國立政治大學財務管理研究所

博士論文

Two Essays of the Market Friction Effects on Asset Prices: Evidence from Syndicated Loan and Futures Markets

指導教授：張元晨　博士

研究生：吳偉劭　撰

中華民國一百零一年七月

謝辭

一個人一個故事，每個生命的背後都有她自己獨特的故事，我也努力的寫著屬於自己的故事，不求故事有多精彩，只求可以感動自己，感動那些我看重的人，於是我不停筆的寫著，我知道，在寫這則故事的過程中我得到很多，也失去了不少，我都知道！故事寫到這，我暫時停筆，因為我想停下腳步來感謝一些人！

猶記得十年前自己最常思考的問題就是“我到底想要做什麼？”這個問題在十年前困擾自己很深，就在靈光一現的過程中，我循著上天的指引，辭去銀行的工作，一步無悔的走進研究所，從碩士班開始，轉眼間博士學位也要完成了；對自己而言，取得博士學位是一種事業的選擇，而不是為了享有博士的虛榮，但現實環境總是嚴苛無情的考驗著每個人的意志力，如果不小心忘了初衷，很容易就會迷失、迷惘，所幸世間有惡必有善，上天在不斷考驗你的同時，也派了許多天使來引導你，對於這些天使，我要表達自己最深切的感恩！

首先，我要感謝在上天一直看顧我的父母，我把對你們的思念全部化為勇往直前的力量，再艱難的挫折，我都因你們而勇敢！我要感謝我大哥、大姊、二姊、三姊，你們是我最強有力的後盾，謝謝你們對我任何的決定都是那麼的支持，沒有你們的支持我走不到今天，我任何的榮耀，都是屬於你們的。

我要感謝我的指導老師張元晨老師、劉玉珍老師、李怡宗老師，你們不但在做學問上讓我得到最多的學習，在生活上也給我極大的支持，我在你們身上學到的，早已超越學術的範疇，你們是我將來有機會為人師表最佳的典範，將來我必以相同的胸襟對待自己的學生；我要感謝杜化宇老師，謝謝杜老師對我的信任與提攜，這份恩情我會永遠銘記在心；我還要回過頭感謝我的碩士班指導老師陳樹衡老師，謝謝陳老師在我碩士階段的照顧，在陳老師身上學到的精神永植我心！

我要感謝文謙學長，除了平常學問的切磋之外，也在我挫折最多的時候給我最大的安慰，沒有你，我不曉得該如何排解那一段苦悶；我要感謝我的博士班同學肇榮，猶記得那一年我們一起準備學科考，何媽媽的便當讓我免於除夕夜挨餓

的窘境；感謝曉梅，我們總是可以掏心掏肺的聊天；感謝健偉，你總能用你開朗的天性驅散別人的憂鬱；感謝胤哲總是友善的面對每個人、感謝崇齡，女中豪傑的最佳典範，不論大家最後選擇了什麼路，那段有你們的時光，是再怎麼艱辛彼此都不孤單的時光；我要感謝系辦的助教郁芳及雪玲，謝謝你們凡事都那麼的幫忙；我要感謝淑華、楚彬、還有玉美學姊，你們給我極多的幫助，雖然我很少表達什麼，但很多事我都感受在心裡；我要感謝所有財管系博士班的成員美菁、苡文、周燕、富德…，謝謝你們這些日子所帶來的歡笑、汗水、甚至淚水，也謝謝你們對我的諸多包容。

我還要感謝我的幾位碩士班同學，惟毅、毓芝、仁傑、家綺、秋媚、顥一、俊緯，謝謝你們義氣相挺，從碩士班一路陪我到博士班；也謝謝以前經研所的學長科宏、秉聰；我要感謝吳經理，謝謝你多年來的關照，跟你在形而上想法的交流，讓我對生命有更深切的體悟，你是我另類的指導老師；我還要感謝更多更多的人…

要感謝的人太多了，這一刻才發現自己是何等渺小，為了完成屬自己個人的階段性任務，竟然需要集結那麼多人的力量，希望自己將來有能力可以加倍回饋於這個社會！

故事寫到這裡，我合上書本，不是結束，而是即將開啟新頁；謝謝在這塊土地上曾經陪伴過我的一切，另一塊土地正在等著我，無論走到哪我都會繼續寫著屬於自己的故事，也希望有機會可以常常回過頭來聽聽大家的故事！

Preface

In this dissertation, two essays are used to illustrate how market friction would influence asset prices. In the first essay, I use the data from the U.S. syndicated loan market to show how communication barriers among lenders would affect their decisions, which in turn affect loan spreads. If potential lenders can't freely communicate with each other, those who make decision after will learn by observing the decisions of their predecessors. Once their own signals are dominated by the information revealed by their predecessors', those who act later will rationally ignore their own signals and follow suit. In economics, this is called "informational cascade." In the first essay, I will show how cascade effect can influence loan spreads and non-price contract terms.

In the second essay, I use the data from Taiwan Futures Exchange to examine the lead-lag relationships between the index futures returns, index futures volume, and spot returns. The lead-lag relationship depends on which venue informed traders will choose to trade and informed traders' choice may depend on short sale constraints, transaction costs, leverage effect, and so on. In my second essay, I will show which market is the venue the informed traders choose to trade and which type of traders tend to be informed traders.

The two essays of this dissertation have been transformed into working papers for conference presentation and journal submission. The first working paper based on Chapter II, The Cascade Effect in the Syndicated Loan Market, has accepted by the 19th Conference on the Theories and Practices of Securities and Financial Markets and is presented on 10th December 2011 in Kaohsiung. The second working paper based on Chapter III, The Relationships between the Futures Returns, Futures Volume, and Spot Returns, has been submitted to Journal of Financial Studies.

Abstract

Two essays are comprised in this dissertation to explore how market friction affects the processes of price formation. The first essay investigates on both theoretical and empirical bases how segmentation of communication amongst potential lenders can influence loan contracts. Two cases are considered. The first one assumes that potential lenders can freely communicate with each other; the second one assumes that each potential lender can only observe the decisions of its predecessors. I show theoretically that the ex post observed interest rate will be higher and the probability of syndication failure will be lower if the potential lenders cannot communicate freely with each other. These predictions are confirmed by my empirical work. Using a novel proxy, relational distance, for the segmentation of communication, I show that the larger the relational distance, the higher is the loan spread and the lower is the probability of syndication failure. In addition, the relational distance is positively correlated with the probability of the existence of non-price contract terms, such as the requirement for collateral and guarantees. My conclusions are found to be robust to endogeneity issues, potentially omitted variables and alternative model specifications.

The second essay focuses on the informational effects between futures market and its spot market. Intraday data are used to investigate the lead-lag relationship between the TX returns, the TX trading activity and the TAIEX stock index returns. I focus on the transmission direction and the source of information and find that there are specific lead-lag relationships between futures returns and spot returns, in addition to the contemporaneous relationship predicted by carry-cost theory and efficient market theory. The results show that futures returns significantly lead spot returns, which suggests that informed trades occur in the futures market and makes information flows from the futures market to the spot market. By distinguishing

different types of futures traders and using private information, net open buy, as a proxy for futures trading activity, I found that the major source of informed trades is foreign institutional traders because their trading activity have predictive power for future movements in both spot and futures prices. In contrary, traders in the other categories carry no information about the directional changes in both spot and futures prices.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

In a frictionless market, price should perfectly aggregate information held by market participants and reflect an asset's true value. Moreover, in two related markets such as a spot market and its derivative market should impound information simultaneously so that there is no lead-lag relationship between their price adjustments. However, market friction such as transaction costs, communication barriers, trading limitations, illiquidity, and market structure often affects the process of price formation and price discovery. This dissertation uses two markets, the U.S. syndicated loan market and Taiwan Futures market, to illustrate the effects of market friction on asset prices. There are two essays in this dissertation. In the first essay, I extend Welch (1992)'s model to the syndicated loan market to show how segmentation of communication can affect lending conditions, especially loan spreads. Two cases are considered. The first case assumes all potential lenders can freely share their information with each other. In the second one, I assume each potential lender can only observe the decisions of her predecessors. By extending Welch's model, I show that if potential lenders can freely share their information about the borrower, the ex-post observed loan spreads are lower and probability of syndication failures is higher. The intuition is that in the second case the lead bank will increase the interest rate to elicit a positive cascade and make failure become impossible. Using relational distance as a proxy for segmentation of communication, the model's predictions are confirmed by my empirical work.

The first essay contributes to the extant literature in three ways. First, to my best knowledge, this is the first study to explore cascade effect in the syndicated loan market. Although some studies have examined the herd behavior in banks' investment decisions (e.g., Jain and Gupta, 1987; Nakagawa and Uchida, 2003; Uchida and Nakagawa, 2007; Acharya and Yorulmazer, 2008), they did not explore how herd behavior affects loan pricing.

Second, several distance measures have been proposed and associated with economic decisions, for example, physical distance (Mian, 2006; and Giannetti and Yafeh, 2012), distance of specialization (Cai et al., 2011), and cultural distance (Giannetti and Yafeh, 2012). Also, many studies suggest that a tight relationship between economic agents favors communication and the dissemination of information (e.g., Baum et al., 2004; Hochberge et al., 2007; Cohen et al., 2008; and Meuleman et al., 2009). We extend these ideas to propose a novel proxy, relational distance. It is difficult to capture relationships amongst economic agents and to quantify the relational distance. We overcome these obstacles by conducting a social network analysis.

Third, to my best knowledge, this essay is also the first study to empirically test the determinants of syndication failures. It is an important issue because syndication failures are costly to borrowers and lenders and may impair investment activities. Understanding the causes of syndication failures helps to improve the success of the syndicated loan market.

In the second essay, I study the lead-lag relationship between Taiwan futures market and its spot market. According to the carry-cost theory, if the markets are frictionless, contemporaneous returns in the futures market and spot market would be perfectly correlated (Stoll and Whaley, 1990). However, transaction costs (such as fees and taxes), trading limitations (such as limitation on short sales), or the transaction characteristics of the asset itself (such as leverage effect) may cause one market to react faster to information than the other. In this essay, I show that the futures returns strongly lead its spot returns and this phenomenon is driven by the trading activity of foreign institutional traders. By distinguishing the trading volumes of different trader types, I found the net open buy of foreign institutional traders in futures market have predictive power for not only the spot returns but also the futures returns. In contrary, traders in the other categories make no directional prediction on both spot and futures returns.

The main contributions of the second essay come from three aspects. First, unlike the previous literature simply discusses the lead-lag relationship between futures price changes and spot price changes or between futures price changes and futures trading volume, I simultaneously explores the information content regards to the futures returns and spot returns, futures trading activitiy and spot returns, and futures returns and futures trading activity. It is important to analyse the relationships between the futures trading activity, futures returns, and spot returns at the same time. Only considering the relation between the two market returns, we cannot determine who tend to be informed traders. Simply examining the relation between the futures trading activity and futures returns, it is hard to conclude that informed traders do choose futures market to trade in first.

Second, past literatures specifically explored the relationship between the options trading activity and spot price changes. Under the situation where it is impossible to distinguish the types of trader, they mainly use the overall market trading activity as their analysis basis. Because it is impossible that every trade is informed trade, this choice will affects the credibility of the result, or even inconsistent explanations appear (e.g. EOS and CCF). In view of this, I not only study the relationship between the overall futures market trading activity and the spot price changes, but also further explores whether different identities of traders have different information content.

Third, in the research of trading activity, most literatures use the total trading volume directly (e.g. Though Sadath and Kamaiah, 2009; Anthony, 1988; Stephan and Whaley, 1990), or divide trading into buyer-initiated and seller- initiated (e.g. EOS; CCF). Pan and Poteshman (2006), however, found that observing the signal quality of these variables is inferior to dividing the opening and closing trades. Therefore, I uses non-public information, net open buy (open-buy volume minus open-sell volume), as the analysis variable of the futures trading activity, which can better capture the informed trades on the market.

The remainder of this dissertation is structured as follows: Chapter II investigates how cascade effect can happen in syndicated loan market and how the effect can affect the loan contract terms and the probability of syndication failure. Chapter III examines the lead-lag relationships between the price movements of futures market and its spot market and the volumes of different classes of futures traders. Chapter IV summarizes the results and concludes this dissertation.

Chapter II

The Cascade Effect in the Syndicated Loan Market

1. Introduction

The imitation in economic decision-making is prevalent in human society. It is found in choice of restaurants, schools, political candidates, research topics, investments, and initial public offerings (Banerjee, 1992; Shiller, 1995). However, it does not necessarily imply that imitators are irrational. When people make decisions in sequence and are subject to imperfect signals, they often take account of the information revealed by the actions of others. If their own signals are dominated by information stemming from the actions of their predecessors, those who act later may rationally ignore their own signals to follow suit. This phenomenon is the so-called informational cascade or herd behavior.

Welch (1992) developed a theoretical model to illustrate the role of information in generating cascades. Using the example of the IPO market, Welch showed that with perfect communication amongst all investors, all successful offerings by a risk-neutral uninformed issuer are underpriced. In the absence of perfect communication from early to late investors but allowing for late investors to observe the decisions of early investors, this can give rise to either positive or negative cascades. Given the high likelihood of a negative cascade originating, in which there is no buyer independent of the signal received, there is a tendency for sellers to under price so that all goods are sold. The model of Welch has been applied to the insurance market for special risks by D’Arcy and Oh (1997). In this paper, I extend Welch’s model to the U.S. syndicated loan market. My goal is to show that the cascade effect also affects loan pricing and the probability of syndication failure.

A syndicated loan is a loan that is provided by a group of lenders and is structured by one or several commercial or investment banks known as lead banks. In the syndication process, once the mandate has been awarded, the lead bank and the borrower establish a relationship and become “partners” (Sufi, 2007; Focarelli et al., 2008). After negotiating lending conditions and contract terms with the borrower, the lead bank turns to approach potential lenders. If potential lenders accept (or reject) the invitation nonsynchronously, then informational cascades become possible.

Whether informational cascades happen or not in the syndication process depends on the scenarios of interaction amongst potential lenders. This is the first issue I want to investigate. By extending Welch’s model, I show that if potential lenders can freely share their information about the borrower, the probability of syndication failure is always positive. This is because the loan spread which the lead bank sets may not be the same as one which reflects information held by all potential lenders. The probability that the former rate is lower than the latter is always positive. However, if each potential lender can only observe the decisions of potential lenders which were previously approached by the lead bank, the syndication is unlikely to fail. The intuition is that in this case the rational lead bank will set a higher loan spread to elicit a positive cascade in seeking to avoid costly failure. Accordingly, it is also reasonable to expect that the loan spread in this scenario will be higher than that where information is freely shared amongst potential lenders.

The second issue I want to explore is whether my model’s predictions are supported in practice. I use two proxies, physical distance and relational distance, to empirically capture the segmentation of communication amongst potential lenders. Physical distance measures the average distance between the cities in which the principal executive offices of the syndicate members are located. Long physical distances may impair communication. However, when it is examined, I find no evidence to support my model's predictions. Indeed, progress and innovations in technology, transportation, and communication may make physical distance an unbinding constraint. Giannetti and Yafeh (2012) tested the effect of physical distance between the lead bank and the borrower on loan cost and their results are also insignificant.

Relational distance captures the close relationship amongst potential lenders arising from their collaboration over the preceding five years. Three variables, average path length, clustering coefficient, and density, are used to measure relational distance. All of them are used in social network analysis. As stressed by Hochberg et al. (2007, 2010) and Cohen et al. (2008), networks facilitate the sharing of information. A tight relationship favors communication and the dissemination of information. My empirical results confirm this hypothesis and support the model's predictions. I find that relational distance affects not only loan spreads but also non-price contract terms. The longer the relational distance, the higher is the loan spread and the probability of collateral or guarantees being offered. In addition, relational distance also affects the probability of syndication failure. The longer the relational distance, the lower is the probability that a syndication fails.

My paper is related to several strands of literature. The first one is the literature on herd behavior in the investment decisions of banks. Jain and Gupta (1987) tested the lending behavior of U.S. banks of different sizes and found weak evidence of herding in international lending decisions. Nakagawa and Uchida (2003) and Uchida and Nakagawa (2007) both demonstrated the existence of herd behavior between different types of Japanese banks. The evidence is affirmative. Acharya and Yorulmazer (2008) developed a model to show that the likelihood of information contagion can induce bank owners to herd with other banks in investment decisions. While these papers investigate the herd behavior of banks in the choice of regions, industries, or portfolio, I study the decisions of banks as to whether they will join a lending syndicate. My results provide evidence of how the probability of cascade occurrence affects loan conditions and syndication failures.

This paper is also related to the literature on financial networks. There are widespread examples of financial networks such as the coinvestment networks of venture capital firms (Hochberg et al., 2007, 2010), the links between mutual fund managers and corporate board members (Cohen et al., 2008), and the co-underwriting networks of investment banks in security offerings (Baum et al., 2004). Here, I focus on syndication networks of the loan market. I am not the first to study the networks in the syndicated loan market. Based on the lead-participant relationship between banks, Godlewski et al. (2012) constructed the networks in the French syndicated loan market. They used three network centrality measures, betweenness, closeness, and degree, to proxy lenders' experience and reputation, and demonstrate that they play a significant role in reducing loan spread. Cai et al. (2010) explored how networks are formed, that is, how lead banks choose their syndicate partners. They found lead banks are more likely to choose banks that have similar lending expertise and give these banks more senior roles in the syndicate. My paper is different from previous literature in several ways. While Godlewski et al. (2012) are concerned with the experience and reputation of lenders and use centrality measures as proxies, I am interested in how cascade effect can happen and how it can affect loan prices. I use the average path length, clustering coefficient, and density as proxies for segmented communication amongst lenders. Unlike Cai et al. (2010) who examined how distance of lending expertise can affect the future collaboration of lenders and loan spreads, I study how the past collaborative relationships can affect the efficiency of sharing information, which in turn affects loan spreads.

My paper is mainly related to the growing literature on syndicated loans. Many studies emphasize the effect of information asymmetry on loan spreads (e.g., Focarelli et al., 2008; Ivashina, 2009; Nandy and Shao, 2010) and the structure of loan syndicates (e.g., Sufi, 2007; Lee and Mullineaux, 2004; Esty and Megginson, 2003; Bosch and Steffen, 2011). I contribute to this literature by providing additional evidence that the cascade effect also plays an important role in resolving contract terms.

Ivashina and Sun (2010), Nandy and Shao (2010), and Nini (2008) paid attention to the significant increase in institutional investors' demand for loans and examined the relationship between loan spread and institutional investment. Their evidence shows that because of information asymmetry institutional loans have higher loan spreads than bank loans (Nandy and Shao, 2010; Nini, 2008). In addition, the strong demand in institutional investors puts downward pressure on loan spreads (Ivashina and Sun, 2010). In this paper, I also investigate loan pricing and lenders' behavior in subscriptions, but my research does not restrict itself to institutional demand.

Finally, my work is also related to the article of Giannetti and Yafeh (2012). They studied the effect of cultural distance between the lead bank and borrower on lending contracts. The motivation is that cultural differences between contracting parties may impede communication and cause friction. Their results show that the lead bank is prone to offer better loan conditions to culturally similar borrowers. My paper complements Giannetti and Yafeh's work in at least two respects. First, while Giannetti and Yafeh concentrate on how cultural distance between lead bank and borrower can affect economic interaction, I focus on how relational distance amongst lenders can segment communication. Second, relational distance can be viewed as a contributing factor which can impair communication even when there is no cultural difference.

The remainder of this paper is organized as follows. Section 2 extends Welch's model to the syndicated loan market. Section 3 introduces the empirical methodology. Section 4 describes the data and summary statistics. The empirical results and robust checks are presented in Section 5. I conclude this paper in Section 6.

2. The Model

Suppose a firm intends to invest in a project and decides to raise capital from the syndicated loan market. It has chosen a lead bank. Assume the lead bank will approach *n* other banks (potential lenders) to inquire whether they are willing to participate in the deal. If the return of the project is r^p (e.g., IRR) then the borrower will not accept the loan interest rate, *r*, which is higher than r^p. On one hand, the interest rate *r* is the borrower's financing cost and on the other hand, it reflects the borrower's credit quality. Although the borrower's credit quality is unobservable, each bank knows its range, which lies somewhere between r^g and r^b with a uniform distribution, and $r^b > r^g$. The higher the borrower's credit quality the lower the interest rate will be. Since the borrower never accepts an interest rate larger than r^p, I also assume $r^b = r^p$.

Now suppose each potential lender can observe its own signal $s \in \{g, b\}$ independent of other banks' signal. Here, g and b denote a good and bad signal, respectively. Define θ as a linear transformation of $[r^g, r^b]$ such that $r = (1-\theta)r^g + \theta r^b$. θ can be viewed as the probability that a lender receives a bad signal, b. Since *r* and θ have an one-to-one mapping, the loan interest rate can be expressed not only in terms of *r* but also in terms of θ. For example, in the case of $r^b = r^p$, I can express r^p in terms of θ, that is $\theta^p = 1$.

2.1. Perfect Communication

Consider the first case in which all potential lenders can fully communicate with each other and share information amongst themselves. Let *k* be the number of *b* signals under *n* possible lenders. The lowest interest rate, where each lender would accept given *k* bad signals for *n* potential lenders, is equal to the posterior expected value of θ. If θ is uniformly distributed in [0,1], the posterior expected value of θ given *k* bad signals will be:

$$E(\theta \mid k,n) = \frac{k+1}{n+2} \tag{1}$$

This is the Lemma 1 of Welch (1992). Equation (1) is the lowest interest rate which each potential lender is willing to accept given *k* bad signals amongst *n* potential lenders.

According to the Lemma 2 of Welch (1992), the ex ante probability of observing *k bad* signals is:

$$prob(k\ b\ signals \mid n\ possible\ lenders) = \frac{1}{n+1} \tag{2}$$

and the ex ante probability of observing *k* or less bad signals is:

$$prob(k\ or\ less\ b\ signals \mid n\ possible\ lenders) = \frac{k+1}{n+1} \tag{3}$$

Using equations (1) to (3), I can transform the borrower's optimization problem over price into optimizing over the number of bad signal, *k*, by minimizing expected cost:

$$\min_{k}\ \left(\frac{k+1}{n+1}\right)\left(\frac{k+1}{n+2}\right)+\left(1-\frac{k+1}{n+1}\right)\theta^{p} \tag{4}$$

Here, I assume that the borrower is risk neutral. The first term of equation (4) is the borrower's financing cost multiplied by the probability that the borrower receives the loan. The second term is the opportunity cost of giving up the project, multiplied by the probability that the syndication fails. According to the first order condition, the optimal *k* is:

$$k^* = \left(\frac{n}{2}+1\right)\theta^p - 1 \tag{5}$$

Due to $\theta^p = 1$, it follows that $k^* = n/2$. According to equation (1), the optimal interest rate is $r^* = 1/2$. In order to safeguard its reputation and to establish a good relationship with the borrower, I assume that the lead bank will choose this interest rate. Indeed, as noted by Focarelli et al. (2008), once the lead bank has been mandated, it will partner with the borrower to obtain a satisfactory result from the placement.

2.2. Informational Cascades

Now I consider the case where the lead bank approaches potential lenders sequentially. Each potential lender who is approached has to make a decision about participating in the loan syndication and its decision is irreversible. In addition, each lender is fully informed of its own signal and previous lenders' decision but not their signals. In this situation, the decision made by the subsequent lender is dependent upon the action of earlier lenders. Once two consecutive lenders make the same decision, so will all subsequent lenders. That is, a cascade happens whenever a lender ignores its own signal and relies only on the decisions of the earlier lenders. To see how this could happen, I illustrate with the following examples.

Figure 2.1 provides the decision rule of each consecutive lender. Assume the lead bank sets a loan price at r =1/2. If the first lender approached by the lead bank holds a good signal (g), according to equation (2), the lowest interest rate for which the lender will accept is 1/3, so this bank will choose to participate in the loan syndication. If the second lender has a bad signal (b) and the lowest interest rate the second lender is willing to accept is 1/2, the bank can choose to participate or not participate in the deal. Suppose the bank chooses to participate in the loan syndication. Given that the two consecutive lenders have joined in the loan syndication, the third lender and all subsequent lenders will accept the deal no matter what their signals are.

This gives rise to a positive cascade in lending. If the second lender's signal is good (*g*), the lowest interest rate the bank is willing to accept is 1/4. In this case, the second lender will accept the lead bank's invitation to join in the loan syndication and a positive cascade ensues.

[*Insert Figure 2.1*]

Now consider another scenario where the first lender observes a bad (*b*) signal. The lower bound of the interest rate for which the first lender could accept is 2/3 hence the deal will be rejected. If the second lender also has a bad (*b*) signal, the deal will be accepted by this lender if and only if the loan interest rate is equal to or higher than 3/4. In this scenario, the second lender chooses to reject the deal. Given that two consecutive lenders have rejected the deal, all subsequent lenders will also reject the deal regardless of their signals. This outcome results in a negative cascade. If the second lender observes a good (*g*) signal instead of a bad (*b*) one, she will price the lending rate at 1/2 and will be indifferent to participating or not being a participant of the loan syndication. Suppose the second lender decides to accept the deal, now that the first lender is not willing to finance the project and the second lender chooses to finance it, the cascade effect will not take place. If the third lender's signal is good (*g*), it will accept the deal and a positive cascade will occur. However, if the signal is bad (*b*) the third lender will not participate in the loan syndication and, again, the cascade effect will not occur.

In sum, whether a cascade happens or not depends on the ordering of signals of the consecutive lenders which the lead bank approaches. However, for the case where interest rate is higher or equal to 2/3, a positive cascade will take place with certainty. Following Theorem 5 of Welch (1992), a risk-neutral lead bank will optimally choose the interest rate of 2/3 which is the lowest interest rate a positive cascade will undoubtedly occur and the probability that the loan syndication will fail is zero.

It is easy to see that the interest rate under the case of perfect communication is lower than the one under the cascade case. This always holds if the borrower is risk-neutral or moderately risk-averse. The intuition is that loan failures are costly for both borrowers and lead banks. Under the cascade case, the lead bank can completely rule out the possibility of syndication failure by increasing interest rates. This is impossible under the case of perfect communication because the probability of the syndication failure is always positive under the distributional assumption. However, I should note that even though interest rate is lower under the case of perfect communication than cascade cases, the expected cost of the former is in fact higher. It can be seen from equation (4) that the expected cost is $(3n+2)/(4n+4)$ for the case of perfect communication. This implies that for n larger than 1, the expected cost will be equal to or larger than 2/3. The bigger n is, the larger is the expected cost. In summary, I conclude that the ex post financing cost is lower under perfect communication than under cascade, although the ex ante cost of financing is higher due to some probability of failure.

3. Empirical Framework

3.1. Proxies for the degree of segmented communication

To empirically test the model's prediction, proxies that measure segmented communication amongst lenders are required. Two proxies, namely physical and relational distance, are used. Welch (1992) proposed the use of "locality" as a segmentation measure of communication. Following this idea, my first proxy is the physical distance amongst lenders. For a given facility (or tranche), I calculate the physical distance between pairs of lenders based on the cities of their principal executive offices. After calculating all possible pairs, I then average the distance across all pairs and obtain the mean distance for lenders of the facility. The intuition is that the farther the physical distance amongst lenders, the more difficult they communicate with each other. One problem with the use of this proxy is that it may not truly reflect the ease of communication amongst lenders. In this day and age, innovations in technology, transportation, and communication may attenuate the constraint of physical distance, thus communication between two lenders that are located far apart may not necessarily be hindered by their physical distance. To circumvent the limitation of this proxy, I proposed a second proxy, which captures the relational distance amongst lenders.

I assume that lenders who have collaborated in past syndications would be prone to communicating and sharing their information with each other. To measure the extent of collaborative relationships amongst lenders, I turn to social network analysis. It is widely acknowledged that the structure of financial networks has important implications for information dissemination (e.g., Baum et al., 2004; Hochberg et al., 2007; Cohen et al., 2008). I apply graph theory, a mathematical discipline widely used to solve the problems of networks, and construct collaboration networks that loan syndication gives rise to. Specifically, the relational distance is quantified via the use of average path length, clustering coefficient, and density.

3.2. Loan syndication networks

A network is a set of items (called nodes or vertices) with connections (called links or edges) between them (Newman, 2003). In this paper, nodes represent lenders and links represent the existence of a leader-participant relationship between them. Following Baum et al. (2004), Hochberg et al. (2007, 2010), and Godlewski et al. (2012), for any given year, I construct a syndication network using the data from the preceding five years. Since there are very little contacts amongst participant banks in the same syndicate, I only consider the relationship between lead-participant banks (Godlewski et al., 2012).[1] For instance, suppose there are four syndicate members in a deal, banks A, B, C, and D. Bank A is the lead bank while banks B, C, and D are the participating banks. In this case, there are four nodes which represent the four banks. Bank A has three links which are directed towards the other three. But there is no link between banks B, C, and D.

Having constructed the syndication network, I then compute the relational distance using facility level data. For any given facility, I extract a subgraph (subnetwork) from the preceding-five-year network based on the members of the facility. If some lenders are not in the network, the equivalent number of isolated nodes is added to the subgraph. I consider a facility in 1995 to illustrate the computation of the relational distance. The result is shown in Figure 2.2. In this example, in order to derive the subgraph for the facility, I first construct a network using syndicated loan data from 1990 to 1994. I then extract the subgraph from the network using the list of facility members. It can be seen that there is a bank that did not appear in the original network, therefore, an isolated node (node 11) is added to the subgraph. In Figure 2.2, each node represents a lender, and there are twelve lenders in the facility which are denoted by number 0 to 11. Any line linking the two

[1] Following Cai et al. (2010) and Godlewski et al. (2012), I treat a bank as a lead bank if it has the following lender roles: Admin agent, Agent, Arranger, Bookrunner, Co-agent, Co-arranger, Co-lead arranger, Co-lead manager , Collateral agent, Co-manager, Co-syndications agent, Coordinating arranger, Documentation agent, Joint arranger, Joint lead manager, Lead arranger, Lead bank, Lead manager, Managing agent, Mandated arranger, Senior co-arranger, Senior co-lead manager, Senior co-manager, Senior lead manager, Senior manager, Senior managing agent, Syndications agent, Underwriter.

nodes indicates that the two banks had a lead-participant relationship in the period 1990 to 1994.

[*Insert Figure 2.2*]

To gauge the relational distance three measures, average path length, clustering coefficient, and density, are used. The first two measures are important indicators to examine small-world networks. Typically, short average path length and large clustering coefficient are two primary characteristics of small-world networks.[2] Average path length is calculated as the average shortest path of all pairs of nodes for a network. Larger average path length indicates that the relational distance amongst lenders is longer. Clustering coefficient measures the transitivity of the links. In other word, if node 1 is connected to nodes 2 and 3, clustering coefficient measures the probability that nodes 2 and 3 are also connected. It captures the idea of local density (cliques). Density is the proportion of links in a given network relative to the total number possible. In contrast to clustering coefficient, density captures the idea of global-level density. Smaller clustering coefficient or density implies that the relational distance amongst lenders is farther. Using the example in Figure 2.2, the average path length is 3.18, the clustering coefficient is 0.7, and the density is 0.48.

To summarise, I expect that the larger is the average path length (clustering and density), the more difficult information is disseminated amongst lenders. Consequently, there is a higher probability that a cascade will take place, and this in turn implies that the loan interest rates will be higher and the probability of failure will be lower.

---

[2] A network is a small-world network if its average path length is lower than a regular network and its clustering coefficient is bigger than a random network with equal size.

3.3. Hypotheses

My theoretical model generates some testable hypotheses to guide my empirical investigation. First, the model predicts for a positive cascade to occur in the syndicated loan market, the interest rate needs to be set higher under cascade than under perfect communication. This leads to the following hypothesis.

H1: The ex post observed interest rate is higher if communication is more segmented amongst potential lenders.

When communication amongst potential lenders is more segmented, it is harder for lenders to obtain information about the borrower's credit quality. The potential lenders' information cannot be aggregated efficiently in the market. As a result, cascade is more likely to occur and the ex post observed interest rate is higher. Corollary, I expect that the loan spread will be higher when a facility has longer physical or relational distance amongst lenders.

Although increasing interest rate may result in a positive cascade and prevent a syndication failure, this is not the only means of inducing a positive cascade. In some situations, the lead bank may use non-price contract terms to increase potential lenders' interest to participate in the syndicate because non-price contract terms can reduce borrower's moral hazard behavior and to lower lenders' loss when default occurs. This scenario leads to the second hypothesis:

H2: The probability of using non-price contract terms is higher if communication is more segmented amongst potential lenders.

Under *H2*, I expect that the probability of setting non-price contract terms is higher when a facility has longer physical or relational distance amongst lenders.

The model of informational cascade shows that a positive cascade will occur with certainty should the interest rate be higher than a certain threshold. Consequently, a rational lead bank will choose this lower-bound interest rate to ensure that a positive cascade will happen. This implies that the probability of syndication failure is zero. However, the probability of failure will always be positive under the case of perfect communication. My model, therefore, predicts that the probability of failure is lower under cascade case than under perfect communication case. This leads to my third hypothesis.

H3: The probability of the loan failure is lower if communication is more segmented amongst potential lenders.

Under *H3*, I expect that the probability of failure is higher when a facility has longer physical or relational distance amongst lenders. In addition, if lead bank can choose the syndication size, I will expect that the number of lenders will be smaller under perfect communication than under cascade. This is because the larger the size is, the higher the ex ante cost will be as shown in section 2.2. This further suggests that failure cases will concentrate more on low lender number loans.

3.4. Empirical specifications

To empirically test the aforementioned hypotheses, I use the following specification:

Dependent variable=f (*Distance, Control variables*) (6)

Different dependent variables are used to test my hypotheses. For hypothesis one, the dependent variable is all-in-drawn spread which is interest margin paid over LIBOR. For hypothesis two, I use three dependent variables to capture the contractual features. They are denoted by Covenant, Secured, and Guaranteed. Covenant takes the value of one if the contract has financial covenants and zero otherwise. Secured takes the value of one if the loan has been secured and zero otherwise. Guaranteed takes the value of one if the contract requires guarantees and zero otherwise. For hypothesis three, the dependent variable is also a dummy variable which is equal to one if the loan status is

“Cancelled” or “Suspended” as reported in the data and zero otherwise. To test hypothesis one, ordinary least squares (OLS) is used to estimate the regression parameters. And probit models are used to test hypotheses two and three.

Distance is the key variable to proxy for segmentation of communication. Four proxies, physical distance, average path length, clustering coefficient, and density, are used in this paper. I also include a large number of control variables, such as loan and borrower characteristics, in the regressions. Prior research suggests that larger loans carry lower spreads (Carey and Nini, 2007); long-term loans may be required liquidity premiums by lenders (Graham et al., 2008); the presence of collateral and covenant could moderate adverse selection and moral hazard in the loan syndicate (Ivashina, 2009); the number of lenders has a significant impact on loan pricing (Giannetti and Yafeh, 2012; Ivashina and Kovner, 2011). I, therefore, control loan characteristics which include logarithmic facility amount, time to maturity, secured/unsecured, guaranteed/nonguaranteed, financial covenant indicator, and lender number in my analyses.

Borrower’s quality (credit risk) also has an important influence on loan contract terms. The differences in credit risk may reflect on borrower’s characteristics. By referring to prior research (e.g., Focarelli, et al. 2008; Ivashina, 2009; Ivashina and Kovner, 2011), I control public and rated firm indicator, public and unrated firm indicator, logarithmic sales, market to book ratio, return on asset, and total debt to total asset ratio (leverage). The first two variables are used as proxies for borrower’s transparency and the others are financial variables which respectively represent a firm’s size, growth opportunity, profitability, and capital structure.

In addition, macroeconomic conditions could affect loan pricing (Graham et al., 2008). To control macroeconomic factors, I follow Giannetti and Yafeh (2012) and Graham et al. (2008) to include GDP per capita, credit spread, and term spread in the regressions.[3] Credit spread is the difference between BAA and AAA corporate boand yields. It tends to widen in recessions. Term spread is the difference between 10-year and 2-year Treasury yields. It tends to widen if economic prospects are good. Physical distance between borrower and lead bank may also influence loan contract terms (Giannetti and Yafeh, 2012), hence, I control location dummy variable which is equal to 1 if the borrower and lead bank have the same 3-digit zip code and 0 otherwise. Definitions of aforementioned variables are provided in Appendix B, Table B1.

Finally, various fixed effects (2-digit SIC fixed effect, loan type fixed effect, or loan purpose fixed effect) are also controlled in the regressions. Standard errors are clustered at the year level throughout my analysis.[4]

4. Data and Descriptive Statistics

Data on syndicated loans are collected from Reuters Loan Pricing Corporation (LPC) DealScan database, which provides detailed information on contract terms, borrowers, and lenders. I focus on the U.S. market in the period from January 1990 to August 2010. The initial data consists of 60,237 deals or 86,904 facilities.

Facilities with only one lender are excluded from the analyses because the physical and relational distance cannot be calculated. I use lenders' geographic information to compute their physical distance. The cities of the lenders' principal executive offices are used to calculate the average physical distance on the syndicated loan facility level between 1990 and 2010. I am able to identify the geographic data for 23,188 deals or 31,938 facilities (hereafter referred to as Sample I).

---

[3] My yearly GDP per capita data is from World Bank's World Development Indicators. The credit spread and term spread are monthly data and they are from Federal Reserve Board of Governors.

[4] Results are stronger if standard errors are clustered at the borrower or the loan type level. I, therefore, report the most conservative outcomes.

The information of lenders' roles (leader or participant) is used to construct the syndication networks between 1990 and 2009. Because the networks are not static, following previous literature, I use overlapping, moving five-year windows to construct collaboration networks amongst lenders (Baum et al., 2004, Hochberg et al., 2007, 2010, and Godlewski et al., 2012). Sixteen moving windows (i.e. sixteen networks) are used to compute the relational distance for the members of each facility from 1995 to 2010. There are 44,327 deals or 65,390 facilities in this sample (hereafter referred to as Sample II).

Descriptive statistics of the variables are provided in Table 2.1. Panel A of Table 2.1 shows that the average loan size is 172 (265) million with a 223 (227) bps spread over LIBOR rate for sample I (sample II). While the mean (median) of loan amount is larger in Sample I than in Sample II, the mean (median) spreads are similar for both samples. The fraction of secured loans is high in both samples. For Sample I, 82% of the loans are secured. In contrast, 78% of the loans are secured for Sample II. Due to the fact that there are many missing values for the variable Secured in the DealScan database, I treat these missing values as unsecured facility throughout my analyses, except in the case when Secured is a dependent variable. My treatment of missing values as unsecured facility clearly may give rise to a bias estimate on the coefficient of the Secured variable. To ameliorate the effect of this bias, I define a dummy variable that takes the value one if the variable Secured is missing, and zero otherwise, and I include this variable in my regression to control for the effect of my treatment on the missing values.

[*Insert Table 2.1*]

The cases of syndication failures consist of 0.6% (0.5%) of the total observations in Sample I (Sample II). The borrowers are divided into three types: private, public and unrated, and public and rated. They are used as proxies for the borrowers' informational opacity (Sufi, 2007, 2009; Dennis and Mullineaux, 2000; Lee and Mullineaux, 2004; Bosch and Steffen, 2011). I match my sample with Compustat to obtain borrowers' financial variables and their descriptive statistics are shown in Panel B. The average firm size in terms of sales is about 3.2 billion in Sample I and about 4.9 billion in Sample II. There are more missing values for the market to book value (MB) ratio than other financial variables. I am able to retrieve 11,359 (24,742) observations of MB ratio in Sample I (Sample II).

Panel C reports the four distance measures. The average physical distance between any two lenders is 2,571 kilometers and the standard deviation is 2,433. In the formal analysis, I standardize the physical distance to facilitate the interpretation and expression of the regression coefficient. The mean of the average path length is 1.68 which is much smaller than 6.9, the average lender number. The means of the clustering coefficient and density are 0.84 and 0.79, respectively.

5. Empirical Results

5.1. Results for Physical Distance

I first examine my hypotheses using physical distance as the proxy for segmented communication amongst lenders. The regression results are reported in Table 2.2. It is seen from column (1) that the coefficient of physical distance is -2.58. The sign is inconsistent with my model's prediction although it is statistically insignificant. The result should be interpreted with caution because physical distance is sensitive to the existence of foreign lenders. The presence of foreign lenders in a facility will lengthen the physical distance, which can bias my results (Mian, 2006; Haselmann and Wachtel, 2011). When I divide my sample to those with and without foreign lender, and re-run the regression, the coefficient on physical distance for without-foreign-lender subsample become positive although it is still statistically insignificant.[5]

In columns (2) to (4), the relationship between physical distance and other non-price contract terms are either insignificant or of the wrong sign. When I divide my sample into those with and without foreign lender, the conclusion does not change. These results show that the effects of physical distance on syndicated loan contract terms are weak. In addition, column (5) shows that physical distance seems to lack explanatory power in predicting loan failures.

[*Insert Table 2.2*]

One possible reason that my model's predictions are not supported by physical distance may be that using lenders' principal executive offices to calculate the distance is questionable. For example, if lending decision is made by banks' branches, my calculation of physical distance may be biased. To disentangle this concern, I assume that the lending decision is made by principal executive offices as loan size is large, and it is made by branches as loan size is small. I, therefore, divide my sample into large loans and small loans according to the sample median of the ratio of facility amount to borrower's total asset and re-run the regression. However, the pattern of the results is unchanged.[6] In summary, none of my three hypotheses are supported if physical distance is used to proxy for segmentation of communication amongst lenders.

The coefficients on control variables are mostly consistent with those reported in the literature. In column (1), loan spread tend to decrease with loan amount and maturity (e.g., Ivashina, 2009; Ivashina and Sun, 2010; Cai et al., 2010). The loan

---

[5] The results are provided in Appendix B, Tables B2 and B3.

[6] The results are provided in Appendix B, Tables B4 and B5.

spreads on secured/guaranteed loans are significantly higher than those on unsecured/non-guaranteed loans (e.g., Focarelli et al., 2007; Ivashina, 2009; Nandy and Shao, 2010). The coefficient of the "missing" variable is positive and significant. This means I may have underestimated the coefficient of Secured/Guaranteed because I treat the missing value as unsecured/non-guaranteed. The results also show that macroeconomic factors and borrowers' size, growth opportunity, profitability, and capital structure are all important determinants of loan spread.

In column (2), the coefficients on control variables reveal some interesting patterns. The loan amount and the size of borrowers have a statistically significant and negative effect on the existence of financial covenants. Somewhat surprisingly, the coefficients on ROA and leverage are statistically significant but their signs are different from those reported in column (1). Intuitively, if the borrower's profitability is high and leverage is low, then I would expect the financial covenants to be less binding. My results suggest that borrowers who are less financially constrained are more likely to use financial covenants to attract potential lenders.

In column (5), I confine my analysis to data with lender numbers between two and ten.[7] The coefficients of the control variables indicate that the probability of syndication failure is higher with larger loan amount or longer maturity. Loans with

---

[7] As noted earlier, my model predicts that failures will concentrate more on loans with low number of lenders. The frequency distributions reported in Table B6 of Appendix B support the prediction. In Table B6, the number of syndication failures decreases as the number of lenders increases. However, the number of failures is falling at a faster rate than the increase in observations thus giving rise to a falling proportion of failures. This is with the exception of lender numbers eight and nine for both samples. Given this pattern in lender number and syndication failures, when *H3* are tested, I confine my analysis to data with lender numbers between two and ten.

financial covenants are less likely to fail. Borrowers' financial characteristics seem to exert little influence on failure probability.

5.2. Results for Relational Distance

Results in the previous section suggest that communication amongst lenders does not seem to be affected by physical distance. In this section, I consider another proxy for communication segmentation, the relational distance, and test its effect on syndicated loan contracts.

5.2.1. Relational distance and loan spreads

First, I test the relationship between relational distance and loan spread. The regressions are performed using equation (6) to test hypothesis one. The results are presented in Table 2.3. In all specifications, my proxies for relational distance have statistically significant effects on loan spreads. Column (1) shows that the coefficient of average path length (Path) is positive and statistically significant at the 0.1% level. Columns (2) and (3) show that the coefficients of clustering coefficient (Clustering) and density (Density) are negative and statistically significant at the 0.1% level. These results lend support to my hypothesis that the further the relational distance, the higher is the probability of a cascade occurring, hence the larger is the spread.

[*Insert Table 2.3*]

The effect of relational distance on loan spreads is also economically significant. A one-standard deviation increase (decrease) in average path length (clustering coefficient/density) increases the all-in-drawn spread by 7.7 (11.5/15.0) basis points, which is approximately 3.4 (5.1/6.7) percent of the sample median spread of 225 basis points.

The coefficients of the control variables are by and large similar to those reported in column (1) of Table 2.2. In column (1), loan spreads significantly decrease with loan amount. The reason may be due to the effect of economies of scale (Graham et al., 2008; and Godlewski et al., 2012) or because larger borrowers, which have greater transparency and lower default risk, typically issue larger loans (Focarelli et al., 2008; and Carey and Nini, 2007). The negative coefficient on maturity seems to be consistent with 'credit-quality hypothesis' – banks limit their exposure by lending riskier borrowers shorter loans (see, for example, Gottesman and Roberts, 2004). The loan spreads on secured/guaranteed loans are significantly higher than those on unsecured/non-guaranteed loans. This result seems to support "observed-risk hypothesis" – banks charge riskier borrowers larger loan spreads and require more collateral (see, for example, Godlewski and Weill, 2011). The coefficient of the "missing" variable is positive and significant. This means I may have underestimated the coefficient of Secured/Guaranteed because I treat the missing value as unsecured/non-guaranteed.

Loans with financial covenants seem to have lower spreads. The intuition may be that covenants is a contract design that can work as an ex ante monitoring device to moderate moral hazard, which in turn reduces the loan spread (Ivashina, 2009). Lee and Mullineaux (2004) and Bosch and Steffen (2011) showed that syndicate size is larger when the borrower has higher transparency and lower credit risk. This finding is consistent with my negative coefficient on lender number. I also find that borrowers' characteristics such as larger size, higher growth opportunities, and greater profitability are negatively associated with loan spreads. Borrowers with higher leverage bear higher loan spreads. GDP per capita and credit spread have significant influence on loan spread, which is consistent with the findings of Giannetti and Yafeh (2012) and Graham et al. (2008).

5.2.2. Relational distance and non-price requirements

In this section, I run probit regressions to test my second hypothesis. The results are presented in Table 2.4. In Panel A of Table 2.4, I examine whether the probability of setting financial covenants increases with relational distance measures. The results show that the coefficients of my variables of interest, Path, Clustering, and Density, are statistically insignificant at conventional levels of significance. Relational distance does not seem to exert any influence on the probability of having financial covenants. As for the coefficients of the control variables, they are qualitatively the same as those reported in column (2) of Table 2.2. Again I find that financially less constrained firms typically characterised by higher ROA and lower leverage are more likely to use financial covenants.

[*Insert Table 2.4*]

Panel B of Table 2.4 reports the regression results for collateral requirements. In column (1), the coefficient of average path length is positive and statistically significant with a value of 0.11. In columns (2) and (3) the coefficients of the clustering coefficient and density are negative and statistically significant at the 1% level. These results suggest that the longer is the relational distance amongst lenders, the higher is the probability of requiring collateral, which is consistent with hypothesis two. In addition, the coefficients of the control variables show that loans with higher amount are less likely to require collateral. The probability of requiring collateral is lower for public firms than for private firms. Large sized borrowers tend to have lower probability of offering collateral. Unlike the results for financial covenants, there is evidence to suggest that borrowers that are more financially constrained, as characterised by lower profitability and higher leverage, are more likely to offer collateral. Finally, macroeconomic factors such as GDP per capita and credit spread also have a significant impact on requiring collateral.

Panel C of Table 2.4 provides the regression results of equation (6). Column (1) shows that the average path length has a strong effect on the likelihood of guarantees requirements. The coefficient of Path is 0.03, which has the expected positive sign and is statistically significant at the 0.1% level. The same strong effect can also be found for the coefficient of Density which also has the expected sign and is significant at the 0.1% level. On the contrary, the coefficient of Clustering is insignificant at the 5% level. Given these results, I argue that the evidence in Panel C supports my hypothesis that the longer is the relational distance, the higher is the probability of guarantee requirements. Results in Panel C also show that loans with larger amount or with financial covenants have higher probability to require guarantees. Secured loans also exhibit higher probability of setting up guarantees. It is worth noting that given the negative and statistically significant coefficient of the variable Missing, I may overestimate the effect of secured loans on the probability of requiring guarantees. Public firms, again, have lower probability to offer guarantees than private firms. Value firms are more likely to offer guarantees than growth firms. Finally, the probability of requiring guarantees increases in borrowers' leverage, GDP per capita, and term spread.

In summary, my evidence suggests that when a cascade is more likely to happen, the loan contracts have a greater tendency to include non-price agreements to attract potential lenders. This is especially true for non-price contract terms like collateral and guarantees but not for financial covenants. This observation may be in line with the intuition that even though financial covenants can prevent borrowers' moral hazard, only collateral and guarantees can reduce lender's damage once default occurs.

Therefore, contract terms on collateral and guarantees seem to be more attractive for lenders compared to financial covenants.

5.2.3. Relational distance and syndication failures

To test hypothesis three, I focus on sample with lender numbers between two and ten. The results are reported in Table 2.5. The coefficient of average path length is negative and significant at the 5% level, and the coefficient of density is positive and significant at the 5% level. The results from columns (1) and (3) suggest that the longer is the relational distance amongst lenders, the lower is the probability of syndication failure. However, I do not find support for the clustering coefficient. In column (2) of Table 2.5, the coefficient of Clustering is marginally significant but the sign is inconsistent with my model's prediction.

[*Insert Table 2.5*]

The effect of control variables on the probability of syndication failures is similar to the results in column (5) of Table 2.2. Loans with larger amount or longer maturity have higher probability to fail. Contracts with financial covenants are less likely to fail. The probability of syndication failures is lower in expansions, as characterised by higher GDP per capita and term spread.

Due to the large difference between the number of failure cases and the number of success cases in my sample, I also test my third hypothesis using a matching sample method of failure and success syndication observations.[8] I obtain 224 pairs of such observations. I find that there are significant differences for the three relational distance measures and the patterns are similar to those reported in Table 2.5. The

[8] I find in Table 2.5 that the probability of syndication failure is significantly affected by loan amount, maturity, financial covenants, and ROA, so I use these four variables as criteria to match the sample. Results of matched sample are provided in Appendix B, Table B7.

average path length and density have the expected signs, which support my third hypothesis.

5.3. Robustness checks

The results presented above suggest that using relational distance as a proxy for segmented communication lend supports to my model's predictions. In order to identify the presence of information cascades and distinguish them from other possible explanations for the results, I conduct a number of additional analyses to gauge the robustness of my empirical findings. I first conduct robustness checks by repeating my analysis separately with interaction terms between relational distance measures and previous lead bank and borrower relationship, investment grades, covenant violation and foreign bank participation variables. I examine the robustness of the relationship between relational distance and loan spread in each of these subsamples. Finally, I examine the potential endogeneity between spread and relational distance and the impact of omitted variables.

5.3.1. Alternative explanations and model specifications

The first concern of my results is that it is information asymmetry between lead bank and participants instead of cascade effect which drives my results. The information asymmetry effect gives rise to adverse selection and moral hazard problems (e.g., Focarelli et al., 2008; Ivashina, 2009; Sufi, 2007, 2009; Lee and Mullineaux, 2004), and enable participants to require a higher spread to compensate for their informational disadvantage. To disentangle this alternative explanation, I first check the relationships between my relational distance measures and the adverse selection effect. Adverse selection happens when lead bank has private information on the borrower that is unknown to participants. Following Sufi (2007), I use previous lending relationships between the lead bank and the borrower to capture the lead bank's information advantage. If the relational distance measures mainly capture the adverse selection effect, we should expect that the impact of these measures on loan spreads is stronger when there are previous relationships between the lead bank and borrower.

The results are shown in Table 2.6.[9] If my relational distance measures mainly capture adverse selection effect, I shall expect to see significant interaction terms, P_Previous_Relation, C_Previous_Relation, and D_Previous_Relation. As shown in Panel A of Table 2.6, none of the three interactions is statistically significant. It suggests that my results are not driven by adverse selection effect. The negative coefficients on the dummy variable, Previous_Relation, are consistent with Ivashina and Kovner (2011), although they are statistically insignificant.

[*Insert Table 2.6*]

An additional concern is that my results may be driven by moral hazard effect rather than cascade effect. I conduct a similar exercise that classifies borrowers into two categories, investment grade and non-investment grade, according to their S&P senior debt rating. The borrowers' credit rating is used as a proxy for their informational opacity. Borrowers with non-investment grade require more intense monitoring and due diligence, hence moral hazard problem for the lead bank is more severe. If my relational distance measures are driven by the moral hazard effect, we should expect that the impact of the three measures on loan spreads will be stronger when borrowers are informationally opaque. The evidence in Panel B of Table 2.6 show that the interaction terms, P_Investment_Grade, C_Investment_Grade, and

[9] I do not report the results of control variables for brevity. However, the full tables be found in Appendix B, Tables B8 to B11.

D_Investment_Grade, are statistically insignificant, which suggests that my results are not driven by moral hazard effect.

Another worry is that the relational distance amongst lenders may imply difficulties in ex post renegotiation, which should be priced ex ante. I use ex post violations of financial covenants as a proxy for expected probability of renegotiation.[10] I assume that loans with covenant violations have higher ex ante probability of renegotiation. When it is difficult to execute renegotiation and the expected probability of renegotiation is high, the loan spread may be higher to reflect the renegotiation costs. I incorporate the covenant violation dummy and the interaction term of the dummy variable with my relational distance measures in the regression and the results are reported in Panel C of Table 2.6. If my measures of relational distance mainly capture renegotiation costs, we should expect to see significant coefficients on the interaction terms, P_Covenant_Violation, C_Covenant_Violation, and D_Covenant_Violation. Results in Panel C of Table 2.6 show that covenant violation dummy, Covenant_Violation, is significantly positive which means that the loans which the borrowers ex post violate financial covenants have ex ante higher spreads. However, most of the interaction terms are insignificant. The evidence does not support the relational distance amongst lenders may imply difficulties in ex post renegotiation.

One may also question that behaviour may differ amongst different types of lenders, e.g., foreign banks and domestic banks, and my results may not be

[10] To match my sample with covenant violation provided by Nini et al. (2011), I drop observations with facility end date before July 1, 1997 and the facility start date after March 31, 2008. However, my results are not sensitive to this sample adjustment. For example, the results are qualitatively the same if I drop the observations with facility end date before January 1, 1997 and facility start date after June 30, 2008.

generalized to different lender classes. To answer this question, I test whether my relational distance measures show different patterns between foreign lenders and domestic lenders. The results are presented in Panel D of Table 2.6. The coefficients on the foreign lender dummy, Foreign_Lender, is positive which indicates that loans with foreign participants have lower spreads, but most of them are statistically insignificant. The interaction terms between the dummy variable and the three distance measures, P_Foreign_Lender, C_Foreign_Lender, and C_Foreign_Lender, are mostly insignificant. The evidence show that my results can be generalized to different lender classes.

5.3.2. Endogeneity

In Section 5.2, the results of my statistical analyses suggest that the further the relational distance, the higher is the probability of a cascade occurring, hence the larger is the spread. However, relational distance may also be endogenously determined, to some extent, by the spread of the loan contract. It may therefore lead to inconsistent estimation results if I only use ordinary least square regression to estimate. I address this potential endogeneity issue with instrumental variables to conduct a two-stage regression. In so doing, I also satisfy the exclusion restriction in both economical and statistical terms.

In general, there are two criteria in selecting instrumental variables: (1) instrument relevance; and (2) instrument exogeneity.[11] In my framework, I should select those instrumental variables that are highly correlated with the endogenous variable (i.e., relational distance), but at the same time uncorrelated with the error term or the dependent variable. The instrumental variable I consider is the aggregate

---

[11] Detailed illustrations of the selection and estimation of instrumental variables can be found in Wooldridge (2002) and Greene (2008).

amount of inactive participants (denoted as Inactive_num). That is, for each facility I calculate the number of nonlead-bank members who had not joined any loan syndicate in the previous three years. In general, the loan spread is influenced by the lead bank's characteristics, but not those of participants. While the characteristics of participants may only have minimal effect on the spread, those inactive participants will have significant impact on network links, which in turn affect the relational distance. It therefore satisfies economically the appropriateness of selecting the number of inactive participants in the syndicate as our instrumental variable.

To assess the degree of relevance between my instrumental variable and the endogenous variable, I check the statistical significance of the coefficients of the instrumental variables in the "first-stage" regression and by conducting the tests of Staiger and Stock (1997). A second group of tests are conducted to justify the "instrument exogeneity condition." To confirm my instrumental variables are in fact uncorrelated with the error term or my dependent variable, I conduct the exogeneity test of Hausman (1978).

The results are shown in Panel A of Table 2.7.[12] Let's start with the first stage regression results. Confirming my expectation that the instrumental variable is highly related to the relational distance measures, the coefficients are statistically significant at the 0.1% level. I conduct Staiger and Stock (1997) test to examine the significance of my instrument after adding the instrument in the first stage regression. Results show that my instrument is statistically significant at the 0.1% level. It thus rejects the null hypothesis that the instruments are invalid and confirms the instrument is indeed related to relational distance measures. The results of instrument exogeneity test of

[12] Table 2.7 only reports part results of the two-stage regression. The complete results are provided in Appendix B, Tables B12 and B13.

Hausman (1978) show mixed results. It is seen that the insignificant statistic for the Path variable support the exogeneity of the instrument, while the result for the Clustering and Density variable is marginally significant. Finally, I examine the results of the second-stage regressions in order to verify the relation between the spread and relational distance measures based on my instrumental variable. Here, I regress loan spreads against fitted value of relational distance measures while controlling for other explanatory variables previously considered in Table 2.3. As before, Panel B of Table 2.7 shows a positive and statistically significant relation between relational distance and loan spread. In general, the findings are similar to those reported in Table 2.3 confirming the robustness of my conclusions.

[*Insert Table 2.7*]

5.3.3. Omitted variables

In addition to the characteristics of the loan contracts already controlled for in my analysis, other forms of issuing firm or lender heteroskedasticity may also affect the loan spread. To check the robustness of my conclusions to the potential omission of other variables, I execute a regression specification with more control variables such as the frequency that the borrower borrows from the syndicated loan markets (Borrower_Experience), the average frequency that the lenders participate in the syndicated loan markets (Lender_Experience), whether the lead bank belongs to the top five lenders in the league table (Top5),[13] borrower's Z-score of Altman (1968) (Z-score), and the four dummy variables in Table 2.6. Results in Table 2.8 show that coefficients on my relational distance measures are still statistically significant with expected signs. In comparison with the results in Table 2.3, the magnitude of the

[13] I calculated each lead bank's ranking in the league table according to the data provided by Loan Price Corporation.

estimated coefficients of the relational distance measures is slightly smaller than that obtained when I do not control for these effects. The relative degrees of statistical significance are similar to those reported in Table 2.3, suggesting my empirical results are robust to unobservable firm-level or lender-level heteroskedasticity.

[*Insert Table 2.8*]

6. Conclusions

The interaction of economic agents is complicated and no single rule can capture all situations. The complicated interaction also forms complex networks. This phenomenon has inspired me to construct my theoretical and empirical work in this paper. I use two cases to model lenders' interaction in the syndicated loan market. My purpose is to explore how informational cascade could happen in this market and how it would affect loan pricing. I also empirically test the model's predictions.

The first case is my benchmark where I assume all potential lenders can freely share information held by them. The second one is the cascade case in which I assume each potential lender can only observe the decisions of its predecessors. If potential lenders only have imperfect signals, the actions of their predecessors are important information for evaluating a loan. My model shows that if the lead bank is rational and risk-neutral, the probability of syndication failure is always positive in the benchmark but is zero in the case of cascade. This results in lower ex ante financing cost under the cascade but the ex post interest rate will be higher. The intuition is that the lead bank will increase the interest rate to elicit a positive cascade and ensure that financing is obtained.

To empirically test the models' predictions, physical distance and relational distance are used to proxy for segmentation of communication amongst lenders. I use average path length, clustering coefficient, and density, which are both taken from the analysis of syndication networks, to gauge the relational distance. The longer distance signifies that communication or information is more segmented amongst lenders. The results show that the physical distance does not support the predictions, but the relational distance does. I argue that the physical distance is not a good proxy due to innovations in technology, transportation, and communication. The relational distance is a good proxy because the influence of network structures on information dissemination and transmission is well-established in the literature.

As the relational distance is used as a proxy for segmented communication, the results show that the relational distance is positively correlated with loan spread and is negatively correlated with the probability of syndication failure. In addition, the higher relational distance also results in more non-price contract terms, especially the requirements for collateral and guarantees. These findings confirm my model's predictions. The cascade effect does matter for lending conditions.

My study contributes to the literature in three aspects. First, to my best knowledge, this is the first study to explore cascade effect in the syndicated loan market. As noted in section 1, although some studies have examined the herd behavior in banks' investment decisions, they did not focus on the syndicated loan market and did not explore how herd behavior affects loan pricing.

Second, several distance measures have been proposed and associated with economic decisions, for example, physical distance (Mian, 2006; Giannetti and Yafeh, 2012), distance of specialization (Cai et al., 2010), and cultural distance (Giannetti and Yafeh, 2012). Also, many researches have proven the existence of relationship lending (e.g., Elyasiani and Goldberg, 2004; Champagne and Kryzanowski, 2007). I extend these ideas to propose a novel proxy, relational distance. It is difficult to capture relationships amongst economic agents and to quantify the relational distance. I overcome these obstacles by conducting a social network analysis.

Finally, there have been plenty of papers that investigate the determinants of loan prices. In this paper, I analyze not only the factors that affect loan price but also those that cause syndication failures. To my best knowledge, this is also the first study to empirically test the determinants of syndication failures. It is an important issue because syndication failures are costly to borrowers and lenders and may impair investment activities. Understanding the causes of syndication failures helps to improve the success of the syndicated loan market.

Figure 2.1: An example of the decision rule of the cascade case

This figure illustrates an example of the decision rule of the cascade case. The lowest interest rate for which the potential lender is willing to accept is reported in parentheses.

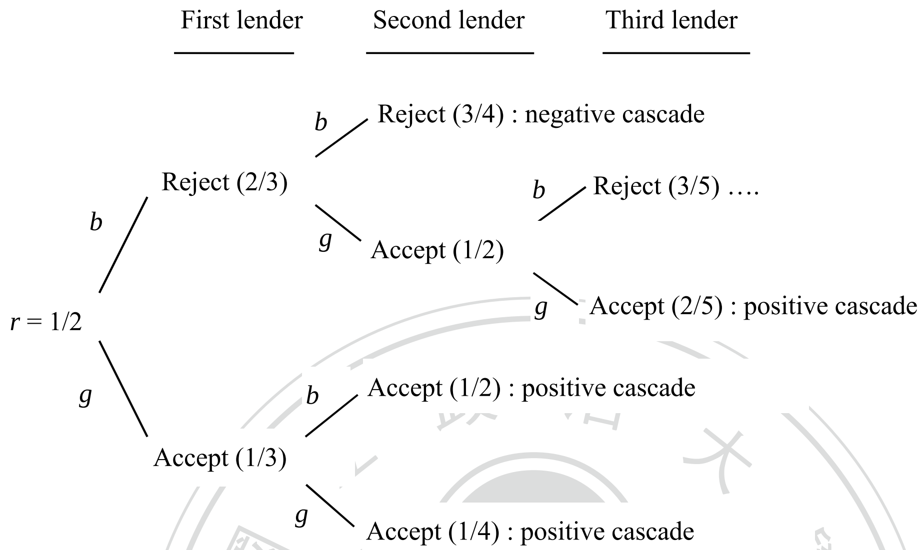

Figure 2.2: An example of a syndication network.

This figure uses a facility in 1995 as an example to show what the syndication network look like. Each node represents a lender and any line linking two nodes indicates that the two banks had a lead-participant relationship in the period 1990 to 1994.

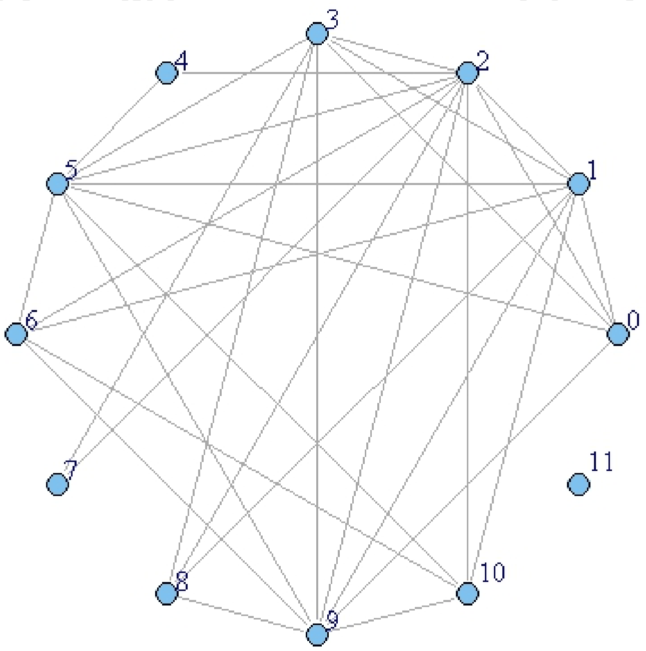

Table 2.1: Descriptive statistics of regression variables

Data are collected form DealScan database and for the period from January 1990 to August 2010. Only facilities which have lender number that is larger than 1 are included. Sample I is the sample used in the analysis for physical distance. Sample II is the sample used in the analysis for relational distance. Definitions and units of the variables are provided in Appendix B, Table B1.

	Sample I				Sample II			
Variable	Obs.	Mean	Std.	Median	Obs.	Mean	Std.	Median
Panel A: Loan characteristics								
Allindrawn	26,083	223.11	144.33	215	56,973	227.24	151.82	225
Amt	31,938	171.50	400.77	66	65,390	265.10	650.60	100
Maturity	27,750	46.71	147.07	46	59,528	47.92	102.60	50
Secured	11,346	0.82	0.38	1	29,080	0.78	0.42	1
Guaranteed	31,938	0.02	0.14	0	65,390	0.05	0.22	0
Covenant	31,938	0.20	0.40	0	65,390	0.33	0.47	0
Lender_num	31,938	4.25	3.76	3	65,390	6.90	7.52	4
Failed	31,938	0.006	0.08	0	65,390	0.005	0.07	0
Panel B: Borrower characteristics								
Pub_unrated	24,507	0.23	0.42	0	50,592	0.20	0.40	0
Pub_rated	24,507	0.27	0.44	0	50,592	0.32	0.47	0
Sales	13,794	3,235.24	11,391.40	596.75	30,518	4,881.90	16,288.59	932.06
MB	11,359	3.01	25.94	1.87	24,742	7.38	409.59	1.97
ROA	13,794	1.16	47.54	2.91	30,511	1.59	50.64	3.09
Leverage	13,750	36.70	26.92	33.67	30,513	37.20	27.85	34.08
Panel C: Proxies for the segmentation of communication								
Distance	31,938	2,570.99	2,432.64	1,580.41				
Path					65,326	1.68	2.15	1.07
Cluster					46,508	0.84	0.26	0.94
Density					65,326	0.79	0.29	0.93

Table 2.2: Effects of physical distance on loan contract terms and syndication failures

The dependent variable in column (1) is all-in-drawn spread. The dependent variable in column (2) is an indicator variable which takes the value 1 if a loan contract includes at least one financial covenant and 0 otherwise. The dependent variable in column (3) is an indicator variable which takes the value 1 if the loan has been secured and zero otherwise. The dependent variable in column (4) is an indicator variable which takes the value 1 if the contract requires guarantees and zero otherwise. In column (5), the dependent variable is an indicator variable which takes the value 1 if the loan failed and 0 otherwise. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Distance is standardized throughout the sample. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and year fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	Loan spread (1)		Covenant (2)		Secured (3)		Guaranteed (4)		Failure (5)	
Distance	-2.5845	(1.5614)	-0.0730**	(0.0283)	0.0607	(0.0339)	-0.1030*	(0.0481)	-0.1386	(0.1063)
Log (Amt)	-18.0157***	(1.7925)	-0.0348	(0.0263)	-0.2392***	(0.0355)	0.0087	(0.0290)	0.3367***	(0.0951)
Maturity	-0.4291***	(0.0972)	-0.0013	(0.0012)	0.0109***	(0.0022)	-0.0050**	(0.0018)	0.0108***	(0.0025)
Secured/Guaranteed	63.3905***	(3.6353)	-0.1438*	(0.0655)					0.3766	(0.2078)
Guaranteed					0.1010	(0.1065)				
Secured							0.1310	(0.0999)		
Missing	12.9322**	(3.4319)	-1.6385***	(0.1223)			-0.4298*	(0.1828)	-0.1348	(0.2148)
Covenant	-8.1614	(5.1872)			-0.1942**	(0.0652)	1.1363***	(0.1573)	-0.7089***	(0.1377)
Lender_num	-0.6503	(0.3545)	0.0433***	(0.0106)	-0.0301***	(0.0086)	0.0024	(0.0113)		
Pub_unrated	2.2010	(3.2079)	-0.0139	(0.0753)	-0.1226*	(0.0581)	-0.0560	(0.1017)	0.0952	(0.1531)
Pub_rated	-1.6349	(2.9906)	-0.0156	(0.1004)	-0.0262	(0.0620)	-0.1903*	(0.0916)	0.3298	(0.1730)
Log (Sales)	-9.9506***	(1.3652)	-0.1562***	(0.0298)	-0.2443***	(0.0494)	0.0641	(0.0343)	0.0470	(0.0446)
MB	-0.0995***	(0.0258)	0.0009	(0.0012)	0.0002	(0.0008)	-0.0011**	(0.0004)	0.0007	(0.0004)
ROA	-1.8984***	(0.2806)	0.0046**	(0.0018)	-0.0296***	(0.0065)	-0.0024	(0.0033)	-0.0061	(0.0069)
Leverage	0.7969***	(0.0895)	-0.0017	(0.0011)	0.0129***	(0.0024)	0.0029	(0.0018)	-0.0003	(0.0020)
GDP_PC	27.5905***	(5.4388)	0.9412***	(0.1997)	0.3932***	(0.0679)	0.9142***	(0.1106)	0.0153	(0.1675)
Credit_Spread	37.4990***	(6.6034)	-0.2989	(0.3437)	0.1474	(0.0973)	-0.2305	(0.1650)	0.2779	(0.2454)
Term_Spread	13.4258**	(4.6859)	-0.2341	(0.1536)	0.0613	(0.0330)	0.2233***	(0.0504)	-0.1435	(0.1453)
Location	3.7076	(3.2702)	-0.1041	(0.0851)	-0.0025	(0.0823)	-0.0734	(0.0856)	-0.6003*	(0.2614)
Fix effect:										
Loan type	Yes		Yes		No		No		No	
Loan purpose	Yes		Yes		No		No		No	
2-digit SIC	Yes		Yes		Yes		Yes		Yes	
N	8,605		9,559		5,524		9,559		8,487	
Adj. (Pseudo) R-sq	0.5369		0.4073		0.2670		0.3384		0.3162	

Table 2.3: Effect of relational distance on loan spreads

The dependent variable is all-in-drawn spread. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	3.5792***	(0.6606)				
Clustering			-44.3250***	(7.4740)		
Density					-51.6100***	(8.9697)
Log (Amt)	-21.2732***	(1.4399)	-19.1894***	(1.4915)	-21.0105***	(1.4879)
Maturity	-0.0057	(0.0039)	-0.0049	(0.0031)	-0.0056	(0.0039)
Secured/Guaranteed	59.3465***	(3.9918)	60.0370***	(4.2133)	57.9215***	(3.9096)
Missing	8.1895*	(3.3282)	8.3715*	(3.5560)	8.6066*	(3.3020)
Covenant	-11.8981**	(3.4480)	-8.4844*	(3.6945)	-13.0771**	(3.6638)
Lender_num	-0.7946*	(0.2765)	-0.2291	(0.1819)	-0.5435*	(0.2093)
Pub_unrated	-1.9282	(2.8232)	-1.4963	(2.2395)	-2.1985	(2.8676)
Pub_rated	-3.1289	(3.7527)	-4.3647	(3.0567)	-3.0795	(3.7890)
Log (Sales)	-6.4798***	(1.4194)	-5.0598**	(1.4370)	-6.0614***	(1.3628)
MB	-0.0016***	(0.0003)	-0.0013**	(0.0004)	-0.0016***	(0.0003)
ROA	-2.2445***	(0.2952)	-2.0234***	(0.2799)	-2.2100***	(0.3103)
Leverage	0.6966***	(0.0860)	0.6394***	(0.0796)	0.6857***	(0.0833)
GDP_PC	15.8162*	(6.8117)	13.1458	(7.4952)	18.1406*	(6.8772)
Credit_Spread	57.0536***	(10.5777)	58.5774***	(12.9945)	56.3230***	(10.9258)
Term_Spread	21.0268**	(6.9659)	19.8088*	(7.4205)	20.3506*	(7.0597)
Location	-2.6788	(2.5588)	-4.9733	(2.4562)	-2.5183	(2.5595)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	20,170		17,342		20,170	
Adj. R-sq	0.5380		0.5605		0.5429	

Table 2.4: Effects of relational distance on non-price contract terms

The dependent variable in Panel A is an indicator variable which takes the value 1 if a loan contract includes at least one financial covenant and 0 otherwise. The dependent variable in Panel B is an indicator variable which takes the value 1 if a loan is secured and 0 otherwise. The dependent variable in Panel C is an indicator variable which takes the value 1 if a loan is guaranteed and 0 otherwise. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Panel A:Results for financial covenants						
Path	0.0024	(0.0210)				
Clustering			-0.1619	(0.0937)		
Density					-0.1143	(0.0928)
Log (Amt)	-0.0630*	(0.0256)	-0.0934***	(0.0278)	-0.0617*	(0.0263)
Maturity	-0.0020*	(0.0008)	-0.0034***	(0.0010)	-0.0020*	(0.0008)
Secured/Guaranteed	-0.0621	(0.0618)	-0.0510	(0.0673)	-0.0712	(0.0621)
Missing	-1.6189***	(0.1106)	-1.6203***	(0.1093)	-1.6184***	(0.1125)
Lender_num	0.0332***	(0.0065)	0.0295***	(0.0055)	0.0329***	(0.0055)
Pub_unrated	0.0775	(0.0479)	0.1160*	(0.0575)	0.0829	(0.0461)
Pub_rated	0.1372*	(0.0569)	0.1852**	(0.0583)	0.1426*	(0.0556)
Log (Sales)	-0.1240***	(0.0192)	-0.1176***	(0.0221)	-0.1217***	(0.0190)
MB	-0.0000*	(0.0000)	-0.0000*	(0.0000)	-0.0000*	(0.0000)
ROA	0.0072***	(0.0013)	0.0063***	(0.0015)	0.0072***	(0.0013)
Leverage	-0.0019*	(0.0009)	-0.0021*	(0.0010)	-0.0019*	(0.0009)
GDP_PC	0.1406	(0.1659)	0.1626	(0.1578)	0.1465	(0.1638)
Credit_Spread	0.1528	(0.0941)	0.1423	(0.0941)	0.1500	(0.0928)
Term_Spread	0.0388	(0.0521)	0.0218	(0.0537)	0.0381	(0.0518)
Location	-0.0082	(0.0926)	-0.0096	(0.0868)	-0.0032	(0.0929)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	21,567		18,318		21,567	
Pseudo R-sq	0.3826		0.3869		0.3815	
Panel B:Results for collateral						
Path	0.1104***	(0.0097)				
Clustering			-1.2202***	(0.1794)		
Density					-0.9384***	(0.0924)
Log (Amt)	-0.2388***	(0.0233)	-0.2383***	(0.0257)	-0.2364***	(0.0231)
Maturity	0.0001	(0.0001)	0.0001	(0.0001)	0.0001	(0.0001)
Guaranteed	0.1325*	(0.0534)	0.1543**	(0.0519)	0.1270*	(0.0559)
Covenant	-0.2302*	(0.0900)	-0.2540**	(0.0920)	-0.2430**	(0.0907)

Lender_num	-0.0055	(0.0042)	0.0033	(0.0048)	-0.0010	(0.0036)
Pub_unrated	-0.3078**	(0.1052)	-0.3356**	(0.1195)	-0.3088**	(0.1093)
Pub_rated	-0.2791***	(0.0708)	-0.3176***	(0.0800)	-0.2771***	(0.0767)
Log (Sales)	-0.2801***	(0.0226)	-0.2866***	(0.0229)	-0.2748***	(0.0224)
MB	-0.0002*	(0.0001)	-0.0002**	(0.0001)	-0.0003**	(0.0001)
ROA	-0.0306***	(0.0042)	-0.0307***	(0.0049)	-0.0312***	(0.0043)
Leverage	0.0162***	(0.0010)	0.0164***	(0.0012)	0.0162***	(0.0010)
GDP_PC	0.4408***	(0.0509)	0.4797***	(0.0586)	0.4750***	(0.0510)
Credit_Spread	-0.0634*	(0.0256)	-0.0872***	(0.0259)	-0.0779**	(0.0271)
Term_Spread	-0.0132	(0.0241)	-0.0136	(0.0293)	-0.0227	(0.0252)
Location	-0.1013	(0.0624)	-0.1108*	(0.0494)	-0.0933	(0.0593)
Fix effect:						
Loan type	No		No		No	
Loan purpose	No		No		No	
2-digit SIC	Yes		Yes		Yes	
N	14,418		12,358		14,418	
Pseudo R-sq	0.2765		0.2723		0.2815	
Panel C: Results for guarantees						
Path	0.0317***	(0.0088)				
Clustering			-0.0126	(0.1180)		
Density					-0.3428***	(0.0779)
Log (Amt)	0.0185	(0.0185)	0.0018	(0.0216)	0.0144	(0.0190)
Maturity	-0.0019*	(0.0008)	-0.0017*	(0.0009)	-0.0020*	(0.0008)
Secured	0.1946***	(0.0536)	0.2124***	(0.0508)	0.1926***	(0.0527)
Missing	-0.4381***	(0.0951)	-0.4019***	(0.0934)	-0.4289***	(0.0937)
Covenant	0.8845***	(0.0671)	0.9194***	(0.0639)	0.8770***	(0.0673)
Lender_num	-0.0080**	(0.0029)	0.0003	(0.0024)	-0.0027	(0.0022)
Pub_unrated	-0.2310***	(0.0260)	-0.2333***	(0.0489)	-0.2294***	(0.0277)
Pub_rated	-0.2510***	(0.0501)	-0.2560***	(0.0531)	-0.2531***	(0.0512)
Log (Sales)	0.0336	(0.0240)	0.0224	(0.0271)	0.0337	(0.0241)
MB	-0.0002	(0.0001)	-0.0001	(0.0001)	-0.0002	(0.0001)
ROA	-0.0001	(0.0014)	-0.0009	(0.0019)	0.0002	(0.0014)
Leverage	0.0007	(0.0007)	0.0002	(0.0009)	0.0006	(0.0007)
GDP_PC	0.6774***	(0.1208)	0.7058***	(0.1170)	0.7001***	(0.1185)
Credit_Spread	-0.0944	(0.1252)	-0.0766	(0.1237)	-0.0971	(0.1227)
Term_Spread	0.2102**	(0.0641)	0.2034**	(0.0648)	0.2055**	(0.0632)
Location	0.0409	(0.0355)	0.0326	(0.0476)	0.0412	(0.0350)
Fix effect:						
Loan type	No		No		No	
Loan purpose	No		No		No	
2-digit SIC	Yes		Yes		Yes	
N	21,567		18,318		21,567	
Pseudo R-sq	0.1909		0.1918		0.1916	

Table 2.5: Effect of relational distance on syndication failures

The dependent variable is a binary variable which takes the value 1 if the loan failed and 0 otherwise. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. Only samples with lenders number between two and five are used. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	-0.3081*	(0.1423)				
Clustering			-0.7203*	(0.3537)		
Density					0.8519*	(0.3885)
Log (Amt)	0.3027***	(0.0873)	0.3366**	(0.1100)	0.3044***	(0.0875)
Maturity	0.0123***	(0.0017)	0.0120***	(0.0016)	0.0120***	(0.0016)
Secured/Guaranteed	0.2997	(0.1537)	0.0172	(0.1343)	0.3122*	(0.1527)
Missing	-0.0759	(0.2062)	-0.2438	(0.2051)	-0.0723	(0.2071)
Covenant	-0.8569***	(0.1434)	-0.8504***	(0.2078)	-0.8470***	(0.1399)
Pub_unrated	0.0887	(0.1230)	0.1691	(0.1864)	0.1038	(0.1244)
Pub_rated	0.2775*	(0.1364)	0.2979	(0.1987)	0.2874*	(0.1407)
Log (Sales)	0.0034	(0.0361)	0.0428	(0.0455)	-0.0019	(0.0356)
MB	-0.0000	(0.0001)	-0.0001	(0.0001)	-0.0001	(0.0002)
ROA	-0.0059	(0.0035)	-0.0068*	(0.0030)	-0.0056	(0.0033)
Leverage	0.0029	(0.0019)	0.0032	(0.0034)	0.0028	(0.0019)
GDP_PC	-0.4569*	(0.1811)	-0.5941**	(0.2414)	-0.4609*	(0.1779)
Credit_Spread	0.3051	(0.2258)	0.5890**	(0.2202)	0.3137	(0.2217)
Term_Spread	-0.0395	(0.1046)	-0.0569	(0.1309)	-0.0430	(0.1041)
Location	-0.7130*	(0.2885)	-0.4634*	(0.2349)	-0.7180*	(0.2901)
Fix effect:						
Loan type	No		No		No	
Loan purpose	No		No		No	
2-digit SIC	Yes		Yes		Yes	
N	14,619		11,376		14,619	
Pseudo R-sq	0.3091		0.3517		0.3069	

Table 2.6: Robustness check: alternative explanations and model specifications

This table presents robustness tests to explore alternative explanations for the relationship between relational distance and loan price. The dependent variable is all-in-drawn spread in all panels. Previous_Relation, Investment_Grade, Covenant Violation, and Foreign_Lender, are indicator variables whose definitions are provided in Appendix B, Table B1. P_Previous_Relation, C_Previous_Relation, and D_Previous_Relation are interactions between Previous_Relation and Path, Clustering, and Density, respectively. The similar notations and definitions can be applied to the other panels. Estimates for the intercepts and control variables are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Panel A: Previous relationship between lead bank and borrower						
Path	4.3131**	(1.2442)				
Clustering			-41.3556***	(8.7480)		
Density					-54.7697***	(10.7821)
Previous_Relation	-5.1585	(4.1042)	-0.9542	(9.1590)	-16.2604	(7.8482)
P_Previous_Relation	-1.3678	(1.3015)				
C_Previous_Relation			-6.0565	(11.4687)		
D_Previous_Relation					11.9728	(10.0375)
Panel B: Borrower's rating						
Path	3.1386***	(0.7530)				
Clustering			-46.9178***	(8.8689)		
Density					-49.1490***	(9.0554)
Investment_Grade	-41.2863***	(4.4285)	-43.6701**	(14.6073)	-49.7238***	(10.1636)
P_Investment_Grade	1.2634	(0.8460)				
C_Investment_Grade			4.8189	(13.5304)		
D_Investment_Grade					12.1345	(8.3259)
Panel C: Covenant violation						
Path	4.4761**	(1.4665)				
Clustering			-35.0436*	(12.3289)		
Density					-39.0086*	(13.2394)
Covenant_Violation	23.4100***	(2.7215)	26.9637**	(8.0773)	46.0926***	(9.6531)
P_Covenant_Violation	1.2044	(1.2861)				
C_Covenant_Violation			-8.0228	(9.6935)		
D_Covenant_Violation					-26.9496*	(12.0530)
Panel D: Foreign participants						
Path	7.8838**	(2.6224)				
Clustering			-41.8129***	(6.4784)		
Density					-54.7658***	(10.9528)
Foreign_Lender	-1.0971	(4.0324)	0.5974	(9.7382)	-16.0555	(8.3753)
P_Foreign_Lender	-5.1485	(2.6087)				
C_Foreign_Lender			-4.1282	(10.6922)		
D_Foreign_Lender					8.0208	(7.9902)

Table 2.7: Robustness check: endogeneity

This table reports the results of two-stage instrumental variable regressions. Panel A is the first-stage results and Panel B is the second-stage results. All control variables and fix effects in Table 3 are included in both stages but not reported. The dependent variables in column (1), (2), and (3) of Panel A are Path, Clustering, and Density, respectively. The dependent variable in Panel B is all-in-drawn spread. Inactive_num is the instrumental variable and defined as the number of nonlead-bank members who had not joined any loan syndicate in the previous three years. Definitions of the other variables are provided in Appendix B, Table B1. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Panel A: First stage						
Inactive _num	1.7375***	(0.0994)	-0.0297***	(0.0034)	-0.0547***	(0.0073)
N	20,170		17,342		20170	
Adj. R-sq	0.8205		0.1648		0.2023	
Test of instrument strength (Ho: the instrument is weak)						
F-statistic	305.3780***		76.4786***		56.1779***	
Test of exogeneity (Ho: the instrument is exogenous)						
F-statistic	0. 0376		7.1267*		5.1933*	
Panel B: Second stage						
Path	3.6828***	(0.9323)				
Clustering			-167.2799**	(55.9041)		
Density					-116.8941**	(39.5805)
N	20,170		17,342		20,170	
Adj. R-sq	0.5380		0.5249		0.5311	

Table 2.8: Robustness check: omitted variables

The dependent variable is all-in-drawn spread. Borrower_Experience is the number of previous loans that the borrower borrows from the syndicated loan markets. Lender_Experience is the average number of previous loans that the lenders of a given facility participate in the syndicated loan market. Top5 is a dummy variable, which indicates whether the lead bank belongs to the top five lenders in the league table. Z-score is Altman's (1968) Z-score. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	3.6654**	(1.0971)				
Clustering			-34.6429**	(9.5440)		
Density					-38.6990**	(10.0943)
Previous_Relation	1.3912	(0.8791)	2.3584*	(1.0526)	1.3491	(0.8990)
Investment_Grade	-45.5459***	(9.0027)	-43.0530***	(7.3032)	-45.1702***	(9.0547)
Covenant_Violation	4.7711***	(0.7752)	4.2724***	(0.8658)	4.8635***	(0.7511)
Foreign_Lender	-12.4417**	(3.9321)	-5.9076*	(2.6491)	-13.6973**	(4.0234)
Borrower_Experience	0.0346	(0.3418)	0.3504	(0.2975)	0.0602	(0.3441)
Lender_Experience	-0.0085***	(0.0014)	-0.0063**	(0.0019)	-0.0067***	(0.0014)
Top5	-0.8537	(4.0508)	1.9458	(3.7815)	0.1499	(3.8500)
Z_Score	-0.6531*	(0.2472)	-2.4235*	(0.8742)	-0.6376*	(0.2469)
Log (Amt)	-20.4164***	(2.0093)	-19.6463***	(1.7956)	-20.6941***	(1.9692)
Maturity	-0.9595***	(0.1749)	-0.9168***	(0.1809)	-0.9451***	(0.1820)
Secured/Guaranteed	63.3064***	(4.5449)	64.6131***	(5.4718)	62.8998***	(4.5512)
Missing	23.8258***	(4.7392)	25.7356***	(3.6585)	23.5859***	(4.5572)
Covenant	-15.5452**	(4.7237)	-8.6916*	(3.2024)	-15.7647**	(4.6035)
Lender_num	-1.3320*	(0.4565)	-0.5351	(0.2675)	-0.8750*	(0.3198)
Pub_unrated	4.0972	(6.6157)	7.3356	(5.3142)	3.9213	(6.5292)
Pub_rated	16.8857*	(6.9468)	15.8912*	(6.6389)	16.4534*	(6.9543)
Log (Sales)	-1.3783	(2.7821)	-1.1280	(2.3908)	-1.0359	(2.7370)
MB	0.0445	(0.0497)	0.0721	(0.0444)	0.0493	(0.0535)
ROA	-1.7446***	(0.3679)	-1.7158***	(0.3382)	-1.7329***	(0.3804)
Leverage	0.9090***	(0.0989)	0.6273***	(0.1312)	0.8865***	(0.0975)
GDP_PC	34.0044***	(5.2237)	22.2191***	(5.1405)	32.7333***	(5.3619)
Credit_Spread	54.2489**	(13.0833)	55.6013**	(16.9992)	53.3515**	(13.3535)
Term_Spread	12.9894*	(4.9710)	10.7183	(5.4424)	12.6624*	(4.9237)
Location	6.0115	(5.0293)	-0.4282	(4.8717)	5.9940	(5.0546)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	5,842		4,740		5,842	
Adj. R-sq	0.4902		0.5232		0.4920	

Appendix A

Proof of equation (1):

$$P(\theta|k,n)=\frac{P(k,n|\theta)P(\theta)}{P(k,n)}=\frac{\binom{n}{k}\theta^k(1-\theta)^{n-k}f(\theta)}{\int_{\theta=0}^{1}\binom{n}{k}\theta^k(1-\theta)^{n-k}f(\theta)d\theta}$$

$$=\frac{\theta^k(1-\theta)^{n-k}}{\int_{\theta=0}^{1}\theta^k(1-\theta)^{n-k}d\theta}$$

$$E(\theta|k,n)=\int_{\theta=0}^{1}\theta P(\theta|k,n)d\theta=\frac{\int_{\theta=0}^{1}\theta^{k+1}(1-\theta)^{n-k}d\theta}{\int_{\theta=0}^{1}\theta^k(1-\theta)^{n-k}d\theta} \quad \text{(A.1)}$$

Consider the denominator. By performing integration by parts:

$$\int_{\theta=0}^{1}\theta^k(1-\theta)^{n-k}d\theta=\frac{1}{k+1}\theta^{k+1}(1-\theta)^{n-k}\Big|_0^1+\int_{\theta=0}^{1}\frac{1}{k+1}\theta^{k+1}(n-k)(1-\theta)^{n-k-1}d\theta$$

Continuing the process of integration by parts:

$$\int_{\theta=0}^{1}\theta^k(1-\theta)^{n-k}d\theta=\int_{\theta=0}^{1}\frac{(n-k)!}{(k+1)(k+2)...n}\theta^n(1-\theta)^0d\theta$$

$$=\frac{(n-k)!k!}{n!}\int_{\theta=0}^{1}\theta^n d\theta=\frac{1}{(n+1)\binom{n}{k}} \quad \text{(A.2)}$$

The same procedure can be applied to the numerator in (A.1):

$$\int_{\theta=0}^{1}\theta^{k+1}(1-\theta)^{n-k}d\theta=\int_{\theta=0}^{1}\frac{n-k}{k+2}\theta^{k+2}(1-\theta)^{n-k-1}d\theta$$

$$=\int_{\theta=0}^{1}\frac{(n-k)!}{(k+2)(k+3)...n(n+1)}\theta^{n+1}(1-\theta)^0d\theta$$

$$=\frac{k+1}{(n+1)\binom{n}{k}}\int_{\theta=0}^{1}\theta^{n+1}d\theta$$

$$=\frac{k+1}{(n+2)(n+1)\binom{n}{k}} \quad \text{(A.3)}$$

Dividing expression (A.2) by (A.3) yields

$$E(\theta \mid k,n) = \frac{k+1}{(n+2)(n+1)\binom{n}{k}} \bigg/ \frac{1}{(n+1)\binom{n}{k}} = \frac{k+1}{n+2} \qquad \text{(QED)}$$

Proof of equations (2) and (3):

Using the result of equation (A.2):

$$p(k \mid n) = \int_{\theta=0}^{1} \binom{n}{k} \theta^{k} (1-\theta)^{n-k} f(\theta) d\theta$$

$$= \binom{n}{k} \frac{1}{(n+1)\binom{n}{k}} = \frac{1}{n+1}$$

$$p(x \le k \mid n) = \sum_{x=0}^{k} \int_{\theta=0}^{1} \binom{n}{k} \theta^{x} (1-\theta)^{n-x} f(\theta) d\theta$$

(QED)

$$= \frac{k+1}{n+1}$$

Proof of equation (4):

$$\min_{k} \left(\frac{k+1}{n+1}\right)\left(\frac{k+1}{n+2}\right) + \left(1 - \frac{k+1}{n+1}\right)\theta^{p}$$

$$= \frac{k^{2}+2k+1}{(n+1)(n+2)} + \left(1 - \frac{k+1}{n+1}\right)\theta^{p}$$

The first order condition is

$$\frac{2k+2}{(n+1)(n+2)} + \left(\frac{-1}{n+1}\right)\theta^{p} = 0$$

$$k^{*} = \left(\frac{n}{2}+1\right)\theta^{p} - 1$$

(QED)

Appendix B

Table B1: Variable definitions

Variable	Definition	Units
Panel A: Loan characteristics		
Amt	Facility amount	Million US$
Maturity	Loan maturity	Months
Secured	1 if the loan is secured, 0 otherwise	0 or 1
Guaranteed	1 if the loan is guaranteed, 0 otherwise	0 or 1
Covenant	1 if the loan has financial covenants, 0 otherwise	0 or 1
Lender_num	Lender number of a given facility	
Failed	1 if the deal status is "Cancelled" or "Suspended", 0 otherwise	0 or 1
Panel B: Borrower characteristics		
Pub_unrated	1 if the borrower is public and not rated, 0 otherwise	
Pub_rated	1 if the borrower is public and rated, 0 otherwise	
Sales	Borrower's sales at the end of the year prior to the loan active	
MB	Borrower's market to book ratio at the end of the year prior to the loan active	
ROA	Borrower's return on assets at the end of the year prior to the loan active	
Leverage	Borrower's total debt to total assets ratio at the end of the year prior to the loan active	
Panel C: Proxies for the segmentation of communication		
Distance	Average physical distance among lenders for a given facility	kilometer
Path	Average shortest path among lenders for a given facility	
Cluster	The member's clustering coefficient for a given facility	
Density	The proportion of links in a given network relative to the total possible links	
Panel D: other variables		
GDP_PC	GDP per capita (annual)	US$10,000
Credit_Spread	The difference between the yields of BAA and AAA corporate bonds	%
Term_Spread	The difference between the yields of 10-year and 2-year Treasury bonds	%
Location	1 if the borrower's principal executive office is the same as the lead bank's	0 or 1
Previous_Relationship	1 if there exists a previous relationship between the lead bank and borrower	0 or 1
Investment_Grade	1 if the borrower's credit rating is investment grade	0 or 1
Covenant_Violation	1 if the borrower violate a financial covenant	0 or 1
Foreign_Lender	1 if there is at least one foreign lender in a given facility	0 or 1

Table B2: Effects of physical distance on loan contract terms and syndication failures: Facilities without foreign lender

The dependent variable in column (1) is all-in-drawn spread. The dependent variable in column (2) is an indicator variable which takes the value 1 if a loan contract includes at least one financial covenant and 0 otherwise. The dependent variable in column (3) is an indicator variable which takes the value 1 if the loan has been secured and zero otherwise. The dependent variable in column (4) is an indicator variable which takes the value 1 if the contract requires guarantees and zero otherwise. In column (5), the dependent variable is an indicator variable which takes the value 1 if the loan failed and 0 otherwise. Only facilities without foreign number are included in this analysis. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Distance is standardized throughout the sample. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	Loan spread		Covenant		Secured		Guaranteed		Failure	
	(1)		(2)		(3)		(4)		(5)	
Distance	7.1233	(4.8900)	-0.0532	(0.0828)	0.2209*	(0.0951)	-0.0573	(0.1137)	0.1413	(0.4053)
Log (Amt)	-20.3930***	(2.6681)	0.0150	(0.0341)	-0.1965***	(0.0527)	-0.0213	(0.0484)	0.5563***	(0.1639)
Maturity	-0.4934**	(0.1290)	-0.0010	(0.0016)	0.0063*	(0.0029)	-0.0042	(0.0027)	0.0205***	(0.0032)
Secured/Guaranteed	63.8779***	(6.2246)	-0.2752**	(0.0913)					0.4746	(0.4131)
Guaranteed					0.1886	(0.1799)				
Secured							0.1875	(0.1476)		
Missing	13.6002*	(5.9867)	-1.7747***	(0.1359)			-0.3967	(0.2540)	0.1219	(0.3598)
Covenant	-13.9823	(7.6858)			-0.2146*	(0.0882)	1.0953***	(0.1446)	-0.6054	(0.3381)
Lender_num	-3.2634*	(1.1920)	0.1038***	(0.0216)	0.0071	(0.0190)	0.0430	(0.0238)		
Pub_unrated	-2.3848	(6.8450)	0.0332	(0.1010)	-0.1407	(0.1034)	0.0301	(0.1256)	-0.0402	(0.3014)
Pub_rated	-8.2764	(6.6067)	0.0405	(0.1373)	0.1577	(0.1327)	-0.2277	(0.1940)	0.3179	(0.2054)
Log (Sales)	-7.3977*	(2.6924)	-0.1937***	(0.0414)	-0.2703***	(0.0422)	0.1105**	(0.0357)	0.1078	(0.1237)
MB	-0.1301***	(0.0265)	0.0015	(0.0029)	-0.0006	(0.0022)	-0.0014**	(0.0005)	-0.0023	(0.0053)
ROA	-2.7511***	(0.3248)	0.0048	(0.0027)	-0.0459***	(0.0059)	-0.0026	(0.0038)	-0.0047	(0.0078)
Leverage	1.0061***	(0.1042)	-0.0029	(0.0017)	0.0149***	(0.0027)	0.0064***	(0.0014)	-0.0003	(0.0041)
GDP_PC	29.5473***	(6.2287)	0.9372***	(0.2150)	0.3438***	(0.0841)	0.8839***	(0.1298)	0.0094	(0.2985)
Credit_Spread	26.4965***	(4.9958)	-0.1610	(0.3000)	0.0933	(0.1633)	-0.3294*	(0.1560)	-0.9971	(0.8672)
Term_Spread	13.4920**	(4.0776)	-0.1738	(0.1417)	0.0380	(0.0524)	0.2216***	(0.0580)	-0.1262	(0.1617)
Location	1.8572	(4.8759)	-0.0912	(0.1089)	0.0408	(0.1213)	0.0240	(0.1333)		
Fix effect:										
Loan type	Yes		Yes		No		No		No	
Loan purpose	Yes		Yes		No		No		No	
2-digit SIC	Yes		Yes		Yes		Yes		Yes	
N	3,823		4,274		2,619		4,274		4,263	
Adj. (Pseudo) R-sq	0.5010		0.4232		0.2531		0.3205		0.5369	

Table B3: Effects of physical distance on loan contract terms and syndication failures: Facilities with foreign lenders

The dependent variable in column (1) is all-in-drawn spread. The dependent variable in column (2) is an indicator variable which takes the value 1 if a loan contract includes at least one financial covenant and 0 otherwise. The dependent variable in column (3) is an indicator variable which takes the value 1 if the loan has been secured and zero otherwise. The dependent variable in column (4) is an indicator variable which takes the value 1 if the contract requires guarantees and zero otherwise. In column (5), the dependent variable is an indicator variable which takes the value 1 if the loan failed and 0 otherwise. Only facilities with at least one foreign number are included in this analysis. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Distance is standardized throughout the sample. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	Loan spread		Covenant		Secured		Guaranteed		Failure	
	(1)		(2)		(3)		(4)		(5)	
Distance	-0.5539	(1.7834)	-0.1604***	(0.0467)	-0.0055	(0.0497)	-0.0970	(0.0944)	-0.0325	(0.1195)
Log (Amt)	-14.3074***	(1.6015)	-0.1345***	(0.0388)	-0.3237***	(0.0478)	0.0416	(0.0367)	0.1923	(0.0996)
Maturity	-0.2788*	(0.1154)	-0.0032	(0.0018)	0.0152***	(0.0021)	-0.0072***	(0.0021)	0.0065*	(0.0030)
Secured/Guaranteed	62.8617***	(3.9781)	-0.1181	(0.0870)					0.1717	(0.1756)
Guaranteed					-0.0455	(0.1190)				
Secured							-0.0009	(0.1165)		
Missing	11.9394**	(4.0893)	-1.6441***	(0.1419)			-0.5153*	(0.2090)	-0.6331**	(0.1963)
Covenant	-0.7128	(4.9334)	0.0452**	(0.0145)	-0.1872**	(0.0715)	1.3665***	(0.2794)	-0.7964***	(0.2184)
Lender_num	-0.6799*	(0.3237)	-0.0189	(0.1078)	-0.0272**	(0.0102)	-0.0146	(0.0130)		
Pub_unrated	6.3894	(5.7656)	-0.0284	(0.1203)	-0.0979	(0.1058)	-0.2594	(0.2089)	0.2757	(0.2658)
Pub_rated	2.3779	(3.7648)	-0.1300***	(0.0318)	-0.1575	(0.0925)	-0.2626	(0.1789)	0.3675	(0.3035)
Log (Sales)	-10.9876***	(1.6849)	0.0007	(0.0017)	-0.2357***	(0.0673)	0.0161	(0.0535)	0.0390	(0.0349)
MB	-0.0204	(0.0804)	0.0036	(0.0023)	0.0006	(0.0012)	0.0011	(0.0013)	0.0031***	(0.0009)
ROA	-1.2953***	(0.2832)	-0.0005	(0.0012)	-0.0206*	(0.0090)	-0.0030	(0.0040)	-0.0076	(0.0047)
Leverage	0.6720***	(0.0962)	1.0395***	(0.2075)	0.0114***	(0.0029)	-0.0050	(0.0033)	-0.0012	(0.0022)
GDP_PC	24.2665**	(6.9139)	-0.5462	(0.4519)	0.4926***	(0.1035)	1.0315***	(0.1371)	0.0129	(0.1553)
Credit_Spread	52.4664**	(14.1253)	-0.2982	(0.1585)	0.2649	(0.2144)	-0.0223	(0.2536)	0.6078**	(0.2072)
Term_Spread	13.6303*	(5.8556)	-0.0994	(0.0906)	0.0986	(0.0573)	0.2677**	(0.0849)	-0.1343	(0.1762)
Location	4.3726	(5.1728)	-1.6441***	(0.1419)	-0.0092	(0.1221)	-0.2041	(0.1841)		
Fix effect:										
Loan type	Yes		Yes		No		No		No	
Loan purpose	Yes		Yes		No		No		No	
2-digit SIC	Yes		Yes		Yes		Yes		Yes	
N	4,782		5,285		2,905		5,285		4,224	
Adj. (Pseudo) R-sq	0.5798		0.4248		0.2990		0.4282		0.2968	

Table B4: Effects of physical distance on loan contract terms and syndication failures: Big loans

The dependent variable in column (1) is all-in-drawn spread. The dependent variable in column (2) is an indicator variable which takes the value 1 if a loan contract includes at least one financial covenant and 0 otherwise. The dependent variable in column (3) is an indicator variable which takes the value 1 if the loan has been secured and zero otherwise. The dependent variable in column (4) is an indicator variable which takes the value 1 if the contract requires guarantees and zero otherwise. In column (5), the dependent variable is an indicator variable which takes the value 1 if the loan failed and 0 otherwise. Only facilities with facility amount equal to or larger than sample median are included in this analysis. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Distance is standardized throughout the sample. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	Loan spread (1)		Covenant (2)		Secured (3)		Guaranteed (4)		Failure (5)	
Distance	-1.4422	(1.9979)	-0.1088***	(0.0311)	0.0667	(0.0543)	-0.0858	(0.0665)	-0.0071	(0.1216)
Log (Amt)	-20.6907***	(2.9711)	-0.2473***	(0.0453)	-0.2331***	(0.0637)	-0.0588	(0.0565)	0.4317***	(0.1202)
Maturity	-0.4893***	(0.1010)	0.0005	(0.0014)	0.0073***	(0.0022)	-0.0039	(0.0025)	0.0159***	(0.0036)
Secured/Guaranteed	58.8881***	(4.5661)	-0.1550*	(0.0728)					0.5276*	(0.2599)
Guaranteed					0.1442	(0.1299)				
Secured							0.1367	(0.1377)		
Missing	14.0458*	(4.9664)	-1.6101***	(0.1156)			-0.5018	(0.2943)	0.5276*	(0.2599)
Covenant	-12.1862*	(5.4745)			-0.2384**	(0.0830)	1.1737***	(0.2128)	-0.3260*	(0.1446)
Lender_num	-0.5198	(0.5259)	0.0532***	(0.0118)	-0.0179	(0.0106)	-0.0203	(0.0200)		
Pub_unrated	3.7690	(3.2682)	0.0104	(0.0834)	-0.2232*	(0.0993)	-0.1284	(0.1824)	-0.0321	(0.3303)
Pub_rated	6.4561	(5.4597)	0.0787	(0.1272)	0.0552	(0.0973)	-0.1282	(0.1274)	0.0461	(0.3206)
Log (Sales)	-11.9162***	(2.2766)	0.0124	(0.0491)	-0.3235***	(0.0701)	0.1403	(0.0717)	0.0939	(0.0744)
MB	-0.1224***	(0.0293)	0.0052	(0.0035)	-0.0004	(0.0010)	-0.0040	(0.0053)	0.0011	(0.0006)
ROA	-1.3588***	(0.2555)	0.0021	(0.0018)	-0.0209**	(0.0076)	-0.0007	(0.0038)	-0.0083	(0.0057)
Leverage	0.8273***	(0.1222)	-0.0013	(0.0010)	0.0129***	(0.0029)	0.0008	(0.0020)	-0.0001	(0.0038)
GDP_PC	32.3359***	(54.0221)	1.0560***	(2.3876)	0.4847***	(0.6783)	0.9274***	(1.0385)	-0.1013	(2.2041)
Credit_Spread	23.4693*	(8.5410)	-0.3862	(0.4290)	0.0984	(0.0956)	-0.2416	(0.2457)	-1.5341	(1.0255)
Term_Spread	13.0015**	(4.1991)	-0.3746*	(0.1876)	-0.0190	(0.0381)	0.2147***	(0.0560)	-0.5663*	(0.2510)
Location	7.8228	(3.9638)	-0.2045	(0.1046)	0.0681	(0.0702)	0.0468	(0.1405)		
Fix effect:										
Loan type	Yes		Yes		No		No		No	
Loan purpose	Yes		Yes		No		No		No	
2-digit SIC	Yes		Yes		Yes		Yes		Yes	
N	4,303		4,555		3,185		4,086		3,628	
Adj. (Pseudo) R-sq	0.5092		0.4043		0.2345		0.3278		0.4085	

Table B5: Effects of physical distance on loan contract terms and syndication failures: Small loans

The dependent variable in column (1) is all-in-drawn spread. The dependent variable in column (2) is an indicator variable which takes the value 1 if a loan contract includes at least one financial covenant and 0 otherwise. The dependent variable in column (3) is an indicator variable which takes the value 1 if the loan has been secured and zero otherwise. The dependent variable in column (4) is an indicator variable which takes the value 1 if the contract requires guarantees and zero otherwise. In column (5), the dependent variable is an indicator variable which takes the value 1 if the loan failed and 0 otherwise. Only facilities with facility amount smaller than sample median are included in this analysis. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Distance is standardized throughout the sample. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	Loan spread (1)		Covenant (2)		Secured (3)		Guaranteed (4)		Failure (5)	
Distance	-3.0203	(1.9309)	-0.0421	(0.0427)	0.0431	(0.0545)	-0.0853	(0.0543)	-0.1667	(0.1203)
Log (Amt)	-20.1509***	(2.0173)	0.0055	(0.0330)	-0.2895***	(0.0536)	0.0506	(0.0377)	0.2974***	(0.0863)
Maturity	-0.3403*	(0.1399)	-0.0038	(0.0022)	0.0137***	(0.0028)	-0.0065*	(0.0031)	0.0100***	(0.0023)
Secured/Guaranteed	63.1141***	(4.1315)	-0.1449	(0.0983)					0.4370	(0.2871)
Guaranteed					0.0604	(0.1563)				
Secured							0.1404	(0.1389)		
Missing	10.8555*	(4.1559)	-1.6807***	(0.1513)			-0.4000	(0.2244)	0.0217	(0.4083)
Covenant	-3.4217	(7.9233)			-0.1817	(0.1152)	1.0604***	(0.1554)	-0.4364	(0.2480)
Lender_num	-0.5820	(0.3835)	0.0356***	(0.0104)	-0.0437***	(0.0085)	0.0263*	(0.0126)		
Pub_unrated	2.2029	(6.5403)	-0.0708	(0.0933)	0.0830	(0.1445)	-0.0291	(0.1419)	0.0337	(0.1857)
Pub_rated	-5.3388	(5.4471)	-0.1401	(0.1017)	-0.0529	(0.1058)	-0.2584**	(0.0991)	0.4129	(0.2235)
Log (Sales)	-4.0421	(1.9845)	-0.2016***	(0.0407)	-0.1274*	(0.0593)	0.0056	(0.0551)	-0.0454	(0.0539)
MB	-0.0711	(0.0418)	-0.0001	(0.0005)	-0.0003	(0.0011)	-0.0007	(0.0004)	0.0008	(0.0005)
ROA	-3.6654***	(0.2564)	0.0082**	(0.0028)	-0.0546***	(0.0067)	-0.0041	(0.0045)	-0.0115	(0.0089)
Leverage	0.7695***	(0.0855)	-0.0017	(0.0014)	0.0159***	(0.0030)	0.0060**	(0.0020)	-0.0025	(0.0018)
GDP_PC	25.2598***	(61.4475)	0.8941***	(1.7132)	0.3091*	(1.2559)	0.9485***	(1.3914)	0.1031	(1.4462)
Credit_Spread	47.6353***	(7.6926)	-0.2446	(0.2663)	0.2525	(0.1469)	-0.2372	(0.1737)	0.3380	(0.1872)
Term_Spread	12.5416*	(5.1986)	-0.1222	(0.1189)	0.1577*	(0.0653)	0.2160**	(0.0782)	-0.0344	(0.1319)
Location	0.1475	(4.8529)	0.0394	(0.0754)	-0.1109	(0.1492)	-0.1992	(0.1426)		
Fix effect:										
Loan type	Yes		Yes		No		No		No	
Loan purpose	Yes		Yes		No		No		No	
2-digit SIC	Yes		Yes		Yes		Yes		Yes	
N	4,302		4,928		2,305		4,586		3.737	
Adj. (Pseudo) R-sq	0.5828		0.4112		0.3458		0.3388		0.1983	

Table B6: Distribution of loan failures

This table presents the distribution of loan failures. Sample I is the sample used in the analysis for physical distance. Sample II is the sample used in the analysis for relational distance. Data are divided into nine groups according to syndicate size measured by lender numbers. The number of facilities (Obs) and failure cases (Failures) in each group are reported. Prop is Failures divided by Obs.

	Sample I			Sample II		
Lender number	Obs	Failures	Prop (%)	Obs	Failures	Prop (%)
2	12,364	88	0.71	15,955	120	0.75
3	6,844	56	0.82	10,802	104	0.96
4	3,896	21	0.54	7,287	37	0.51
5	2,503	12	0.48	5,527	26	0.47
6	1,652	2	0.12	4,449	14	0.31
7	1,104	3	0.27	3,256	6	0.18
8	821	14	1.71	2,653	15	0.57
9	542	6	1.11	2,184	14	0.64
10 and more	2,212	3	0.14	13,277	21	0.16

Table B7: t-test of paired sample of syndication failure and success.

This table presents the t-test of matched sample of syndication failure and success. The means of the variables from sample of syndication success are reported in column (1). The means of the variables from sample of syndication failure are reported in column (2). Column (3) is the difference between column (1) and (2). Definitions of the variables are provided in Table B1 of Appendix B. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	success	failure	
	(1)	(2)	(1)-(2)
Path	1.8209	1.3290	0.4919**
Clustering	0.8705	0.7971	0.0735**
Density	0.8282	0.8887	-0.0605**
Log(Amt)	19.5620	19.5757	-0.0137
Maturity	62.2455	62.2455	0.0000
Covenant	0.2232	0.2232	0.0000
ROA	1.0750	-0.3256	1.4006

Table B8: Complete results of Table 2.6, Panel A

This table presents the complete results of Table 2.6, Panel A. The dependent variable is all-in-drawn spread in all panels. P_Previous_Relation, C_Previous_Relation, and D_Previous_Relation are interactions between Previous_Relation and Path, Clustering, and Density, respectively. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	4.3131**	(1.2442)				
Clustering			-41.3556***	(8.7480)		
Density					-54.7697***	(10.7821)
Previous_Relation	-5.1585	(4.1042)	-0.9542	(9.1590)	-16.2604	(7.8482)
P_Previous_Relation	-1.3678	(1.3015)				
C_Previous_Relation			-6.0565	(11.4687)		
D_Previous_Relation					11.9728	(10.0375)
Log (Amt)	-21.1372***	(1.4358)	-19.0741***	(1.4871)	-20.8668***	(1.4808)
Maturity	-0.0058	(0.0040)	-0.0050	(0.0030)	-0.0057	(0.0039)
Secured/Guaranteed	59.2858***	(3.9322)	60.0239***	(4.1783)	57.9770***	(3.8248)
Missing	8.3448*	(3.2965)	8.5538*	(3.4850)	8.7447*	(3.2453)
Covenant	-11.8383**	(3.4132)	-8.3260*	(3.5993)	-12.9149**	(3.5654)
Lender_num	-0.7613*	(0.2654)	-0.2242	(0.1815)	-0.5118*	(0.2025)
Pub_unrated	-1.9168	(2.8368)	-1.4071	(2.2636)	-2.1408	(2.8799)
Pub_rated	-2.5393	(3.8803)	-3.7874	(3.2491)	-2.5084	(3.8958)
Log (Sales)	-6.3450***	(1.4478)	-4.9325**	(1.4430)	-5.9766***	(1.3919)
MB	-0.0017***	(0.0004)	-0.0014**	(0.0005)	-0.0017***	(0.0004)
ROA	-2.2352***	(0.2992)	-2.0191***	(0.2819)	-2.2003***	(0.3131)
Leverage	0.6983***	(0.0862)	0.6422***	(0.0804)	0.6888***	(0.0842)
GDP_PC	15.9593*	(6.3303)	13.0220	(6.8587)	18.1706*	(6.4044)
Credit_Spread	56.4903***	(10.4361)	57.9523***	(12.8531)	55.9733***	(10.7863)
Term_Spread	21.2656**	(6.8812)	20.0082*	(7.3447)	20.6068*	(6.9933)
Location	-2.6272	(2.5516)	-4.8739	(2.4428)	-2.4836	(2.5482)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	20,170		17,342		20,170	
Adj. R-sq	0.5388		0.5611		0.5435	

Table B9: Complete results of Table 2.6, Panel B

This table presents the complete results of Table 2.6, Panel B. The dependent variable is all-in-drawn spread in all panels. P_Investment_Grade, C_Investment_Grade, and D_Investment_Grade are interactions between Investment_Grade and Path, Clustering, and Density, respectively. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	3.1386***	(0.7530)				
Clustering			-46.9178***	(8.8689)		
Density					-49.1490***	(9.0554)
Investment_Grade	-41.2863***	(4.4285)	-43.6701**	(14.6073)	-49.7238***	(10.1636)
P_Investment_Grade	1.2634	(0.8460)				
C_Investment_Grade			4.8189	(13.5304)		
D_Investment_Grade					12.1345	(8.3259)
Log (Amt)	-20.5322***	(1.3846)	-17.6027***	(1.2563)	-20.0926***	(1.3325)
Maturity	-0.5456**	(0.1350)	-0.5390***	(0.1242)	-0.5352**	(0.1361)
Secured/Guaranteed	55.0322***	(4.2397)	54.2632***	(4.8996)	53.4929***	(4.2422)
Missing	6.7536*	(3.1346)	6.4632	(3.1849)	6.7421	(3.1522)
Covenant	-12.0778**	(3.6034)	-9.1752*	(3.6782)	-12.8142**	(3.7019)
Lender_num	-0.5415*	(0.2287)	-0.1027	(0.1691)	-0.3831	(0.1974)
Pub_unrated	-1.9843	(2.3761)	-2.6935	(1.5702)	-2.4693	(2.3347)
Pub_rated	7.4613	(3.6705)	5.9000	(3.3057)	7.4088	(3.6651)
Log (Sales)	-4.2593*	(1.4600)	-3.4631*	(1.4105)	-3.8278*	(1.4360)
MB	-0.0048	(0.0507)	0.0425	(0.0465)	-0.0029	(0.0512)
ROA	-2.2069***	(0.2313)	-2.0524***	(0.2206)	-2.1778***	(0.2471)
Leverage	0.6453***	(0.0917)	0.5839***	(0.0886)	0.6349***	(0.0899)
GDP_PC	12.2534*	(4.8726)	7.7676	(4.9547)	14.0787*	(4.8918)
Credit_Spread	58.7377***	(5.2513)	63.8297***	(8.0266)	58.3370***	(5.6931)
Term_Spread	13.3311*	(5.2087)	10.8083*	(5.0023)	12.6991*	(5.2481)
Location	-1.1940	(2.3444)	-2.0679	(2.4396)	-1.0124	(2.3761)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	17,107		14,669		17,107	
Adj. R-sq	0.5561		0.5856		0.5609	

Table B10: Complete results of Table 2.6, Panel C

This table presents the complete results of Table 2.6, Panel C. The dependent variable is all-in-drawn spread in all panels. P_Covenant_Violation, C_Covenant_Violation, and D_Covenant_Violation are interactions between Covenant_Violation and Path, Clustering, and Density, respectively. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	4.4761**	(1.4665)				
Clustering			-35.0436*	(12.3289)		
Density					-39.0086*	(13.2394)
Covenant_Violation	23.4100***	(2.7215)	26.9637**	(8.0773)	46.0926***	(9.6531)
P_Covenant_Violation	1.2044	(1.2861)				
C_Covenant_Violation			-8.0228	(9.6935)		
D_Covenant_Violation					-26.9496*	(12.0530)
Log (Amt)	-20.0596***	(2.0782)	-19.2278***	(1.7764)	-20.1235***	(2.0316)
Maturity	-0.9635***	(0.1934)	-0.8819***	(0.1796)	-0.9500***	(0.1956)
Secured/Guaranteed	67.4459***	(5.4620)	69.8604***	(5.5523)	66.5506***	(5.3898)
Missing	25.5554***	(5.2443)	29.5174***	(4.4348)	25.8698***	(5.1084)
Covenant	-12.3530*	(5.3188)	-6.3780	(3.9805)	-13.9379*	(5.3049)
Lender_num	-1.4699*	(0.4995)	-0.3252	(0.2553)	-0.8972*	(0.3397)
Pub_unrated	2.8631	(6.5823)	8.4315	(4.2798)	3.7515	(6.2813)
Pub_rated	10.9163	(6.8213)	13.1736*	(5.6840)	11.4272	(6.6313)
Log (Sales)	-5.8152	(2.7269)	-3.7828	(2.2199)	-5.1458	(2.5752)
MB	0.0147	(0.0636)	0.0491	(0.0600)	0.0195	(0.0699)
ROA	-1.7708**	(0.4774)	-1.7260**	(0.4612)	-1.7367**	(0.5018)
Leverage	0.8830***	(0.0947)	0.7227***	(0.1079)	0.8651***	(0.0925)
GDP_PC	23.0087**	(6.2877)	13.9483*	(5.8235)	24.6190**	(6.5949)
Credit_Spread	51.0797***	(11.4552)	51.4561**	(14.2765)	50.8042**	(12.1406)
Term_Spread	16.1486**	(4.7822)	14.6420*	(5.3815)	15.4050**	(4.8980)
Location	5.6840	(5.3965)	-2.5256	(4.1455)	5.8481	(5.0592)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	7,244		5,902		7,244	
Adj. R-sq	0.4703		0.4974		0.4747	

Table B11: Complete results of Table 2.6, Panel D

This table presents the complete results of Table 2.6, Panel D. The dependent variable is all-in-drawn spread in all panels. P_Foreign_Lender, C_Foreign_Lender, and D_Foreign_Lender are interactions between Foreign_Lender and Path, Clustering, and Density, respectively. Secured/Guaranteed takes the value 1 if a loan is secured or guaranteed. Missing is a missing value indicator which equals 1 if Secured/Guaranteed is missing and 0 otherwise. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	7.8838**	(2.6224)				
Clustering			-41.8129***	(6.4784)		
Density					-54.7658***	(10.9528)
Foreign_Lender	-1.0971	(4.0324)	0.5974	(9.7382)	-16.0555	(8.3753)
P_Foreign_Lender	-5.1485	(2.6087)				
C_Foreign_Lender			-4.1282	(10.6922)		
D_Foreign_Lender					8.0208	(7.9902)
Log (Amt)	-20.9967***	(1.5064)	-19.0194***	(1.5374)	-20.5001***	(1.4906)
Maturity	-0.0056	(0.0038)	-0.0048	(0.0030)	-0.0054	(0.0037)
Secured/Guaranteed	59.1693***	(3.9937)	59.9559***	(4.2505)	57.9062***	(3.9422)
Missing	8.2731*	(3.3303)	8.3787*	(3.5805)	8.6224*	(3.3089)
Covenant	-12.0754**	(3.5183)	-8.4355*	(3.7207)	-12.9075**	(3.6753)
Lender_num	-0.6026*	(0.2587)	-0.2065	(0.1749)	-0.4049	(0.2021)
Pub_unrated	-2.6123	(2.8686)	-1.6901	(2.2818)	-2.7795	(2.9269)
Pub_rated	-3.2357	(3.7440)	-4.3118	(3.0291)	-2.9795	(3.7845)
Log (Sales)	-6.1990***	(1.4077)	-4.9635**	(1.4381)	-5.7676***	(1.3628)
MB	-0.0016***	(0.0003)	-0.0013*	(0.0005)	-0.0016***	(0.0004)
ROA	-2.2318***	(0.2979)	-2.0233***	(0.2800)	-2.1960***	(0.3131)
Leverage	0.7009***	(0.0844)	0.6420***	(0.0788)	0.6940***	(0.0832)
GDP_PC	16.8746*	(6.7997)	13.8314	(7.4573)	18.7267*	(6.8861)
Credit_Spread	56.8397***	(10.6679)	58.4761***	(13.0152)	56.1629***	(10.9779)
Term_Spread	21.0421**	(7.0088)	19.8051*	(7.4271)	20.4118*	(7.0929)
Location	-2.8508	(2.5023)	-5.0244	(2.4492)	-2.7852	(2.5194)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	20,170		17,342		20,170	
Adj. R-sq	0.5395		0.5606		0.5437	

Table B12: The first-stage results of instrumental regression

This table reports the complete first-stage results of two-stage instrumental variable regressions. The dependent variables in column (1), (2), and (3) are Path, Clustering, and Density, respectively. Inactive_num is the instrumental variable and defined as the number of nonlead-bank members who had not joined any loan syndicate in the previous three years. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	Path		Clustering		Deinsit	
	(1)		(2)		(3)	
Inactive_num	1.7285***	(0.0971)	-0.0299***	(0.0033)	-0.0547***	(0.0073)
Log (Amt)	-0.0658**	(0.0201)	0.0247***	(0.0017)	0.0201***	(0.0031)
Maturity	-0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)
Secured/Guaranteed	0.1322***	(0.0287)	-0.0291***	(0.0050)	-0.0449***	(0.0067)
Missing	0.0729**	(0.0238)	-0.0088*	(0.0037)	-0.0035	(0.0061)
Covenant	-0.0162	(0.0433)	-0.0107	(0.0061)	-0.0184*	(0.0081)
Lender_num	0.0395***	(0.0087)	-0.0005	(0.0006)	-0.0025**	(0.0007)
Pub_unrated	-0.0150	(0.0330)	0.0013	(0.0066)	-0.0033	(0.0109)
Pub_rated	-0.0319	(0.0324)	0.0087	(0.0067)	0.0065	(0.0118)
Log (Sales)	-0.0357**	(0.0088)	0.0058*	(0.0020)	0.0139***	(0.0019)
MB	-0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)
ROA	0.0003	(0.0010)	0.0007*	(0.0003)	0.0007	(0.0005)
Leverage	-0.0002	(0.0006)	-0.0001	(0.0001)	-0.0001	(0.0001)
GDP_PC	-0.0795	(0.0288)	0.0462***	(0.0052)	0.0488***	(0.0061)
Credit_Spread	0.0473	(0.0395)	-0.0281*	(0.0105)	-0.0229	(0.0175)
Term_Spread	0.0403	(0.0203)	-0.0077*	(0.0030)	-0.0130*	(0.0054)
Location	-0.0568	(0.0394)	0.0054	(0.0047)	0.0137*	(0.0061)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	21,567		18,318		20,170	
Adj. R-sq	0.8173		0.1658		0.2023	

Table B3: The second-stage results of instrument regression

This table reports the complete second-stage results of two-stage instrumental variable regressions. The dependent variable is all-in-drawn spread. Path, Clustering, and Density are predicted value from the first-stage results. Definitions of the other variables are provided in Appendix B, Table B1. Estimates for the intercepts are not reported in this table. All regressions include loan type, loan purpose, and 2-digit SIC fixed effects. Robust standard errors allowing for clustering by year are presented in parentheses. ***, **, and * denote statistical significance at the 0.1%, 1%, and 5%, respectively.

	(1)		(2)		(3)	
Path	3.6828***	(0.9323)				
Clustering			-167.2799**	(55.9041)		
Density					-116.8941**	(39.5805)
Log (Amt)	-21.2379***	(1.2984)	-15.5168***	(2.0707)	-19.1369***	(1.6001)
Maturity	-0.0057	(0.0038)	-0.0049	(0.0027)	-0.0055	(0.0036)
Secured/Guaranteed	59.3092***	(3.8718)	55.9231***	(4.7761)	54.4923***	(4.4620)
Missing	8.1634*	(3.2634)	7.0159*	(3.5732)	7.9926*	(3.2980)
Covenant	-11.8867***	(3.3025)	-9.6229*	(3.9427)	-14.0718***	(3.7120)
Lender_num	-0.8115*	(0.3175)	-0.5627	(0.3269)	-0.9615*	(0.3755)
Pub_unrated	-1.9232	(2.7173)	-1.2167	(2.2848)	-2.3196	(2.9898)
Pub_rated	-3.1152	(3.6203)	-3.0504	(3.0405)	-2.4181	(3.7383)
Log (Sales)	-6.4674***	(1.3762)	-4.1634**	(1.5077)	-4.9913***	(1.4016)
MB	-0.0016***	(0.0003)	-0.0012**	(0.0004)	-0.0015***	(0.0004)
ROA	-2.2444***	(0.2849)	-1.9427***	(0.2792)	-2.1608***	(0.3173)
Leverage	0.6968***	(0.0833)	0.6350***	(0.0762)	0.6795***	(0.0754)
GDP_PC	16.9258**	(6.4750)	19.6168**	(7.3281)	22.3354**	(6.7895)
Credit_Spread	57.0335***	(10.1880)	54.4742***	(13.0681)	54.5201***	(11.0432)
Term_Spread	21.0311**	(6.7197)	19.1623**	(7.1963)	19.6803**	(6.9266)
Location	-2.6542	(2.4957)	-3.9618	(2.2288)	-1.2420	(2.5884)
Fix effect:						
Loan type	Yes		Yes		Yes	
Loan purpose	Yes		Yes		Yes	
2-digit SIC	Yes		Yes		Yes	
N	20,170		17,342		20,170	
Adj. R-sq	0.5380		0.5249		0.5311	

Chapter III

The Relationships between Futures Returns, Futures Volume, and Spot Returns: Evidence from Taiwan Futures Market

1. Introduction

Hayek (1945) proposed that prices function as a form of information aggregation. From this point of view, futures and spot markets reflect the same fundamental information. If markets are frictionless and complete, both futures and spot prices should react to new information at the same speed. In other words, price changes in the two markets should only occur as a contemporaneous relationship, and not a lead-lag relationship.

According to carry-cost theory, in an efficient and frictionless market with no transaction costs, contemporaneous price changes in the futures and spot markets will be perfectly correlated. However, transaction costs (such as fees and taxes), trading limitations (such as limitations on short sales), or the transaction characteristics of the asset itself (such as the leverage effect) may cause one market to react faster to information than the other.

The lead-lag relationship among different markets has been and continues to be a topic of widespread interest. In general, the exploration of lead-lag relationship helps us to understand what direction the information flow is, the role of price discovery among different markets, and where the informed traders choose to trade.

In much of the related research, futures price changes are found to consistently lead spot price changes, that is, changes in futures prices have predictive power for future movements in the spot prices [e.g., French (1986), Kawaller et al. (1987), De

Jong and Nijman (1997), De Jong and Donders (1998) , and Pati and Rajib (2011)]. A number of other studies have found evidence of a (asymmetric) feedback relationship between price changes in the two markets [e.g., Stoll and Whaley (1990), Chan (1992), Kawaller et al. (1993), Chu et al. (1999), and In and Kim (2006)]. Yang et al. (2012) found just the opposite. They investigated the information content in returns and volatilities between the stock index futures and its spot markets in China and found that the spot market plays a more dominant role in the price discovery process. One possible explanation they suggested is that high barriers to entry in the futures market may exclude many informed traders and lead to weakened ability of price discovery.

Except for Yang et al. (2012), most of the above studies tend to agree that price changes in futures markets are more informative than spot markets' price changes. Those results imply that informed traders may prefer to trade in futures markets rather than spot markets. If informed traders do actually choose to trade in futures markets, we should not only observe that futures price changes lead the changes in spot prices, but also find evidence that futures trading activity lead the spot price changes. That is, both futures price changes and trading activity can be used as a predictor of forthcoming movements in spot prices. However, the past literature has not simultaneously tested the relationships between the futures trading activity, and futures and spot price changes.

Some studies have focused on the relationship between the trading activity and price movements within the futures market. For example, Chou and Wang (2009) investigated the order placement strategies of different types of traders in Taiwan Futures market and found that foreign institutional traders and futures proprietary firms are more likely to split their orders and seem to be better informed. Lin et al. (2008) explored the interaction between the trading flows of different type of traders and the futures returns using Taiwan futures trading data, and their results showed that while individual traders are contrarians, institutional traders are positive feedback traders. However, these studies do not analyse the interaction between futures trading activity and spot price changes.

In a recent study, Roll et al. (2011) empirically analysed the joint time-series of volumes on the S&P 500 index and its four contingent claims, the options, the traditional futures, the E-mini futures, and the ETF. They found that the contingent claims volume leads the volume on spot index and predicts the spot returns and volatility around major macroeconomic announcements. However, their analysis is based on total volume and did not distinguish between good news and bad news trades. In this case, the volume series will make no directional prediction on the movements of spot prices.

The purpose of this paper is to study the informational role of the futures trading activity. Based on the Taiwan Stock Exchange Capitalization Weighted Stock Index futures (TX) and its corresponding stock index, the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), I simultaneously examine the lead-lag relationships between the futures trading activity, futures returns, and spot returns. My aims are to better understand whether informed traders do choose to trade in the futures market and to determine which type(s) of traders tend to be informed traders. To achieve these aims, it is important to analyse the relationships between the futures trading activity, futures returns, and spot returns at the same time. Only considering the relation between the two market returns, we cannot determine who tend to be informed traders. Simply examining the relation between the futures trading activity and futures returns, it is hard to conclude that informed traders do choose futures market to trade in first.

Although little research has examined the relationship between spot price changes and futures trading activity, a number of studies have investigated similar relationships in options markets [e.g., Easley et al. (1998) (hereafter EOS), Chan et al. (2002) (hereafter CCF), Pan and Poteshman (2006), Chen et al. (2005), Chan et al. (2009), and Chang et al. (2009)]. Accordingly, even though my analyses focus on the futures market, I occasionally refer to these studies of options markets.

This paper makes three distinct contributions to the field. First, as mentioned above, this paper simultaneously explores the information content of the futures price changes and trading activity for future movements of the spot price. As few studies have been conducted in this area, the relationship between futures trading activity and spot price changes remains unclear.

Second, there exist many studies which tend to support that institutional investors may be informed traders [e.g., Sias and Starks (1997) and Chakravarty (2001) for equity markets, Han et al. (2010) and Chang et al. (2009) for options markets, Lin et al. (2007) for IPOs]. Furthermore, some of the studies show that foreign institutional investors possess more information, yet some papers find that domestic institutional investors tend to be better informed [Chen et al. (2009)]. According to these literatures, we also divide the futures traders into four classes—foreign institutional traders, futures dealers, domestic institutional traders, and individual traders—and explore whether these different classes of traders have access to different information. Our study adds evidence that foreign institutional traders seem to be better informed about both spots and futures price movements.

Third, numerous studies on trading activity directly use the total trading volume [e.g., Roll et al. (2011), Sadath and Kamaiah (2009), Anthony (1988), Stephan and Whaley (1990)] or distinguish between buyer-initiated and seller-initiated trades [e.g., Easley et al. (1998), Chan et al. (2002). However, EOS indicated that the former approach does not distinguish good information from bad one, and Pan and Poteshman (2006) argued that the signal quality of the latter approach is inferior to when trading is divided into the opening and closing trades because buyer-initiated and seller-initiated are derived from the Lee-Ready Algorithm using public and observable trading data, which inevitably contains errors of classification. Therefore, this paper uses net open buy (open-buy volume minus open-sell volume), which is non-public information and can better capture the informed trades, as a proxy for futures trading activity.

Two main results are reported in this paper. First, while the futures returns have predictive power in relation to the spot returns, the overall (unclassified) futures trading activity contains no directionally predictive power for the spot returns, even if the analysis is based on private trading information (net open buy). Second, the futures trading activities of different trader types contain different information regarding the spot and futures returns. Foreign institutional traders are the only class with directionally predictive power for the spot and futures returns. The trades of futures dealers, domestic institutional traders, and individual traders all lag behind the spot and futures returns.

The remainder of this paper is organized as follows. Section 2 illustrates the empirical specifications of the study. Section 3 describes the characteristics and sources of the data used. Section 4 outlines the main empirical findings. Section 5 offers concluding remarks.

2. Empirical Specifications

Two empirical models are used to explore the lead-lag relationship between the futures price changes, futures net open buy, and spot price changes. First, drawing on Stephan and Whaley (1990), Stoll and Whaley (1990), and Chan (1992), the following multiple time series model is adopted:[1]

$$R_{s,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{f,t-i} + \sum_{i=-3}^{3} \theta_{-i} NB_{f,t-i} + e_t \tag{1}$$

where $R_{s,t}$ is the spot return at time *t*, $R_{f,t}$ is the futures return at time *t*, $NB_{f,t}$ represents the futures' net open buy at time *t*. Motivated by Wang (2004) Pan and Poteshman (2006) I use net open buy, defined as open-buy volume minus open-sell volume, as a proxy for futures trading activity. Wang (2004) used net positions (long open interest less short open interest) as a proxy for futures trading activity and transformed it into a sentiment index to test the relationship between the sentiment measure and returns in the foreign exchange futures markets. He found that speculator sentiment is positively correlated with future returns, whereas hedger sentiment is negatively related to future returns.

My work is focused on information content of futures trading, which is a flow concept. Net position, a stock measure, cannot be used directly. To modify Wang's proxy, the most intuitive way is to use the change in net position (which is equal to net buy volume). However, the aggregate net buy in futures market is zero. Even if we divide the traders into several classes so that the aggregate net buy within each class is not equal to zero, their correlations will be pretty high. To deal with this problem, we follow Pan and Poteshman (2006) to classify the transactions into four types: open buy, open sell, close buy, and close sell. Instead of using net buy volume, we introduce net open-buy volume to resolve the zero-sum issue and to mitigate the collinearity problem.[2] Using net open buy helps to more effectively capture informed

---

[1] The lag length 3 is selected according to the Bayesian information criterion (BIC) in equation (1) and in the VAR models.

[2] Following Pan and Poteshman (2006) and Chang et al. (2009), my analyses focus on opening trades.

trades in the futures market because it is private information. I also further divide the overall market into the following four classes: foreign institutional traders, futures dealers, domestic institutional traders, and individual traders.

According to equation (1), if the futures price changes lead the spot price changes, this indicates that at least one of β_{-1}, β_{-2}, β_{-3} is significantly greater than 0. Otherwise, if the spot price changes lead the futures price changes, then at least one of β_1, β_2, β_3 is significantly greater than 0. Similarly, if the futures trading activity lead the spot price changes, then it means one or more of θ_{-1}, θ_{-2}, θ_{-3} is significantly greater than 0. Otherwise, if the spot price changes lead the future trading activity, then one or more of θ_1, θ_2, θ_3 is significantly greater than 0.

Second, a reduced-form VAR is proposed. The model can be used as robust checks. Also, when the futures traders are divided into four classes, equation (1) cannot show the lead-lag relationships between the four trader classes. Therefore, a reduced-form VAR model is chosen as a second model:

$$R_{s,t} = \varphi_1 + \sum_{i=1}^{3} \gamma_{1,-i} R_{s,t-i} + \sum_{i=1}^{3} \delta_{1,-i} R_{f,t-i} + \sum_{i=1}^{3} \lambda_{1,-i} NB_{f,t-i} + \mathrm{e}_{1t} \tag{2}$$

$$R_{f,t} = \varphi_2 + \sum_{i=1}^{3} \gamma_{2,-i} R_{s,t-i} + \sum_{i=1}^{3} \delta_{2,-i} R_{f,t-i} + \sum_{i=1}^{3} \lambda_{2,-i} NB_{f,t-i} + \mathrm{e}_{2t} \tag{3}$$

$$NB_{f,t} = \varphi_3 + \sum_{i=1}^{3} \gamma_{3,-i} R_{s,t-i} + \sum_{i=1}^{3} \delta_{3,-i} R_{f,t-i} + \sum_{i=1}^{3} \lambda_{3,-i} NB_{f,t-i} + \mathrm{e}_{3t} \tag{4}$$

The reduced form VAR model clearly shows the lead-lag relationships between the variables, and can be easily used to conduct causality tests. It is capable of rapidly deciding the information content of variables.

According to the VAR model, if the futures price changes lead the spot price changes, it indicates at least one of $\delta_{1,-1}$, $\delta_{1,-2}$, $\delta_{1,-3}$ is significantly greater than 0. Otherwise, if the spot price changes lead the futures price changes, then it means at least one of $\gamma_{2,-1}$, $\gamma_{2,-2}$, $\gamma_{2,-3}$ is significantly greater than 0. If the futures trading activity lead the spot price changes, this indicates that at least one of $\lambda_{1,-1}$, $\lambda_{1,-2}$, $\lambda_{1,-3}$ is significantly greater than 0. Otherwise, if the spot price changes lead the future trading activity, then it means at least one of $\gamma_{3,-1}$, $\gamma_{3,-2}$, $\gamma_{3,-3}$ is significantly greater than 0. In addition, the VAR model can also detect the relationship between the futures price changes and futures trading activity. If the futures price changes lead the futures trading activity, then one or more of $\delta_{3,-1}$, $\delta_{3,-2}$, $\delta_{3,-3}$ is significantly greater than 0. Otherwise, if the futures trading activity lead the future price changes, then one or more of $\lambda_{2,-1}$, $\lambda_{2,-2}$, $\lambda_{2,-3}$ is significantly greater than 0.

Equations (1) to (4) can be extended to the trading activities of the four different classes of traders in the futures market (see latter equation (5) to equation (12)). This is done to understand whether the futures trading activities of the different classes of traders are informative of the spot price and/or the futures price changes.

3. Data and Preliminary Analysis

3.1 Data

This study focuses on the lead-lag relationships between the TX price changes, TX net open buy, and TAIEX Index changes and examines whether the different classes of traders have access to different information. The sample period is from April, 2004 to July, 2008. The Taiwan Futures Exchange (TAIFEX) and Taiwan Stock Exchange (TWSE) both employ an electronic automated trading system, rather than the market maker's quoting mechanism. The TX's price and trading information, which is in the form of tick-by-tick data, is supplied by the TAIFEX. It includes trading date, trading time, type of contract, trading price, trading volume, trading classification (buy or sell), opening/closing trading mark, trader identity code, etc. The TAIEX information is contained in the Taiwan Economic Journal (TEJ), which records the index level per minute in a trading day.

Within the sample period, the nearby contract is the most active contract, which is approximately 89.75% of the total TX trading volume (number of contracts). Hence, only nearby contracts are used. In addition, the futures returns, spot returns, and futures net open buy are calculated at an interval of 30 minutes.[3] Net open buy is defined as the number of open buy contracts minus the number of open sell contracts. While trading time on the TWSE begins at 9:00 a.m. and ends at 1:30 p.m., the TAIFEX trades between 8:45 a.m. and 1:45 p.m. Thus, this study only considers the price changes and trading activity during the period from 9:00 a.m. to 1:30 p.m. A total of 9,684 time series observations can be gathered during the period. In the formal analysis, the net open buy of each trader class is standardized in advance, to facilitate the expression of the regression coefficient and to enable comparisons between the trader classes.

3.2 Preliminary Analysis

Table 3.1 shows the summary data of the nearby TX transactions for the various types of trades and classes of traders. Panel A shows that open buy constitutes 27.71% of the total trading volume during the sample period, while open sell makes up another 25.36%, for a combined 53.07%. These trading data are used in later analysis. The rest of the trading volume is made up of close buy and close sell, which constitute 22.28% and 24.65% of the total, respectively. For the various classes of traders, the market shares (in terms of trading volume) for foreign institutional traders, futures

---

[3] In most cases, researchers do not divide traders by class, and use 5 minutes as the data frequency (e.g. EOS; CCF; Stephan & Whaley, 1990). On the other hand, when traders are categorized by class, the data frequency is daily (e.g. Pan & Poteshman, 2006; Chang et al., 2009). Based on the thin trading situation of each trader class and with reasonable amounts of missing value, this article uses 30 minutes as the data frequency for the empirical analysis.

dealers, domestic institutional traders, and individual traders, are 7.48%, 18.84%, 2.8%, and 70.88%, respectively. So, individual traders and futures dealers are the major players in the Taiwan Index Futures Market. These two classes of traders account for 89.29% of the total market trading volume. If only opening trades (i.e. open buy and open sell) are considered, the market shares for foreign institutional traders, futures dealers, domestic institutional traders, and individual traders are 4.12%, 9.13%, 1.31%, and 38.51%, respectively.

[*Insert Table 3.1*]

Panel B lists the number of nearby TX trading contracts for each trader class at every 30-minute interval. The average 30-minute trading volume of the overall market is 7,131 contracts. Open buy and open sell are 1,990 and 1,842 contracts, respectively. The average open buy for foreign institutional traders is 133 contracts, 335 contracts for futures dealers, 31 contracts for domestic institutional traders, and 1,492 contracts for individual traders. Similarly, the average open sell for foreign institutional traders is 154 contracts, 311 contracts for futures dealers, 63 contracts for domestic institutional traders, and 1,314 contracts for individual traders. In sum, the most active trader class on the TX contracts is the individual traders, and followed by the futures dealers. As foreign and domestic institutional traders are the least frequent traders.

Panel C lists the average trading frequency of each trader class for every 30 minutes. The overall trading frequency of the market reaches a total of approximately 3,596 trades, of which 1,019 are open buy transactions, and 899 are open sell transactions. The rest are close buy and close sell trades. According to Panel B and C of Table 3.1, the most active trader class on the Taiwan Index Futures Market is the individual traders, and followed by the futures dealers. As foreign and domestic institutional traders are the least frequent traders, they could easily give rise to the thin trading problem.[4]

In sum, foreign and domestic institutional traders trade more in open sell and close buy than in open buy and close sell. This implies these two classes of traders tend to conduct more short-sale related transactions. The possible cause may be due to hedging demand, or because of the restrictions on short selling on the spot market. When private or public bad news exists in the market, foreign and domestic institutional traders can only choose to trade on the futures market. Conversely, individual traders and futures dealers trade more in open buy and close sell than in open sell and close buy. This may be due to an awareness of transaction costs and leverage effects, which lead individual traders and futures dealers to choose to trade on the futures market rather than the spot market.

Table 3.2 lists the autocorrelations of the spot index returns and the index futures returns. The problems relating to spot index non-synchronous trading and thin trading on the futures market can be effectively mitigated using the 30-minute time interval. The results from Table 3.2 show that the autocorrelations of the spot index returns and the index futures returns are maintained at a reasonable small value, and are mostly positive, which indicates that the bid-ask bounce effect is not a problem. After lag one, the autocorrelations of the spot index returns are mostly greater than those of the index futures returns. For example, the lag-one to lag-three serial correlations of the

---

[4] I calculate the proportion of no transactions in every 30 minutes in the sample period for each trader class. The results show that the proportions for futures dealers and individual traders are both 0, and the proportions for foreign institutional traders and domestic institutional traders are 0.65% and 0.09%, respectively. Though, on average, foreign institutional traders' trading volume and trading frequency both outweigh the domestic institutional traders, the no trade proportion is relatively higher. This is because the foreign institutional traders' trading time is more clustered compared with domestic institutional traders.

spot index returns are 0.041, 0.023, and 0.024, respectively. And the corresponding figures are 0.069, 0.019, and 0.007 for the index futures returns. This may be due to the fact the TAIEX covers almost all listed stocks, including a small number of very infrequently trading stocks. When these stocks respond to information at different speeds, the index returns of the current period also partially reflect information from the previous period, which induces some degree of positive correlation. Fortunately, because these stocks usually have a relatively small market value and weight very low in the TAIEX, their influence is relatively limited. Accordingly, the autocorrelations generated can be maintained at a reasonable low level.

[*Insert Table 3.2*]

Finally, the augmented Dickey-Fuller unit root test is used to examine the stationarity of the seven time series, the TX returns, TAIEX returns, overall net open buy, and net open buys of foreign institutional traders, futures dealers, domestic institutional traders, and individual traders. The results are reported in Appendix A, Table A1. All of the tests reject the existence of a unit root. That is, all the time series I use are stationary, and it is unlikely to incur spurious regression results in my later analyses.

4. Empirical Results

4.1. The lead-lag relationship between index returns, futures returns, and overall net open buy

I first perform a regression analysis based on equation (1). The results are listed in Table 3.3. The results of regressing the spot returns only on the futures returns are shown in Model 1. When the overall TX net open buy is the only independent variable, the results are shown in Model 2. Model 3 regresses the spot returns on both the futures returns and futures net open buy.

[*Insert Table 3.3*]

The results from Model 1 in Table 3.3 clearly show that the most significant connection between the TX and TAIEX returns is the contemporaneous relationship. The largest relationship is β_0, which reaches 0.7669, and the *t*-statistic is as high as 150.82. Similar strong contemporaneous relationships can be found in Stoll and Whaley (1990), Stephan and Whaley (1990), and Chan (1992). This outcome seems to support the prediction of carry-cost theory. However, the coefficients of lags 1 and 3 are 0.0715 and 0.0269 and they are both statistically significant at 1% levels. The remaining coefficients are not significant, which indicates that the futures returns exert a one-way lead over the spot returns. That is, the futures market reflects new information faster than the spot market. This conclusion is similar to that found by Kawaller et al. (1987), De Jong and Donders (1998), and Pati and Rajib (2011).

Model 2 examines the relationship between the TX net open buy and TAIEX returns. The table shows that the contemporaneous correlation is also the most significant. The contemporaneous coefficient of the TX net open buy, θ_0, is 0.0585, which has the highest *t*-statistic of 14.64. The lead-one to lead-two coefficients are significantly greater than 0, at 0.0095, 0.0093 and 0.0278, respectively. The positive coefficient provides directional prediction, which represents the spot returns lead the TX net open buy. The coefficients of the lag one to lag three are also significantly different from 0, but have opposite signs (-0.0431, -0.0147, and -0.0126, respectively). The negative coefficients contradict the prediction of the information effect. A similar puzzle also appears in the analysis of stock options conducted by EOS and CCF. EOS believes this is due to the fact the behavior of trading volume presents some degree of complexity, and that information is not the only factor that influences the short-term movement relationship between the two markets.

Model 3 puts futures returns and futures transactions into the same regression equation. Except for a slightly lower level of significance for each coefficient, there is no fundamental difference in the conclusion. From an information-based perspective, these results indicate that the futures market leads the spot market and that the leading relationship is mainly present in the futures returns, rather than the futures volume. This conclusion is similar to that found in CCF's analysis of the option market.

The alternative methodology is the VAR model, which is represented in equations (2) to (4). According to the model's results,[5] the lag TX returns have significantly positive effects on the TAIEX returns, but the lag TAIEX returns have no significant effects on the TX returns. This indicates a one-way leading relationship between the TX returns and the TAIEX returns. The outcome is the same as that found in the general regression. In addition, none of the lags of TX net open buy have a significant effect on the TAIEX returns. Thus, by merely observing the overall TX net open buy, I am unable to perceive the information effect on the spot market.

4.2. The differences in information content of the four trader classes

In the previous analysis, I found that the futures returns do contain information on the spot returns, but I could not observe similar information content arising from futures net open buy. According to CCF, because informed traders are less aggressive in submitting orders (e.g., they only submit limit orders instead of market orders), they are able to influence the quoted price but not drive the trades. Alternatively, another plausible reason may be that the leading relationship of trading activity can only be observed on the transactions conducted by specific traders. To verify this, in the following analyses the futures traders are divided into four classes: foreign institutional traders, futures dealers, domestic institutional traders, and individual

[5] The results are reported in Appendix A, Table A2.

traders.

4.2.1 The spot market

I first study the relationship between the spot returns and the net open buy of the four trader classes in the futures market. Equation (1) is revised as follows:

$$R_{s,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{f,t-i} + \sum_{j} \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t \tag{5}$$

where $j = FI,\ D,\ DI,$ and I, represents foreign institutional traders, futures dealers, domestic institutional traders, and individual traders, respectively, and $NB_{j,t}$ represents the net open buy of trader class *j*. The regression results are presented in Table 3.4. Panel A shows the results of four regressions based on the net open buy of each trader class. In Panel B, the net open buy of all trader classes have been put into one regression.

[*Insert Table 3.4*]

The results in Panel A confirm my predictions. The lag one coefficient (0.0054) of foreign institutional traders is significantly positive but the coefficients of leads one and two (0.0091 and 0.0044) are also significantly positive. These results imply that there is a feedback relationship between the TAIEX returns and the net open buy of foreign institutional traders. The lag-two coefficient of futures dealers and the lag-one coefficient of individual traders are significantly positive. However, the magnitude of the positive coefficients almost be canceled out by the negative coefficients. From an information-based viewpoint, only the transactions of foreign institutional traders carry information on the spot market; the net open buys of futures dealers, domestic institutional traders, and individual traders have no directionally predictive power for the TAIEX returns.

Panel B shows a similar outcome to Panel A. The lead-one, lead-two, and lag-one net open buy of foreign institutional traders are positive and significant (coefficient

0.0051, 0.004 and 0.0047, respectively), which indicates a feedback relationship between the TAIEX returns and the net open buy of foreign institutional traders. The lagged net open buys of the other three trader classes are either negative or insignificant. Only the coefficient (0.005) of lag-one net open buy of individual traders is significantly positive. However, the magnitude of the positive coefficient is canceled out by the negative coefficients. The evidence again proves that the net open buy of futures dealers, domestic institutional traders, and individual traders have no directionally predictive power for the spot returns.

While Chou and Wang (2009) found that foreign institutional traders and futures dealers seem to be better informed about the TX price movements than the others in Taiwan futures market, our results provide evidence that only foreign institutional traders seem to be better informed about the spot price movements. Our finding is more consistent with that of Chang et al. (2009). How could domestic traders have less information about their own market? One possible explanation suggested by Huang and Shiu (2009) and Chou and Wang (2009) is that foreign institutional traders may have superior technological, financial, or human expertise, experience, or resources. Furthermore, these advantages are more evident when domestic traders are from emerging markets. Another possible reason mentioned by EOS is that hedging demand may also be an important motivation for trading futures. For example, futures dealers may trade exchange-traded funds (ETFs), construct stock portfolios, or trade TAIEX options based on information and hedge them in futures market. Especially, some futures dealers have qualified as market makers in the TAIEX option market, their hedging demand may be stronger. As noted by Fahlenbrach and Sandas (2003), the cheapest way to hedge delta risk of index options is to use the nearby index futures. In this situation, the trading activity of futures dealers may appear to be inversely related to the spot returns. This may partly explain why the coefficient on the lag-one net open buy of futures dealers is strongly negative in Table 3.4.

In sum, our results show that the leading relationship of futures trading activity can only be observed from a specific trader class and this class is the foreign institutional traders.[6] However, within the sample period, the foreign institutional traders' open-buy and open-sell volume is only 7.76% of the overall TX's open-buy and open-sell volume. Because the proportion is very low, the results are easily dominated by the trading activities of the other trader classes if overall volume is used. That's the reason why we observe that the futures returns informationally lead the spot returns but the overall futures trading activity informationally lags the spot returns as shown in Table 3.3.

4.2.2 The futures market

In addition to analyzing cross-market information, this article also briefly discusses price-quantity relationships solely within the futures market. The regression model is similar to equation (5); the dependent variable becomes TX returns, and the lead, contemporaneous, and lag spot returns are controlled:

$$R_{f,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{s,t-i} + \sum_{j} \sum_{i=-3}^{3} \theta_{-i} NB_{j,t-i} + e_t \tag{6}$$

where $j = FI$, D, DI, and I.

The regression results are shown in Table 3.5. Panel A reports the results of four regressions based on the net open buy of each trader class. Panel B presents the result of putting the net open buy of all trader classes into one regression.

[*Insert Table 3.5*]

According to Table 3.5, there exists a feedback relationship between the spot returns and TX returns because the lag-one, contemporaneous, and lead-one

---

[6] In the literature of options markets, Chang et al. (2009) used daily data to examine the one-way relation between the option market and spot market and found similar results to ours.

coefficients on the spot returns are all significantly positive. The lag-three coefficient of individual traders is positively significant at the 10% level. However, the magnitude of the positive coefficient is smaller than the negative one. None of the lag net open buys of the other trader types shows significantly positive. This implies that all of the trading activities of the four trader types have no predictive power for TX returns.

If effect, if I do not control the spot returns, the results of Panel A becomes that the foreign institutional traders' ($j = FI$) TX net open buy leads the index futures returns. The lag-two and lag-three coefficients are significantly positive (coefficients 0.0093 and 0.009, and the *t*-statistics are 2.46 and 2.64, respectively). The lead coefficients are all negative. These results suggest that the net open buy of foreign institutional traders informationally leads the TX returns. The lagged net open buys of the other three classes of traders appear to be significantly negatively correlated with the TX returns, and their lead net open buys appear to be significantly positively correlated with the TX returns. Therefore, the trading activities of futures dealers, domestic institutional traders, and individual traders all informationally lag behind the TX returns. These results imply that the foreign institutional traders are the only class whose trading activity have information content relating to the futures returns.

4.2.3 Robustness check

As mentioned previously, information may not be the only factor that pushes traders to trade in futures market. Traders may trade futures contracts based on other purposes such as hedging or arbitrage. To mitigate the noise from non-information-based trading, I drop the transactions from arbitrager and re-run equations (5) and (6). That is if a trader trade finance sector index futures (TF) or/and electronic sector index futures (TE) and TX at the same day, it may be an arbitrage-based transaction and I exclude it from the sample. The regression results are shown in Appendix A, Tables A3 and A4.

After purging the effect of arbitrage, the new regression results show no conclusive change. The patterns reported in Tables A3 and A4 are quite similar to those in Tables 3.4 and 3.5, which confirm the robustness of my conclusions.

4.3 VAR Results

In the final analysis, I augment the VAR model presented by equations (2) to (4). I put the net open buy of the various trader classes into the model:

$$R_{s,t} = \varphi_1 + \sum_{i=1}^{3}\gamma_{1,-i}R_{s,t-i} + \sum_{i=1}^{3}\delta_{1,-i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{1,-i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{1,-i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{1,-i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{1,-i}NB_{I,t-i} + \mathrm{e}_{1t} \quad (7)$$

$$R_{f,t} = \varphi_2 + \sum_{i=1}^{3}\gamma_{2,-i}R_{s,t-i} + \sum_{i=1}^{3}\delta_{2,-i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{2,-i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{2,-i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{2,-i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{2,-i}NB_{I,t-i} + \mathrm{e}_{2t} \quad (8)$$

$$NB_{FI,t} = \varphi_3 + \sum_{i=1}^{3}\gamma_{3,-i}R_{s,t-i} + \sum_{i=1}^{3}\delta_{3,-i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{3,-i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{3,-i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{3,-i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{3,-i}NB_{I,t-i} + \mathrm{e}_{3t} \quad (9)$$

$$NB_{D,t} = \varphi_4 + \sum_{i=1}^{3}\gamma_{4,-i}R_{s,t-i} + \sum_{i=1}^{3}\delta_{4,-i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{4,-i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{4,-i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{4,-i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{4,-i}NB_{I,t-i} + \mathrm{e}_{4t} \quad (10)$$

$$NB_{DI,t} = \varphi_5 + \sum_{i=1}^{3}\gamma_{5,-i}R_{s,t-i} + \sum_{i=1}^{3}\delta_{5,-i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{5,-i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{5,-i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{5,-i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{5,-i}NB_{I,t-i} + \mathrm{e}_{5t} \quad (11)$$

$$NB_{I,t} = \varphi_6 + \sum_{i=1}^{3}\gamma_{6,-i}R_{s,t-i} + \sum_{i=1}^{3}\delta_{6,-i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{6,-i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{6,-i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{6,-i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{6,-i}NB_{I,t-i} + \mathrm{e}_{6t} \quad (12)$$

All the notations have the same definitions as previously stated.

The spot returns have a high degree of contemporaneous correlation with the futures returns. The correlation coefficient reaches 0.838. Hence, two additional VAR models are constructed to check robustness. One is to exclude the futures returns, and the other is to exclude the spot returns from the model. I only point out the important differences, and do not report the detailed results of the two VAR models.[7]

The results of the augmented VAR parameter estimation are shown in Table 3.6. The estimation of equation (7) shows that both the lag-one to lag-three TX returns (coefficients 0.3685, 0.1054, and 0.039, respectively) and the lag-two net open buy of foreign institutional traders (coefficient 0.0081) have significantly positive correlation with the TAIEX returns. The coefficients of lag net open buys of futures dealers, domestic institutional traders, and individual trader are either insignificant or negative. These again confirm that, while futures returns and the net open buy of foreign institutional traders carry information about spot returns, the other three trader classes do not.

[*Insert Table 3.6*]

The estimation of equation (8) shows the lag-two coefficient (0.0421) of the spot returns appears to be significantly positively correlated with the TX returns. This implies the spot returns also informationally lead the futures returns. As a result, both the TX and the spot returns have predictive power for each other, and take on an asymmetrical feedback relationship. The effect the futures returns have in leading the spot returns is stronger than the inverse leading relationship. In addition, the lag-two net open buy of foreign institutional traders appears to be significantly positively correlated with the futures returns (coefficient 0.0147). The coefficient of lag-two net open buy of futures dealers is also positive and marginal significant (coefficient 0.0059), but its magnitude is canceled out by the negative coefficients. That is, the foreign institutional traders are still the only class whose trades have information content relating to the futures returns.

The estimation result of equation (9) shows that the spot returns (the futures

[7] The tables are provided in Appendix A, Tables A5 and A6.

returns) positively (negatively) lead the net open buy of the foreign institutional traders. The coefficients are all statistically significant at the 1% level. When the spot returns or the futures returns are excluded from the model, only the coefficient of lag-one TX returns on foreign institutional traders' net open buy is significant negative. When the futures returns are excluded from the model, only the coefficient of the lag-three spot returns is significantly positive. Neither of the other two lag coefficients is significant. The results are consistent with Table 3.4 and Panel A of Table 3.5.

The estimation result of equation (10) shows that the lag-one to lag-three spot returns are negatively correlated with the net open buy of the futures dealers (coefficients -0.2664, -0.1508, and -0.1768, respectively). The lag-one and lag-two coefficients of the futures returns, on the other hand, are significantly positively correlated with the net open buy of the futures dealers (coefficients 0.6825 and 0.106, respectively). When the spot returns are excluded from the model, the coefficient of the lag-one futures returns is significantly positive but the coefficient of lag-three futures returns is significantly negative. Because the positive coefficient is larger than the negative, the conclusion is unchanged. Similarly, when the futures returns are excluded from the model, the coefficient of the lag-one spot returns is positive (0.3574) and reaches the 1% significance level. The coefficient of the lag-three spot returns is negative and is also significant at the 1% level (coefficients 0.1591). Because the positive coefficient is greater than the negative, the futures dealers' net open buy informationally lags behind both the spot and futures returns.

The estimation of equation (11) shows that the lag-one coefficient of the spot returns is significantly positive (0.1718). When I exclude the futures returns from the model, the result is unchanged. When I exclude the spot returns from the model, the coefficient of lag-two TX returns becomes significantly positive. This indicates the net open buy of domestic institutional traders lags behind both spot and futures returns.

The estimation of equation (12) shows that the lag-one and lag-three spot returns are significantly negatively correlated with the net open buy of individual traders (coefficient -0.3405 and -0.1247). On the other hand, the lag-one to lag-three TX returns are significantly positively correlated with the net open buy of individual traders (coefficient 0.2258, 0.182 and 0.1308, respectively). When the spot returns are excluded from the model, the coefficient of the lag-two futures returns is significantly positive. When the futures returns are excluded from the model, the coefficient of the lag-one spot returns is still significantly negative, and the lag-two coefficient becomes significantly positive. The magnitude of these two coefficients is almost identical. These results show that the net open buy of the individual traders is less informative of futures returns.

In addition to the relationship between the TAIEX index returns, TX returns and the net open buy of the four TX trader classes, further insights are contained in Table 3.6. To begin with, it is useful to observe the interaction between the different classes of traders. Because the net open buy of each trader class is affected by its own past behaviors, the persistence of the behavior exists over time. Besides, each trader class also interacts with the other classes to some degree. However, the direction of these interactions is complex. I only present the results that have relatively clear direction. First, the lag-one and lag-three net open buy of the individual traders is positively correlated with that of the futures dealers (coefficients 0.0510 and 0.0252). Symmetrically, the lag-one and lag-three net open buy of the futures dealers is also positively correlated with that of the individual traders (coefficients 0.0340 and 0.0424). This shows that the trading behavior between the two groups appears to be a feedback relation. Second, domestic institutional traders have only minimal interaction with the other groups. Especially, their trading activity seems to exert no influence on those of the other groups.

In the unreported results,[8] I conducted separate Granger tests for each of the three VAR models (the augmented VAR, the VAR excluding spot returns, and the VAR excluding the futures returns). The results fully confirm the above conclusions.

Thus far, I have explained my empirical results by focusing on the information-based effect. However, other perspectives such as trading strategies may also partly interpret our results. For example, Table 3.6 shows that the lags of TX returns are negatively related to the net open buy of foreign institutional traders and positively related to the net open buys of the other three trader classes. This suggests that foreign institutional traders may adopt contrarian strategies, whereas the other three classes of traders may adopt momentum strategies. These findings are not totally consistent with those of Lin et al. (2008) which found that foreign institutional traders and dealers (both futures and securities dealers) are positive feedback traders and individual traders are contrarians. The possible reasons may have two. First, while we use intraday net open buy as a proxy for trading activity, their proxy is daily net buy volume. Second, our sample period is from April 2004 to July 2008, but theirs is from January 2001 to December 2002.

In summary, the main conclusions of this section are as follows. After classifying the futures traders, the futures returns still strongly lead the spot returns, although this relationship is not unidirectional. The spot returns also lead the futures returns, but with a weak significance. This suggests an asymmetric feedback relationship. The net open buy of foreign institutional traders is the only trade to inform changes in both the futures and spot prices. This information effect is one-way in the futures market, but a

[8] The results are provided in Appendix A, Tables A7 to A9.

two-way feedback relationship in the spot market. The net open buys of the futures dealers, domestic institutional traders, and the individual traders have no directionally predictive power for both futures and spot prices.

5. Conclusions

This study uses a unique dataset to explore the intraday information-based relation between the futures market and the spot market. I examine the lead-lag relationship between the TX returns, TX net open buy, and TAIEX returns in the overall market. Also, I categorize the futures traders into foreign institutional traders, futures dealers, domestic institutional traders, and individual traders. Following this, I conduct a detailed analysis of whether the trades of the four trader classes contain different information for the spot and futures markets.

For the overall market, the results indicate that the futures market leads the spot market. However, this leading relationship of futures market is only reflected in futures returns, and not in the overall futures trading activity. When the different sources of trading are not distinguished, observing the overall TX net open buy has no informationally leading effect for either futures prices or spot prices.

After dividing the futures traders into four classes, the TX returns still lead the TAIEX returns. The leading relationship is an asymmetric feedback relationship. That is, the TX returns strongly lead the TAIEX returns, and the TAIEX returns weakly lead the TX returns. In addition, the net open buy of foreign institutional traders have directionally predictive power for both the TX returns and the TAIEX returns. The net open buy of the foreign institutional traders has a one-way leading relationship with the TX returns, but a two-way feedback relationship with the TAIEX returns. The net open buy of the futures dealers, domestic institutional traders, and the individual traders all lag behind the futures and spot returns.

In sum, this paper reveals that informed traders do choose to trade in the futures market. In particular, foreign institutional traders tend to be the informed traders. However, due to the low market share foreign institutional traders have in the futures market (less than 8%), the leading relation is difficult to be discerned from the overall market trading activity. This implies that analyses based on overall market trading volume may produce inaccurate results.

Table 3.1: Summary statistics of the TX transactions

This table reports the summary statistics of the TX trading activity. The transactions are divided into open buy, open sell, close buy, and close sell. The traders are categorized as foreign institutional traders, futures dealers, domestic institutional traders, and individual traders. The statistics in Panel A represent the market share in terms of trading volume for each trader class according to the different transaction types. Panel B shows the average 30-minutes volume traded by each trader class according to the different transaction types. The statistics in Panel C show that, during the sample period, the average trading frequency of each trader class according to the different transaction types within a 30-minute interval.

	Open buy	Open sell	Close buy	Close sell
Panel A: Market share (%)				
All	27.71	25.36	22.28	24.65
Foreign institutional traders	1.92	2.20	1.85	1.51
Futures dealers	4.73	4.40	4.76	4.95
Domestic institutional traders	0.43	0.88	1.13	0.36
Individual traders	20.62	17.89	14.54	17.83
Panel B: Average volume (30 minutes)				
All	1,990	1,842	1,575	1,724
Foreign institutional traders	133	154	130	106
Futures dealers	335	311	339	352
Domestic institutional traders	31	63	82	24
Individual traders	1,492	1,314	1,024	1,242
Panel C: Average frequency (30 minutes)				
All	1,019	899	788	890
Foreign institutional traders	49	53	50	40
Futures dealers	135	123	143	147
Domestic institutional traders	11	25	36	9
Individual traders	823	699	559	695

Table 3.2: Autocorrelations of the spot returns and futures returns

This table shows series correlations of the TAIEX and TX returns. $R_{s,t}$ represents the 30-minute TAIEX returns. $R_{f,t}$ represents the 30-minute TX returns. AC represents autocorrelation. *t*-stat. represents the *t*-statistic. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	$R_{s,t}$		$R_{f,t}$	
Lag	AC	t-stat.	AC	t-stat.
1	0.041	(4.08)***	0.069	(7.52)***
2	0.023	(2.21)**	0.019	(2.08)**
3	0.024	(2.31)**	0.007	(0.80)
4	-0.026	(-2.56)**	-0.011	(-1.22)
5	0.023	(2.29)**	0.006	(0.61)
6	0.032	(3.13)**	0.007	(0.73)
7	0.023	(2.23)**	0.019	(2.05)**
8	-0.006	(-0.62)	0.023	(2.48)**
9	-0.045	(-4.36)***	0.008	(0.82)
10	-0.018	(-1.77)*	0.003	(0.31)
11	-0.007	(-0.68)	-0.046	(-4.94)***
12	-0.002	(-0.22)	-0.025	(-2.66)***

Table 3.3: Regression results of the spot returns on the futures returns and net open buy

This table presents the estimated parameters of the regression and the t-statistic. The regression model is:

$$R_{s,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{f,t-i} + \sum_{i=-3}^{3} \theta_{-i} NB_{f,t-i} + e_t$$

$R_{s,t}$ represents the 30-minute TAIEX returns. $R_{f,t}$ represents the 30-minute TAIEX returns. $NB_{f,t}$ represents the 30-minute net open buy of the TX. e_t is the residuals. Model 1 regresses the TAIEX returns on the TX returns. Model 2 regresses the TAIEX returns on the TX net open buy. Model 3 regresses the TAIEX returns on both the TX returns and the TX net open buy. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. The Adj. R^2 in the table represents the adjusted R-square. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Model 1		Model 2		Model 3	
$R_{f,t+3}$	0.0046	(0.87)			0.0036	(0.67)
$R_{f,t+2}$	0.0060	(1.06)			0.0047	(0.82)
$R_{f,t+1}$	-0.0079	(-1.44)			-0.0047	(-0.83)
$R_{f,t}$	0.7669***	(150.82)			0.7628***	(146.44)
$R_{f,t-1}$	0.0715***	(13.83)			0.0680***	(12.81)
$R_{f,t-2}$	0.0073	(1.35)			0.0077	(1.39)
$R_{f,t-3}$	0.0269***	(5.04)			0.0269***	(4.98)
$NB_{f,t+3}$			0.0095***	(2.69)	0.0014	(0.72)
$NB_{f,t+2}$			0.0093**	(2.27)	0.0009	(0.40)
$NB_{f,t+1}$			0.0278***	(6.62)	-0.0075***	(-3.19)
$NB_{f,t}$			0.0585***	(14.64)	0.0143***	(6.34)
$NB_{f,t-1}$			-0.0431***	(-10.95)	-0.0048**	(-2.18)
$NB_{f,t-2}$			-0.0147***	(-3.77)	-0.0017	(-0.78)
$NB_{f,t-3}$			-0.0126***	(-3.63)	-0.0011	(-0.56)
Adj. R^2	0.7095		0.0526		0.7108	

Table 3.4: Regression results of the spot returns on the net open buy of the four trader classes

This table presents the t-statistic of the parameter estimation of the regression and the estimated parameters from regressing the TAIEX returns on the net open buy of the four trader classes. Panel A uses one of the four trader classes to perform the regressions. Panel B uses the net open buy of four trader classes in the same regression to conduct the parameter estimation. $R_{s,t}$ represents the 30-minute TAIEX returns. $NB_{j,t}$ represents the 30-minute net open buy of the trader class j in the futures market. FI, D, DI, and I represent the foreign institutional traders, the futures dealers, the domestic institutional traders, and the individual traders, respectively. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. The Adj. R^2 in the table represents the adjusted R-square. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	$j=FI$		$j=D$		$j=DI$		$j=I$	
Panel A: $R_{s,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{f,t-i} + \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{f,t+3}$	0.0052	(0.99)	0.0031	(0.56)	0.0048	(0.90)	0.0047	(0.88)
$R_{f,t+2}$	0.0081	(1.45)	0.0081	(1.37)	0.0064	(1.13)	0.0062	(1.10)
$R_{f,t+1}$	-0.0070	(-1.28)	0.0004	(0.08)	-0.0084	(-1.53)	-0.0071	(-1.30)
$R_{f,t}$	0.7664***	(151.88)	0.7474***	(138.86)	0.7669***	(150.31)	0.7685***	(151.22)
$R_{f,t-1}$	0.0705***	(13.76)	0.0726***	(13.33)	0.0712***	(13.73)	0.0723***	(13.97)
$R_{f,t-2}$	0.0049	(0.92)	0.0149***	(2.65)	0.0077	(1.43)	0.0061	(1.12)
$R_{f,t-3}$	0.0261***	(4.93)	0.0255***	(4.63)	0.0271***	(5.06)	0.0248***	(4.66)
$NB_{j,t+3}$	0.0017	(0.95)	0.0021	(1.05)	-0.0006	(-0.36)	-0.0014	(-0.71)
$NB_{j,t+2}$	0.0044**	(2.06)	-0.0046**	(-2.00)	-0.0032	(-1.54)	0.0006	(0.25)
$NB_{j,t+1}$	0.0091***	(4.22)	-0.0079***	(-3.39)	0.0054***	(2.61)	-0.0099***	(-4.38)
$NB_{j,t}$	-0.0252***	(-13.52)	0.0312***	(14.38)	-0.0022	(-1.27)	0.0145***	(7.03)
$NB_{j,t-1}$	0.0054***	(2.83)	-0.0265***	(-12.28)	0.0009	(0.53)	0.0074***	(3.58)
$NB_{j,t-2}$	0.0028	(1.51)	0.0040*	(1.85)	-0.0013	(-0.76)	-0.0059***	(-2.86)
$NB_{j,t-3}$	-0.0003	(-0.19)	0.0011	(0.58)	0.0010	(0.60)	-0.0027	(-1.46)
Adj. R^2	0.7150		0.7176		0.7098		0.7122	
Panel B: $R_{s,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{f,t-i} + \sum_{j} \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{f,t+3}$	0.0037	(0.67)						
$R_{f,t+2}$	0.0097	(1.64)						
$R_{f,t+1}$	0.0020	(0.34)						
$R_{f,t}$	0.7495***	(138.83)						
$R_{f,t-1}$	0.0722***	(13.17)						
$R_{f,t-2}$	0.0128**	(2.28)						
$R_{f,t-3}$	0.0237***	(4.33)						
$NB_{j,t+3}$	0.0017	(0.89)	0.0021	(1.00)	-0.0007	(-0.43)	0.0003	(0.14)
$NB_{j,t+2}$	0.0040*	(1.80)	-0.0042*	(-1.79)	-0.0024	(-1.17)	-0.0001	(-0.05)
$NB_{j,t+1}$	0.0051**	(2.25)	-0.0091***	(-3.83)	0.0034*	(1.69)	-0.0095***	(-3.93)
$NB_{j,t}$	-0.0179***	(-8.74)	0.0301***	(13.42)	0.0005	(0.30)	0.0121***	(5.30)
$NB_{j,t-1}$	0.0047**	(2.30)	-0.0251***	(-11.31)	0.0013	(0.75)	0.0050**	(2.19)
$NB_{j,t-2}$	0.0022	(1.10)	0.0034	(1.52)	-0.0019	(-1.07)	-0.0029	(-1.28)
$NB_{j,t-3}$	-0.0011	(-0.61)	0.0014	(0.69)	0.0008	(0.47)	-0.0035*	(-1.74)
Adj. R^2				0.7236				

Table 3.5: Regression results of the TX returns on the net open buy of the four trader classes

This table presents the t-statistic of the parameter estimation of the regression and the estimated parameters from regressing the TX returns on the net open buy of the four trader classes. Panel A uses one of the four trader classes to perform the regressions. Panel B uses the net open buy of four trader classes in the same regression to conduct the parameter estimation. $R_{f,t}$ represents the 30-minute TX returns. $NB_{j,t}$ represents the 30-minute net open buy of the trader class j in the futures market. *FI*, *D*, *DI*, and *I* represent the foreign institutional traders, the futures dealers, the domestic institutional traders, and the individual traders, respectively. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. The Adj. R^2 in the table represents the adjusted R-square. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	j=FI		*j=D*		*j=DI*		*j=I*	
Panel A: $R_{f,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{s,t-i} + \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{s,t+3}$	0.0023	(0.39)	0.0055	(0.92)	0.0017	(0.29)	0.0021	(0.36)
$R_{s,t+2}$	0.0048	(0.81)	0.0090	(1.48)	0.0060	(1.02)	0.0050	(0.84)
$R_{s,t+1}$	0.1170***	(19.85)	0.1034***	(16.60)	0.1196***	(20.22)	0.1179***	(19.99)
$R_{s,t}$	0.9084***	(154.01)	0.8845***	(139.80)	0.9026***	(152.63)	0.9040***	(153.37)
$R_{s,t-1}$	0.0237***	(4.02)	0.0165***	(2.67)	0.0234***	(3.96)	0.0207***	(3.52)
$R_{s,t-2}$	0.0059	(1.00)	0.0068	(1.11)	0.0053	(0.89)	0.0046	(0.78)
$R_{s,t-3}$	-0.0112*	(-1.91)	-0.0067	(-1.10)	-0.0111*	(-1.87)	-0.0092	(-1.57)
$NB_{j,t+3}$	0.0010	(0.51)	-0.0055***	(-2.63)	-0.0006	(-0.33)	0.0034	(1.62)
$NB_{j,t+2}$	-0.0065***	(-2.79)	-0.0003	(-0.13)	0.0039*	(1.73)	0.0037	(1.55)
$NB_{j,t+1}$	-0.0096***	(-4.12)	0.0230***	(9.24)	-0.0019	(-0.84)	0.0071***	(2.93)
$NB_{j,t}$	0.0208***	(10.30)	0.0039*	(1.67)	0.0068***	(3.57)	-0.0157***	(-7.12)
$NB_{j,t-1}$	-0.0041**	(-1.98)	-0.0019	(-0.81)	-0.0057***	(-3.02)	-0.0082***	(-3.71)
$NB_{j,t-2}$	-0.0003	(-0.17)	-0.0056**	(-2.40)	0.0022	(1.14)	0.0034	(1.53)
$NB_{j,t-3}$	0.0008	(0.46)	-0.0055***	(-2.67)	-0.0018	(-1.02)	0.0034*	(1.69)
Adj. R^2	0.7186		0.7190		0.7159		0.7179	
Panel B: $R_{f,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{s,t-i} + \sum_{j} \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{f,t+3}$	0.0056	(0.94)						
$R_{f,t+2}$	0.0044	(0.72)						
$R_{f,t+1}$	0.0980***	(15.72)						
$R_{f,t}$	0.8844***	(139.47)						
$R_{f,t-1}$	0.0138**	(2.23)						
$R_{f,t-2}$	0.0051	(0.84)						
$R_{f,t-3}$	-0.0063	(-1.04)						
$NB_{j,t+3}$	0.0009	(0.47)	-0.0052**	(-2.45)	-0.0001	(-0.04)	0.0025	(1.11)
$NB_{j,t+2}$	-0.0043*	(-1.76)	0.0004	(0.18)	0.0043*	(1.95)	0.0030	(1.18)
$NB_{j,t+1}$	-0.0032	(-1.31)	0.0240***	(9.37)	-0.0002	(-0.11)	0.0089***	(3.41)
$NB_{j,t}$	0.0203***	(9.14)	0.0063**	(2.56)	0.0061***	(3.20)	-0.0061**	(-2.48)
$NB_{j,t-1}$	-0.0087***	(-3.92)	-0.0049**	(-2.07)	-0.0074***	(-3.94)	-0.0138***	(-5.64)
$NB_{j,t-2}$	-0.0008	(-0.37)	-0.0043*	(-1.82)	0.0022	(1.16)	0.0017	(0.72)
$NB_{j,t-3}$	0.0005	(0.23)	-0.0057***	(-2.73)	-0.0013	(-0.73)	0.0021	(0.97)
Adj. R^2				0.7244				

Table 3.6: Vector autoregression (VAR) results

This table presents the estimated parameters and the t-statistic of vector autoregression (VAR). The model is:

$$R_{s,t}=\varphi_1+\sum_{i=1}^{3}\gamma_{1,-i}R_{s,t-i}+\sum_{i=1}^{3}\delta_{1,-i}R_{f,t-i}+\sum_{i=1}^{3}\eta_{1,-i}NB_{FI,t-i}+\sum_{i=1}^{3}\mu_{1,-i}NB_{D,t-i}+\sum_{i=1}^{3}\pi_{1,-i}NB_{DI,t-i}+\sum_{i=1}^{3}\omega_{1,-i}NB_{I,t-i}+\mathrm{e}_{1t}$$

$$R_{f,t}=\varphi_2+\sum_{i=1}^{3}\gamma_{2,-i}R_{s,t-i}+\sum_{i=1}^{3}\delta_{2,-i}R_{f,t-i}+\sum_{i=1}^{3}\eta_{2,-i}NB_{FI,t-i}+\sum_{i=1}^{3}\mu_{2,-i}NB_{D,t-i}+\sum_{i=1}^{3}\pi_{2,-i}NB_{DI,t-i}+\sum_{i=1}^{3}\omega_{2,-i}NB_{I,t-i}+\mathrm{e}_{2t}$$

$$NB_{FI,t}=\varphi_3+\sum_{i=1}^{3}\gamma_{3,-i}R_{s,t-i}+\sum_{i=1}^{3}\delta_{3,-i}R_{f,t-i}+\sum_{i=1}^{3}\eta_{3,-i}NB_{FI,t-i}+\sum_{i=1}^{3}\mu_{3,-i}NB_{D,t-i}+\sum_{i=1}^{3}\pi_{3,-i}NB_{DI,t-i}+\sum_{i=1}^{3}\omega_{3,-i}NB_{I,t-i}+\mathrm{e}_{3t}$$

$$NB_{D,t}=\varphi_4+\sum_{i=1}^{3}\gamma_{4,-i}R_{s,t-i}+\sum_{i=1}^{3}\delta_{4,-i}R_{f,t-i}+\sum_{i=1}^{3}\eta_{4,-i}NB_{FI,t-i}+\sum_{i=1}^{3}\mu_{4,-i}NB_{D,t-i}+\sum_{i=1}^{3}\pi_{4,-i}NB_{DI,t-i}+\sum_{i=1}^{3}\omega_{4,-i}NB_{I,t-i}+\mathrm{e}_{4t}$$

$$NB_{DI,t}=\varphi_5+\sum_{i=1}^{3}\gamma_{5,-i}R_{s,t-i}+\sum_{i=1}^{3}\delta_{5,-i}R_{f,t-i}+\sum_{i=1}^{3}\eta_{5,-i}NB_{FI,t-i}+\sum_{i=1}^{3}\mu_{5,-i}NB_{D,t-i}+\sum_{i=1}^{3}\pi_{5,-i}NB_{DI,t-i}+\sum_{i=1}^{3}\omega_{5,-i}NB_{I,t-i}+\mathrm{e}_{5t}$$

$$NB_{I,t}=\varphi_6+\sum_{i=1}^{3}\gamma_{6,-i}R_{s,t-i}+\sum_{i=1}^{3}\delta_{6,-i}R_{f,t-i}+\sum_{i=1}^{3}\eta_{6,-i}NB_{FI,t-i}+\sum_{i=1}^{3}\mu_{6,-i}NB_{D,t-i}+\sum_{i=1}^{3}\pi_{6,-i}NB_{DI,t-i}+\sum_{i=1}^{3}\omega_{6,-i}NB_{I,t-i}+\mathrm{e}_{6t}$$

$R_{s,t}$ represents the 30-minute TAIEX returns. $R_{f,t}$ represents the 30-minute TX returns. $NB_{FI,t}$, $NB_{D,t}$, $NB_{DI,t}$, and $NB_{I,t}$ represent the 30-minute TX net open buy of the foreign institutional traders, the futures dealers, the domestic institutional traders, and the individual traders, respectively. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	$R_{s,t}$		$R_{f,t}$		$NB_{FI,t}$		$NB_{D,t}$		$NB_{DI,t}$		$NB_{I,t}$	
$R_{s,t-1}$	-0.3056***	(-15.73)	-0.0150	(-0.70)	0.3617***	(5.62)	-0.2664***	(-4.54)	0.1718***	(2.60)	-0.3405***	(-5.74)
$R_{s,t-2}$	-0.0836***	(-4.16)	0.0421*	(1.89)	0.3299***	(4.95)	-0.1508**	(-2.49)	-0.0262	(-0.38)	-0.0620	(-1.01)
$R_{s,t-3}$	-0.0221	(-1.15)	0.0260	(1.22)	0.2970***	(4.68)	-0.1768***	(-3.06)	0.0130	(0.20)	-0.1247**	(-2.13)
$R_{f,t-1}$	0.3685***	(21.11)	0.0818***	(4.23)	-0.3450***	(-5.96)	0.6825***	(12.95)	0.0261	(0.44)	0.2258***	(4.23)
$R_{f,t-2}$	0.1054***	(5.74)	-0.0153	(-0.75)	-0.2855***	(-4.69)	0.1060*	(1.91)	0.0635	(1.02)	0.1820***	(3.24)
$R_{f,t-3}$	0.0390**	(2.19)	-0.0243	(-1.23)	-0.2226***	(-3.76)	0.0103	(0.19)	0.0326	(0.54)	0.1308**	(2.40)
$NB_{FI,t-1}$	-0.0007	(-0.19)	-0.0014	(-0.38)	0.3230***	(28.61)	-0.0216**	(-2.10)	0.0071	(0.62)	-0.0055	(-0.53)
$NB_{FI,t-2}$	0.0081**	(2.27)	0.0147***	(3.73)	0.1062***	(9.00)	0.0055	(0.51)	-0.0203*	(-1.68)	-0.0327***	(-3.00)
$NB_{FI,t-3}$	0.0006	(0.18)	0.0035	(0.93)	0.0536***	(4.77)	0.0236**	(2.30)	-0.0183	(-1.59)	0.0315***	(3.04)
$NB_{D,t-1}$	-0.0037	(-1.01)	0.0037	(0.91)	-0.0288**	(-2.38)	0.3673***	(33.36)	0.0051	(0.41)	0.0510***	(4.58)
$NB_{D,t-2}$	-0.0010	(-0.27)	-0.0009	(-0.21)	-0.0023	(-0.18)	0.1250***	(10.68)	-0.0335**	(-2.54)	-0.0164	(-1.38)
$NB_{D,t-3}$	0.0046	(1.31)	0.0026	(0.65)	-0.0170	(-1.44)	0.1152***	(10.77)	-0.0146	(-1.21)	0.0252**	(2.33)
$NB_{DI,t-1}$	-0.0018	(-0.59)	-0.0055	(-1.63)	0.0124	(1.23)	-0.0084	(-0.92)	0.2536***	(24.60)	-0.0033	(-0.36)
$NB_{DI,t-2}$	0.0027	(0.86)	0.0059*	(1.71)	-0.0041	(-0.39)	0.0030	(0.32)	0.1095***	(10.35)	-0.0084	(-0.89)
$NB_{DI,t-3}$	-0.0017	(-0.56)	-0.0028	(-0.84)	-0.0194*	(-1.94)	-0.0133	(-1.46)	0.0956***	(9.31)	0.0018	(0.20)
$NB_{I,t-1}$	0.0010	(0.26)	-0.0101**	(-2.48)	0.0335***	(2.75)	0.0340***	(3.06)	-0.0082	(-0.65)	0.3846***	(34.18)
$NB_{I,t-2}$	-0.0107***	(-2.71)	-0.0072*	(-1.65)	0.0131	(1.00)	-0.0225*	(-1.89)	-0.0224*	(-1.67)	0.1014***	(8.38)
$NB_{I,t-3}$	0.0004	(0.10)	0.0028	(0.69)	-0.0391***	(-3.21)	0.0424***	(3.82)	0.0033	(0.27)	0.1481***	(13.19)

Appendix A

Table A1: Unit root tests

This table presents the results of the augmented Dickey-Fuller unit root tests using the following equation:

$$\Delta x_t = \alpha x_{t-1} + \sum_{i=1}^{3} \beta_i \Delta x_{t-i} + e_t$$

Δ is the differencing operator. x_t is one of the following variables: the TX returns ($R_{f,t}$), TAIEX returns ($R_{s,t}$), overall net open buy ($NB_{f,t}$), net open buy of foreign institutional traders ($NB_{FI,t}$), net open buy of futures dealers ($NB_{D,t}$), net open buy of domestic institutional traders ($NB_{DI,t}$), net open buy of individual traders ($NB_{I,t}$). Only the coefficient α is reported in this table. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Estimated parameter (α)	t-statistics
$R_{f,t\text{-}1}$	-0.9150***	(-53.26)
$R_{s,t\text{-}1}$	-0.9441***	(-48.70)
$NB_{f,t\text{-}1}$	-0.3121***	(-32.10)
$NB_{FI,t\text{-}1}$	-0.4767***	(-39.66)
$NB_{D,t\text{-}1}$	-0.3488***	(-34.05)
$NB_{DI,t\text{-}1}$	-0.4906***	(-38.53)
$NB_{I,t\text{-}1}$	-0.3305***	(-32.58)

Table A2: VAR results – unclassified data

This table presents the estimated parameters and the t-statistic of vector autoregression (VAR). The model is:

$$R_{s,t} = \varphi_1 + \sum_{i=1}^{3} \gamma_{1i} R_{s,t-i} + \sum_{i=1}^{3} \delta_{1i} R_{f,t-i} + \sum_{i=1}^{3} \lambda_{1i} NB_{f,t-i} + \mathrm{e}_{1t}$$
$$R_{f,t} = \varphi_2 + \sum_{i=1}^{3} \gamma_{2i} R_{s,t-i} + \sum_{i=1}^{3} \delta_{2i} R_{f,t-i} + \sum_{i=1}^{3} \lambda_{2i} NB_{f,t-i} + \mathrm{e}_{2t}$$
$$NB_{f,t} = \varphi_3 + \sum_{i=1}^{3} \gamma_{3i} R_{s,t-i} + \sum_{i=1}^{3} \delta_{3i} R_{f,t-i} + \sum_{i=1}^{3} \lambda_{3i} NB_{f,t-i} + \mathrm{e}_{3t}$$

$R_{s,t}$ represents the 30-minute TAIEX returns. $R_{f,t}$ represents the 30-minute TX returns. $NB_{f,t}$ represent the 30-minute overall TX net open buy. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	$R_{s,t}$		$R_{f,t}$		$NB_{f,t}$	
$R_{s,t-1}$	-0.3108***	(-16.22)	-0.0209	(-0.98)	0.1692***	(2.60)
$R_{s,t-2}$	-0.0980***	(-4.94)	0.0253	(1.15)	-0.0375	(-0.56)
$R_{s,t-3}$	-0.0221	(-1.17)	0.0214	(1.02)	0.0108	(0.17)
$R_{f,t-1}$	0.3716***	(21.48)	0.0930***	(4.85)	0.0297	(0.51)
$R_{f,t-2}$	0.1138***	(6.27)	0.0001	(0.00)	0.0469	(0.76)
$R_{f,t-3}$	0.0418**	(2.37)	-0.0191	(-0.98)	-0.0004	(-0.01)
$NB_{f,t-1}$	-0.0017	(-0.56)	-0.0042	(-1.26)	0.2561***	(25.27)
$NB_{f,t-2}$	0.0045	(1.48)	0.0075**	(2.20)	0.1146***	(11.02)
$NB_{f,t-3}$	-0.0014	(-0.49)	-0.0027	(-0.83)	0.0970***	(9.59)

Table A3: Robustness check of Table 3.4

This table presents the t-statistic of the parameter estimation of the regression and the estimated parameters from regressing the TAIEX returns on the net open buy of the four trader classes. If a trader trade TF (or TE) and TX at the same day, then the transaction is excluded from the sample. Panel A uses one of the four trader classes to perform the regressions. Panel B uses the net open buy of four trader classes in the same regression to conduct the parameter estimation. $R_{s,t}$ represents the 30-minute TAIEX returns. $NB_{j,t}$ represents the 30-minute net open buy of the trader class j in the futures market. *FI*, *D*, *DI*, and *I* represent the foreign institutional traders, the futures dealers, the domestic institutional traders, and the individual traders, respectively. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. The Adj. R^2 in the table represents the adjusted R-square. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	j=FI		*j=D*		*j=DI*		*j=I*	
Panel A: $R_{s,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{f,t-i} + \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{f,t+3}$	0.0062	(1.18)	0.0072	(1.35)	0.0061	(1.16)	0.0057	(1.07)
$R_{f,t+2}$	0.0063	(1.13)	0.0057	(1.00)	0.0052	(0.93)	0.0049	(0.87)
$R_{f,t+1}$	-0.0042	(-0.77)	-0.0023	(-0.42)	-0.0045	(-0.81)	-0.0045	(-0.81)
$R_{f,t}$	0.7626***	(151.11)	0.7614***	(148.33)	0.7645***	(150.27)	0.7700***	(150.66)
$R_{f,t-1}$	0.0700***	(13.68)	0.0689***	(13.24)	0.0710***	(13.77)	0.0760***	(14.60)
$R_{f,t-2}$	0.0087	(1.63)	0.0106**	(1.96)	0.0097*	(1.80)	0.0084	(1.55)
$R_{f,t-3}$	0.0265***	(5.01)	0.0287***	(5.35)	0.0273***	(5.13)	0.0237***	(4.43)
$NB_{j,t+3}$	0.0022	(1.27)	-0.0034*	(-1.79)	-0.0002	(-0.14)	-0.0021	(-1.08)
$NB_{j,t+2}$	-0.0015	(-0.71)	0.0006	(0.29)	-0.0037*	(-1.78)	0.0005	(0.25)
$NB_{j,t+1}$	0.0083***	(3.92)	-0.0046**	(-2.19)	0.0043**	(2.13)	-0.0081***	(-3.63)
$NB_{j,t}$	-0.0204***	(-11.27)	0.0117***	(5.86)	-0.0024	(-1.33)	0.0151***	(7.41)
$NB_{j,t-1}$	0.0047**	(2.57)	-0.0046**	(-2.27)	-0.0001	(-0.05)	0.0075***	(3.68)
$NB_{j,t-2}$	0.0031*	(1.72)	-0.0022	(-1.06)	0.0000	(0.00)	-0.0059***	(-2.87)
$NB_{j,t-3}$	-0.0010	(-0.62)	-0.0024	(-1.30)	0.0001	(0.06)	-0.0026	(-1.39)
Adj. R^2	0.7122		0.7095		0.7085		0.7112	
Panel B: $R_{s,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{f,t-i} + \sum_{j} \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{f,t+3}$	0.0075	(1.41)						
$R_{f,t+2}$	0.0071	(1.25)						
$R_{f,t+1}$	-0.0025	(-0.45)						
$R_{f,t}$	0.7647***	(147.98)						
$R_{f,t-1}$	0.0732***	(13.93)						
$R_{f,t-2}$	0.0106*	(1.93)						
$R_{f,t-3}$	0.0262***	(4.86)						
$NB_{j,t+3}$	0.0020	(1.12)	-0.0035*	(-1.87)	-0.0007	(-0.42)	-0.0013	(-0.66)
$NB_{j,t+2}$	-0.0023	(-1.07)	0.0001	(0.05)	-0.0038*	(-1.84)	-0.0006	(-0.26)
$NB_{j,t+1}$	0.0071***	(3.31)	-0.0044**	(-2.10)	0.0038*	(1.85)	-0.0069***	(-3.08)
$NB_{j,t}$	-0.0180***	(-9.75)	0.0109***	(5.51)	-0.0005	(-0.30)	0.0114***	(5.44)
$NB_{j,t-1}$	0.0065***	(3.51)	-0.0040**	(-1.99)	0.0012	(0.66)	0.0092***	(4.36)
$NB_{j,t-2}$	0.0028	(1.52)	-0.0023	(-1.16)	-0.0011	(-0.62)	-0.0045**	(-2.15)
$NB_{j,t-3}$	-0.0014	(-0.85)	-0.0030	(-1.63)	-0.0002	(-0.11)	-0.0028	(-1.44)
Adj. R^2				0.7150				

Table A4: Robustness check of Table 3.5

This table presents the t-statistic of the parameter estimation of the regression and the estimated parameters from regressing the TX returns on the net open buy of the four trader classes. If a trader trade TF (or TE) and TX at the same day, then the transaction is excluded from the sample. Panel A uses one of the four trader classes to perform the regressions. Panel B uses the net open buy of four trader classes in the same regression to conduct the parameter estimation. $R_{f,t}$ represents the 30-minute TX returns. $NB_{j,t}$ represents the 30-minute net open buy of the trader class j in the futures market. *FI*, *D*, *DI*, and *I* represent the foreign institutional traders, the futures dealers, the domestic institutional traders, and the individual traders, respectively. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. The Adj. R^2 in the table represents the adjusted R-square. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	j=FI		*j=D*		*j=DI*		*j=I*	
Panel A: $R_{f,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{s,t-i} + \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{f,t+3}$	0.0003	(0.06)	0.0010	(0.16)	0.0008	(0.14)	0.0017	(0.29)
$R_{f,t+2}$	0.0066	(1.10)	0.0077	(1.29)	0.0061	(1.03)	0.0047	(0.80)
$R_{f,t+1}$	0.1163***	(19.61)	0.1141***	(19.04)	0.1173***	(19.73)	0.1153***	(19.49)
$R_{f,t}$	0.9099***	(153.03)	0.9018***	(150.14)	0.9049***	(152.27)	0.9000***	(152.38)
$R_{f,t-1}$	0.0221***	(3.72)	0.0239***	(3.99)	0.0225***	(3.78)	0.0160***	(2.69)
$R_{f,t-2}$	0.0037	(0.62)	0.0044	(0.74)	0.0039	(0.65)	0.0046	(0.77)
$R_{f,t-3}$	-0.0099*	(-1.67)	-0.0095	(-1.59)	-0.0105*	(-1.77)	-0.0075	(-1.26)
$NB_{j,t+3}$	-0.0001	(-0.07)	0.0004	(0.20)	-0.0012	(-0.61)	0.0067***	(3.28)
$NB_{j,t+2}$	0.0014	(0.62)	-0.0028	(-1.22)	0.0043*	(1.93)	0.0025	(1.04)
$NB_{j,t+1}$	-0.0075***	(-3.27)	0.0113***	(4.97)	-0.0010	(-0.46)	0.0059**	(2.48)
$NB_{j,t}$	0.0146***	(7.42)	0.0024	(1.13)	0.0062***	(3.25)	-0.0275***	(-12.65)
$NB_{j,t-1}$	-0.0033*	(-1.69)	-0.0063***	(-2.89)	-0.0044**	(-2.29)	-0.0029	(-1.30)
$NB_{j,t-2}$	-0.0009	(-0.47)	0.0008	(0.39)	0.0006	(0.34)	0.0041*	(1.86)
$NB_{j,t-3}$	0.0013	(0.73)	-0.0003	(-0.13)	-0.0004	(-0.23)	0.0051**	(2.52)
Adj. R^2	0.7156		0.7149		0.7144		0.7197	
Panel B: $R_{f,t} = \alpha_0 + \sum_{i=-3}^{3} \beta_{-i} R_{s,t-i} + \sum_{j} \sum_{i=-3}^{3} \theta_{j,-i} NB_{j,t-i} + e_t$								
$R_{f,t+3}$	0.0019	(0.33)						
$R_{f,t+2}$	0.0051	(0.86)						
$R_{f,t+1}$	0.1106***	(18.46)						
$R_{f,t}$	0.8980***	(149.32)						
$R_{f,t-1}$	0.0158***	(2.62)						
$R_{f,t-2}$	0.0047	(0.79)						
$R_{f,t-3}$	-0.0073	(-1.24)						
$NB_{j,t+3}$	0.0011	(0.61)	0.0009	(0.44)	0.0001	(0.05)	0.0067***	(3.14)
$NB_{j,t+2}$	0.0029	(1.25)	-0.0027	(-1.20)	0.0046**	(2.07)	0.0028	(1.18)
$NB_{j,t+1}$	-0.0058**	(-2.54)	0.0111***	(4.89)	-0.0006	(-0.29)	0.0057**	(2.35)
$NB_{j,t}$	0.0098***	(4.89)	0.0022	(1.04)	0.0028	(1.48)	-0.0252***	(-11.24)
$NB_{j,t-1}$	-0.0050**	(-2.49)	-0.0063***	(-2.92)	-0.0052***	(-2.72)	-0.0051**	(-2.27)
$NB_{j,t-2}$	-0.0008	(-0.40)	0.0018	(0.83)	0.0014	(0.73)	0.0036	(1.61)
$NB_{j,t-3}$	0.0019	(1.03)	0.0004	(0.22)	0.0003	(0.17)	0.0053**	(2.56)
Adj. R^2				0.7215				

Table A5: VAR results – excluding the futures returns

This table presents the estimated parameters and the t-statistic of vector autoregression (VAR). The model is:

$$R_{s,t} = \varphi_1 + \sum_{i=1}^{3}\gamma_{1i}R_{s,t-i} + \sum_{i=1}^{3}\eta_{1i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{1i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{1i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{1i}NB_{I,t-i} + e_{1t}$$

$$NB_{FI,t} = \varphi_2 + \sum_{i=1}^{3}\gamma_{2i}R_{s,t-i} + \sum_{i=1}^{3}\eta_{2i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{2i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{2i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{2i}NB_{I,t-i} + e_{2t}$$

$$NB_{D,t} = \varphi_3 + \sum_{i=1}^{3}\gamma_{3i}R_{s,t-i} + \sum_{i=1}^{3}\eta_{3i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{3i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{3i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{3i}NB_{I,t-i} + e_{3t}$$

$$NB_{DI,t} = \varphi_4 + \sum_{i=1}^{3}\gamma_{4i}R_{s,t-i} + \sum_{i=1}^{3}\eta_{4i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{4i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{4i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{4i}NB_{I,t-i} + e_{4t}$$

$$NB_{I,t} = \varphi_5 + \sum_{i=1}^{3}\gamma_{5i}R_{s,t-i} + \sum_{i=1}^{3}\eta_{5i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{5i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{5i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{5i}NB_{I,t-i} + e_{5t}$$

$R_{s,t}$ represents the 30-minute TAIEX returns. $NB_{FI,t}$, $NB_{D,t}$, $NB_{DI,t}$, and $NB_{I,t}$ represent the 30-minute TX net open buy of the foreign institutional traders, the futures dealers, the domestic institutional traders, and the individual traders, respectively. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	$R_{s,t}$		$NB_{FI,t}$		$NB_{D,t}$		$NB_{DI,t}$		$NB_{I,t}$	
$R_{s,t-1}$	0.0364***	(3.35)	0.0210	(0.60)	0.3574***	(11.04)	0.2020***	(5.58)	-0.1182***	(-3.63)
$R_{s,t-2}$	0.0206*	(1.89)	0.0462	(1.30)	-0.0439	(-1.35)	0.0342	(0.94)	0.1178***	(3.61)
$R_{s,t-3}$	0.0182*	(1.70)	0.0899***	(2.58)	-0.1591***	(-4.98)	0.0435	(1.22)	-0.0024	(-0.08)
$NB_{FI,t-1}$	0.0052	(1.50)	0.3186***	(28.25)	-0.0103	(-1.00)	0.0073	(0.64)	-0.0026	(-0.25)
$NB_{FI,t-2}$	0.0063*	(1.73)	0.1046***	(8.88)	0.0006	(0.06)	-0.0195	(-1.61)	-0.0318***	(-2.93)
$NB_{FI,t-3}$	0.0027	(0.79)	0.0503***	(4.48)	0.0273***	(2.65)	-0.0182	(-1.58)	0.0335***	(3.24)
$NB_{D,t-1}$	0.0027	(0.73)	-0.0365***	(-3.03)	0.3783***	(34.24)	0.0062	(0.50)	0.0561***	(5.05)
$NB_{D,t-2}$	-0.0006	(-0.15)	-0.0066	(-0.51)	0.1243***	(10.56)	-0.0326**	(-2.48)	-0.0138	(-1.17)
$NB_{D,t-3}$	0.0029	(0.79)	-0.0154	(-1.31)	0.1121***	(10.40)	-0.0149	(-1.23)	0.0241**	(2.23)
$NB_{DI,t-1}$	0.0005	(0.16)	0.0100	(0.99)	-0.0042	(-0.46)	0.2538***	(24.63)	-0.0018	(-0.19)
$NB_{DI,t-2}$	0.0006	(0.20)	-0.0033	(-0.32)	-0.0013	(-0.14)	0.1097***	(10.38)	-0.0090	(-0.95)
$NB_{DI,t-3}$	-0.0010	(-0.32)	-0.0197**	(-1.97)	-0.0117	(-1.27)	0.0955***	(9.31)	0.0020	(0.22)
$NB_{I,t-1}$	0.0001	(0.02)	0.0329***	(2.69)	0.0317***	(2.83)	-0.0078	(-0.62)	0.3849***	(34.19)
$NB_{I,t-2}$	-0.0147***	(-3.64)	0.0169	(1.29)	-0.0297**	(-2.47)	-0.0229*	(-1.70)	0.0988***	(8.17)
$NB_{I,t-3}$	0.0019	(0.50)	-0.0374***	(-3.07)	0.0463***	(4.15)	0.0028	(0.22)	0.1472***	(13.11)

Table A6: VAR results – excluding the spot returns

This table presents the estimated parameters and the t-statistic of vector autoregression (VAR). The model is:

$$
\begin{aligned}
R_{f,t} &= \varphi_1 + \sum_{i=1}^{3}\delta_{1i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{1i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{1i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{1i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{1i}NB_{I,t-i} + e_{1t} \\
NB_{FI,t} &= \varphi_2 + \sum_{i=1}^{3}\delta_{2i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{2i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{2i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{2i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{2i}NB_{I,t-i} + e_{2t} \\
NB_{D,t} &= \varphi_3 + \sum_{i=1}^{3}\delta_{3i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{3i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{3i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{3i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{3i}NB_{I,t-i} + e_{3t} \\
NB_{DI,t} &= \varphi_4 + \sum_{i=1}^{3}\delta_{4i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{4i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{4i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{4i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{4i}NB_{I,t-i} + e_{4t} \\
NB_{I,t} &= \varphi_5 + \sum_{i=1}^{3}\delta_{5i}R_{f,t-i} + \sum_{i=1}^{3}\eta_{5i}NB_{FI,t-i} + \sum_{i=1}^{3}\mu_{5i}NB_{D,t-i} + \sum_{i=1}^{3}\pi_{5i}NB_{DI,t-i} + \sum_{i=1}^{3}\omega_{5i}NB_{I,t-i} + e_{5t}
\end{aligned}
$$

$R_{f,t}$ represents the 30-minute TX returns. $NB_{FI,t}$, $NB_{D,t}$, $NB_{DI,t}$, and $NB_{I,t}$ represent the 30-minute TX net open buy of the foreign institutional traders, the futures dealers, the domestic institutional traders, and the individual traders, respectively. Estimates for the intercepts are not reported in this table. t-statistics are presented in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	$R_{f,t}$		$NB_{FI,t}$		$NB_{D,t}$		$NB_{DI,t}$		$NB_{I,t}$	
$R_{f,t-1}$	0.0709***	(6.60)	-0.0713**	(-2.21)	0.4814***	(16.42)	0.1544***	(4.68)	-0.0299	(-1.01)
$R_{f,t-2}$	0.0152	(1.40)	-0.0093	(-0.28)	-0.0282	(-0.95)	0.0569*	(1.70)	0.1091***	(3.63)
$R_{f,t-3}$	-0.0015	(-0.14)	0.0301	(0.94)	-0.1371***	(-4.69)	0.0418	(1.27)	0.0291	(0.99)
$NB_{FI,t-1}$	-0.0009	(-0.25)	0.3206***	(28.42)	-0.0193*	(-1.88)	0.0047	(0.41)	-0.0016	(-0.16)
$NB_{FI,t-2}$	0.0140***	(3.55)	0.1047***	(8.88)	0.0052	(0.48)	-0.0183	(-1.52)	-0.0349***	(-3.22)
$NB_{FI,t-3}$	0.0035	(0.92)	0.0508***	(4.53)	0.0257**	(2.51)	-0.0191*	(-1.66)	0.0337***	(3.26)
$NB_{D,t-1}$	0.0029	(0.73)	-0.0240**	(-1.99)	0.3630***	(33.16)	0.0093	(0.75)	0.0441***	(3.97)
$NB_{D,t-2}$	0.0003	(0.08)	-0.0039	(-0.30)	0.1283***	(11.07)	-0.0380***	(-2.92)	-0.0096	(-0.81)
$NB_{D,t-3}$	0.0024	(0.61)	-0.0153	(-1.31)	0.1133***	(10.69)	-0.0135	(-1.13)	0.0227**	(2.12)
$NB_{DI,t-1}$	-0.0054	(-1.62)	0.0134	(1.33)	-0.0089	(-0.98)	0.2538***	(24.61)	-0.0038	(-0.41)
$NB_{DI,t-2}$	0.0060*	(1.73)	-0.0021	(-0.21)	0.0018	(0.19)	0.1099***	(10.38)	-0.0096	(-1.01)
$NB_{DI,t-3}$	-0.0027	(-0.81)	-0.0193*	(-1.92)	-0.0132	(-1.45)	0.0953***	(9.28)	0.0023	(0.25)
$NB_{I,t-1}$	-0.0106***	(-2.60)	0.0352***	(2.88)	0.0320***	(2.88)	-0.0060	(-0.48)	0.3811***	(33.89)
$NB_{I,t-2}$	-0.0070	(-1.60)	0.0149	(1.14)	-0.0231*	(-1.94)	-0.0227*	(-1.69)	0.1013***	(8.39)
$NB_{I,t-3}$	0.0032	(0.80)	-0.0388***	(-3.18)	0.0426***	(3.84)	0.0020	(0.16)	0.1495***	(13.32)

Table A7: Granger test I

This table presents the Granger test of VAR model in Table 3.6.

Dependent variable	Null hypothesis	Chi-square
$R_{s,t}$	$\delta_{1i}=0$, for i=1, 2, 3	445.82^{***}
	$\eta_{1i}=0$, for i=1, 2, 3	6.65^{*}
	$\mu_{1i}=0$, for i=1, 2, 3	2.37
	$\pi_{1i}=0$, for i=1, 2, 3	1.08
	$\omega_{1i}=0$, for i=1, 2, 3	10.36^{**}
	$\delta_{1i}=\eta_{1i}=\mu_{1i}=\pi_{1i}=\omega_{1i}=0$, for i=1, 2, 3	495.66^{***}
$R_{f,t}$	$\gamma_{2i}=0$, for i=1, 2, 3	5.74
	$\eta_{2i}=0$, for i=1, 2, 3	20.72^{***}
	$\mu_{2i}=0$, for i=1, 2, 3	1.89
	$\pi_{2i}=0$, for i=1, 2, 3	4.91
	$\omega_{2i}=0$, for i=1, 2, 3	16.16^{***}
	$\gamma_{2i}=\eta_{2i}=\mu_{2i}=\pi_{2i}=\omega_{2i}=0$, for i=1, 2, 3	65.61^{***}
$NB_{FI,t}$	$\gamma_{3i}=0$, for i=1, 2, 3	55.95^{***}
	$\delta_{3i}=0$, for i=1, 2, 3	52.85^{***}
	$\mu_{3i}=0$, for i=1, 2, 3	14.73^{***}
	$\pi_{3i}=0$, for i=1, 2, 3	5.16
	$\omega_{3i}=0$, for i=1, 2, 3	16.35^{***}
	$\gamma_{3i}=\delta_{3i}=\mu_{3i}=\pi_{3i}=\omega_{3i}=0$, for i=1, 2, 3	100.20^{***}
$NB_{D,t}$	$\gamma_{4i}=0$, for i=1, 2, 3	28.14^{***}
	$\delta_{4i}=0$, for i=1, 2, 3	169.95^{***}
	$\eta_{4i}=0$, for i=1, 2, 3	9.15^{**}
	$\pi_{4i}=\delta_{3i}=0$, for i=1, 2, 3	3.58
	$\omega_{4i}=0$, for i=1, 2, 3	31.51^{***}
	$\gamma_{4i}=\delta_{4i}=\eta_{4i}=\pi_{4i}=\omega_{4i}=0$, for i=1, 2, 3	367.85^{***}
$NB_{DI,t}$	$\gamma_{5i}=0$, for i=1, 2, 3	8.06^{**}
	$\delta_{5i}=0$, for i=1, 2, 3	1.15
	$\eta_{5i}=0$, for i=1, 2, 3	7.97^{**}
	$\mu_{5i}=0$, for i=1, 2, 3	14.46^{***}
	$\omega_{5i}=0$, for i=1, 2, 3	5.91
	$\gamma_{5i}=\delta_{5i}=\eta_{5i}=\mu_{5i}=\omega_{5i}=0$, for i=1, 2, 3	59.59^{***}
$NB_{I,t}$	$\gamma_{6i}=0$, for i=1, 2, 3	36.44^{***}
	$\delta_{6i}=0$, for i=1, 2, 3	25.36^{***}
	$\eta_{6i}=0$, for i=1, 2, 3	15.01^{***}
	$\mu_{6i}=0$, for i=1, 2, 3	35.49^{***}
	$\pi_{6i}=0$, for i=1, 2, 3	1.20
	$\gamma_{6i}=\delta_{5i}=\eta_{5i}=\mu_{5i}=\pi_{5i}=0$, for i=1, 2, 3	114.93^{***}

Table A8: Granger test II

This table presents the Granger test of VAR model in Table A5.

Dependent variable	Null hypothesis	Chi-square
$R_{s,t}$	$\eta_{1i}=0$, for i=1, 2, 3	11.47***
	$\mu_{1i}=0$, for i=1, 2, 3	1.86
	$\pi_{1i}=0$, for i=1, 2, 3	0.14
	$\omega_{1i}=0$, for i=1, 2, 3	18.66***
	$\eta_{1i}=\mu_{1i}=\pi_{1i}=\omega_{1i}=0$, for i=1, 2, 3	47.65***
$NB_{FI,t}$	$\gamma_{2i}=0$, for i=1, 2, 3	8.87**
	$\mu_{2i}=0$, for i=1, 2, 3	22.07***
	$\pi_{2i}=0$, for i=1, 2, 3	4.86
	$\omega_{2i}=0$, for i=1, 2, 3	16.32***
	$\gamma_{2i}=\mu_{2i}=\pi_{2i}=\omega_{2i}=0$, for i=1, 2, 3	47.10***
$NB_{D,t}$	$\gamma_{3i}=0$, for i=1, 2, 3	145.68***
	$\eta_{3i}=0$, for i=1, 2, 3	8.05**
	$\pi_{3i}=0$, for i=1, 2, 3	2.55
	$\omega_{3i}=0$, for i=1, 2, 3	30.91***
	$\gamma_{3i}=\eta_{3i}=\pi_{3i}=\omega_{3i}=0$, for i=1, 2, 3	194.49***
$NB_{DI,t}$	$\gamma_{4i}=0$, for i=1, 2, 3	33.96***
	$\eta_{4i}=0$, for i=1, 2, 3	7.59*
	$\mu_{4i}=\delta_{3i}=0$, for i=1, 2, 3	13.92***
	$\omega_{4i}=0$, for i=1, 2, 3	6.15
	$\gamma_{4i}=\eta_{4i}=\mu_{4i}=\omega_{4i}=0$, for i=1, 2, 3	58.43***
$NB_{I,t}$	$\gamma_{5i}=0$, for i=1, 2, 3	26.32***
	$\eta_{5i}=0$, for i=1, 2, 3	14.98***
	$\mu_{5i}=0$, for i=1, 2, 3	42.30***
	$\pi_{5i}=0$, for i=1, 2, 3	1.12
	$\gamma_{5i}=\eta_{5i}=\mu_{5i}=\pi_{5i}=0$, for i=1, 2, 3	89.34***

Table A9: Granger test III

This table presents the Granger test of VAR model in Table A6.

Dependent variable	Null hypothesis	Chi-square
$R_{f,t}$	$\eta_{1i}=0$, for i=1, 2, 3	19.31***
	$\mu_{1i}=0$, for i=1, 2, 3	1.78
	$\pi_{1i}=0$, for i=1, 2, 3	4.86
	$\omega_{1i}=0$, for i=1, 2, 3	16.44***
	$\eta_{1i}=\mu_{1i}=\pi_{1i}=\omega_{1i}=0$, for i=1, 2, 3	59.83***
$NB_{FI,t}$	$\delta_{2i}=0$, for i=1, 2, 3	5.78
	$\mu_{2i}=0$, for i=1, 2, 3	11.61***
	$\pi_{2i}=0$, for i=1, 2, 3	5.05
	$\omega_{2i}=0$, for i=1, 2, 3	17.45***
	$\delta_{2i}=\mu_{2i}=\pi_{2i}=\omega_{2i}=0$, for i=1, 2, 3	43.99***
$NB_{D,t}$	$\delta_{3i}=0$, for i=1, 2, 3	289.20***
	$\eta_{3i}=0$, for i=1, 2, 3	9.53**
	$\pi_{3i}=0$, for i=1, 2, 3	3.84
	$\omega_{3i}=0$, for i=1, 2, 3	29.84***
	$\delta_{3i}=\eta_{3i}=\pi_{3i}=\omega_{3i}=0$, for i=1, 2, 3	338.72***
$NB_{DI,t}$	$\delta_{4i}=0$, for i=1, 2, 3	27.04***
	$\eta_{4i}=0$, for i=1, 2, 3	7.77*
	$\mu_{4i}=\delta_{3i}=0$, for i=1, 2, 3	16.04***
	$\omega_{4i}=0$, for i=1, 2, 3	5.62
	$\delta_{4i}=\eta_{4i}=\mu_{4i}=\omega_{4i}=0$, for i=1, 2, 3	51.49***
$NB_{I,t}$	$\delta_{5i}=0$, for i=1, 2, 3	15.26***
	$\eta_{5i}=0$, for i=1, 2, 3	16.43***
	$\mu_{5i}=0$, for i=1, 2, 3	29.36***
	$\pi_{5i}=0$, for i=1, 2, 3	1.54
	$\delta_{5i}=\eta_{5i}=\mu_{5i}=\pi_{5i}=0$, for i=1, 2, 3	78.20***

Chapter IV

Conclusions

In this dissertation I investigate the effects of market friction on the asset prices. Two topics are included in this dissertation. In the first topic I study the impact of communication barriers among potential lenders on the contract terms in the U.S. syndicated loan market. In the second topic I examine market microstructure effects on lead-lag relationships between Taiwan futures returns, futures volume, and spot returns.

In the first topic, two cases are used to model lenders' interactions. My purpose is to explore how informational segmentation among lenders would result in cascade effect, which in turn affects the loan spreads and other contract terms. I also empirically test the model's predictions.

The first case is the benchmark case where it assumes all potential lenders can freely share their information with each other. The second one is the cascade case in which it assumes each potential lender can only observe the decisions of her predecessors. If potential lenders only have imperfect signals, the actions of their predecessors are important information for evaluating a loan. The model shows that if lead bank is rational and risk-neutral, the probability of syndication failures is always positive under the benchmark case but is zero under the cascade case. This results in lower ex-ante financing cost under the cascade case. However, the ex-post interest rate will be higher under this case. The intuition is that in the cascade case the lead bank will increase the interest rate to elicit a positive cascade and make failure become impossible.

To empirically test the models' predictions, physical distance and relational distance are used to proxy for segmentation of communication among lenders. I use average path length and clustering coefficient, which are both from the analysis of syndication networks, to gauge the relational distance. The larger distance represents the communication or information is more segmented among lenders. The results show that the physical distance doest not support the predictions, but the relational distance does. I argue that the physical distance is not a good proxy due to innovations in technology, transportation, and communication. The relational distance is a good proxy because the influence of network structures on information dissemination and transmission is well-known.

In the second topic, I explore the intraday information-based relation between the futures market and the spot market. I not only observe the lead-lag relationship between the TX returns, TX trading activity, and TAIEX returns in the overall market, but also categorize the traders in the futures market into foreign institutional traders, futures dealers, domestic institutional traders, and individual traders, and conduct a deep research of whether there exists different information content among four types of traders for the spot and the futures markets.

In the overall market, I discover that futures market leads the spot market. This leading relationship, however, is only reflected in the price movements, but not in the overall trading activity of the futures contracts. That is, the TX returns one-way lead the TAIEX returns, but the overall TX net open buy lags the TAIEX returns. If the types of the trading source are not distinguished, there is no information leading effect in both futures price and spot price only through observing the overall TX net open buy.

When the futures traders are partitioned into four categories, I find that the TX returns still lead the TAIEX returns. The leading relationship, however, is no longer a one-way relationship, but an asymmetric feedback relationship, that is, the TX returns strongly lead the TAIEX return, and the TAIEX returns weakly lead the TX returns.

In addition, in the four types of the futures traders, I find that the foreign institutional traders' trading activity has predictive power for both the TX returns and the TAIEX returns, The net open buy of the foreign institutional traders has a one-way leading relationship with the TX returns, but appears two-way feedback relationship with the TAIEX returns. There only exists information lagged effect and no leading effect from the trading activity of the futures dealers, domestic institutional traders, and the individual traders to both the futures and the spot returns.

The results suggest that informed traders do choose to trade in the futures market, which makes the information flow from the futures market to the spot market. In particular, the foreign institutional traders more tend to be the informed traders in the futures market. Due to the market share of the foreign institutional traders is low in the futures market (7.76%), even though the informed trades in the futures market can easily be seen from the leading relation of the futures price changes, it can hardly be seen from the overall market trading activity. This implies that using the overall market's trading volume as the analysis basis may result in an inaccurate result.

Reference

Acharya, V. V. and Yorulmazer, T., 2008. Information Contagion and Bank Herding. Journal of Money, Credit and Banking 40, 215-231.

Altman, E. I., 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance 23, 489-609.

Anthony, J. H., 1988. The Interrelation of Stock and Option Market Trading-Volume Data. Journal of Finance 43, 949-964.

Banerjee, A. V., 1992. A Simple Model of Herd Behavior. Quarterly Journal of Economics 107, 796-817.

Baum, J. A. C., Rowley, T. J., and Shipilov, A. V., 2004. The Small World of Canadian Capital Markets: Statistical Mechanics of Investment Bank syndicate Networks, 1952-1989. Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de I'Administration 21, 307-325.

Bosch, O. and Steffen, S., 2011. On Syndicate Composition, Corporate Structure and the Certification Effect of Credit Ratings. Journal of Banking and Finance 35, 290-299.

Cai, J., Saunders, A., and Steffen, S., 2010. Diversification or Specialization? An Analysis of distance and Collaboration in Loan Syndication Networks. Working paper, University of Mannheim.

Carey, M. and Nini, G., 2007. Is the Corporate Loan Market Globally Integrated? A Pricing Puzzle. Journal of Finance 62, 2969-3007.

Chakravarty, S., 2001. Stealth-trading: Which Traders' Trades Move Stock Prices? Journal of Financial Economics 61, 29-307.

Champagne, C. and Kryzanowski, L., 2007. Are Current Syndicated Loan Alliances Related to Past Alliances? Journal of Banking and Finance 31, 3145-3161.

Chan, K., 1992. A Further Analysis of the Lead-Lag Relationship between the Cash Market and Stock Index Futures Market. Review of Financial Studies 5, 123-152.

Chan, K., Chung, Y. P., and Fong, W. M., 2002. The Informational Role of Stock and Option Volume. Review of Financial Studies 15, 1049-1075.

Chan, K. C., Chang, Y. C., and Lung, P. P., 2009. Informed Trading under Different Market Conditions and Moneyness: Evidence from TXO Options. Pacific-Basin Finance Journal 17, 189-208.

Chang, C. C., Hsieh, P. F., and Lai, H. N., 2009. Do Informed Option Investors Predict Stock Returns? Evidence from the Taiwan Stock Exchange. Journal of Banking and Finance 33, 757-764.

Chen, C. R., Lung, P. P., and Tay, N. S. P., 2005. Information Flow between the Stock and Option Markets: Where do Informed Traders Trade? Review of Financial Economics 14, 1-23.

Chen, L. W., Johnson, S. A., Lin, J. C., and Liu, Y. J., 2009. Information, Sophistication, and Foreign versus Domestic Investors' Performance. Journal of Banking and Finance 33, 1636-1651.

Chou, R. K. and Wang,Y.-Y., 2009. Strategic Order Splitting, Order Choice, and Aggressiveness: Evidence from the Taiwan Futures Exchange. Journal of Futures Markets 29, 1102-1129.

Chu, Q. C., Hsieh, W. G., and Tse, Y., 1999. Price Discovery on the S&P 500 Index Markets: An Analysis of Spot Index, Index Futures, and SPDRs. International Review of Financial Analysis 8, 21-34.

Cohen, L., Frazzini, A., and Malloy, C., 2008. The Small World of Investing: Board Connections and Mutual Fund Returns. Journal of Political Economy 116, 951-979.

D'Arcy, S. P. and Oh, P., 1997. The Cascade Effect in Insurance Pricing. Journal of Risk and Insurance 64, 465-480.

De Jong, F. and Donders, M. W. M., 1998. Intraday Lead-Lag Relationships between the Futures, Options and Stock Market. European Finance Review 1, 337-359.

De Jong, F. and Nijman, T., 1997. High Frequency Analysis of Lead-Lag Relationships between Financial Markets. Journal of Empirical Finance 4, 259-277.

Dennis, S. A. and Mullineaux, D. J., 2000. Syndicated Loans. Journal of financial Intermediation 9, 404-426.

Easley, D., O'Hara, M., and Srinivas, P. S., 1998. Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. Journal of Finance 53, 431-465.

Elyasiani, E. and Goldberg, L. G., 2004. Relationship Lending: A Survey of the Literature. Journal of Economics and Business 56, 315-330.

Esty, B. C. and Megginson, W. L., 2003. Creditor Rights, Enforcement, and Debt Ownership Structure: Evidence from the global Syndicated Loan Market. Journal of Financial and Quantitative Analysis 38, 37-59.

Fahlenbrach, R. and Sandas, P., 2003. Bid-Ask Spreads and Inventory Risk: Evidence from the FTSE-100 Index Options Market. Working paper, University of Pennsylvania.

Focarelli, D., Pozzolo, A. F., and Casolaro, L., 2008. The Pricing Effect of Certification on Syndicated Loans. Journal of Monetary Economics 55, 335-349.

French, K. R., 1986. Detecting Spot Price Forecasts in Futures Prices, Journal of Business 59, 39-54.

Giannetti, M. and Yafeh, Y., 2012. Do Cultural Differences Between Contracting Parties Matter? Evidence from syndicated Bank Loans. Management Science 58, 365-383.

Godlewski, C. J., Sanditov, B., and Burger-Helmchen, T., 2012. Bank Lending Networks, Experience, Reputation, and Borrowing Costs. Journal of Business Finance & Accounting 39, 113-140.

Godlewski, C. J. and Weill, L., 2011. Does Collateral Help Mitigate Advere Selection? A Cross-Country Analysis. Journal of Financial Services Research 40, 49-78.

Gottesman, A. A. and Roberts , G. S., 2004. Maturity and Corporate Loan Pricing. Financial Review 39, 55-77.

Graham, J. R., Li, S., and Qiu, J., 2008. Corporate Misreporting and Bank Loan Contracting. Journal of Financial Economics 89, 44-61.

Green, W. H., 2008. Econometric Analysis. Prentice Hall, New Jersey.

Han, B., Lee, Y. T., and Liu, Y. J., 2009. Investor Trading Behavior and Performances: Evidence from Taiwan Stock Index Options, Working Paper, University of Texas at Austin.

Haselmann, R. and Wachtel, P., 2011. Foreign Banks in Syndicated loan Markets. Journal of Banking and Finance, forthcoming.

Huang, R. D. and Shiu, C. Y., 2009. Local Effects of Foreign Ownership in an Emerging Financial Market: Evidence from Qualified Foreign Institutional Investors in Taiwan. Financial Management 38, 567-602.

Hausman, J. A., 1978. Specification Tests in Econometrics. Econometrica 46, 1251–1271.

Hochberg, Y. V., Ljungqvist, A., and Lu, Y., 2007. Whom You Know Matters: Venture Capital Networks and Investment Performance. Journal of Finance 62, 251-301.

Hochberg, Y. V., Ljungqvist, A., and Lu, Y., 2010. Networking as a Barrier to Entry and the Competitive Supply of Venture Capital. Journal of Finance 65, 829-859.

In, F. and kim, S., 2006. The Hedge Ratio and the Empirical Relationship between the Stock and Futures Markets: A new Approach Using Wavelet Analysis. Journal of Business 79, 799-820.

Ivashina, V., 2009. Asymmetric Information Effects on Loan Spreads. Journal of Financial Economics 92, 300-319.

Ivashina, V. and Kovner, A., 2011. The Private Equity Advantage: Leveraged Buyout Firms and Relationship Banking. Review of Financial Studies 24, 2462-2498.

Ivashina, V. and Sun, Z., 2010. Institutional Demand Pressure and the Cost of Corporate Loans. Journal of Financial Economics 99, 500-522.

Jain, A. K. and Gupta, S., 1987. Some Evidence on "Herding" Behavior of U.S. Banks. Journal of Money, Credit, and Banking 19, 78-89.

Kawaller, I. G., Koch, P. D., and Koch, T. W., 1987. The Temporal Price Relationship between S&P 500 Futures and the S&P 500 Index. Journal of Finance 42, 1309-1329.

Kawaller, I. G., P. D. Koch, and T. W. Koch, 1993. Intraday Market Behavior and the Extent of Feedback between S&P 500 Futures Prices and the S&P 500 Index, The Journal of Financial Research 16, 107-121.

Lee, S. W. and Mullineaux, D. J., 2004. Monitoring, Financial Distress, and the Structure of Commercial Lending Syndicates. Financial Management 33, 107-130.

Lin, C. H., Hsu, H. N., Li, H. C., 2008. The Dynamics of Major Type of Traders in Taiwan Futures Market. Journal of Financial Studies 16, 149-172.

Lin, J. C., Lee, Y. T., and Liu, Y. J., 2007. IPO Auctions and Private Information. Journal of Banking and Finance 31, 1483-1500.

Mian, A., 2006. Distance Constraints: The Limits of foreign Lending in Poor Economies. Journal of Finance 61, 1465-1505.

Nakagawa, R. and Uchida, H., 2003. Herd Behavior in the Japanese Loan Market: Evidence from Semi-Macro Data. Working paper, Hiroshima University of Economics.

Nandy, D. and Shao, P., 2010. Institutional Investment in Syndicated Loans. Working paper, York University.

Newman, M. E. J., 2003. The Structure and Function of Complex Networks. SIAM Review 45, 167-256.

Nini, G., 2008. How Non-Banks Increased the Supply of Bank Loans: Evidence from Institutional Term Loans. Working paper, University of Pennsylvania.

Nini, G., Smith, D. C., and Sufi, A., 2011. Creditor Control Rights, Corporate Governance, and Firm Value. Working paper, University of Pennsylvania.

Pan, J. and Poteshman, A. M., 2006. The Information in Option Volume for Future Stock Prices. Review of Financial Studies 19, 871-908.

Pati, P. C. and Rajib, P., 2011. Intraday Return Dynamics and Volatility Spillovers between NSE S&P CNX Nifty Stock Index and Stock Index Futures. Applied Economics Letters 18, 567-574.

Roll, R., Schwartz, E. and Subrahmanyam, A., 2011. Volume in Redundant Assets. Working paper, University of California at Los Angeles.

Sadath, A. and Kamaiah, B., 2009. Evidence of Informed Trading in Single Stock Futures Market in the National Stock Exchange of India Ltd. Indian Journal of Economics and Business 8, 387-394.

Shiller, R. J., 1995. Conversation, Information, and Herd Behavior. American Economic Review 85, 181-185.

Sias, R. W. and Starks, L. T., 1997. Return Autocorrelation and Institutional Investors. Journal of Financial Economics 46, 103-131.

Staiger, D. and Stock, J. H., 1997. Instrumental Variables Regression with Weak Instruments. Econometrica 65, 557–586.

Stephan, J. A. and Whaley, R. E., 1990. Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets. Journal of Finance 45, 191-220.

Stoll, H. R. and Whaley, R. E., 1990. The Dynamics of Stock Index and Stock Index Futures Returns. Journal of Financial and Quantitative Analysis 25, 441-468.

Sufi, A., 2007. Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans. Journal of Finance 62, 629-668.

Sufi, A., 2009. The Real Effects of Debt Certification: Evidence from the Introduction of Bank Loan Ratings. Review of Financial Studies 22, 1659-1691.

Uchida, H. and Nakagawa, R., 2007. Herd Behavior in the Japanese Loan Market: Evidence from Bank Panel Data. Journal of Financial Intermediation 16, 555-583.

Wang, C., 2004. Futures Trading Activity and Predictable Foreign Exchange Market Movements. Journal of Banking and Finance 28, 1023-1041.

Welch, I., 1992. Sequential Sales, Learning, and Cascades. Journal of Finance 47, 695-732.

Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Massachusetts.

Yang, J., Yang, Z., and Zhou, Y., 2012. Intraday Price Discovery and Volatility Transmission in Stock Index and Stock Index Futures Markets: Evidence from China. Journal of Futures Markets 32, 99-121.