.5 Conclusion

In this chapter, we empirically investigate the relationship between the interaction of firm centrality and business group size and firm performance. Specifically, we find a positive relationship between firm centrality and performance in small business groups, which however is not always significant in larger business groups. We also find a positive relationship between firm performance and group size for the peripheral subsidiaries, and the group size premium for them is larger than for the central firms. The findings suggest that there can be a trade-off effect on the performance for firms at central positions in large business groups. Though they can facilitate the access to funds and information, they possibly face higher coordination costs than in small groups. But for firms located in the periphery of larger groups' ownership network, the benefit of the reputation effect and the availability to diverse resources can be much larger than that in small groups. Especially during the crisis, peripheral subsidiaries of a large business group are more resilient to the exogenous financial shocks compared with similar ones in small groups. The findings implicate that it is beneficial for the peripheral subsidiaries

Table 11: Regression Results of Ownership level by Business Group Size

Dependent variable: 3-years average sales growth rate					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Size 3-5	Size 6-7	Size 8-13	Size 14-50	Size 50+
ownership level	-0.00588	-0.0347*	-0.0160	0.00969	0.00962
	(0.00737)	(0.0187)	(0.0144)	(0.0116)	(0.0100)
PI	0.0639***	0.139**	0.145*	0.0784	1.105**
	(0.00874)	(0.0646)	(0.0840)	(0.150)	(0.470)
(log) No. employees	0.196***	0.162***	0.185***	0.149***	0.175***
	(0.00537)	(0.0135)	(0.0145)	(0.0118)	(0.0124)
(log) capital intensity	0.0181***	0.00446	0.0179***	0.0167***	0.0192***
	(0.00177)	(0.00533)	(0.00480)	(0.00447)	(0.00470)
(log) sales 2011	-0.211***	-0.186***	-0.210***	-0.170***	-0.174***
	(0.00478)	(0.0121)	(0.0139)	(0.0126)	(0.0111)
industry growth	0.0954***	0.162***	0.0625*	0.0890***	0.108**
	(0.0202)	(0.0504)	(0.0357)	(0.0328)	(0.0485)
mne	0.0609***	0.0266	0.0773***	0.0195	0.0648
	(0.00928)	(0.0202)	(0.0213)	(0.0304)	(0.0732)
age	-0.00356***	-0.00398***	-0.00298***	-0.00282***	-0.00371***
	(0.000222)	(0.000635)	(0.000996)	(0.000562)	(0.000567)
sectors	Yes	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes	Yes
Constant	1.205***	1.177***	1.303***	1.016***	0.871***
	(0.0313)	(0.0787)	(0.0931)	(0.0810)	(0.0968)
Observations	54,472	6,505	6,363	6,433	5,971
R-squared	0.137	0.133	0.148	0.140	0.116

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

in small groups and the standalone firms to be merged into larger business groups.

Our contribution to the literature on ownership network of business group is twofold. First, we provide a new measure of centrality. Considering that the ownership level measure used in Belenzon et al. (2017) has a high correlation with group size, we adopt the normalized harmonic centrality measure to reduce the size effect. Consequently, we make centrality more comparable across groups of different sizes. Second, though the effect of individual-level centrality and group-level complexity on firm performance has been separately discussed in literature, the relation between their interaction and firm performance is underexploited. By performing a detailed analysis with multiple robustness checks, we shed light on the role of a firm's network position in groups of different sizes on its performance.

The current study is subject to several limitations. First, we have only cross-section data of the ownership structure, which is standard in the literature since it is difficult to trace its change over time. Due to this restriction, we cannot test the issue of causality. In our analysis, we find that the peripheral firms in large business groups are much more productive than the standalone ones. However, we cannot tell whether these firms are already productive before being merged into the business group or they improve their performance after the take-over.

Second, we have limited information about the shareholders in the current data. As a result, a business group defined in our work may also include its ultimate beneficial owner who controls the parent company. Though it is necessary to distinguish different types of ownership, how to deal with the individual or family ultimate beneficial owners requires more attention. If firms are directly controlled by the same ultimate beneficial owner(s), they can be viewed as members of a group rather than independent since their business can be related in some way. Cainelli and Iacobucci (2011) and Belenzon et al. (2017) define these firms as members of a business group. In the case of Italy, business groups controlled by family are quite common (Corbetta and Tomaselli, 1996; Cucculelli and Micucci, 2008) and this case of business groups should be taken more attention.

Finally, in our data we have only the balance sheet data of the Italian domestic firms but not for the foreign firms in these business groups. The existing studies also control the average performance of all the firms within a business group. We will add these group-level measures in the future once the data is available. Moreover, with the financial data of firms in other countries, we can also extend our analysis to the global scale and explore the difference of firm performance in different countries.

2.6 Appendix

2.6.1 Variables Definition

Table 12: Variables Definition

Variables	Definition
3-years sales growth	$(\sum_{i=2011}^{2013}(sales_{i+1} - sales_i)/sales_i)/3$.
centrality	A measure of firm's centrality in a business group.
BG size	The number of firms that belong to a business group.
diameter	The longest of all the shortest paths in a network.
labour productivity	$value\ added/No.\ employees$ in 2014.
sales per employee	$sales/No.\ employees$ in 2014.
capital intensity	$fixed\ assets/No.\ employees$ in 2014.
mne	A dummy variable equal to 1 if a firm belongs to a multinational group, otherwise 0.
ownership level	The number of subsidiaries separating a focal subsidiary from parent company plus 1.
PI	The pyramidal index defined by Belenzon et al. (2013).
sectors	A set of dummy variables that equal to 1 if a firm belongs to a certain sector, and sector 4 is used as benchmark in the regression. 1: Agriculture, forestry and fishing. 2: Manufacturing, mining and quarrying and other industry. 3: Construction. 4: Wholesale and retail trade, transportation and storage, accommodation and food service activities. 5: Information and communication. 6: Financial and insurance activities. 7: Real estate activities (includeing imputed rents of owner-occupied dwellings). 8: Professional, scientific, technical, administration and support service activities. 9: Public administration, defence, education, human health and social work activities. 10: Other services.
regions	A set of dummy variables that takes value of 1 if a firm belongs to a certain region and region 2 is used as benchmark in the regression. 1: NordOvest. 2: NordEst. 3: Centro. 4: Mezzogiorno.

2.6.2 Robustness Check 1: Using Labor Productivity as Dependent Variable

Table 13: Regression Results of Centrality and Different BG Size Measures

Dependent variable: log labor productivity				
VARIABLES	(1)	(2)	(3)	(4)
centrality	0.244***	0.239***	0.114***	0.111***
	(0.00855)	(0.00863)	(0.0142)	(0.0143)
(log) No. firms	0.104***		0.0552***	
	(0.0136)		(0.0149)	
diameter		0.0530***		0.0248***
		(0.00682)		(0.00688)
PI			0.0749***	0.0605***
			(0.0106)	(0.0103)
(log) No. employees	0.188***	0.188***	0.126***	0.126***
	(0.00247)	(0.00238)	(0.00385)	(0.00373)
(log) capital intensity	0.157***	0.157***	0.152***	0.153***
	(0.00106)	(0.00105)	(0.00187)	(0.00186)
age	0.00303***	0.00301***	0.00153***	0.00148***
	(0.000168)	(0.000168)	(0.000273)	(0.000271)
mne	0.151***	0.180***	0.195***	0.223***
	(0.0119)	(0.0108)	(0.0150)	(0.0130)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	2.254***	2.273***	2.556***	2.578***
	(0.00986)	(0.00807)	(0.0211)	(0.0178)
Observations	285,309	285,309	88,070	88,070
R-squared	0.239	0.238	0.199	0.199

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Regression Results of Subsamples by Business Group Sizes

	Dependent variable: (log) labor productivity				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	BG size 3-5	BG size 6-7	BG size 8-13	BG size 14-50	BG size 50+
centrality	0.156***	0.150***	-0.0694	-0.0509	-0.239*
	(0.0159)	(0.0432)	(0.0485)	(0.0568)	(0.122)
PI	0.0656***	0.0856	0.0123	0.310	3.283***
	(0.0106)	(0.0655)	(0.0935)	(0.207)	(0.870)
(log) No. employees	0.163***	0.0688***	0.0452***	0.0281***	0.119***
	(0.00360)	(0.00917)	(0.00935)	(0.0103)	(0.0121)
(log) capital intensity	0.163***	0.148***	0.139***	0.129***	0.112***
	(0.00206)	(0.00606)	(0.00630)	(0.00675)	(0.0100)
age	0.00175***	0.000838	-0.000339	0.00151*	0.00338***
	(0.000305)	(0.000811)	(0.000598)	(0.000846)	(0.000956)
mne	0.127***	0.154***	0.195***	0.209***	0.391***
	(0.0145)	(0.0299)	(0.0290)	(0.0521)	(0.121)
sectors	Yes	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes	Yes
Constant	2.485***	2.922***	3.116***	3.285***	2.745***
	(0.0142)	(0.0504)	(0.0580)	(0.0830)	(0.143)
Observations	61,744	6,905	6,622	6,557	6,242
R-squared	0.194	0.179	0.166	0.145	0.154

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: Regression Results of Subsamples by Centrality Levels

Dependent variable: (log) labor productivity				
	(1)	(2)	(3)	(4)
VARIABLES	Periphery	Central	Periphery	Central
(log) BG size	0.0370**	0.0611***		
	(0.0144)	(0.0175)		
BG size 4-5			0.119***	0.0832***
			(0.00995)	(0.0199)
BG size 6-7			0.230***	0.166***
			(0.0168)	(0.0313)
BG size 8-13			0.314***	0.0693*
			(0.0188)	(0.0362)
BG size 14-50			0.349***	0.0912**
			(0.0255)	(0.0428)
BG size 50+			0.210***	0.154*
			(0.0379)	(0.0844)
PI	0.118***	-0.177***	0.0947***	-0.149***
	(0.0137)	(0.0218)	(0.0136)	(0.0230)
(log) No. employees	0.145***	0.0752***	0.138***	0.0758***
	(0.00437)	(0.00610)	(0.00390)	(0.00612)
(log) capital intensity	0.152***	0.156***	0.150***	0.155***
	(0.00215)	(0.00414)	(0.00212)	(0.00415)
age	0.00134***	0.00240***	0.00139***	0.00246***
	(0.000318)	(0.000513)	(0.000325)	(0.000512)
mne	0.260***	0.0927***	0.180***	0.0914***
	(0.0194)	(0.0191)	(0.0167)	(0.0191)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	2.523***	2.976***	2.523***	3.005***
	(0.0212)	(0.0455)	(0.0145)	(0.0410)
Observations	69,432	13,495	69,432	13,495
R-squared	0.193	0.189	0.201	0.191

2.6.3 Robustness Check 2: Using Sales per Employee as Dependent Variable

Table 16: Regression Results of Centrality and Different BG Size Measures

Dependent variable: log sales per employee				
VARIABLES	(1)	(2)	(3)	(4)
centrality	0.299***	0.295***	0.156***	0.156***
	(0.00949)	(0.00957)	(0.0153)	(0.0154)
(log) No. firms	0.0953***		0.0412***	
	(0.0130)		(0.0119)	
diameter		0.0503***		0.0201***
		(0.00673)		(0.0056)
PI			0.0466***	0.0363***
			(0.0115)	(0.0113)
(log) No. employees	0.0849***	0.0848***	0.104***	0.104***
	(0.00264)	(0.00261)	(0.00379)	(0.00375)
(log) capital intensity	0.0932***	0.0934***	0.0912***	0.0913***
	(0.00112)	(0.00111)	(0.00195)	(0.00194)
age	0.00373***	0.00370***	0.00228***	0.00224***
	(0.000184)	(0.000184)	(0.000315)	(0.000314)
mne	0.204***	0.226***	0.217***	0.234***
	(0.0120)	(0.0110)	(0.0145)	(0.0129)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	4.142***	4.158***	4.250***	4.263***
	(0.00958)	(0.00807)	(0.0199)	(0.0177)
Observations	299,871	299,871	92,800	92,800
R-squared	0.162	0.162	0.153	0.153

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 17: Regression Results of Subsamples by Business Group Sizes

Dependent variable: (log) sales per employee					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Size 3-5	Size 6-7	Size 8-13	Size 14-50	Size 50+
centrality	0.198***	0.143***	-0.00595	0.0729	-0.218
	(0.0178)	(0.0478)	(0.0490)	(0.0622)	(0.142)
PI	0.0427***	-0.00752	-0.0221	0.212	3.393***
	(0.0118)	(0.0731)	(0.108)	(0.211)	(0.780)
(log) No. employees	0.117***	0.0778***	0.0584***	0.0579***	0.116***
	(0.00405)	(0.0106)	(0.0107)	(0.0106)	(0.0117)
(log) capital intensity	0.0929***	0.0912***	0.0750***	0.0732***	0.0919***
	(0.00221)	(0.00652)	(0.00639)	(0.00652)	(0.00843)
age	0.00276***	0.000828	0.000195	0.00310***	0.00288***
	(0.000331)	(0.000923)	(0.000841)	(0.000932)	(0.000960)
mne	0.166***	0.154***	0.250***	0.225***	0.367***
	(0.0156)	(0.0325)	(0.0338)	(0.0525)	(0.124)
sectors	Yes	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes	Yes
Constant	4.247***	4.439***	4.658***	4.604***	4.290***
	(0.0163)	(0.0609)	(0.0647)	(0.0964)	(0.144)
Observations	64,505	7,399	7,237	7,080	6,579
R-squared	0.140	0.120	0.124	0.119	0.190

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18: Regression Results of Subsamples by Centrality Levels

Dependent variable: (log) sales per employee				
	(1)	(2)	(3)	(4)
VARIABLES	Periphery	Central	Periphery	Central
(log) BG size	0.0341***	0.0567***		
	(0.0119)	(0.0197)		
BG size 4-5			0.0848***	0.0520**
			(0.0111)	(0.0226)
BG size 6-7			0.186***	0.107***
			(0.0195)	(0.0344)
BG size 8-13			0.262***	0.0440
			(0.0214)	(0.0375)
BG size 14-50			0.296***	0.120**
			(0.0262)	(0.0490)
BG size 50+			0.191***	0.0239
			(0.0339)	(0.115)
PI	0.0830***	-0.141***	0.0678***	-0.130***
	(0.0151)	(0.0244)	(0.0152)	(0.0256)
(log) No. employees	0.111***	0.0770***	0.106***	0.0776***
	(0.00445)	(0.00693)	(0.00415)	(0.00695)
(log) capital intensity	0.0862***	0.116***	0.0838***	0.116***
	(0.00225)	(0.00429)	(0.00221)	(0.00430)
age	0.00212***	0.00240***	0.00220***	0.00245***
	(0.000373)	(0.000585)	(0.000383)	(0.000583)
mne	0.240***	0.159***	0.172***	0.161***
	(0.0185)	(0.0216)	(0.0177)	(0.0217)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	4.245***	4.515***	4.253***	4.558***
	(0.0205)	(0.0523)	(0.0162)	(0.0468)
Observations	72,618	14,663	72,618	14,663
R-squared	0.145	0.179	0.149	0.179

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.6.4 Robustness Check 3: Matching the Peripheral firms with the Standalone Ones

Table 19: Regression of Peripheral Firms' Subsamples by BG Sizes after CEM

	Dependent variable: 3-years average sales growth rate					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	BG size 2-3	BG size 4-5	BG size 6-7	BG size 8-13	BG size 14-50	BG size 50+
treatment	0.020***	0.104***	0.143***	0.162***	0.175***	0.195***
	(0.00269)	(0.00612)	(0.0101)	(0.00988)	(0.00949)	(0.00970)
(log) No. employees	0.217***	0.181***	0.180***	0.182***	0.179***	0.180***
	(0.00159)	(0.00234)	(0.00239)	(0.00229)	(0.00227)	(0.00238)
(log) capital intensity	0.0136***	0.0147***	0.0158***	0.0162***	0.0172***	0.0144***
	(0.000585)	(0.000860)	(0.000894)	(0.000873)	(0.000875)	(0.000883)
(log) sales 2011	-0.227***	-0.188***	-0.187***	-0.183***	-0.187***	-0.185***
	(0.00104)	(0.00153)	(0.00157)	(0.00151)	(0.00153)	(0.00157)
Constant	1.159***	0.958***	0.952***	0.922***	0.956***	0.946***
	(0.00555)	(0.00782)	(0.00799)	(0.00766)	(0.00786)	(0.00804)
Observations	275,637	104,169	94,395	96,129	95,051	93,933
R-squared	0.150	0.129	0.133	0.134	0.137	0.131

Notes: the perfipheral firms here are restricted to those with centrality value of 0.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.6.5 Robustness Check 4: Ownership Level

Table 20: Regression Results of Ownership level

Dependent variable: sales growth subsidiary - sales growth standalone				
VARIABLES	(1)	(2)	(3)	(4)
ownership level	0.0272***	0.0207***	0.0243***	0.0123*
	(0.00420)	(0.00570)	(0.00430)	(0.00674)
(log) BG size		0.0107*		0.0169**
		(0.00643)		(0.00753)
PI			0.0295***	0.0452***
			(0.00768)	(0.00927)
(log) No. employees	0.187***	0.187***	0.187***	0.186***
	(0.00413)	(0.00414)	(0.00413)	(0.00414)
(log) capital intensity	0.0184***	0.0182***	0.0181***	0.0176***
	(0.00142)	(0.00142)	(0.00142)	(0.00143)
(log) sales 2011	-0.199***	-0.199***	-0.199***	-0.199***
	(0.00378)	(0.00381)	(0.00378)	(0.00383)
industry growth	0.106***	0.104***	0.107***	0.105***
	(0.0149)	(0.0150)	(0.0149)	(0.0150)
mne	0.0829***	0.0687***	0.0867***	0.0661***
	(0.00772)	(0.00794)	(0.00762)	(0.00812)
age	-0.00318***	-0.00319***	-0.00319***	-0.00321***
	(0.000181)	(0.000181)	(0.000181)	(0.000181)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	0.917***	0.927***	0.922***	0.940***
	(0.0257)	(0.0274)	(0.0258)	(0.0286)
Observations	79,688	79,688	79,684	79,684
R-squared	0.126	0.126	0.126	0.126

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21: Regression Results of Ownership level by Business Group Size

Dependent variable: sales growth subsidiary - sales growth standalone					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Size 3-5	Size 6-7	Size 8-13	Size 14-50	Size 50+
ownership level	-0.00479	-0.0348*	-0.0180	0.0113	0.0115
	(0.00738)	(0.0186)	(0.0147)	(0.0132)	(0.0145)
PI	0.0626***	0.140**	0.146*	0.0800	1.190**
	(0.00875)	(0.0647)	(0.0835)	(0.152)	(0.481)
(log) No. employees	0.196***	0.161***	0.188***	0.149***	0.176***
	(0.00537)	(0.0135)	(0.0145)	(0.0118)	(0.0125)
(log) capital intensity	0.0180***	0.00449	0.0197***	0.0167***	0.0197***
	(0.00177)	(0.00534)	(0.00475)	(0.00447)	(0.00478)
(log) sales 2011	-0.211***	-0.186***	-0.210***	-0.171***	-0.178***
	(0.00479)	(0.0121)	(0.0139)	(0.0126)	(0.0110)
industry growth	0.0987***	0.163***	0.0550	0.0908***	0.102**
	(0.0202)	(0.0508)	(0.0356)	(0.0327)	(0.0502)
mne	0.0598***	0.0248	0.0752***	0.0167	0.0625
	(0.00928)	(0.0202)	(0.0213)	(0.0304)	(0.0747)
age	-0.00319***	-0.00364***	-0.00413***	-0.00254***	-0.00318***
	(0.000226)	(0.000646)	(0.000589)	(0.000584)	(0.000590)
sectors	Yes	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes	Yes
Constant	1.081***	1.118***	1.230***	0.868***	0.730***
	(0.0388)	(0.0963)	(0.105)	(0.0880)	(0.110)
Observations	54,469	6,502	6,359	6,420	5,934
R-squared	0.131	0.122	0.146	0.136	0.111

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.6.6 Robustness Check 5: A New Algorithm to Identify the Individual Shareholders and Results

Though there is no information on the ownership type in the original data, we make some further attempts to identify the possible ownership links by the non-companies such as individual, family or state through a new algorithm. In the main text, we do not differentiate between the ultimate beneficial owners and the parent companies, which results in that some groups belonging to type (a) in Figure 7 are treated as type (c). The new algorithm first identifies all the nodes with no inward link in the groups where there is at least one Italian company or non-company.[17] If they have shareholders data in the original ORBIS database but with only minority shareholders, or their BvD ID number can be found in AIDA, we can confirm that they are the parent companies. Otherwise, we check the number of outward links they have. If they have more than 1 outward link, we assume that they are individual or family ultimate beneficial owners who control multiple companies (see type (b) in Figure 7), thus the group size of type (b) is 1 less than the previous measure. If they have only 1 outward link, we assume that they are the ultimate controlling shareholder of the parent company (see type (a)). We then remove their links to the parent company and recompute the centrality and the other network measures for the firms belonging to groups of type (a). Some groups of size 2 are thus divided into two disconnected nodes, and the firms used to be part of such groups are excluded from the sample. We finally derive a new sample of 183,177 Italian firms that belong to business groups. Table 22 presents some descriptive statistics. Nevertheless, the main regression results of the new algorithm and the previous one in this chapter are generally consistent (see details in Table 23, Table 24 and Table 25). It should be noticed that this new algorithm may overly drop some ownership links actually by companies or institutions whose shareholders data are not covered by ORBIS. The results reported here only aim to provide further support for our findings in this chapter.

[17] The BvD ID number in ORBIS provides the nationality information of an entity or individual.

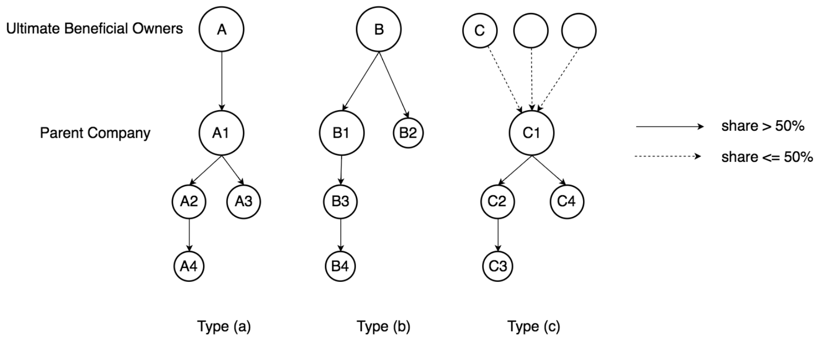

Figure 7: Three Cases of Business Groups' Ownership Structure

Table 22: Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) p10	(5) p50	(6) p90
3-years sales growth	92,177	0.915	23.88	-0.189	0.0128	0.535
centrality	183,177	0.197	0.375	0	0	1
No. firms in BG	183,177	34.89	175.3	2	3	22
diameter	183,177	2.673	2.223	1	2	5
employees	154,975	26.49	265.3	1	2	36
fixed assets	155,089	3,644	127,495	0	72	3,903
capital intensity	154,617	725.3	8,852	0	10	1,061
value added	155,450	1,989	34,599	-17	91	2,266
labour productivity	154,975	72.95	1,102	-13	29	122
age	183,142	16.43	13.92	4	12	35
mne	183,177	0.174	0.379	0	0	1
sales 2014	155,395	9,407	177,714	0	305	9,237
sales per labour 2014	154,920	354.5	3,623	0	88.50	540.5
sales 2013	111,294	12,718	219,270	26	709	13,564
sales 2012	107,151	13,203	228,039	22	741	13,842
sales 2011	102,349	13,640	213,122	18	787	14,921

Table 23: Regression Results of Centrality and Business Group Size

Dependent variable: 3-years average sales growth rate				
VARIABLES	(1)	(2)	(3)	(4)
centrality	0.0624***	0.0611***	0.106***	0.103***
	(0.00763)	(0.00778)	(0.0111)	(0.0112)
(log) No. firms	0.0267***		0.0306***	
	(0.00663)		(0.00720)	
diameter		0.0118***		0.0123***
		(0.00318)		(0.00334)
PI			0.0886***	0.0742***
			(0.0166)	(0.0169)
(log) No. employees	0.194***	0.194***	0.193***	0.193***
	(0.00435)	(0.00435)	(0.00489)	(0.00489)
(log) capital intensity	0.0188***	0.0190***	0.0197***	0.0200***
	(0.00151)	(0.00151)	(0.00169)	(0.00170)
(log) sales 2011	-0.218***	-0.218***	-0.217***	-0.217***
	(0.00407)	(0.00407)	(0.00459)	(0.00459)
mne	0.0656***	0.0785***	0.0587***	0.0784***
	(0.00909)	(0.00830)	(0.0102)	(0.00910)
age	-0.00388***	-0.00390***	-0.00402***	-0.00405***
	(0.000234)	(0.000235)	(0.000284)	(0.000285)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	1.286***	1.297***	1.267***	1.280***
	(0.0239)	(0.0237)	(0.0268)	(0.0266)
Observations	91,196	91,196	71,207	71,207
R-squared	0.127	0.127	0.128	0.127

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24: Regression Results of Subsamples by Business Group Sizes

	Dependent variable: 3-years average sales growth rate					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Size 3-4	Size 5-7	Size 8-11	Size 12-22	Size 23-84	Size 85+
centrality	0.0603***	0.137***	0.0563	0.240***	0.0599	0.177
	(0.0140)	(0.0291)	(0.0414)	(0.0667)	(0.0628)	(0.273)
PI	0.0402**	0.00876	-0.0269	0.259	0.0437	1.763
	(0.0191)	(0.0545)	(0.0900)	(0.183)	(0.267)	(1.143)
(log) No. employees	0.198***	0.173***	0.175***	0.184***	0.153***	0.173***
	(0.00897)	(0.0130)	(0.0183)	(0.0173)	(0.0166)	(0.0137)
(log) capital intensity	0.0232***	0.00880*	0.0216***	0.0243***	0.0184**	0.0206***
	(0.00295)	(0.00499)	(0.00580)	(0.00629)	(0.00757)	(0.00556)
(log) sales 2011	-0.221***	-0.213***	-0.207***	-0.241***	-0.173***	-0.184***
	(0.00832)	(0.0123)	(0.0170)	(0.0208)	(0.0171)	(0.0132)
age	-0.00376***	-0.00371***	-0.00406***	-0.00442***	-0.00440***	-0.00370***
	(0.000382)	(0.00104)	(0.000730)	(0.000843)	(0.000685)	(0.000653)
mne	0.0421***	-0.000948	0.0523**	0.000832	0.116***	-
	(0.0147)	(0.0180)	(0.0263)	(0.0368)	(0.0420)	-
sectors	Yes	Yes	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.333***	1.416***	1.383***	1.489***	1.132***	1.101***
	(0.0505)	(0.0764)	(0.109)	(0.128)	(0.125)	(0.0869)
Observations	21,952	9,261	4,245	4,153	4,484	4,567
R-squared	0.127	0.133	0.134	0.177	0.118	0.106

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25: Regression Results of Subsamples by Centrality Levels

Dependent variable: 3-years average sales growth rate				
	(1)	(2)	(3)	(4)
VARIABLES	Periphery	Central	Periphery	Central
(log) BG size	0.0259***	0.0712***		
	(0.00721)	(0.0145)		
BG size 5-7			0.0828***	0.0859***
			(0.0125)	(0.0203)
BG size 8-11			0.0979***	0.0863***
			(0.0174)	(0.0318)
BG size 12-22			0.129***	0.158***
			(0.0185)	(0.0463)
BG size 23-84			0.140***	0.0946***
			(0.0188)	(0.0351)
BG size 85+			0.126***	0.122
			(0.0234)	(0.0851)
PI	0.117***	0.0150	0.0753***	0.0182
	(0.0258)	(0.0246)	(0.0265)	(0.0244)
(log) No. employees	0.196***	0.178***	0.196***	0.178***
	(0.00550)	(0.0123)	(0.00548)	(0.0123)
(log) capital intensity	0.0199***	0.0240***	0.0194***	0.0234***
	(0.00191)	(0.00397)	(0.00189)	(0.00396)
(log) sales 2011	-0.221***	-0.203***	-0.223***	-0.204***
	(0.00509)	(0.0124)	(0.00510)	(0.0125)
age	-0.00432***	-0.00309***	-0.00427***	-0.00307***
	(0.000365)	(0.000495)	(0.000358)	(0.000494)
mne	0.0702***	0.0124	0.0437***	0.0117
	(0.0123)	(0.0183)	(0.0119)	(0.0184)
sectors	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	1.293***	1.268***	1.328***	1.355***
	(0.0294)	(0.0827)	(0.0298)	(0.0876)
Observations	57,886	8,511	57,886	8,511
R-squared	0.129	0.131	0.131	0.132

Notes: 1. Firms belonging to groups of size 3 and 4 are considered as the benchmark in the specifications in columns (3) and (4).
2. Robust standard errors in parentheses.
3. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Chapter 3

Foreign Ownership and Firm Performance: Evidence from Italy

3.1 Introduction

A vast literature has discussed the advantage of multinational enterprises (MNEs) over the domestic firms (Antras and Helpman, 2004; Dunning, 1988; Helpman et al., 2008; Kohler et al., 2012; Tomiura, 2007). Based on the location of the ultimate owner, firms involved in a MNE in a certain country can be generally divided into two types: foreign-owned subsidiaries and firms in a domestic-owned MNE. A widely discussed advantage of foreign ownership is the spillover effect (Bernstein and Mohnen, 1998; Blomström and Sjöholm, 1999; Markusen, 1995). The foreign-owned subsidiaries can not only receive financial capital but also superior knowledge and managerial skills, which is especially crucial for firms in developing countries (Arnold and Javorcik, 2005). As for firms that invest abroad, opening subsidiaries in foreign countries can develop new market and increase their profitability because of economies of scale (Dunning, 1989; Lovelock and Yip, 1996). Some MNEs also invest in offshore financial centres and special purpose entities for tax reasons (Desai et al., 2004, 2006; UNCTAD, 2015).

UNCTAD (2016) points out that 41% of the foreign affiliates all over the world are owned through complex hierarchical chains with multiple cross-border links involving on average three jurisdictions. Since corporate structures have become increasingly complex, it is important to gain deep insights into the impact of foreign ownership on domestic firms from various perspectives such as its direction, multiple cross-border links, organizational or geographical distance, etc. This chapter is related to the literature on MNE. In particular, we investigate the relation between foreign ownership and firm performance by exploiting Italian firms data that are used in Chapter 2. Through regression analysis, we find that the firms making part of MNEs are more productive than those in the domestic business groups. Moreover, among the multinational firms, the foreign-owned Italian subsidiaries have higher productivity than the Italian firms in the domestic-owned MNEs.

Then we focus on the foreign-owned Italian subsidiaries. Interestingly, we find that the Italian subsidiaries with shorter organizational or geographical distance from their foreign owners are on average more productive. Furthermore, subsidiaries with multiple cross-border links in the upstream ownership chain are found to be more productive.

The remainder of the chapter is structured as follows. In Section 3.2 we review the existing literature. Then Section 3.3 provides the definition to classify firms and introduce the econometric specifications, together with some descriptive statistics of the data. In Section 3.4, we provide the estimation results and a thorough analysis of them. Finally, Section 3.5 concludes the work.

3.2 Literature Review

Based on the location of the ultimate owners, MNEs in a certain country can be generally divided into two types: foreign-owned subsidiaries and firms in a domestic-owned MNE. A large number of studies have empirically compared the performance of either of these two types of MNEs with the purely domestic business groups and most of them demonstrate that MNEs have a better performance over the domestic groups.

Some literature focuses on the foreign-owned subsidiaries. The MNEs of this type are found to have higher productivity or profitability in both developed countries such as Canada (Globerman et al., 1994), UK (Girma et al., 2001), Italy (Bentivogli et al., 2016) and developing countries such as Indonesia (Arnold and Javorcik, 2005) and China (Greenaway et al., 2014) since they can benefit from the transfer of financial capital, superior knowledge and managerial skills. Another strand of works presents evidence that firms that invest abroad improve their performance in terms of output (Desai et al., 2005; Hijzen et al., 2007), productivity (De La Potterie and Lichtenberg, 2001; Navaretti and Castellani, 2005) and home employment (Bruno and Falzoni, 2003; Hijzen et al., 2007; Navaretti et al., 2010) because of reducing production costs, increasing competitiveness and gaining market shares (Navaretti et al., 2004).

Some studies also compare the performance of these two types of MNEs, but the results highly depend on the location of the ultimate owner. For example, Doms and Jensen (1998) find that the foreign MNEs are less productive than the US-owned MNEs in the United States. Castellani and Zanfei (2003) find that both the crucial innovative activities including R&D, product innovation, patenting and technological cooperation with local firms are more likely in Italian MNEs than in foreign-owned firms in Italy. Temouri et al. (2008) find that the German domestic firms are less productive than MNE, but there is no significant difference between the domestic German MNEs and the foreign-owned subsidiaries. Criscuolo and Martin (2009) find that the MNEs in UK are significantly more productive than the domestic firms. The US-owned subsidiaries are on average more productive than all the other MNEs. Furthermore, the US MNEs tend to take over plants that are already more productive prior to acquisition.

A recent work by UNCTAD (2016) has further discussed the foreign ownership from the perspective of ownership complexity and investor nationality. They find that 41% of the foreign subsidiaries worldwide are owned through complex hierarchical chains with their direct owners and ultimate owners located in different jurisdictions, and these mismatch nationality cases account for 50% in terms of revenue. Their findings have shed light on the investor nationality conundrum, which has important implications for national and international investment policies.

3.3 Method

3.3.1 Empirical Method

In this section we discuss the empirical method to explore the relation between foreign ownership and firm performance.

First we investigate the impact of the direction of the foreign ownership on firm performance. Based on the location of the ultimate owner[1], MNEs in a given country can be generally divided into two types: foreign-owned subsidiaries and firms in domestic-owned MNEs. We use econometric models to compare the difference of their performance together with the purely domestic groups and the standalone ones. The logarithm form of labour productivity, defined as value added per employee, is used to measure firm performance. We also use the logarithm form of sales per employee as a robustness check. The regression model is as follows:

$$\log Y_i = \alpha + \sum_j \beta_j D_{ij} + \boldsymbol{\gamma}' \boldsymbol{X_i} + \epsilon_i \tag{3.1}$$

where the dependent variable Y_i is the measure of firm performance. D_{ij} are the dummy variables of the categories of firm i and the firms in the domestic-owned MNEs are considered as the benchmark group. $\mathbf{X}_i$ are the covariates, including the number of employees[2], capital intensity, firm age, the number of firms in the business group, the dummies of sectors and regions. Since some of them belong to the same business group

[1]As mentioned in Chapter 2, we cannot distinguish the ownership by entities, states, individuals or families due to the restriction of the data. Therefore, we consider the one locates at the top of each corporate ownership structure as the ultimate owner. It can be either a parent company whose shareholders are not provided by ORBIS or all have minority ownership, or the individual who is the ultimate beneficial owner of the business group. We make a further attempt to deal with this issue in 3.6.4.

[2]In the original data, 22.2% of the firms are provided with the number of employees equal to 0. In the main text we shift the variable by adding 1 to all firms in the sample.

and their performance can be correlated, we relax the assumption of the independence of residuals. Instead, we use the cluster-robust standard errors in the estimation by controlling the component they belong to.

Then we restrict our sample to the foreign-owned Italian subsidiaries. Taking into account the location of their direct owners and ultimate owners, we adopt the strategy provided in UNCTAD (2016) and further divide the foreign-owned Italian subsidiaries (FISs) into three types. Type 1 represents the firms with domestic direct owner and foreign ultimate owner. Type 2 represents the firms with direct owner and ultimate owner in two different foreign countries, which is one case of having multiple cross-border links in the upstream ownership chain. Type 3 represents the firms with direct owner and foreign owner in the same foreign country. Figure 8 illustrates the ownership structure of them. The black nodes at the top layer represent the foreign ultimate owners while the white nodes at the bottom layer correspond to the Italian subsidiaries, which are the focus firms in our analysis. The nodes at the intermediate layer represent the direct owners and the grey node means the nationality of the direct owner is non-Italian and different from that of the ultimate owner.

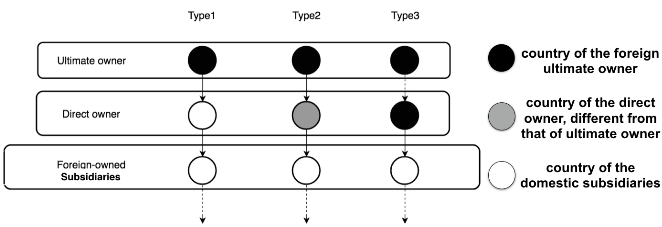

Figure 8: Three Types of Foreign-owned Subsidiaries

In the second model we compare the performance of these three types. Type 3 are used as benchmark. Besides the aforementioned control variables, we consider the foreign ownership in the downstream structure. We add two dummy variables in the model, which are whether they control any domestic affiliate and whether they control directly or indirectly any foreign affiliate.

To generalize the classification in UNCTAD (2016), we adopt an ownership level measure similar to the one in Belenzon et al. (2017) which reflects the organizational distance from the closest foreign owner in the corporate control structure. The definition of our measure is the number of intermediate subsidiaries separating a focal firm from the closest foreign owner. We further develop two dummy variables to investigate the impact of the geographical distance from the closest foreign owner and the existence of multiple cross-border links in the ownership chain on the performance of the focal affiliate. The first one is defined as 1 if a focal affiliate's closest foreign owner is located in the EU countries, and 0 otherwise. The second one is defines as 1 if the nationality of a focal affiliate's closest foreign owner is the same as that of its ultimate owner, and 0 otherwise. We use the following regression model to examine the relation between these measures and firm performance.

$$\log Y_i = \alpha + \beta_1 s_i + \beta_2 d_i + \boldsymbol{\gamma}^{'} \boldsymbol{X_i} + \epsilon_i \tag{3.2}$$

where s_i is the number of steps from the closest foreign owner, d_i is the dummy variable of whether the closest foreign owner is located in the EU countries or whether its nationality is the same as the ultimate owner's.

3.3.2 Data

We derive the ownership data of 17.8 million global firms in 2014 from ORBIS database. Based on their direct shareholding relationship, we construct the global control network by retaining only the majority shares.[3] Then we restrict our sample to Italian firms. We combine the ownership data with the AIDA database and collect all the Italian firms' financial data. Among the 1,164,871 active Italian firms in 2014 in AIDA, we identify 719,144 firms with ownership information in ORBIS.[4] After

[3]See the details in 2.3.1. In this chapter we consider only the majority control and the other two methods to build the network are not used here.

[4]In ORBIS ownership data of 2014, there are 235,309 Italian firms with minority owners, i.e., who hold no larger than 50% of the stakes. In our algorithm, we cannot identify which business groups these firms belong to. To separate them from the other types, we define

removing 320,241 firms with incomplete observations[5] and 279,860 firms with negative financial data, we finally derive a sample of 564,770 Italian firms.

Table 26 presents some summary statistics of the variables in our data.[6] In our sample the standalone firms have on average the lowest labour productivity than other types of firms (see Figure 9). Firms in the purely Italian domestic groups have higher average productivity than the standalone firms but much lower than the firms in multinational business groups. The two types of MNE, i.e., firms in the domestic-owned MNEs and subsidiaries in the foreign-owned MNEs have almost the same level of productivity.[7] To test whether the difference among categories is statistically significant, we perform econometric models in the following analysis.

Table 26: Summary Statistics

VARIABLES	N	mean	sd	p10	p50	p90
labour productivity	564,770	56.76	625.9	6.500	28.17	84
sales per employee	564,547	223.5	1,771	19.43	90.75	394
value added	564,770	852.1	18,077	12	118	1,035
sales	564,547	3,605	67,958	39	400	4,098
No. employees	564,770	14.58	142.5	1	4	21
fixed assets	564,770	1,440	66,810	3	59	1,688
capital intensity	564,770	352.3	4,837	1	11	469.5
age	564,761	16.16	13.77	3	12	35
group size	564,770	7.347	74.08	1	2	3
steps to closest foreign	13,186	1.298	0.601	1	1	2
closest foreign equal uo	13,186	0.844	0.363	0	1	1
closest foreign close Italy	13,186	0.414	0.493	0	0	1

We present the location distribution of the direct and ultimate owners of all the FISs in Figure 10 and Figure 11. The top three frequent jurisdictions of ultimate owners are Luxembourg, Netherlands and Ger-

them as *other*.

[5] There are 4 Italian firms involved in business groups with cross-shareholdings in which no ultimate owner can be identified.

[6] The standalone firms and *other* types here are considered as groups of size 1.

[7] Table 34 presents the frequency of the 5 types of firms in our sample. We also perform the new algorithm described in 2.6.6 to classify the firms into different types of MNEs and FISs. See more details in Appendix 3.6.4.

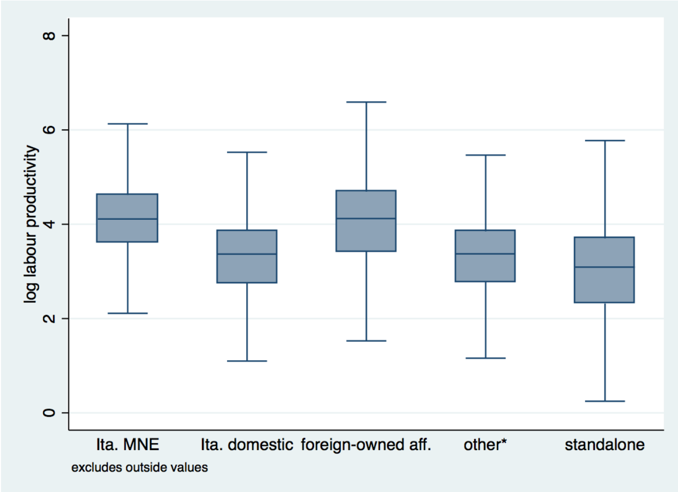

Figure 9: Labour Productivity (log) by Different Categories of Firms

many, which represent respectively 11.8%, 10.9% and 10.4% of the sample. Among the subsidiaries with foreign direct owners (Type 2 and Type 3), the top three foreign countries where the direct owners are located are Germany (12.6%), Switzerland (10.9%) and Luxembourg (10.0%). Figure 13 and Figure 14 presents the box plot of the five most frequent countries of the ultimate and direct owners of the FISs. The rankings of average labour productivity by the location of direct and ultimate owners are shown in Table 35 in Appendix 3.6. On average, subsidiaries owned by firms in Luxembourg, Netherlands and Germany are more productive than in other foreign countries. Though quite a number of subsidiaries are directly or indirectly owned by firms in Switzerland, their productivity is below the average.

Among the 13,186 FISs, 66.0% of them belong to Type 3. Type 1 and Type 2 represent respectively 23.4% and 10.6% of the sample. Figure 12 shows that firms of Type 2 on average are slightly more productive than the other two types. Table 27, 28 and 29 report the top 10 countries where the ultimate owners of the FISs are located and the FISs' average

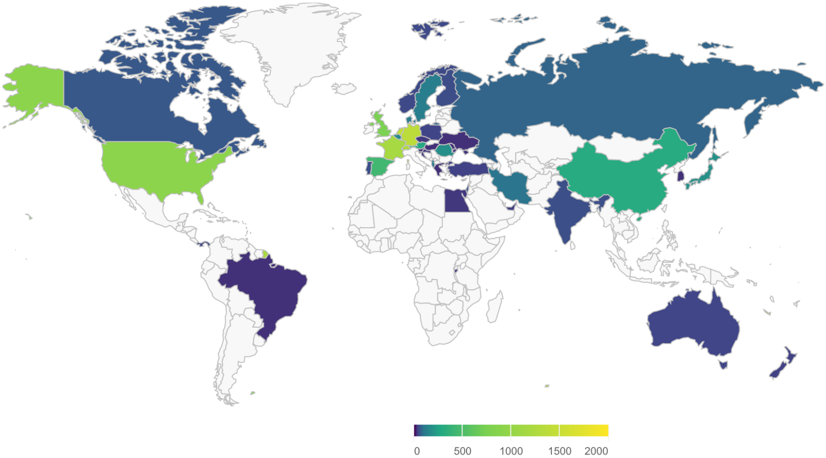

Figure 10: Frequency of Ultimate Owner of FIS by Country

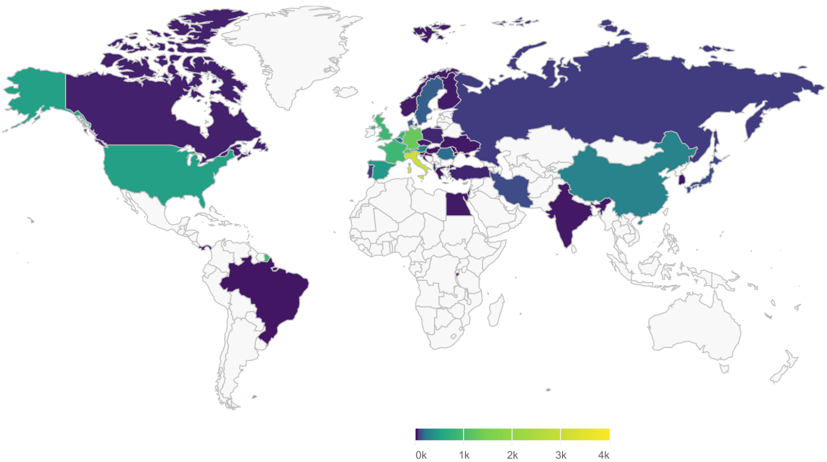

Figure 11: Frequency of Direct Owner of FIS by Country

productivity. We notice that US ultimate owners represent 6.9% of all the three types, but take up to 16.3% of the sample of Type 2, which implies that more of them control Italian subsidiaries through firms in another jurisdiction. On the other hand, Luxembourg ultimate owners represent 11.8% of the overall sample but only 5.8% of the sample of Type 2, which means most of them control directly Italian subsidiaries rather than through firms in a third country.

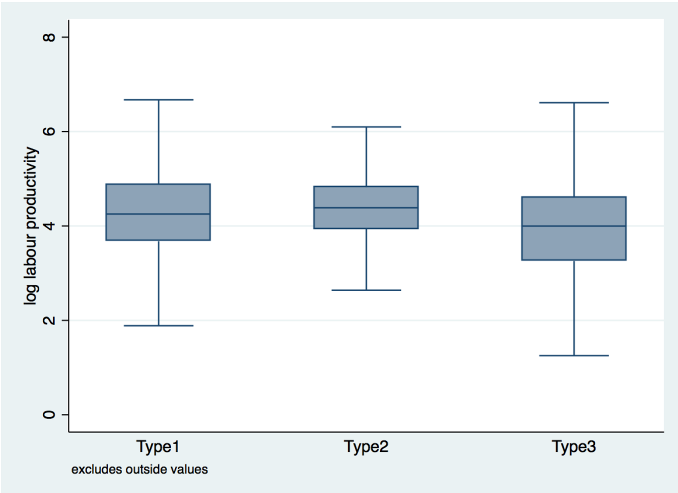

Figure 12: Labour Productivity (log) by Different Types of FIS

We further explore the location of the direct owners of Type 2. For each of the 5 most frequent countries of the ultimate owners, Table 30 presents the 3 most frequent countries of the direct owners. We notice that most of these direct owners are located in Germany and UK.

3.4 Results and Discussion

Table 31 reports the regression results of the comparison of different categories of firms. The group of domestic-owned MNEs is used as

Table 27: Type 1 of FIS's Labour Productivity (log) by Ultimate Owners' Location

Rank	Country	Frequency	Percentage	Mean	Std
1	Luxembourg	585	18.95%	4.505	1.667
2	Netherlands	509	16.49%	4.642	1.381
3	France	500	16.20%	4.294	1.528
4	Switzerland	221	7.16%	4.366	1.539
5	Germany	215	6.96%	4.617	1.397
6	United States	212	6.87%	4.336	1.053
7	Great Britain	170	5.51%	4.006	1.123
8	Spain	75	2.43%	4.081	1.529
9	Austria	32	1.04%	4.124	1.134
10	Japan	24	0.78%	4.388	1.039

Table 28: Type 2 of FIS's Labour Productivity (log) by Ultimate Owners' Location

Rank	Country	Frequency	Percentage	Mean	Std
1	United States	228	16.32%	4.507	0.827
2	Netherlands	147	10.52%	4.500	0.784
3	Germany	90	6.44%	5.068	1.464
4	Luxembourg	81	5.80%	4.398	0.929
5	Great Britain	71	5.08%	5.147	1.500
6	Japan	71	5.08%	4.504	0.571
7	Switzerland	55	3.94%	4.292	1.220
8	France	38	2.72%	4.751	1.355
9	Belgium	24	1.72%	4.838	1.173
10	Austria	19	1.36%	4.361	0.593

Table 29: Type 3 of FIS's Labour Productivity (log) by Ultimate Owners' Location

Rank	Country	Frequency	Percentage	Mean	Std
1	Germany	1065	12.24%	4.262	1.052
2	Switzerland	984	11.31%	3.646	1.161
3	Luxembourg	890	10.23%	4.557	1.581
4	Netherlands	776	8.92%	4.568	1.094
5	France	630	7.24%	4.108	1.044
6	Great Britain	507	5.83%	3.774	1.179
7	United States	474	5.45%	4.196	1.028
8	Spain	358	4.11%	4.032	1.275
9	China	275	3.16%	2.991	1.026
10	Austria	233	2.68%	4.180	1.041

Table 30: Type 2 of FIS's Labour Productivity (log) by Direct Owners' Location

Rank	Ultimate owner	Mean	Direct owner	Frequency	Percentage	Mean	Std
1	United States	4.507	Great Britain	65	28.51%	4.260	0.827
			France	37	16.23%	4.472	0.633
			Germany	27	11.84%	4.706	0.877
2	Netherlands	4.500	Germany	32	21.77%	4.736	0.638
			Great Britain	24	16.33%	4.678	1.191
			France	22	14.97%	4.324	0.761
3	Germany	5.068	Switzerland	25	27.78%	4.300	0.629
			Austria	23	25.56%	5.672	1.715
			Great Britain	14	15.56%	5.932	2.266
4	Luxembourg	4.398	France	24	29.63%	4.545	1.251
			Germany	14	17.28%	4.383	0.4837
			Great Britain	13	16.05%	4.564	0.653
5	Great Britain	5.147	Luxembourg	21	29.58%	6.378	2.148
			Germany	15	21.13%	4.597	0.412
			France	11	15.49%	4.730	0.792
5	Japan	4.504	Germany	23	32.39%	4.518	0.334
			Great Britain	17	23.94%	4.507	0.721
			Belgium	13	18.31%	4.413	0.774

the baseline. The results in column 1 show that the category of foreign-owned subsidiaries is significant and takes a positive sign while that of the purely domestic groups is significant and takes a negative sign. We further restrict the sample to only the manufacturing sector, and the results shown in columns 2 are in line with the previous ones. The findings suggest that the MNEs in Italy are on average more productive than the domestic groups, and the foreign-owned Italian subsidiaries are even more productive than the Italian firms involved in the domestic-owned MNEs. We also make some robustness checks by using the sales per employee as the dependent variable. The results shown in Table 36[8] in Appendix 3.6 are consistent with the findings here.

Table 32 presents the regression results of the comparison of different types of foreign-owned Italian subsidiaries. Using firms of Type 3 as benchmark, we find that in both specifications the coefficient of Type 1 is significant and negative and that of Type 2 is positive, though not significant when restricting the sample to manufacturing firms. The results indicate that the Italian subsidiaries directly owned by another domestic firm are less productive than the ones directly owned by a foreign firm. We make a robustness check by using the sales per employee as the dependent variable, and the results (see Table 37 in Appendix 3.6) are consistent.

Table 33 reports the regression results of model 3.2. The number of steps to the closest foreign owner in the control network is always significant and takes a negative sign in all the specifications, which suggests that the Italian subsidiaries with shorter organizational distance from their foreign owners are on average more productive. What's more, we notice that whether the closest foreign owner in the corporate structure is in the EU countries is significant and positive, implying that shorter geographical distance from the foreign owner is also a premium. We also control the dummy variable whether the ultimate owner is in the

[8] As mentioned before, we shift the number of employees by adding 1 to all firms in the previous analysis. As a robustness check, here we adopt the original value of number of employees in the regression and in defining the labour productivity. We drop the observations with the number of employees smaller than 3 and larger than 10,000. In Table 38 and Table 39, the regression also adopts the original value of number of employees.

Table 31: Comparison of Different Categories of Firms

Dependent variable: log of labour productivity		
	(1)	(2)
VARIABLES	Overall	Manufacturing
foreign-owned subsidiaries	0.178***	0.0769***
	(0.0258)	(0.0186)
purely domestic BGs	-0.0590***	-0.0343***
	(0.0167)	(0.0133)
standalone	-0.153***	-0.134***
	(0.0216)	(0.0166)
other	0.0208	0.0251
	(0.0215)	(0.0163)
log No. employees	0.217***	0.226***
	(0.00147)	(0.00252)
log capital intensity	0.163***	0.128***
	(0.000716)	(0.00154)
age	0.00419***	0.00273***
	(0.000114)	(0.000222)
log group size	0.0909***	0.0524***
	(0.0111)	(0.00599)
sector	Yes	Yes
regions	Yes	Yes
Constant	2.223***	2.332***
	(0.0212)	(0.0187)
Observations	554,890	101,409
R-squared	0.244	0.306

Notes: 1. In the regression we remove the firms in the 1% upper and lower tail of labour productivity and the firms with no sector and region information.
2. Robust standard errors in parentheses.
3. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 32: Comparison of Different Types of FIS

Dependent variable: log of labour productivity		
	(1)	(2)
VARIABLES	Overall	Manufacturing
Type 1	-0.0859***	-0.0988***
	(0.0313)	(0.0337)
Type 2	0.0971***	0.0626
	(0.0350)	(0.0398)
control domestic subsidiaries	0.00552	0.0675
	(0.0314)	(0.0483)
control foreign subsidiaries	-0.119***	-0.060*
	(0.0336)	(0.0359)
log No. employees	0.0851***	0.0835***
	(0.0121)	(0.0125)
log capital intensity	0.108***	0.0632***
	(0.00535)	(0.00902)
age	0.00399***	0.00368***
	(0.000636)	(0.000755)
log group size	0.109***	0.0773***
	(0.0174)	(0.0117)
sector	Yes	Yes
regions	Yes	Yes
Constant	2.856***	3.248***
	(0.0540)	(0.0730)
Observations	12,362	2,856
R-squared	0.201	0.189

Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

EU countries but find it not significant. In addition, the dummy variable of whether the nationality of the closest foreign owner is equal to that of the ultimate owner is found to be significant and negative in the first two specifications. The results reveal that the Italian subsidiaries with multiple cross-border links in the upstream ownership chain are more productive, which further support the previous findings that Type 2 outperforms the other two types.

Table 33: Regression Results Related to the Closest Foreign Owner

VARIABLES	(1) Overall	(2) Manufacturing	(3) Overall	(4) Manufacturing
steps_to_closest_foreign	-0.0795***	-0.0808***	-0.0725***	-0.0717***
	(0.0201)	(0.0232)	(0.0208)	(0.0231)
closest_foreign_in_EU	0.224***	0.186***		
	(0.0210)	(0.0311)		
closest_foreign_equal_uo			-0.0808*	-0.0709**
			(0.0444)	(0.0359)
control domestic subsidiaries	0.00524	0.0671	0.00473	0.0677
	(0.0313)	(0.0476)	(0.0317)	(0.0485)
control foreign subsidiaries	-0.111***	-0.0530	-0.118***	-0.0577
	(0.0331)	(0.0355)	(0.0337)	(0.0361)
log No. employees	0.0802***	0.0766***	0.0843***	0.0827***
	(0.0126)	(0.0127)	(0.0118)	(0.0124)
log capital intensity	0.106***	0.0641***	0.108***	0.0632***
	(0.00529)	(0.00901)	(0.00534)	(0.00902)
age	0.00401***	0.00360***	0.00404***	0.00371***
	(0.000637)	(0.000746)	(0.000638)	(0.000752)
log group size	0.101***	0.0713***	0.109***	0.0747***
	(0.0182)	(0.0117)	(0.0192)	(0.0122)
sector	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	2.832***	3.251***	3.015***	3.388***
	(0.0570)	(0.0787)	(0.0674)	(0.0845)
Observations	12,362	2,856	12,362	2,856
R-squared	0.210	0.200	0.201	0.189

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.5 Conclusion

In this chapter we empirically investigate the relationship between foreign ownership and firm performance from diverse perspectives using the data of Italian firms. We find that the foreign-owned Italian firms are on average more productive than the ones in domestic-owned Italian MNEs. Moreover, we notice that the Italian subsidiaries with shorter organizational or geographical distance from their foreign owners are on average more productive. In addition, subsidiaries with multiple cross-border links in the upstream ownership chain are found to be more productive.

The results of this work are limited to the Italian firms based on cross-section data. As mentioned in the previous chapter, it is difficult to trace the change of ownership structure over time, which limits us to perform more profound econometric analysis. In the future we will collect the financial data of firms in other countries and extend our analysis to the global scale. By doing so we can explore the difference of firm performance in different countries and provide more policy implications for global investors.

3.6 Appendix

3.6.1 Descriptive Statistics

Table 34: Labour Productivity (log) by Different Categories of Firms

category	No. Obs	Freq.	Mean	Std. Dev.
firms in domestic MNE	14,184	2.51%	4.151	1.192
foreign-owned subsidiaries	13,186	2.33%	4.103	1.322
firms in domestic BG	262,109	46.41%	3.281	1.045
stand-alone	133,598	23.66%	2.973	1.192
other	141,693	25.09%	3.276	1.020
overall	564,770	100%	3.248	1.112

Table 35: Labour Productivity (log) by the Location of Direct and Ultimate Owner of FIS

Ultimate owner				Direct owner			
Rank	Country	mean	std	Rank	Country	mean	std
1	Netherlands	4.587	1.179	1	Luxembourg	4.615	1.574
2	Luxembourg	4.529	1.587	2	Netherlands	4.567	1.091
3	Germany	4.371	1.164	3	Italy	4.391	1.452
4	Japan	4.322	0.788	4	Germany	4.334	1.057
5	United States	4.306	0.996	5	Austria	4.317	1.161
6	France	4.208	1.289	6	Belgium	4.236	1.080
7	Austria	4.186	1.026	7	United States	4.200	1.021
8	Sweden	4.150	0.801	8	Denmark	4.200	1.168
9	Belgium	4.140	1.084	9	Sweden	4.187	0.784
10	Denmark	4.103	1.194	10	France	4.185	0.994
11	Spain	4.046	1.319	11	Spain	4.083	1.210
12	Great Britain	3.957	1.263	12	Great Britain	3.905	1.234
13	Switzerland	3.801	1.271	13	Switzerland	3.724	1.137
14	China	3.161	1.210	14	China	3.001	1.030
15	Romania	2.889	0.762	15	Romania	2.891	0.766

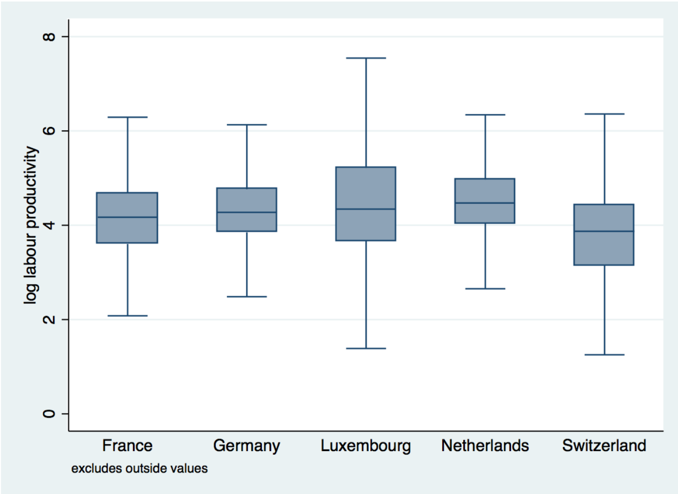

Figure 13: Labour Productivity (log) by Ultimate Owner Nationality of FIS

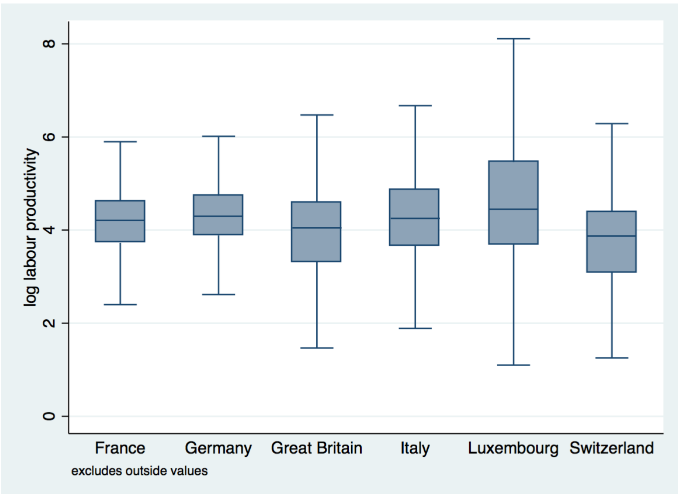

Figure 14: Labour Productivity (log) by Direct Owner Nationality of FIS

3.6.2 Robustness Check 1: Using Sales per Employee as Dependent Variable

Table 36: Comparison of Different Categories of Firms

Dependent variable: log of sales per employee		
	(1)	(2)
VARIABLES	Overall	Manufacturing
foreign-owned subsidiaries	0.0680**	0.0890***
	(0.0302)	(0.0227)
purely domestic BGs	-0.234***	-0.198***
	(0.0188)	(0.0159)
standalone	-0.362***	-0.250***
	(0.0241)	(0.0211)
other	-0.142***	-0.0922***
	(0.0241)	(0.0207)
log No. employees	0.0402***	0.111***
	(0.00182)	(0.00287)
log capital intensity	0.109***	0.131***
	(0.000797)	(0.00177)
age	0.00585***	0.00322***
	(0.000135)	(0.000281)
log group size	0.0973***	0.0846***
	(0.0135)	(0.00936)
sector	Yes	Yes
regions	Yes	Yes
Constant	4.388***	3.929***
	(0.0234)	(0.0228)
Observations	537,537	100,097
R-squared	0.178	0.199

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 37: Comparison of Different Types of FIS

Dependent variable: log of sales per employee		
	(1)	(2)
VARIABLES	Overall	Manufacturing
Type 1	-0.0759**	-0.136***
	(0.0323)	(0.0389)
Type 2	0.0996***	0.0960**
	(0.0342)	(0.0456)
control domestic subsidiaries	0.0572*	0.167***
	(0.0341)	(0.0573)
control foreign subsidiaries	0.0292	-0.0346
	(0.0371)	(0.0411)
log No. employees	0.0588***	0.0454***
	(0.0103)	(0.0138)
log capital intensity	0.0756***	0.0833***
	(0.00498)	(0.00999)
age	0.00575***	0.00324***
	(0.000669)	(0.000882)
log group size	0.0862***	0.0755***
	(0.0145)	(0.0127)
sector	Yes	Yes
regions	Yes	Yes
Constant	4.485***	4.478***
	(0.0578)	(0.0958)
Observations	12,319	2,812
R-squared	0.182	0.133

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.6.3 Robustness Check 2: Using Original Number of Employees

Table 38: Comparison of Different Categories of Firms

Dependent variable: log of labour productivity. Sample: No. employees (original) between 3 and 10000.		
	(1)	(2)
VARIABLES	Overall	Manufacturing
foreign-owned subsidiaries	0.164***	0.0709***
	(0.0215)	(0.0188)
purely domestic BGs	-0.142***	-0.0844***
	(0.0130)	(0.0130)
standalone	-0.253***	-0.170***
	(0.0169)	(0.0169)
other	-0.0667***	-0.0278*
	(0.0167)	(0.0165)
log No. employees	0.0452***	0.105***
	(0.00184)	(0.00284)
log capital intensity	0.131***	0.117***
	(0.000924)	(0.00169)
age	0.00712***	0.00304***
	(0.000172)	(0.000253)
log group size	0.106***	0.0660***
	(0.00934)	(0.00702)
sector	Yes	Yes
regions	Yes	Yes
Constant	2.832***	2.830***
	(0.0168)	(0.0190)
Observations	308,078	81,702
R-squared	0.255	0.244

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 39: Comparison of Different Types of FIS

Dependent variable: log of labour productivity. Sample: No. employees (original) between 3 and 10000.		
	(1)	(2)
VARIABLES	Overall	Manufacturing
Type 1	-0.0951***	-0.108***
	(0.0266)	(0.0331)
Type 2	0.0910***	0.0406
	(0.0254)	(0.0382)
control domestic subsidiaries	0.0343	0.0698
	(0.0294)	(0.0471)
control foreign subsidiaries	0.0123	-0.0246
	(0.0313)	(0.0362)
log No. employees	-0.00279	0.0265**
	(0.00851)	(0.0122)
log capital intensity	0.0515***	0.0500***
	(0.00521)	(0.00868)
age	0.00689***	0.00345***
	(0.000589)	(0.000736)
log group size	0.0976***	0.0789***
	(0.00888)	(0.0112)
Constant	3.310***	3.552***
	(0.0621)	(0.0776)
sector	Yes	Yes
regions	Yes	Yes
Observations	8,642	2,661
R-squared	0.163	0.127

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.6.4 Robustness Check 3: Results of the New Algorithm

Though there is no information on the ownership type in the original data, we perform a new algorithm to further identify the possible links between non-company (individual, family or state) and company (see details in 2.6.6). As illustrated in Figure 7, some firms belonging to type (c) before are classified as type (a) in the new algorithm. Thus, we use the nationality of the parent company A1 instead of that of the ultimate beneficial owner A in the classification of MNEs and FISs. For types (b) and (c), we use respectively the nationality of the ultimate beneficial owner B and the parent company C1 for classification as in the main text. The frequency of the new classification of firms is displayed in Table 40 and Table 41. The main difference of the classification results lie in that a majority of the firms belonging to a group of size 2 before are now classified as standalone ones. Nevertheless, the main regression results of the two algorithms in this chapter are consistent (see Table 42, Table 43 and Table 44). It should be noticed that this new algorithm may overly drop some ownership links actually by companies or institutions whose shareholders data are not covered by ORBIS. The results reported here only aim to provide further support for our findings in this chapter.

Table 40: Labour Productivity (log) by Different Categories of Firms

category	No. Obs	Freq.	Mean	Std. Dev.
firms in domestic MNE	14,429	2.55%	4.151	1.192
foreign-owned subsidiaries	7,922	1.40%	4.350	1.235
firms in domestic BG	80,936	14.33%	3.651	1.131
stand-alone	319,790	56.02%	3.066	1.075
other	141,693	25.09%	3.276	1.020
overall	564,770	100%	3.248	1.112

Table 41: Labour Productivity (log) by Different Categories of Firms

category	No. Obs	Freq.	Mean	Std. Dev.
Type 1	2,318	29.26%	4.389	1.395
Type 2	876	11.06%	4.539	0.999
Type 3	4,728	59.68%	4.296	1.187
overall	7,922	100%	4.350	1.235

Table 42: Comparison of Different Categories of Firms

Dependent variable: log of labour productivity		
	(1)	(2)
VARIABLES	Overall	Manufacturing
foreign-owned subsidiaries	0.280***	0.128***
	(0.0161)	(0.0222)
purely domestic BGs	-0.0718***	0.0138
	(0.0114)	(0.0138)
other	-0.173***	-0.0858***
	(0.0134)	(0.0161)
standalone	-0.283***	-0.191***
	(0.0134)	(0.0160)
log No. employees	0.193***	0.210***
	(0.00131)	(0.00271)
log capital intensity	0.189***	0.137***
	(0.000796)	(0.00174)
age	0.00340***	0.00251***
	(0.000117)	(0.000225)
log group size	0.0744***	0.0251***
	(0.00378)	(0.00548)
sector	Yes	Yes
regions	Yes	Yes
Constant	2.387***	2.445***
	(0.0142)	(0.0190)
Observations	563,193	101,884
R-squared	0.257	0.293

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 43: Comparison of Different Types of FIS

Dependent variable: log of labour productivity		
	(1)	(2)
VARIABLES	Overall	Manufacturing
Type 1	-0.164***	-0.118***
	(0.0259)	(0.0348)
Type 2	0.0976***	0.0760*
	(0.0292)	(0.0423)
control domestic subsidiaries	-0.0132	0.0805
	(0.0357)	(0.0517)
control foreign subsidiaries	-0.0554	-0.0192
	(0.0352)	(0.0433)
log No. employees	0.0388***	0.0447***
	(0.00766)	(0.0133)
log capital intensity	0.0891***	0.0392***
	(0.00558)	(0.0101)
age	0.00314***	0.00257***
	(0.000663)	(0.000843)
log group size	0.0749***	0.0510***
	(0.00663)	(0.0102)
sector	Yes	Yes
regions	Yes	Yes
Constant	3.381***	3.698***
	(0.0639)	(0.0884)
Observations	7,357	1,894
R-squared	0.134	0.086

Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 44: Regression Results Related to the Closest Foreign Owner

VARIABLES	(1) Overall	(2) Manufacturing	(3) Overall	(4) Manufacturing
steps_to_closest_foreign	-0.129***	-0.0854***	-0.118***	-0.0773***
	(0.0179)	(0.0229)	(0.0180)	(0.0230)
closest_foreign_in_EU	0.136***	0.0889**		
	(0.0264)	(0.0378)		
closest_foreign_equal_uo			-0.115***	-0.0723*
			(0.0261)	(0.0376)
control domestic subsidiaries	-0.0189	0.0797	-0.0126	0.0819
	(0.0357)	(0.0513)	(0.0358)	(0.0518)
control foreign subsidiaries	-0.0478	-0.0156	-0.0487	-0.0144
	(0.0353)	(0.0434)	(0.0354)	(0.0438)
log No. employees	0.0352***	0.0419***	0.0365***	0.0437***
	(0.00769)	(0.0135)	(0.00766)	(0.0134)
log capital intensity	0.0876***	0.0392***	0.0890***	0.0392***
	(0.00555)	(0.0101)	(0.00558)	(0.0101)
age	0.00324***	0.00256***	0.00322***	0.00262***
	(0.000661)	(0.000840)	(0.000662)	(0.000842)
log group size	0.0769***	0.0530***	0.0721***	0.0481***
	(0.00651)	(0.0100)	(0.00684)	(0.0106)
sector	Yes	Yes	Yes	Yes
regions	Yes	Yes	Yes	Yes
Constant	3.430***	3.720***	3.626***	3.842***
	(0.0695)	(0.0975)	(0.0716)	(0.101)
Observations	7,357	1,894	7,357	1,894
R-squared	0.136	0.085	0.134	0.084

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Chapter 4

On the Optimal Marketing Aggressiveness Level of C2C Sellers in Social Media: Evidence from China

4.1 Introduction

Built on the Web 2.0 technology, social media is the kind of applications that allow the creation and exchange of User Generated Content (UGC) on the Internet. It includes collaborative projects such as Wikipedia, social networking sites such as Facebook, microblogging platforms such as Twitter, and visual content communities such as YouTube. More recently, Mobile Web 2.0 (i.e., Web 2.0 evolution with mobile devices) expands the scope of social media to an unprecedented scale (Kaplan and Haenlein, 2010). For example, an increasing number of merchants have employed social media as a marketing tool in electronic commerce. They can promote their goods by simply posting a message, usually containing pictures, links, and a short description of them. Compared with the traditional marketing channels, social media has provided a more efficient and economic way for sellers to reach potential customers. What's more, sellers can get feedback in real time and gain a better understanding of customers' demands through their likes and comments, i.e., the popularity of microblogs (De Vries et al., 2012).

Not only sellers benefit from social media marketing, but customers gain gratifications as well. By reading the microblogs, they can get information of the products they need. However, customers use social networks also for social connection and entertainment (Lin and Lu, 2011). While they appreciate a moderate level of marketing communication on products, customers may get annoyed of a marketer if she is too aggressive in promoting her products.

Previous literature has mainly focused on the gratification brought by social media (Kaplan and Haenlein, 2010; Mangold and Faulds, 2009). Some works have pointed out people's negative attitude towards aggressive social media (and online) marketing by making survey (Akar and Topçu, 2011; Grant, 2005) and investigated the factors driving customers' engagement with marketing information (Bauer et al., 2005; Chu and Choi, 2011; Chu and Kim, 2011; Tsai and Men, 2013). However, to the best of our knowledge, no one has empirically provided an optimal marketing aggressiveness level in social media.

This work attempts to fill the gap by analysing the data of Taobao (China's largest C2C e-commerce platform, similar to eBay and Amazon) sellers on Sina Weibo[1] (China's largest microblogging platform, similar to Twitter). In particular, this chapter contributes to the few literature on C2C sellers' behaviour in social media. Prior studies on social media marketing are by and large confined to business-to-customer (B2C) (Kumar and Mirchandani, 2012; Rapp et al., 2013; Taylor et al., 2011), business-to-business (B2B) (Michaelidou et al., 2011; Swani et al., 2014; Wiersema, 2013), and eWOM between customers (Chu and Choi, 2011; Chu and Kim, 2011; King et al., 2014). This is mainly due to the practical difficulty of studying C2C business in social media. On the one hand, unlike firms that can be searched by their brand names, C2C sell-

[1] Some recent studies on Sina Weibo include Guan et al. (2014); He and Song (2015).

ers have little additional information to be found in social media. On the other hand, it is rare that a large number of C2C sellers from the same e-commerce platform use the same social media platform to promote their products. The dispersed distribution of social media platforms they use make their behaviour less comparable.

Thanks to the collaboration between Taobao and Sina Weibo, the difficulty of studying C2C business in social media can be overcome. On August 5, 2013, Sina Weibo collaborated with Taobao, and released a new module specifically designed for Taobao sellers. The new module gives the verified Taobao sellers additional capabilities (compared to other regular Sina Weibo users) to promote their merchandise. More importantly, it grants the verified Taobao sellers an identity of "Tao", which is highlighted in their Sina Weibo profiles.[2] What's more, in January, 2014, Sina Weibo has cooperated with Alipay (Taobao and Alipay are both subsidiaries of Alibaba Group) to launch a new platform called Weibo Payment, making payment much easier for Sina Weibo users. These policies have encouraged more Taobao sellers to create accounts on Sina Weibo and to make full use of this marketing channel. The gathering of Taobao sellers on Sina Weibo gives us a unique opportunity to analyse their marketing behaviour and to provide marketing strategies for them. Such marketing strategies would be of great importance for C2C sellers since they cannot afford to market through the traditional channels such as TV, newspapers and magazines.

We identify the 52,187 Taobao sellers on Sina Weibo and collect their microblogs in November 2014. For the 12,744 sellers who add the links of their Taobao shop, we further track their microblogs from July to October in 2016. For each seller, we use the proportion of her marketing-related microblogs to measure her marketing aggressiveness. To define whether a microblog is about marketing, we employ different machine learning algorithms to a training set of 5,000 manually labelled microblogs. The Multinomial Naive Bayes classifier has the best performance and is thus

[2]The minimum requirement for applying for an identity of "Tao" is that the virtual store owner in Taobao should have the level of credibility of at least "one diamond," see http://help.weibo.com/newtopic/taobao/list/1770/1772 for more details.

used to classify all the rest microblogs. Defining the marketing popularity as the average number of likes a seller receives per marketing-related microblog, we conjecture that the relationship between the marketing aggressiveness level and the marketing popularity can be depicted as an inverted U-shaped curve. By performing different models to explore their relationship, we find the linear regression model using Yeo-Johnson transformation of the number of followers has the best performance. After multiple tests, we empirically confirm that there is an inverted U-shape relationship between the marketing aggressiveness level and the marketing popularity. Specifically, the optimal proportion of marketing microblogs is around 0.3.

The remainder of the chapter is structured as follows. In Section 4.2 we present the conceptual framework and hypothesis. Then Section 4.3 discusses the design of the study, the algorithms to classify microblogs and the models to explore the relation between the marketing aggressiveness level and the marketing popularity. In Section 4.4, we provide the estimation results and analysis of them. Finally, Section 4.5 concludes the work.

4.2 Conceptual Framework

In this section we present the conceptual framework of our study. We define the marketing popularity as the average number of likes that a seller receives per marketing-related microblog over the observation period. The proportion of marketing-related microblogs that a seller posts is used to measure her marketing aggressiveness level. We assume that a too low proportion of marketing microblogs cannot satisfy potential customers' needs of being informed about products whereas a too high level of marketing microblogs may make people annoyed.

4.2.1 Marketing Popularity

The growth of social network sites has supported the increase of UGC social media communication (Gangadharbatla, 2008). New ideas to monetize social networks and UGC are pushing more and more sellers to create public accounts on social media, enabling them with a novel way to market their products. By leveraging the power of electronic word-of-mouth (eWOM), sellers can disseminate the news of their products to a large number of potential customers. Moreover, customers can interact with sellers and publicly state their opinions by liking, commenting, and forwarding (retweet) their microblogs.

Previous studies use the number of likes, comments, and forwards to measure the popularity of microblogs (De Vries et al., 2012; Yu et al., 2011). Since a user can comment or forward a microblog as many times as she wants while she can only like a microblog once on Sina Weibo, we adopt the number of likes as the measure of popularity in our study. More importantly, we consider the likes that sellers receive only for the microblogs related to marketing behaviour because they can directly reflect the purchase intention of potential customers.

We define that the marketing popularity of seller i as the average number of likes received per marketing-related microblog:

$$marketing\,popularity_i = \frac{No.\,likes\,received\,from\,marketing\,microblogs_i}{No.\,marketing\,microblogs_i}$$

Schivinski and Dabrowski (2014) argue that high level of popularity has a positive influence on brand equity and brand attitude, which in turn shows a positive influence on purchase intention. Hence, the popularity of microblogs is of crucial importance to sellers. They must have a good knowledge of the factors that can affect the popularity of their microblogs and find strategy to improve their popularity.

4.2.2 Marketing Aggressiveness Level

Our work is also related to the longstanding literature of uses and gratifications (U&G) theory (Eighmey and McCord, 1998; Katz et al., 1973; Ruggiero, 2000). In principle, potential customers can gain gratifications by interacting with both marketing microblogs (i.e., being informed or educated about products and services) and non-marketing microblogs (i.e., being connected in social life).

We define that the marketing aggressiveness level of seller i as the proportion of marketing microblogs she posts:

$$marketing\ aggressiveness\ level_i = \frac{No.\ marketing\ microblogs_i}{No.\ all\ microblogs_i}$$

Given a moderate aggressive level of marketing, potential customers can gain both social and business benefits and are more likely to interact with sellers. However, if the microblogs are solely focused on business, followers may get annoyed because they use social networks also for social connection and entertainment (Lin and Lu, 2011). On the other hand, the microblogs cannot be solely focused on regular social interactions because at least some of them are supposed to attract potential customers. Therefore, either extreme of the spectrum will likely decrease customers' gratifications and discourage them from liking the marketing microblogs. We state below formally our hypothesis on the relationship between the marketing aggressiveness level and the marketing popularity.

Hypothesis: The relationship between the marketing aggressiveness level and the marketing popularity can be depicted as an inverted U-shape curve and there exists an optimal marketing aggressiveness level to achieve the maximum popularity.

4.2.3 Control Variables

We control other variables that may affect the marketing popularity such as whether the seller's identity has been verified, gender, the number of followers and the average number of pictures posted per marketing microblog.

The verification of identities provides a signal of trust and reputation. The importance of trust in e-commerce has long been emphasized in previous studies (Gefen, 2000; Gefen and Straub, 2004; Hajli, 2014; Hoffman et al., 1999). Trust is a major factor that affects the prosperity and success of e-commerce because in such a virtual environment, participants are usually anonymous and do not engage in direct face-to-face communication (Lu et al., 2010). What's more, the growing number of fraudulent practices further increases customers' concerns to adopt e-commerce (Lek et al., 2001). Hence, trust is a crucial element of e-commerce (Cofta, 2006; Kim et al., 2009; Pentina et al., 2013). Enhancing the degree of trust in an online seller can increase people's intentions to purchase products on that seller's website (Gefen, 2000). Moreover, high levels of trust can help maintain long-term relationships between businesses and customers (Hoffman et al., 1999; Reichheld and Schefter, 2000).

Since its collaboration with Taobao, Sina Weibo grants the verified Taobao sellers an identity of "Tao"[3], which is highlighted in their Sina Weibo profiles. Furthermore, like Twitter, Sina Weibo also supports verification for individuals or entities[4]. After being approved, the verified user will have a "V"[5] identity in the profile. We believe that both the "Tao" and "V" identity can provide a signal of trust, and customers are more inclined to interact more with the verified sellers. Furthermore, since the verification of identities is associated with some thresholds of credibility and prestige[6], it can also be a proxy of other unmeasurable factors such as the quality of service. Hence, we assume that the sellers with verified identities can gain more popularity in social media.

Moreover, we expect that the number of followers have a positive effect on a seller's marketing popularity (Wang and Jin, 2010), so as the average number of pictures posted per marketing microblog (De Vries et al., 2012; Fortin and Dholakia, 2005). Some studies have also suggested the gender differences in online activity on Sina Weibo (Guan et al., 2014; Li et al., 2015) and we control the dummy variable of gender in our analysis as well.

[3] The minimum requirement for applying for an identity of "Tao" is that the virtual store owner in Taobao should have the level of credibility of at least "one diamond," see http://help.weibo.com/newtopic/taobao/list/1770/1772 for more details.

[4] There are a number of requirements such as the number of followers being at least 100 and the number of followees being at least 30, see http://verified.weibo.com/verify/help?fr=home&frpos=leftnav for more details.

[5] Sina Weibo accepts voluntary requests from the elite of 34 categories and 542 professions which include electronic retailers (see http://verified.weibo.com/verify/applystd?fr=home&frpos=morestd for more details).

[6] See footnotes 4 and 5 above.

4.3 Method

4.3.1 Sample

First, we identify the Taobao sellers on the microblog platform Sina Weibo in November 2014 using data scraping techniques. We select this period because November 11 (a.k.a. Singles' Day[7]), the largest online shopping day in the world[8] occurs in this month. We believe that the Taobao sellers have substantial economic incentives to post marketing microblogs during this period. Taking advantage of the search function of Sina Weibo, we find 281,160 profiles including the Chinese characters "Taobao" in their personal labels. However, apart from the 19,309 users with a verified "Tao" badge, we cannot make sure whether all the other users are real Taobao sellers or just fond of shopping on Taobao website. Hence, we retain only the users who add a link of Taobao shop or at least one of the Chinese words meaning Taobao seller[9] in their personal tags. Finally we identify 52,187 Taobao sellers and collected the 465,812 microblogs they posted in November 2014.

As a quality check, we explore the geographical distribution of the Taobao sellers in our sample. It fits our intuition that a good sample can reflect to some extent the local economic situation. As Figure 15 shows, most of the Taobao sellers live in Guangdong, the province with its GDP ranking first in China in 2014.[10] Sellers from Guangdong take up 18.45% of the observations in our sample, while sellers from Xizang, the province with the lowest GDP in China, represent only 0.05% of the sample. We find that the ranking of the number of regional Taobao sellers and the ranking of provincial GDP are highly correlated (see Figure 16), with a correlation coefficient of 0.901. Based on these facts, we believe that our data is a representative sample with regard to geographical

[7]Singles' Day is a day for people who are single, celebrated on November 11 (11/11). The date is chosen for the connection between singles and the number "1". This holiday has become popular in recent years among young Chinese people.

[8]The sales of Alibaba's sites Tmall and Taobao are $9.3 billion on November 11th, 2014.

[9]See some examples in Figure 23.

[10]Data source is from http://www.stats.gov.cn/tjsj/ndsj/2014/zk/html/Z0314e.htm.

distribution.

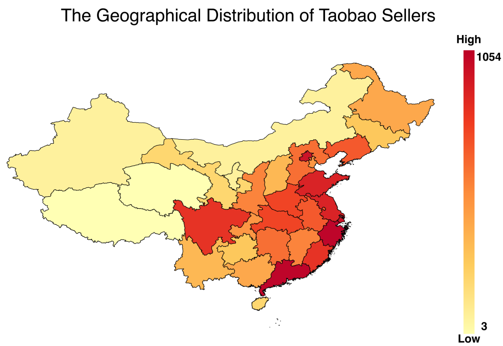

Figure 15: The Geographical Distribution of Taobao sellers on Sina Weibo

Among the 52,187 sellers we identify on Sina Weibo in 2014, only 12,744 sellers add the links of their Taobao shop. We further track these sellers and collect the 308,167 microblogs they posted from July to October in 2016.

4.3.2 Classification of Microblogs

A traditional text classification framework comprises preprocessing, feature extraction, feature selection and classification steps (Allahyari et al., 2017). In the preprocessing step, we first randomly select 5,000 out of the 774,429 microblogs and manually label them as "marketing" and "non-marketing". We employ these microblogs to train the classifiers so as to predict the label of all the microblogs.[11] The second process is tokenization, which is a task of breaking a character sequence and a defined document unit into pieces such as words, phrases, symbols and other elements called tokens (Manning and Schutze., 2008). Taking the

[11]We also adopt a simple classification approach by checking whether a microblog contains at least one of the most frequent words related to marketing. See the details in Appendix 4.6.6.

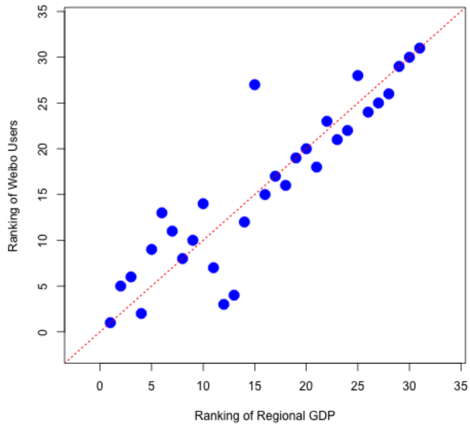

Figure 16: The Rankings of Regional GDP and Number of Taobao Sellers on Sina Weibo

advantage of the package *"jieba"* in python which is designed for Chinese words segmentation, we split each of the 5,000 microblogs into a list of words.

Most machine learning algorithms require numerical feature vectors with a fixed size rather than the raw texts with variable length. Hence, we need to extract numerical features from the contents instead of using the symbols directly. An intuitive way is to assign a weight to each word in a given document. We here adopt the method of *term frequency - inverse document frequency* (TF-IDF), which is a numerical statistic that is intended to reflect how important a word is to a document in a corpus (Leskovec et al., 2014; Salton and Buckley, 1988). In this step we filter out some stop words[12] and the words that appear only once[13]. In our case, each row in the TF-IDF matrix **A** represents a microblog d and

[12]The stop words refer to some extremely common words that would appear to be of little value in helping select documents matching a user's need (Manning and Schutze., 2008).

[13]This step is executed using the *TfidfVectorizer* function in the *sklearn* module in python.

each column corresponds to a word t. The term frequency $tf(t, d)$ is the number of times that word t appears in the microblog d. The inverse document frequency $idf(t)$ is equal to $\log(\frac{N+1}{n_t+1}) + 1$, where N is the number of microblogs and n_t is the number of microblogs containing word t. The corresponding value for word t in microblog d in the matrix $\mathbf{A}$ is defined as $\mathbf{A}(t, d) = tf(t, d) * idf(t)$. Table 45 presents several rows and columns of the TF-IDF matrix of our sample. The TF-IDF matrix will be used as inputs to predict the label of the microblogs in the next procedures.

Table 45: Term Frequency-Inverse Document Frequency Matrix

	Terms	*http*	11	*purchasing agent*	*really*	*new fashion*	...
label	Documents						
0	microblog 1	0.173	0	0	0	0	...
1	microblog 2	0.069	0.182	0	0	0.201	...
1	microblog 3	0	0	0.192	0	0	...
...	...				...		
0	microblog 5000	0	0	0	0	0	...

Notes: we label the "marketing" microblogs as 1 and "non-marketing" ones as 0.

Given that the classification task in our case belongs to supervised learning, we perform the following steps. First, we choose four machine learning algorithms to classify separately the microblogs, including logistic regression (logit), decision tree[14] (CART), random forest (RndFor) and multinomial naive bayes (MNB)[15]. The first two algorithms require that only few features can be introduced into the model. To reduce the dimension, we perform the truncated singular vector decomposition (TSVD) method (Manning and Schutze., 2008). Then we select the main features extracted by TSVD in the two algorithms.[16] When using RndFor and MNB algorithms, we directly use the TF-IDF matrix as inputs.

Second, we randomly select 4,000 out of the 5,000 labelled microblogs as the training set and the rest 1,000 microblogs as the test set. Since the algorithms logit, CART and RndFor require us to determine the parameters, we apply a cross-validation method[17] to optimize them. Using a

[14]We use the GINI criterion in the splitting decision.

[15]See the details of the four algorithms in Baesens (2014); Baesens et al. (2015).

[16]In logit, we further apply the recursive feature selection to remove several less important features.

[17]See the details in James et al. (2014).

10-fold cross-validation approach, we randomly divide the training set into 10 groups of equal size. Each time we use a given algorithm to fit 9 folds and evaluate its performance on the rest 1 fold, which is the validation set. The evaluation is based on the *area under the receiver operating characteristics curve* (AUC) score, which is the most informative and objective indicator of predictive accuracy within a benchmarking context (Lessmann et al., 2008). This procedure is repeated 10 times and each time a different group of observations is treated as the validation set. For each set of parameters, we compute the average of the 10 AUC scores. Then we compare the average AUC score of different parameter settings and select the best performing set of parameters. We adopt the parameters' values suggested in Lessmann et al. (2008)[18], and the ones with the best average AUC score of each algorithm are listed in Table 46.

Table 46: Best Performing Parameters in the 10-fold Cross Validation

	No. components	No. features selected	No. trees	$AUC\_train\_average$
logit	19	14		0.843
CART	20	20		0.758
RndFor			250	0.943
MNB				0.941

Finally, after we determine the parameters for each algorithm, we assess their performance on the test set. The AUC scores of the four algorithms are listed in Table 47. To further understand which algorithm has a statistically significant better performance, we conduct the test in DeLong et al. (1988). From Table 47 we can find that RndFor and MNB are significantly better than logit and CART. But there is no significant difference between RndFor and MNB. Since MNB has the largest AUC value on the test set, we adopt it to classify all the rest microblogs.

[18]The parameters' values suggested in Lessmann et al. (2008) for each classifier are as follows: for logit and CART, we first perform dimension reduction to extract the k main components where k varies from 5 to 20. The 5 main components explain 23.0% of the variance while the 20 main components explain 28.9%. In logit we further perform the recursive feature elimination to select the inputs in the regression model and the number of features selected varies from 5 to k. As for RndFor, the parameter to be tuned is the number of trees, and the suggested values include [10, 50, 100, 250, 500, 1,000].

Table 47: DeLong's Test to Compare Different Models' AUC Scores

	AUC_test	logit	CART	RndFor	MNB
logit	0.854				
CART	0.765	0.000			
RndFor	0.956	0.000	0.000		
MNB	0.960	0.000	0.000	0.496	

Notes: Except the AUC scores, the above listed numbers are p-values.

4.3.3 Modelling and Data

After we classify all the microblogs, we aggregate the data by seller and thus obtain the number of marketing-related microblogs each seller posts in each observation period. Then we compute the proportion of the number of marketing microblogs to the total number of microblogs, which is used to measure marketing aggressiveness. We also count the average number of likes per marketing microblog for each seller, which is used to measure marketing popularity. The sellers whose average number of likes located in the 1% tail of the distribution are excluded from the sample in order to reduce the possible outlier effect. Furthermore, we remove the sellers who posted less than 10 microblogs in each observation period since the possible wrong classification may cause a large error on the value of marketing proportion for sellers who post less. The details of the data in each of the previous steps are presented in Table 48.

Table 48: Statistics by Data Processing Steps

Data of 2014				
Steps	No. sellers	No. microblogs	No. marketing_microblogs	marketing proportion
1. Initial sample	52,187	465,812	152,772	32.8%
2. sellers who post	18,809	465,812	152,772	32.8%
3. sellers who market	9,577	402,533	152,772	38.0%
4. final sample*	5,809	385,911	145,328	37.7%

Data of 2016				
Steps	No. sellers	No. microblogs	No. marketing_microblogs	marketing proportion
1. Initial sample	12,744	308,617	86,115	27.9%
2. sellers who post	4,227	308,617	86,115	27.9%
3. sellers who market	2,262	217,806	86,115	39.5%
4. final sample*	1,812	212,717	84,724	39.8%

* In the final sample we remove the outliers and sellers who post less than 10 microblogs.

In the final sample, the response variable is the average number of likes a seller receives per marketing microblog. The inputs include the marketing aggressiveness, the number of followers, gender, whether the seller has the "V" and the "Tao" identities. For the data of 2016, we manage to extract the average number of pictures a seller posts in each marketing-related microblog as well. The definitions of the variables are shown in Table 49 and their descriptive statistics are shown in Table 50. On average, a seller receives correspondingly 0.604 and 0.641 likes per marketing microblog in 2014 and 2016. The standard deviation of the variable in the two data sets are also quite close, which are respectively 1.107 and 1.150. The average proportion of marketing microblogs is 33.3% for the data of 2014 and 28.5% for 2016.

Table 49: Variable Definitions

Variables	Definitions
$likes\_average\_marketing$	The average number of likes received from all the marketing-related microblogs.
$marketing\_proportion$	The proportion of the number of marketing-related microblogs to the total number of microblogs.
Tao	1 if the seller has a verified "Tao" identity; otherwise 0.
V	1 if the seller's real identity has been verified; otherwise 0.
$female$	1 if the seller is a female; otherwise 0.
$No.\ followers$	The number of followers the seller has.
$No.\ pic$	The average number of pictures the seller posts per marketing-related microblog.

To have a preliminary understanding of the relationship between a Taobao seller's marketing aggressiveness level and marketing popularity, we divide the proportion of marketing microblogs into 5 intervals uniformly. From Table 51 we can find that for both the data of 2014 and 2016, posting 20%-40% microblogs on marketing gains the largest average number of likes whereas posting 80% to 100% microblogs on marketing has the smallest. When the marketing aggressiveness level is beyond 40%, the average marketing popularity decreases. From these facts we gain the impression that a seller's marketing popularity and marketing aggressiveness level have an inverted U-shape relationship, and marketing popularity is maximized when 20% to 40% of her microblogs are focused on marketing.

Table 50: Summary Statistics

Data of 2014					
Variables	N	mean	sd	min	max
$likes\_average\_marketing$	5,809	0.604	1.107	0	7.25
$marketing\_proportion$	5,809	0.333	0.302	0.0034	1
$female$	5,809	0.759	0.428	0	1
$No. followers$	5,809	9,830	46,553	1	1,011,968
Tao	5,809	0.583	0.493	0	1
V	5,809	0.209	0.407	0	1

Data of 2016					
Variables	N	mean	sd	min	max
$likes\_average\_marketing$	1,812	0.641	1.150	0	5
$marketing\_proportion$	1,812	0.285	0.292	0.0009	1
$female$	1,812	0.829	0.376	0	1
$followers$	1,812	14,613	83,851	4	1,866,129
Tao	1,812	0.512	0.500	0	1
V	1,812	0.221	0.415	0	1
$No. pic$	1,812	3.804	2.790	0	9

Table 51: Marketing Popularity by Marketing Aggressiveness Level

	Data of 2014			Data of 2016		
Proportion	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
$(0, 0.2]$	0.618	1.180	2736	0.543	1.078	976
$(0.2, 0.4]$	0.846	1.313	1078	1.022	1.427	320
$(0.4, 0.6]$	0.669	1.054	703	0.943	1.316	195
$(0.6, 0.8]$	0.467	0.793	565	0.521	0.916	143
$(0.8, 1]$	0.234	0.451	727	0.258	0.556	178

We then perform linear regression, multi-layer perceptron neural network (MLP)[19] and random forest to explore the relation between the proportion of marketing microblogs and the average number of likes. In the linear regression, we add the squared term of the marketing proportion to test Hypothesis 1. Moreover, since some inputs may have a non-linear effect on the response variable, we follow the steps introduced in Van Gestel et al. (2006, 2005) to perform the Yeo Johnson transformation (Yeo and Johnson, 2000) for the continuous variables in the linear regression (see details in Appendix 4.6.3). In MLP and RndFor, we use 5-fold cross-validation to select the parameters with the best performance.[20]

4.4 Results and Discussion

Two-thirds of each data set of the two years are randomly selected to train the algorithms and the rest one-third are used to test their performance. We evaluate their performance based on three indicators: R-square, mean square error (MSE) and Pearson correlation coefficient between the predicted value and the true value. As Table 52 shows, the linear regression using Yeo Johnson transformation[21] has the best performance compared with the standard linear regression, MLP[22] and RndFor[23].

Table 53 presents the estimation results of the linear regression using Yeo-Johnson Transformation. We notice that the coefficients of marketing proportion and its quadratic term are both significant, and as expected, the coefficients of the quadratic term are negative for both data sets. To confirm that the relation is an inverted U-shape, we further perform the

[19]See the details in Baesens et al. (2015).

[20]Lessmann et al. (2008) assume that there is a single hidden layer for MLP and the parameters' values of the number of neurons are [4,5,6]. As for RndFor, the number of trees are seleced from [10, 50, 100, 250, 500, 1,000].

[21]We perform the Yeo Johnson transformation for all the continuous variables, but only the transformation of the number of followers significantly improves the performance of the model. The estimated parameters for the number of followers are $\lambda = -2, c = -0.5$ for the data of 2014 and $\lambda = -2, c = 0$ for data of 2016.

[22]In MLP, the optimal number of neurons is 4 for both the data of 2014 and 2016.

[23]In RndFor, the optimal number of trees is 250 for the data of 2014 and 1000 for 2016.

Table 52: Performance of Different Models on Test Set

	2014			2016		
Models	R^2	MSE	Pearson	R^2	MSE	Pearson
Linear Regression	0.081	1.085	0.298	0.220	1.039	0.469
Yeo Johnsson	0.247	0.889	0.498	0.369	0.841	0.608
MLP	0.038	1.136	0.230	0.111	1.184	0.341
RndFor	0.134	1.023	0.421	0.308	0.922	0.560

Table 53: Regression Results of Yeo-Johnson Transformation (Training Set)

VARIABLES	2014	2016
$marketing\_proportion$	0.567***	1.185***
	(0.189)	(0.315)
$marketing\_proportion^2$	-1.021***	-1.623***
	(0.183)	(0.313)
V	0.0157	0.0133
	(0.0483)	(0.0755)
Tao	0.0271	-0.0786
	(0.0324)	(0.0546)
$(transformed) No.\ followers$	1.296***	6.204***
	(0.0715)	(0.404)
$female$	0.272***	0.0310
	(0.0357)	(0.0640)
$No.pic$		0.0418***
		(0.0105)
Constant	1.488***	0.159**
	(0.0851)	(0.0762)
Observations	3,872	1,208
R-squared	0.241	0.359

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

test in Lind and Mehlum (2010) (see details in Appendix 4.6.4). Table 54 shows that in both data, the slope at the lower bound of the data range is significant and positive and the slope at the upper bound is significant and negative. Moreover, the turning points are respectively 0.278 and 0.365, and their 95% confidence intervals are respectively [0.146, 0.346] and [0.272, 0.416], which are located in the data range (0, 1]. Thus, we can confirm that there is an inverted U-shape relationship between the average number of likes and the proportion of marketing-related microblogs. Figure 17 and Figure 18 show their relationship.[24] The results imply that when the proportion of marketing-related microblogs increases under a certain level, people are inclined to like them more since they convey the information about the products. However, if the proportion continues to increase beyond a certain level, people may get annoyed of their marketing behaviour.

Table 54: Inverted U-shape Test

	2014		2016	
	Lower bound	Upper bound	Lower bound	Upper bound
Interval	0.003	1	0.001	1
Slope	0.560	-1.474	1.182	-2.060
t-value	2.985	-7.939	3.761	-6.304
$P > \lvert t \rvert$	0.001	0.000	0.000	0.000
Overall test	t-value: 2.98		t-value: 3.76	
	$P > \lvert t \rvert$: 0.001		$P > \lvert t \rvert$: 0.000	
Turning point	0.278		0.365	
95% CI	[0.146, 0.346]		[0.272, 0.416]	

We make some further robustness checks by performing other specifications such as using the cubic, logarithm and exponential forms of the $marketing\_proportion$. As Table 55 shows, the performance of these specifications are almost the same. Though we cannot exclude other possibilities, what we empirically find is that when the proportion of marketing-related microblogs goes beyond a certain level, sellers receive fewer likes from their followers.

[24]In the graphs, the average numer of likes are computed given the mean value of the transformed number of followers and the number of pictures, and female equal to 1 for the data of 2014. The variables that are not significant are ignored in the calculation.

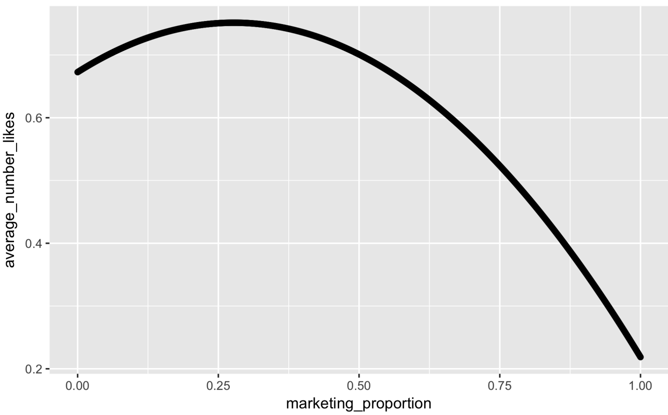

Figure 17: Optimal Proportion 2014

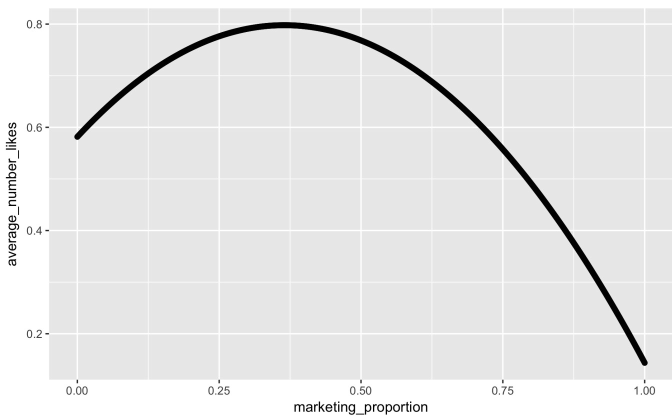

Figure 18: Optimal Proportion 2016

More interestingly, we notice that the non-linear transformation of the number of followers significantly improves the prediction accuracy of the number of likes. The relationship between the number of likes and followers of the two years data is depicted in Figure 19 and Figure 20.[25] It can be seen that an increase of 10,000 followers from 10,000 to 20,000 has a much larger effect on the number of likes than an increase from 100,000 to 110,000. After the number of followers reaches around 100,000, a saturation effect occurs.

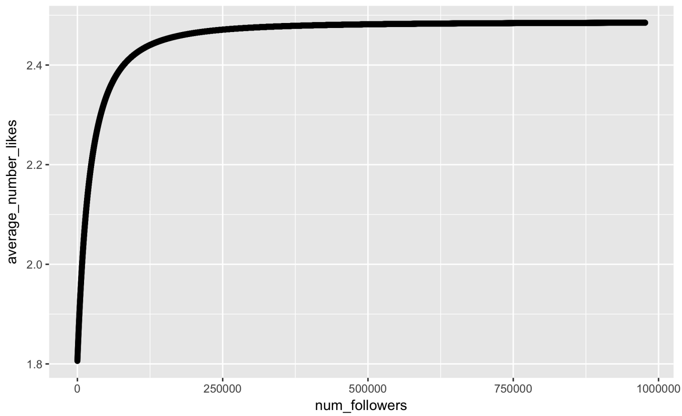

Figure 19: Likes and Followers 2014

4.5 Conclusion and Extension

In this chapter, we empirically investigate the relationship between sellers' marketing aggressiveness level and their marketing popularity in social media. In particular, by analysing the microblogs of Taobao sellers on Sina Weibo, we find that the relationship between the propor-

[25]In the graphs, the average numer of likes are computed given the mean value of the number of pictures, the proportion of marketing microblogs is equal to 0.3, and female equal to 1 for the data of 2014. The variables that are not significant are ignored in the calculation.

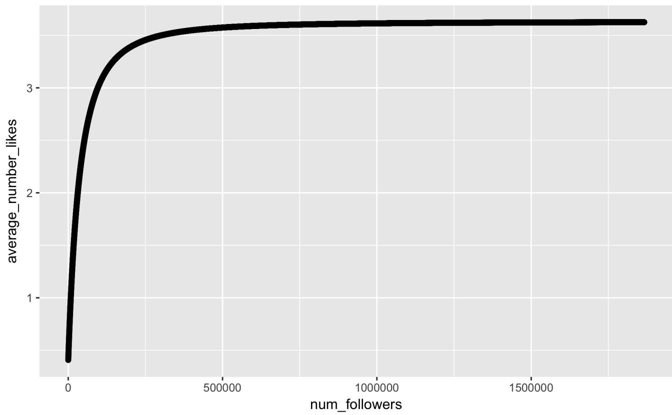

Figure 20: Likes and Followers 2016

tion of marketing-related microblogs and the average number of likes a seller receives per marketing microblog can be depicted as an inverted U-shaped curve. The results imply that when a seller is too aggressive in marketing on social media, she will receive fewer likes as feedback. Specifically, the optimal proportion of marketing microblogs is around 0.3.

Our contribution to the literature of social media marketing is at least twofold. First, we perform a quantitative analysis of the relationship between the level of marketing aggressiveness and the marketing popularity, and obtain a limited range of empirical optimal marketing aggressiveness level to maximize the marketing popularity. Second, our research presents a first attempt to study the C2C sellers' marketing behaviour in social media by collecting and analysing a unique data set. We also employ the text mining techniques to explore a huge amount of microblogs.

In our future work we will investigate the impact of social media marketing on the C2C sellers' sales by collecting data from both Taobao and SinaWeibo. Although the popularity on social media can reflect to a certain extent the purchasing intention of customers, it is not clear how much the popularity is finally converted into sales. If the popularity is highly correlated with the sales, then we must highlight the importance of the marketing behaviour on social media and provide better strategies for the sellers. Moreover, by incorporating popularity in the forecast of sales, sellers can manage their operations better and thus improve profit margin. If there is no significant correlation between popularity and sales, the sellers have to decide whether to continue the investment on the current marketing channel.

Besides the microblogs data we have already analysed in the present work, we have also collected from the Taobao website the data of the 12,744 Taobao sellers from August to November, 2016. As shown in Figure 21, the variables include monthly sales, credit level, scores of service quality, category of products, whether they have a guarantee policy, shop age and region, etc. Considering that the marketing popularity of microblogs has an influence on sales but is also affected by other features such as number of followers, marketing aggressiveness level and so on, we plan to use the structural equation modelling in the future analysis.

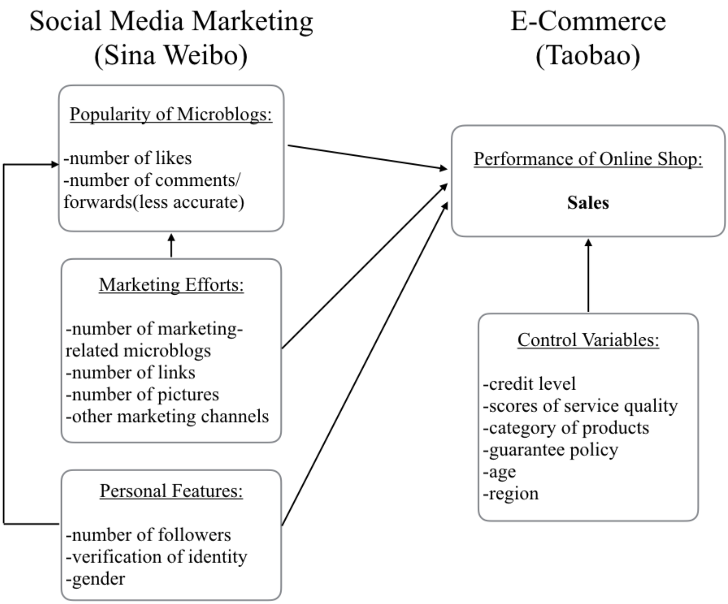

Figure 21: Framework of the Study on Taobao and Weibo

4.6 Appendix

4.6.1 Illustration of Profile Page

Figure 22: An Example of a Taobao Seller's Sina Weibo Account

4.6.2 Variants of "Taobao Sellers" in Chinese

淘宝卖家	淘宝店主	淘宝掌柜	淘宝商户

Figure 23: Variants of "Taobao Sellers" in Chinese

4.6.3 Estimation of Yeo-Johnson Transformation

Yeo and Johnson (2000) has proposed a transformation which is of the same form as Box-Cox transformations and is also valid for negative values. The transformation function is defined as follows:

$$f(x;\lambda) = \begin{cases} ((1+x)^{\lambda}-1)/\lambda, & \lambda \neq 0, x \geq 0, \\ \log(x+1), & \lambda = 0, x \geq 0, \\ -((1-x)^{(2-\lambda)}-1)/(2-\lambda), & \lambda \neq 2, x \leq 0, \\ -\log(-x+1), & \lambda = 2, x \leq 0. \end{cases}$$

We consider the following transformation: $x \mapsto f(x+c,\lambda)$ where c is the location parameter and λ is the transformation parameter. The parameters c and λ are estimated based on the following steps:

Step 1: The variable to be transformed is first normalized to zero median and unit variance.

Step 2: The parameters are estimated using a grid search mechanism. The parameter c varies from -3 to $+3$ and the parameter λ varies from -2 to $+2$. For each parameter combination (c, λ), the model is estimated and MSE is stored.

Step 3: Using a 5-fold cross-validation, the combination (c, λ) with the lowest average MSE is selected. The optimal MSE is compared with the MSE obtained with $\lambda = 1$. When the MSE of the nonlinear model is lower than the MSE of the linear model, the nonlinear transformation is applied.

4.6.4 Test of the Inverted U-shape

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \boldsymbol{\gamma}' \boldsymbol{Z} + \epsilon$$

Haans et al. (2015) framed the U-shape test proposed in Lind and Mehlum (2010) as a three-step procedure. Here we list the steps to test whether the relationship between Y and X is an inverted U-shape:

Step 1: β_2 needs to be significant and negative.

Step 2: The slope must be significantly steep at both ends of the data range $[X_l, X_u]$, where X_l is the minimal of X and X_u is the maximal. To make sure that the inverted U-shape is a phenomenon in the interior of the range of X, the slope at the lower bound $\beta_1 + 2\beta_2 X_l$ should be significant and positive and the slope at the upper bound $\beta_1 + 2\beta_2 X_u$ should be significant and negative. Hence, we need to test whether the combined null hypothesis can be rejected in favour of the alternative hypothesis:

$H_0 : \beta_1 + 2\beta_2 X_l \leq 0$ and/or $\beta_1 + 2\beta_2 X_u \geq 0$

$H_1 : \beta_1 + 2\beta_2 X_l > 0$ and $\beta_1 + 2\beta_2 X_u < 0$

Sasabuchi (1980) provides a test based on the likelihood ratio principle. The rejection areas are as follows:

$$R_\alpha\{(\beta_1, \beta_2) : \frac{\beta_1 + 2\beta_2 X_l}{\sqrt{s_{11} + 4X_l s_{12} + 4X_l^2 s_{22}}} > t_\alpha \& \frac{\beta_1 + \beta_2 2X_u}{\sqrt{s_{11} + 4X_u s_{12} + 4X_u^2 s_{22}}} < -t_\alpha\}$$

where s_{11}, s_{22}, s_{12} are the estimated variance of β_1 and β_2 and their covariance, and t_α is the α percentile of the t-distribution with the appropriate degree of freedom.

Step 3: The turning point and its 95% confidence interval needs to be located within the range of X. The point estimate of the turning point is $X = -\frac{\beta_2}{2\beta_1}$. Fieller (1954) provides how to construct a confidence interval for the ratio of two normally distributed estimates. The lower bound and upper bound of the $(1-2\alpha)$ confidence interval for $-\frac{\beta_2}{2\beta_1}$ are respectively

$$\widetilde{X}_l = \frac{s_{12}t_\alpha^2 - \beta_1\beta_2 - t_\alpha\sqrt{(s_{12}^2 - s_{22}s_{11})t_\alpha^2 + \beta_2^2 s_{11} + \beta_1^2 s_{22} - 2s_{12}\beta_1\beta_2}}{\beta_2^2 - s_{22}t_\alpha^2}$$

$$\widetilde{X}_u = \frac{s_{12}t_\alpha^2 - \beta_1\beta_2 + t_\alpha\sqrt{(s_{12}^2 - s_{22}s_{11})t_\alpha^2 + \beta_2^2 s_{11} + \beta_1^2 s_{22} - 2s_{12}\beta_1\beta_2}}{\beta_2^2 - s_{22}t_\alpha^2}$$

If the confidence interval is located within the range of X, we can make sure that the relationship between Y and X is an inverted U-shape.

4.6.5 Other Specifications of Linear Regression

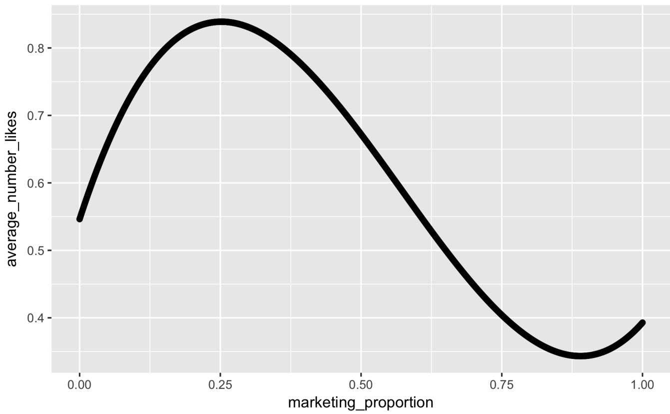

Figure 24: Cubic Specification 2014

4.6.6 Classification Using a Dictionary Approach and the Regression Results

We also adopt a dictionary approach in classifying the microblogs for the data of 2014. After we have collected the 465,812 microblogs, we segment each of them into a list of words, disregarding grammar and word order. Then we count the frequency of all the words and rank them. Starting from the most frequent words, we manually select the top 50 words related to marketing and use them as a dictionary (see Figure 30). We define a microblog is related to marketing if it contains at least one word in the dictionary, otherwise it is non-marketing. As a robustness check, we also create another dictionary using the top 100 words related to marketing (see Figure 31). The dictionary approach is relatively more rough than the machine learning algorithms in classification. Compared

Table 55: Other Specifications of Linear Regression

VARIABLES	(1) 2014_cubic	(2) 2014_log	(3) 2014_exp	(4) 2016_cubic	(5) 2016_log	(6) 2016_exp
$marketing\_proportion$	2.566***			2.026***		
	(0.481)			(0.723)		
$marketing\_proportion^2$	-6.538***			-4.052**		
	(1.152)			(1.816)		
$marketing\_proportion^3$	3.819***			1.710		
	(0.746)			(1.202)		
$(log)\ marketing\ proportion$		-0.0456***			-0.00848	
		(0.0132)			(0.0178)	
$(exp)\ marketing\ proportion$			-0.247***			-0.203***
			(0.0251)			(0.0444)
V	0.0166	0.0298	0.0185	0.0178	0.0130	0.00742
	(0.0482)	(0.0489)	(0.0485)	(0.0752)	(0.0765)	(0.0761)
Tao	0.0224	0.0204	0.0369	-0.0795	-0.0866	-0.0650
	(0.0322)	(0.0325)	(0.0323)	(0.0546)	(0.0551)	(0.0550)
$(transform)\ No.\ followers$	1.294***	1.316***	1.308***	6.169***	6.367***	6.427***
	(0.0712)	(0.0724)	(0.0717)	(0.407)	(0.409)	(0.405)
$female$	0.267***	0.275***	0.286***	0.0344	0.0301	0.0288
	(0.0355)	(0.0358)	(0.0358)	(0.0643)	(0.0640)	(0.0639)
$No.\ pic$				0.0416***	0.0427***	0.0436***
				(0.0105)	(0.0107)	(0.0107)
Constant	1.365***	1.410***	1.825***	0.112	0.204***	0.488***
	(0.0891)	(0.0818)	(0.0902)	(0.0864)	(0.0737)	(0.0902)
Observations	3,872	3,872	3,872	1,208	1,208	1,208
R-squared	0.245	0.228	0.238	0.360	0.343	0.350

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

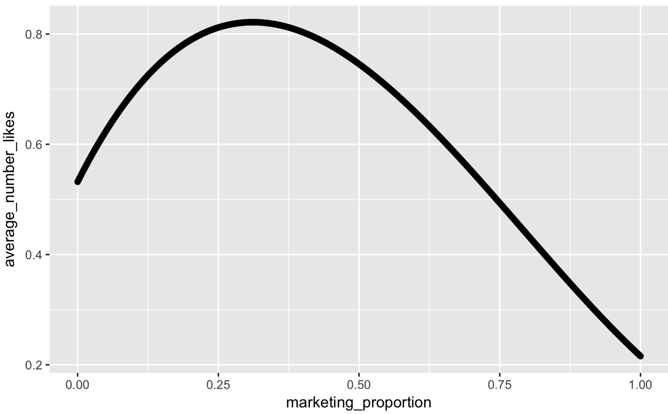

Figure 25: Cubic Specification 2016

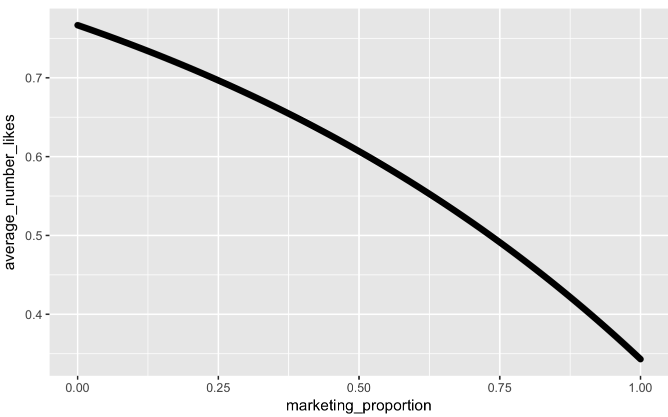

Figure 26: Exponential Specification 2014

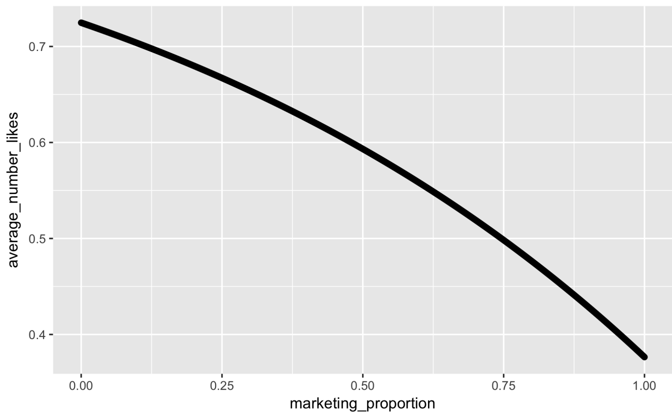

Figure 27: Exponential Specification 2016

with MNB, it results in less microblogs classified as marketing (see details in Table 48 and Table 56).

We use a simple linear regression to explore the relation between the marketing aggressiveness level and the marketing popularity. Similarly, we add the squared term of the marketing proportion to test Hypothesis. Instead of performing the Yeo Johnson transformation, we simply use the logarithm form of the number of followers. The regression results are shown in Table 58. We notice that the coefficient of marketing proportion and its quadratic term are both significant. As expected, the coefficient of the quadratic term is negative in both specifications, using respectively the two dictionaries in the classification of microblogs. Moreover, the average number of likes is maximized when the proportion of marketing microblogs is respectively 37.9% and 36.9%. The relationships between the marketing popularity and marketing aggressiveness level of the two specifications are respectively displayed in Figure 28 and Figure 29, which is similar to the previous ones. Hence, the results of the dictionary approach further support our hypothesis.

Table 56: Statistics by Data Processing Steps

Steps	No. sellers	No. microblogs	No. marketing_microblogs	marketing proportion
1. Initial sample	52,187	465,812	-	-
2. sellers who post	18,809	465,812	-	-
		Using the 50 most frequent words		
3. sellers who market	10,851	416,804	138,071	33.1%
4. final sample*	6,215	397,358	129,484	32.6%
		Using the 100 most frequent words		
3. sellers who market	11,381	423,321	152,519	36.0%
4. final sample*	6,424	402,699	143,082	35.5%

* In the final sample we remove the outliers and sellers who post less than 10 microblogs.

Table 57: Correlation Matrix

Using the 50 most frequent words

Variables	N	mean	sd	1	2	3	4	5	6	7
1. $likes\_average\_marketing$	6,215	0.573	1.046	1.000						
2. $marketing\_proportion$	6,215	0.299	0.248	0.026	1.000					
3. $marketing\_proportion^2$	6,215	0.151	0.218	-0.014	0.954	1.000				
4. $female$	6,215	0.744	0.437	0.095	0.067	0.060	1.000			
5. $\log No.\ followers$	6,215	7.113	1.838	0.445	0.105	0.074	0.020	1.000		
6. Tao	6,215	0.571	0.495	0.072	0.102	0.078	-0.067	0.147	1.000	
7. V	6,215	0.213	0.409	0.146	0.047	0.031	-0.177	0.316	0.127	1.000

Using the 100 most frequent words

Variables	N	mean	sd	1	2	3	4	5	6	7
1. $likes\_average\_marketing$	6,424	0.591	1.082	1.000						
2. $marketing\_proportion$	6,424	0.321	0.259	0.015	1.000					
3. $marketing\_proportion^2$	6,424	0.171	0.234	-0.020	0.958	1.000				
4. $female$	6,424	0 .744	0.436	0.091	0.074	0.071	1.000			
5. $\log No.\ followers$	6,424	7.100	1.836	0.444	0.111	0.081	0.021	1.000		
6. Tao	6,424	0.565	0.496	0.060	0.109	0.084	-0.073	0.145	1.000	
7. V	6,424	0.211	0.408	0.149	0.044	0.027	-0.175	0.316	0.128	1.000

Table 58: Regression Results Using the Dictionary Approach

VARIABLES	Dictionary 50	Dictionary 100
$marketing\_proportion$	1.073***	0.975***
	(0.147)	(0.152)
$marketing\_proportion^2$	-1.414***	-1.320***
	(0.156)	(0.157)
$\log No.\ followers$	0.245***	0.254***
	(0.00920)	(0.00953)
Tao	0.0183	-0.00139
	(0.0241)	(0.0246)
V	0.0562	0.0671*
	(0.0360)	(0.0369)
$female$	0.219***	0.222***
	(0.0273)	(0.0277)
Constant	-1.460***	-1.480***
	(0.0647)	(0.0653)
Observations	6,215	6,424
R-squared	0.214	0.212

Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

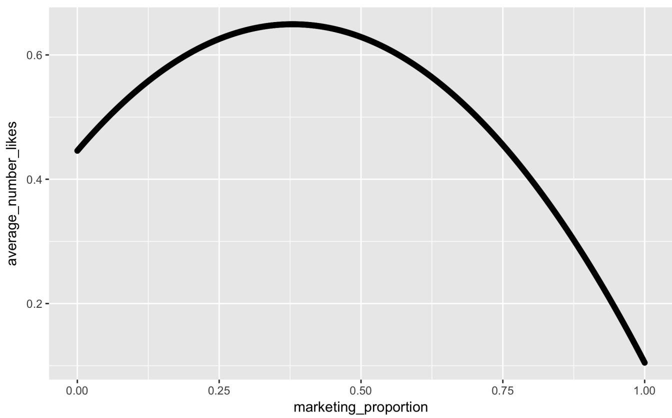

Figure 28: The Relationship between Likes and Followers Using the Dictionary of 50 Words

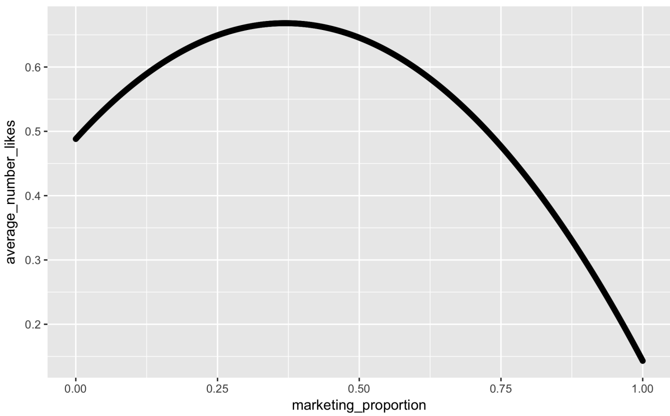

Figure 29: The Relationship between Likes and Followers Using the Dictionary of 100 Words

Chinese	English Translation	Frequency	Chinese	English Translation	Frequency
代购	overseas purchasing agents	17962	促销	promotion	3641
价格	price	13452	大衣	overcoat	3561
包邮	delivery for free	13051	¥	Chinese currency	3547
双11	11 November	12351	原价	original price	3274
淘宝	Taobao	12009	预定	product reservation	3257
面膜	facial mask	11851	限量	limited sales	3229
双十一	11 November	11565	同款	the same style	2932
新款	new style	11004	下单	place order	2783
套装	suit	9220	连衣裙	women's dress	2775
现货	goods in stock	8495	羽绒服	down jacket	2559
产品	product	8410	包包	bags	2541
特价	special offer	8378	订单	order	2536
新品	new arrival	8211	直邮	direct delivery	2252
正品	quality products	7680	爆款	hot sales	2174
店铺	store	7202	本店	our store	2098
专柜	store	6459	抢购	rush to buy	2080
毛衣	sweaters	6324	打折	discounts	1897
优惠	discounts	5845	11.11	11 November	1896
发货	delivery	5614	付款	payment	1879
商品	goods	4792	男装	men's wear	1782
天猫	Tmall shop	4576	1111	11 November	1675
￥	Chinese currency	4573	实惠	affordable	1639
这款	this style	4311	眼霜	eye cream	1639
进口	imported products	4153	女鞋	women's shoes	1621
女装	women's wear	3793	旗舰店	flagship store	1583

Figure 30: The 50 Most Frequent Words Related to Marketing in the Data of 2014

Chinese	English Translation	Frequency	Chinese	English Translation	Frequency	Chinese	English Translation	Frequency
代购	overseas purchasing agents	17962	羽绒服	down jacket	2559	大码	large size	1231
价格	price	13452	包包	bags	2541	出货	delivery	1230
包邮	delivery for free	13051	订单	order	2536	奶粉	milk powder	1226
双11	11 November	12351	直邮	direct delivery	2252	牛仔裤	jeans	1225
淘宝	Taobao	12009	爆款	hot sales	2174	UGG	an American brand	1224
面膜	facial mask	11851	本店	our store	2098	洗面奶	cleanser	1218
双十一	11 November	11565	抢购	rush to buy	2080	彩妆	cosmetics	1202
新款	new style	11004	打折	discounts	1897	黑头	blackhead	1181
套装	suit	9220	11.11	11 November	1896	条纹	stripe	1173
现货	goods in stock	8495	付款	payment	1879	肤色	complexion	1157
产品	product	8410	男装	men’s wear	1782	紧致	skin compactness	1108
特价	special offer	8378	1111	11 November	1675	护手霜	hand cream	1105
新品	new arrival	8211	实惠	affordable	1639	售价	price	1102
正品	quality products	7680	眼霜	eye cream	1639	童装	children’s wear	1093
店铺	store	7202	女鞋	women's shoes	1621	原装	the original	1082
专柜	store	6459	旗舰店	flagship store	1583	棉服	cotton clothing	1076
毛衣	sweaters	6324	购物车	shopping cart	1520	化妆品	cosmetic	1072
优惠	discounts	5845	补货	restock	1499	发售	on sale	1062
发货	delivery	5614	店里	store	1469	光棍节	Singles’ Day	1006
商品	goods	4792	靴子	boots	1438	纯天然	purely natral	1005
天猫	Tmall shop	4576	性价比	quality-price ratio	1415	小店	store	1002
Ｙ	Chinese currency	4573	长款	large size	1414	采购	purchase	989
这款	this style	4311	taobao	Taobao shop	1388	眼影	eye shadow	981
进口	imported products	4153	纯棉	pure cotton	1358	短款	short section	958
女装	women’s wear	3793	网店	online store	1333	大促	big promotion	946
促销	promotion	3641	预订	product reservation	1320	邮费	delivery costs	944
大衣	overcoat	3561	牌子	brand	1312	热卖	hot sales	941
¥	Chinese currency	3547	棉衣	cotton clothing	1304	风衣	windbreaker	927
原价	original price	3274	内衣	underwear	1275	香奈儿	Chanel	905
预定	product reservation	3257	均码	free size	1257	短裤	short pants	905
限量	limited sales	3229	洁面	cleansing	1251	冬装	winter clothing	892
同款	the same style	2932	乳液	lotion	1245	订购	order	890
下单	place order	2783	羊皮	sheepskin	1236			
连衣裙	women’s dress	2775	询价	inquiry of price	1233			

Figure 31: The 100 Most Frequent Words Related to Marketing in the Data of 2014

Chapter 5

Conclusion

The dissertation exploits large datasets in the context of business networks and applies extensive data analysis and computation tools. In particular, we employ econometric models, novel network measures and machine learning algorithms in global inter-firm ownership network and Chinese social media data.

In Chapters 2 and 3 we empirically investigate the link between a firm's performance and its network position within international business groups using the data on Italian firms. In Chapter 2, we find a positive relationship between firm centrality and performance in small business groups. Moreover, we provide evidence that the group size ″premium″ for the peripheral firms is larger than for the central ones. Our contribution to the literature on ownership network of business group is twofold. First, we provide a new measure of centrality, which reduces the high degree of collinearity between component size and traditional measures such as harmonic centrality and ownership level. Our measure makes centrality more comparable across components of different sizes. Second, we discuss the role of a firm's network position in the group on its performance by providing a more detailed analysis of its interaction with business group size, which is underexploited in the literature on organizational network.

In Chapter 3, we investigate the benefit of foreign ownership to domestic firms from diverse perspectives. Our findings reveal that the Italian subsidiaries owned by foreign direct owners are more productive than domestic direct owners. Furthermore, we notice that the Italian subsidiaries with shorter organizational or geographical distance from their foreign owners are more productive. In addition, subsidiaries with multiple cross-border links in the upstream ownership chain are found to have a higher productivity. From the results we gain a deeper insight into the role of a firm's position and foreign ownership in the business group on its performance.

In Chapter 4, we discuss C2C business in social media. Thanks to the collaboration between China's largest C2C e-commerce platform Taobao and China's largest microblogging platform Sina Weibo, we are able to identify the C2C sellers in social media and analyse their marketing behaviour over microblogs. We explore how aggressive the sellers should be when promoting their products over Sina Weibo. Interestingly, we find an inverted U-shape between marketing popularity and marketing aggressiveness level, and the optimal proportion of marketing-related contents is around 30%. To the best of our knowledge, this chapter is the first empirical work to provide an optimal marketing level in the literature of social media marketing. In addition, our research presents a first attempt to study the C2C sellers' marketing behaviour in social media by collecting and analysing this unique data set. Since we have also collected the sellers' sales data, we will investigate the impact of their features and behaviour in social media on their sales performance in our future work. By incorporating these features in the forecast of sales, sellers can manage their operations better and thus improve profit margin.

References

Aghion, P. and J. Tirole (1997). Formal and real authority in organizations. *Journal of political economy 105*(1), 1–29. 19

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly 45*(3), 425–455. 9, 11

Akar, E. and B. Topçu (2011). An examination of the factors influencing consumers' attitudes toward social media marketing. *Journal of Internet Commerce 10*(1), 35–67. 3, 79

Allahyari, M., S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*. 86

Almeida, H. V. and D. Wolfenzon (2006). A theory of pyramidal ownership and family business groups. *The Journal of Finance 61*(6), 2637–2680. 6, 9

Altomonte, C. and A. Rungi (2013). Business groups as hierarchies of firms: determinants of vertical integration and performance. 11, 12

Ansoff, H. I. (1965). *Corporate strategy: business policy for growth and expansion*. McGraw-Hill Book. 15

Antras, P. and E. Helpman (2004). Global sourcing. *Journal of Political Economy 112*(3), 552–580. 51

Arnold, J. M. and B. S. Javorcik (2005). Gifted kids or pushy parents? foreign acquisitions and plant performance in indonesia. 1, 51, 53

Baesens, B. (2014). *Analytics in a big data world: The essential guide to data science and its applications*. John Wiley & Sons. 88

Baesens, B., V. Van Vlasselaer, and W. Verbeke (2015). *Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection*. John Wiley & Sons. 88, 93

Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *science 286*(5439), 509–512. 18, 22

Bauer, H. H., T. Reichardt, S. J. Barnes, and M. M. Neumann (2005). Driving consumer acceptance of mobile marketing: A theoretical framework and empirical study. *Journal of electronic commerce research 6*(3), 181. 79

Belenzon, S. and T. Berkovitz (2010). Innovation in business groups. *Management Science 56*(3), 519–535. 6, 9, 12

Belenzon, S., N. Hashai, and A. Patacconi (2017). The architecture of attention: Group structure and subsidiary autonomy. 6, 18, 30, 31, 34, 56

Belenzon, S., A. Patacconi, and B. Zelner (2013). The nebulous nature of the firm: An empirical study of corporate group structure in 16 developed economies. 6, 11, 19, 36

Bentivogli, C., L. Mirenda, et al. (2016). Foreign ownership and performance: evidence from a panel of italian firms. Technical report, Bank of Italy, Economic Research and International Relations Area. 53

Bernstein, J. I. and P. Mohnen (1998). International r&d spillovers between us and japanese r&d intensive sectors. *Journal of international economics 44*(2), 315–338. 2, 9, 51

Bethel, J. E. and J. P. Liebeskind (1998). Diversification and the legal organization of the firm. *Organization Science 9*(1), 49–67. 6, 11

Blomström, M. and F. Sjöholm (1999). Technology transfer and spillovers: does local participation with multinationals matter? *European economic review 43*(4), 915–923. 2, 6, 51

Bruno, G. S. and A. M. Falzoni (2003). Multinational corporations, wages and employment: Do adjustment costs matter? *Applied Economics 35*(11), 1277–1290. 53

Brush, T. H., P. Bromiley, and M. Hendrickx (2000). The free cash flow hypothesis for sales growth and firm performance. *Strategic Management Journal*, 455–472. 15

Cainelli, G. and D. Iacobucci (2011). Business groups and the boundaries of the firm. *Management Decision 49*(9), 1549–1573. 6, 9, 34

Cainelli, G., D. Iacobucci, and E. Morganti (2006). Spatial agglomeration and business groups: new evidence from italian industrial districts. *Regional Studies 40*(5), 507–518. 6

Caliendo, L. and E. Rossi-Hansberg (2012). The impact of trade on organization and productivity. *The quarterly journal of economics 127*(3), 1393–1467. 6, 11

Castellani, D. and A. Zanfei (2003). Innovation, foreign ownership and multinationality. an empirical analysis on italian manufacturing firms. In *29th EIBA conference" MNC as a Knowing Organization*, pp. 11–13. Citeseer. 53

Chang, S. J. and J. Hong (2000). Economic performance of group-affiliated companies in korea: Intragroup resource sharing and internal business transactions. *Academy of Management Journal 43*(3), 429–448. 11

Chu, S.-C. and S. M. Choi (2011). Electronic word-of-mouth in social networking sites: a cross-cultural study of the united states and china. *Journal of Global Marketing 24*(3), 263–281. 79

Chu, S.-C. and Y. Kim (2011). Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites. *International journal of Advertising 30*(1), 47–75. 79

Cofta, P. (2006). Convergence and trust in ecommerce. *BT Technology Journal 24*(2), 214–218. 84

Collins, C. J. and K. D. Clark (2003). Strategic human resource practices, top management team social networks, and firm performance: The role of human resource practices in creating organizational competitive advantage. *Academy of management Journal 46*(6), 740–751. 15

Corbetta, G. and S. Tomaselli (1996). Boards of directors in italian family businesses. *Family Business Review 9*(4), 403–421. 7, 34

Criscuolo, C. and R. Martin (2009). Multinationals and us productivity leadership: evidence from great britain. *The Review of Economics and Statistics 91*(2), 263–281. 53

Cucculelli, M. and G. Micucci (2008). Family succession and firm performance: Evidence from italian family firms. *Journal of Corporate Finance 14*(1), 17–31. 7, 34

De La Potterie, B. V. P. and F. Lichtenberg (2001). Does foreign direct investment transfer technology across borders? *the Review of Economics and Statistics 83*(3), 490–497. 53

De Vries, L., S. Gensler, and P. S. Leeflang (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing 26*(2), 83–91. 79, 82, 84

Del Prete, D. and A. Rungi (2015). Organizing the global value chain: a firm-level test. 18

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837–845. 89

Desai, M. A., C. F. Foley, and J. R. Hines (2004). Foreign direct investment in a world of multiple taxes. *Journal of Public Economics 88*(12), 2727–2744. 2, 51

Desai, M. A., C. F. Foley, and J. R. Hines (2006). The demand for tax haven operations. *Journal of Public economics 90*(3), 513–531. 2, 52

Desai, M. A., C. F. Foley, and J. R. Hines Jr (2005). Foreign direct investment and domestic economic activity. Technical report, National Bureau of Economic Research. 53

Dess, G. G. and R. B. Robinson (1984). Measuring organizational performance in the absence of objective measures: the case of the privately-held firm and conglomerate business unit. *Strategic management journal 5*(3), 265–273. 15

Doms, M. E. and J. B. Jensen (1998). Comparing wages, skills, and productivity between domestically and foreign-owned manufacturing establishments in the united states. In *Geography and ownership as bases for economic accounting*, pp. 235–258. University of Chicago Press. 53

Dunning, J. H. (1988). The eclectic paradigm of international production: A restatement and some possible extensions. *Journal of international business studies 19*(1), 1–31. 51

Dunning, J. H. (1989). *Transnational corporations and the growth of services: Some conceptual and theoretical issues*. New York: United Nations. 2, 51

Eighmey, J. and L. McCord (1998). Adding value in the information age: Uses and gratifications of sites on the world wide web. *Journal of business research 41*(3), 187–194. 82

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 175–185. 103

Fortin, D. R. and R. R. Dholakia (2005). Interactivity and vividness effects on social presence and involvement with a web-based advertisement. *Journal of business research 58*(3), 387–396. 84

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks 1*(3), 215–239. 11

Gangadharbatla, H. (2008). Facebook me: Collective self-esteem, need to belong, and internet self-efficacy as predictors of the igenerations attitudes toward social networking sites. *Journal of interactive advertising 8*(2), 5–15. 81

Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega 28*(6), 725–737. 83, 84

Gefen, D. and D. W. Straub (2004). Consumer trust in b2c e-commerce and the importance of social presence: experiments in e-products and e-services. *Omega 32*(6), 407–424. 83

Girma, S., D. Greenaway, and K. Wakelin (2001). Who benefits from foreign direct investment in the uk? *Scottish Journal of Political Economy 48*(2), 119–133. 53

Glattfelder, J. B. (2010). Ownership networks and corporate control. 9

Globerman, S., J. C. Ries, and I. Vertinsky (1994). The economic performance of foreign affiliates in canada. *Canadian Journal of Economics*, 143–156. 53

Goto, A. (1982). Business groups in a market economy. *European economic review 19*(1), 53–70. 9

Grant, I. C. (2005). Young peoples' relationships with online marketing practices: An intrusion too far? *Journal of Marketing Management 21*(5-6), 607–623. 3, 79

Greenaway, D., A. Guariglia, and Z. Yu (2014). The more the better? foreign ownership and corporate performance in china. *The European Journal of Finance 20*(7-9), 681–702. 53

Guan, W., H. Gao, M. Yang, Y. Li, H. Ma, W. Qian, Z. Cao, and X. Yang (2014). Analyzing user behavior of the micro-blogging website sina weibo during hot social events. *Physica A: Statistical Mechanics and its Applications 395*, 340–351. 79, 84

Gulati, R. and M. Gargiulo (1999). Where do interorganizational networks come from? 1. *American journal of sociology 104*(5), 1439–1493. 9, 10

Gulati, R., N. Nohria, and A. Zaheer (2000). Strategic networks. *Strategic management journal*, 203–215. 9

Haans, R. F., C. Pieters, and Z.-L. He (2015). Thinking about u: theorizing and testing u-and inverted u-shaped relationships in strategy research. *Strategic Management Journal*. 103

Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research 56*(3), 387–404. 83

He, Y. and L. Song (2015). The study on factors affecting the attraction of microblog. *Computers in Human Behavior*. 79

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics 123*(2), 441–487. 51

Hijzen, A., T. Inui, Y. Todo, et al. (2007). The effects of multinational production on domestic performance: Evidence from japanese firms. *Discussion papers 7006*. 53

Hofer, C. W. and D. Schendel (1978). *Strategy formulation: Analytical concepts*. West Publ. 15

Hoffman, D. L., T. P. Novak, and M. Peralta (1999). Building consumer trust online. *Communications of the ACM 42*(4), 80–85. 83, 84

Iacus, S. M., G. King, and G. Porro (2008). Matching for causal inference without balance checking. *Available at SSRN 1152391*. 26

James, G., D. Witten, and T. Hastie (2014). An introduction to statistical learning: With applications in r. 88

Kali, R. and J. Sarkar (2011). Diversification and tunneling: Evidence from indian business groups. *Journal of Comparative Economics 39*(3), 349–367. 9

Kaplan, A. M. and M. Haenlein (2010). Users of the world, unite! the challenges and opportunities of social media. *Business horizons 53*(1), 59–68. 3, 78, 79

Katz, E., H. Haas, and M. Gurevitch (1973). On the use of the mass media for important things. *American sociological review*, 164–181. 82

Khanna, T. and K. Palepu (2000). Is group affiliation profitable in emerging markets? an analysis of diversified indian business groups. *The Journal of Finance 55*(2), 867–891. 1, 6

Kim, H.-b., T. T. Kim, and S. W. Shin (2009). Modeling roles of subjective norms and etrust in customers' acceptance of airline b2c ecommerce websites. *Tourism management 30*(2), 266–277. 84

King, R. A., P. Racherla, and V. D. Bush (2014). What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *Journal of Interactive Marketing 28*(3), 167–183. 79

Kohler, W., M. Smolka, et al. (2012). Global sourcing: Evidence from spanish firm-level data. *Quantitative Analysis of Newly Evolving Patterns of International Trade*, 139–189. 51

Kostić, Z. (2013). Mergers and acquisitions during the economic-financial crisis. 7

Kumar, V. and R. Mirchandani (2012). Increasing the roi of social media marketing. *MIT sloan management review 54*(1), 55. 79

Kwon, S.-W., J. Haleblian, J. Hagedoorn, et al. (2016). In country we trust? national trust and the governance of international r&d alliances. *Journal of International Business Studies 47*(7), 807–829. 10

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. W. Vishny (1997). Legal determinants of external finance. *Journal of finance*, 1131–1150. 1, 6

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. W. Vishny (1998). Law and finance. *Journal of political economy 106*(6), 1113–1155. 1, 6

Lek, M., B. Anadarajah, N. Cerpa, and R. Jamieson (2001). Data mining prototype for detecting ecommerce fraud. *ECIS 2001 Proceedings*, 60. 84

Leskovec, J., A. Rajaraman, and J. D. Ullman (2014). *Mining of massive datasets*. Cambridge university press. 87

Lessmann, S., B. Baesens, C. Mues, and S. Pietsch (2008). Benchmarking classification models for software defect prediction: A proposed framework and novel findings. *IEEE Transactions on Software Engineering 34*(4), 485–496. 89, 93

Li, Y., H. Gao, M. Yang, W. Guan, H. Ma, W. Qian, Z. Cao, and X. Yang (2015). What are chinese talking about in hot weibos? *Physica A: Statistical Mechanics and its Applications 419*, 546–557. 84

Lin, K.-Y. and H.-P. Lu (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior 27*(3), 1152–1161. 79, 83

Lind, J. T. and H. Mehlum (2010). With or without u? the appropriate test for a u-shaped relationship. *Oxford bulletin of economics and statistics 72*(1), 109–118. 95, 103

Lovelock, C. H. and G. S. Yip (1996). Developing global strategies for service businesses. *California management review 38*(2), 64–86. 2, 51

Lu, Y., L. Zhao, and B. Wang (2010). From virtual community members to c2c e-commerce buyers: Trust in virtual communities and its effect on consumers purchase intention. *Electronic Commerce Research and Applications 9*(4), 346–360. 83

Mahmood, I. P., H. Zhu, and A. Zaheer (2017). Centralization of intragroup equity ties and performance of business group affiliates. *Strategic Management Journal 38*(5), 1082–1100. 11

Mangold, W. G. and D. J. Faulds (2009). Social media: The new hybrid element of the promotion mix. *Business horizons 52*(4), 357–365. 3, 79

Mani, D. and J. Moody (2014). Moving beyond stylized economic network models: The hybrid world of the indian firm ownership network 1. *American Journal of Sociology 119*(6), 1629–1669. 10

Manning, Christopher, P. R. and H. Schutze. (2008). *Introduction to information retrieval.* Cambridge University Press. 86, 87, 88

Markusen, J. R. (1995). The boundaries of multinational enterprises and the theory of international trade. *The Journal of Economic Perspectives 9*(2), 169–189. 1, 6, 9, 51

Michaelidou, N., N. T. Siamagka, and G. Christodoulides (2011). Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium b2b brands. *Industrial Marketing Management 40*(7), 1153–1159. 79

Navaretti, G. B. and D. Castellani (2005). Do italian firms improve their performance at home by investing abroad? 53

Navaretti, G. B., D. Castellani, and A.-C. Disdier (2010). How does investing in cheap labour countries affect performance at home? firm-level evidence from france and italy. *Oxford Economic Papers 62*(2), 234–260. 53

Navaretti, G. B., A. Venables, and F. Barry (2004). *Multinational firms in the world economy*. Princeton University Press. 53

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review 45*(2), 167–256. 1, 6, 10, 16

Owen-Smith, J. and W. W. Powell (2004). Knowledge networks as channels and conduits: The effects of spillovers in the boston biotechnology community. *Organization science 15*(1), 5–21. 10

Patacconi, A. (2009). Coordination and delay in hierarchies. *The RAND Journal of Economics 40*(1), 190–208. 6, 11

Pentina, I., L. Zhang, and O. Basmanova (2013). Antecedents and consequences of trust in a social media brand: A cross-cultural study of twitter. *Computers in Human Behavior 29*(4), 1546–1555. 84

Peterson, M. and R. Rajan (1994). The benefits of lending relationships: Evidence from small business data. *Journal of Finance 49*(1), 3–37. 9

Powell, W. W., K. W. Koput, and L. Smith-Doerr (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative science quarterly*, 116–145. 9, 10, 11

Powell, W. W., K. W. Koput, L. Smith-Doerr, and J. Owen-Smith (1999). Network position and firm performance: Organizational returns to collaboration in the biotechnology industry. *Research in the Sociology of Organizations 16*(1), 129–159. 5

Provan, K. G., A. Fish, and J. Sydow (2007). Interorganizational networks at the network level: A review of the empirical literature on whole networks. *Journal of management 33*(3), 479–516. 9, 11

Rapp, A., L. S. Beitelspacher, D. Grewal, and D. E. Hughes (2013). Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science 41*(5), 547–566. 79

Reichheld, F. F. and P. Schefter (2000). E-loyalty. *Harvard business review 78*(4), 105–113. 84

Rochat, Y. (2009). Closeness centrality extended to unconnected graphs: The harmonic centrality index. In *ASNA*, Number EPFL-CONF-200525. 2, 6, 10, 16

Ruggiero, T. E. (2000). Uses and gratifications theory in the 21st century. *Mass communication & society 3*(1), 3–37. 82

Salton, G. and C. Buckley (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management 24*(5), 513–523. 87

Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika 67*(2), 429–439. 103

Schilling, M. A. and C. C. Phelps (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science 53*(7), 1113–1126. 2, 6, 9, 10

Schivinski, B. and D. Dabrowski (2014). The effect of social media communication on consumer perceptions of brands. *Journal of Marketing Communications*, 1–26. 82

Soh, P.-H. (2003). The role of networking alliances in information acquisition and its implications for new product performance. *Journal of Business Venturing 18*(6), 727–744. 10

Swani, K., B. P. Brown, and G. R. Milne (2014). Should tweets differ for b2b and b2c? an analysis of fortune 500 companies' twitter communications. *Industrial Marketing Management 43*(5), 873–881. 79

Taylor, D. G., J. E. Lewin, and D. Strutton (2011). Friends, fans, and followers: Do ads work on social networks?. *Journal of Advertising Research 51*(1), 258 – 275. 79

Temouri, Y., N. L. Driffield, and D. A. Higón (2008). Analysis of productivity differences among foreign and domestic firms: evidence from germany. *Review of World Economics 144*(1), 32–54. 53

Tomiura, E. (2007). Foreign outsourcing, exporting, and fdi: A productivity comparison at the firm level. *Journal of International Economics 72*(1), 113–127. 51

Tsai, W. (2001). Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. *Academy of management journal 44*(5), 996–1004. 5

Tsai, W. and S. Ghoshal (1998). Social capital and value creation: The role of intrafirm networks. *Academy of management Journal 41*(4), 464–476. 5, 9

Tsai, W.-H. S. and L. R. Men (2013). Motivations and antecedents of consumer engagement with brand pages on social networking sites. *Journal of Interactive Advertising 13*(2), 76–87. 79

UNCTAD (2015). World investment report 2015. reforming international investment governance. 2, 52

UNCTAD (2016). World investment report 2016. investor nationality: Policy challenges. 2, 12, 52, 53, 55, 56

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American sociological review*, 674–698. 5, 10

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative science quarterly*, 35–67. 5, 9, 10

Van Gestel, T., B. Baesens, P. Van Dijcke, J. Garcia, J. A. Suykens, and J. Vanthienen (2006). A process model to develop an internal rating system: Sovereign credit ratings. *Decision Support Systems 42*(2), 1131–1151. 93

Van Gestel, T., B. Baesens, P. Van Dijcke, J. Suykens, J. Garcia, and T. Alderweireld (2005). Linear and nonlinear credit scoring by combining logistic regression and support vector machines. *Journal of credit Risk 1*(4). 93

Vitali, S. and S. Battiston (2011). Geography versus topology in the european ownership network. *New Journal of Physics 13*(6), 063021. 9

Vitali, S., J. B. Glattfelder, and S. Battiston (2011). The network of global corporate control. *PloS one 6*(10), e25995. 9

Wang, R. and Y. Jin (2010). An empirical study on the relationship between the followers' number and influence of microblogging. In *E-Business and E-Government (ICEE), 2010 International Conference on*, pp. 2014–2017. IEEE. 84

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of ?small-world?networks. *nature 393*(6684), 440–442. 18, 22

Wiersema, F. (2013). The b2b agenda: The current state of b2b marketing and a look ahead. *Industrial Marketing Management 4*(42), 470–488. 79

Yeo, I.-K. and R. A. Johnson (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 954–959. 93, 102

Yu, L., S. Asur, and B. A. Huberman (2011). What trends in chinese social media. *arXiv preprint arXiv:1107.3522*. 82

Zaheer, A. and G. G. Bell (2005). Benefiting from network position: firm capabilities, structural holes, and performance. *Strategic management journal 26*(9), 809–825. 5, 9