# Essays on Development Economics: Consumers, Firms, and Financial Institutions

# Essays on Development Economics: Consumers, Firms, and Financial Institutions

A dissertation presented

by

Zhaoning Wang

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2016

*Dissertation Advisor:*
**Professor Michael Kremer**

*Author:*
**Zhaoning Wang**

**Essays on Development Economics:
Consumers, Firms, and Financial Institutions**

# Abstract

This dissertation presents three chapters addressing issues pertaining to consumers, firms, and financial institutions in the developing world. The first chapter, co-authored with Juan Ma and Tarun Khanna, evaluates the effect of voluntary information disclosure in incentivizing firms to deliver high quality in the absence of regulation. We present evidence from two field experiments in China's infant milk powder industry, which is undergoing a serious consumer trust crisis after several safety scandals. Contrary to common beliefs, our results suggest that providing certain positive quality-related information has a significantly negative impact on consumers' purchase decisions and impression of the industry. We explain our findings via the existence of a "reminder effect," where information disclosure triggers recall of and diverts attention to health and safety risks related to certain products. In the second chapter, I study the impact of firms' industrial diversification on their innovation outcomes in China, which has recently been an important topic due to various government initiatives. By exploiting a policy launched by the Chinese government since 2011, I estimate the impact of related and unrelated diversification on corporate innovations measured by the number of patents received after 2012. Overall, I find opposite results for production-oriented and service-oriented firms, with related diversification more effective for the former and unrelated diversification more effective for the latter. One explanation, which is supported by subsequent analysis, is related to the different degrees of transferability of the technology and knowledge required for innovations. In the third chapter, I examine the complementarity of formal and informal finance represented respectively by banks and bidding ROSCAs (Rotating Savings and Credit Associations) in India, a critical source of credit for many impoverished individuals, households, and small businesses. Different from some traditional views, my results suggest that formal financial institutions can benefit the informal ones instead of replacing them. Using auction data from ROSCAs and operational data from banks in Andhra Pradesh, India from 1998 to 2000, I find that the emergence of formal finance, measured by nearby bank openings, increases ROSCA participation, reduces the cost of capital for ROSCA participants, and lowers the amount of ROSCA default.

# Contents

Abstract . . . iii
Acknowledgments . . . xii

**Introduction** **1**

**1 Negative Effects of Voluntary Information Disclosure: Evidence from Two Field Experiments in China's Infant Milk Powder Industry** **4**

1.1 Introduction . . . 4
1.2 Background and Previous Literature . . . 9
1.2.1 The Melamine Scandal . . . 9
1.2.2 Key Investments in Milk Safety . . . 10
1.2.3 Empirical Setting . . . 12
1.2.4 Review of Relevant Theories . . . 13
1.3 Theoretical Framework . . . 16
1.3.1 Basic Setup . . . 16
1.3.2 Separating Equilibrium . . . 19
1.4 Online Study . . . 25
1.4.1 Experiment Design . . . 25
1.4.2 Empirical Methodology . . . 28
1.4.3 Results and Discussions . . . 29
1.5 In-Store Study . . . 36
1.5.1 Experiment Design . . . 36
1.5.2 Empirical Methodology . . . 41
1.5.3 Results and Discussions . . . 43
1.6 Conclusion . . . 54

**2 Industrial Diversification and Corporate Innovations: Evidence from Firms in China** **58**

2.1 Introduction . . . 58

2.2 Background and Previous Literature . . . . . . . . . . . . . . . . . . . . . 64
2.2.1 Government Promotion of Industrial Diversification . . . . . . . . . . 64
2.2.2 Previous Literature on Diversification and Innovation . . . . . . . . . 65
2.2.3 Related Diversification and Unrelated Diversification . . . . . . . . . 67
2.2.4 Diversification and Performance of Chinese Firms . . . . . . . . . . . 68
2.3 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69
2.4 Empirical Methodology . . . . . . . . . . . . . . . . . . . . . . . . . . 70
2.5 Results and Discussions . . . . . . . . . . . . . . . . . . . . . . . . . . 72
2.5.1 All Firms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
2.5.2 Actively Innovating Firms . . . . . . . . . . . . . . . . . . . . . . . 84
2.5.3 Further Discussions . . . . . . . . . . . . . . . . . . . . . . . . . . 90
2.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96

**3 Complementarity of Formal and Informal Finance: Evidence from Bidding ROSCAs in India 100**
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100
3.2 Background and Previous Literature . . . . . . . . . . . . . . . . . . . . 106
3.3 Theoretical Framework . . . . . . . . . . . . . . . . . . . . . . . . . . 108
3.3.1 A Simple Example with Three Participants . . . . . . . . . . . . . . 109
3.3.2 Model for Bidding ROSCAs without Banks . . . . . . . . . . . . . . 110
3.3.3 Model for Bidding ROSCAs with Banks . . . . . . . . . . . . . . . 113
3.4 Data Source . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
3.4.1 Geographical Information of ROSCAs and Bank Branches . . . . . . 117
3.4.2 Dataset on Bidding ROSCAs in Andhra Pradesh . . . . . . . . . . . 121
3.4.3 Limitations of the Existing Data . . . . . . . . . . . . . . . . . . . 122
3.5 Empirical Methodology . . . . . . . . . . . . . . . . . . . . . . . . . . 127
3.5.1 ROSCA Participation . . . . . . . . . . . . . . . . . . . . . . . . . 127
3.5.2 ROSCA Winning Bids . . . . . . . . . . . . . . . . . . . . . . . . . 128
3.5.3 ROSCA Default . . . . . . . . . . . . . . . . . . . . . . . . . . . . 129
3.6 Results and Discussions . . . . . . . . . . . . . . . . . . . . . . . . . . 129
3.6.1 Effect of Bank Openings on ROSCA Participation . . . . . . . . . . . 130
3.6.2 Effect of Bank Openings on Winning Bids . . . . . . . . . . . . . . 135
3.6.3 Effect of Bank Openings on ROSCA Default . . . . . . . . . . . . . 139
3.7 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143

**References 146**

**Appendix A Supplement to Chapter 1** **156**
A.1 Overview of ISO9001 Certification . . . . . . . . . . . . . . . . . . . . . . . . . . 156
A.2 Card Designs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 159
A.3 Supplementary Tables for In-Store Study . . . . . . . . . . . . . . . . . . . . . . 160
A.3.1 Alternative Regression Results using OLS . . . . . . . . . . . . . . . . . 160
A.3.2 Comparison between High-Income Districts and Low-Income Districts 160

**Appendix B Supplement to Chapter 2** **168**
B.1 Supplementary Tables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 168

**Appendix C Supplement to Chapter 3** **170**
C.1 A Detailed Description of ROSCAs . . . . . . . . . . . . . . . . . . . . . . . . . 170
C.1.1 Importance and Prevalence of Chit Funds . . . . . . . . . . . . . . . . . 170
C.1.2 Characteristics of ROSCA Participants . . . . . . . . . . . . . . . . . . . 171
C.1.3 Bidding Mechanisms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 172
C.1.4 Usage of ROSCA Winnings . . . . . . . . . . . . . . . . . . . . . . . . . . 173
C.1.5 Regulation and Registration of Chit Funds . . . . . . . . . . . . . . . . . 174
C.1.6 Possible Failures of Bidding ROSCAs . . . . . . . . . . . . . . . . . . . . 175
C.2 Proofs of Propositions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 176
C.2.1 Proof of Proposition 3.1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 176
C.2.2 Proof of Proposition 3.2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 177
C.2.3 Proof of Proposition 3.3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 177
C.2.4 Proof of Proposition 3.4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 177
C.2.5 Proof of Proposition 3.5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 177
C.2.6 Proof of Proposition 3.6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 178
C.2.7 Proof of Proposition 3.7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 178
C.3 Detailed Dataset Creation Process . . . . . . . . . . . . . . . . . . . . . . . . . . 178
C.3.1 Creating the Dataset of Nearby Banks . . . . . . . . . . . . . . . . . . . 178
C.3.2 Creating the Dataset of ROSCA Branches . . . . . . . . . . . . . . . . . 179

# List of Tables


1.1 Summary Statistics for Online Study: *All Visitors* . . . . . . . . . . . . . . . . 30

1.2 Summary Statistics for Online Study: *Valid Survey Respondents Only* . . . . 31

1.3 Treatment Effect on Online Purchase Decisions . . . . . . . . . . . . . . . . . 32

1.4 Treatment Effect on Consumers' Reported Impression of the Milk Powder Industry (OLS) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

1.5 Treatment Effect on Consumers' Reported Impression of the Milk Powder Industry (Logit) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

1.6 Summary Statistics for In-Store Study: *During the Two-Week Pre-Experiment Period* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44

1.7 Summary Statistics for In-Store Study: *During the Two-Week Experiment Period* 45

1.8 Summary Statistics for In-Store Study: *During the Two-Week Post-Experiment Period* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46

1.9 Treatment Effect on Beingmate Products: *For Period 0 vs. Period 1* . . . . . 47

1.10 Treatment Effect on Beingmate Products: *For Period 1 vs. Period 2* . . . . . 51

1.11 Treatment Effect on Beingmate Products: *For Period 0 vs. Period 2* . . . . . 53

2.1 Summary Statistics for All Firms . . . . . . . . . . . . . . . . . . . . . . . . 73

2.2 Summary Statistics for Actively Innovating Firms Since 2012 . . . . . . . . . 74

2.3 Effect of Diversity on Corporate Innovations in the Production Sector: *Diversity Measure Generated from 4-Digit Industry Codes* . . . . . . . . . . . . . . . . 78

2.4 Effect of Diversity on Corporate Innovations in the Production Sector: *Diversity Measure Generated from 2-Digit Industry Codes* . . . . . . . . . . . . . . . . 79

2.5 Effect of Diversity on Corporate Innovations in the Service Sector: *Diversity Measure Generated from 4-Digit Industry Codes* . . . . . . . . . . . . . . . . 82

2.6 Effect of Diversity on Corporate Innovations in the Service Sector: *Diversity Measure Generated from 2-Digit Industry Codes* . . . . . . . . . . . . . . . . 83

2.7 Effect of Diversity on Innovations among Actively Innovating Firms in the Production Sector: *Diversity Measure Generated from 4-Digit Industry Codes* 85

2.8 Effect of Diversity on Innovations among Actively Innovating Firms in the Production Sector: *Diversity Measure Generated from 2-Digit Industry Codes* 86
2.9 Effect of Diversity on Innovations among Actively Innovating Firms in the Service Sector: *Diversity Measure Generated from 4-Digit Industry Codes* . . 88
2.10 Effect of Diversity on Innovations among Actively Innovating Firms in the Service Sector: *Diversity Measure Generated from 2-Digit Industry Codes* . . 89
2.11 Comparison between High-Technology and Low-Technology Firms in the Production Sector: *Diversity Measure Generated from 4-Digit Industry Codes* 92
2.12 Comparison between High-Technology and Low-Technology Firms in the Production Sector: *Diversity Measure Generated from 2-Digit Industry Codes* 93
2.13 Comparison between More and Less Knowledge-Intensive Firms in the Service Sector: *Diversity Measure Generated from 4-Digit Industry Codes* . . . . . . 94
2.14 Comparison between More and Less Knowledge-Intensive Firms in the Service Sector: *Diversity Measure Generated from 2-Digit Industry Codes* . . . . . . 95

3.1 Payoffs for Each ROSCA Participant in the Example . . . . . . . . . . . . . 109
3.2 Payoffs for Each ROSCA Participant with No Banks . . . . . . . . . . . . . 112
3.3 Payoffs for Each ROSCA Participant with Banks . . . . . . . . . . . . . . . 114
3.4 ROSCA Sustainability with the Introduction of Banks . . . . . . . . . . . 116
3.5 Summary Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 122
3.6 Distribution of the Number of Auctions by Month . . . . . . . . . . . . . . 123
3.7 Impact of Bank Emergence on ROSCA Participation . . . . . . . . . . . . . 131
3.8 Impact of Different Types of Bank Emergence on ROSCA Participation . . . 133
3.9 Impact of Bank Emergence on ROSCA Winning Bids . . . . . . . . . . . . 136
3.10 Impact of Different Types of Bank Emergence on ROSCA Winning Bids . . 138
3.11 Impact of Bank Emergence on ROSCA Default . . . . . . . . . . . . . . . . 140
3.12 Impact of Different Types of Bank Emergence on ROSCA Default . . . . . . 142

A.1 Number of Firms with ISO9001 Certification by Region . . . . . . . . . . . 157
A.2 Treatment Effect on Beingmate Products: *For Period 0 vs. Period 1* . . . . . 161
A.3 Treatment Effect on Beingmate Products: *For Period 1 vs. Period 2* . . . . . 162
A.4 Treatment Effect on Beingmate Products: *For Period 0 vs. Period 2* . . . . . 163
A.5 Comparison between High-Income Districts and Low-Income Districts: *For Period 0 vs. Period 1* . . . . . . . . . . . . . . . . . . . . . . . . . . . . 164
A.6 Comparison between High-Income Districts and Low-Income Districts: *For Period 1 vs. Period 2* . . . . . . . . . . . . . . . . . . . . . . . . . . . . 165

A.7 Comparison between High-Income Districts and Low-Income Districts: *For Period 0 vs. Period 2* . . . . . 166
A.8 Hypothesis Testing Results between High-Income Districts and Low-Income Districts . . . . . 167

B.1 Hypothesis Testing Results between Different Diversity Measures . . . . . 168
B.2 Hypothesis Testing Results between the Production Sector and the Service Sector . . . . . 169
B.3 Hypothesis Testing Results between All Firms and Actively Innovating Firms 169

# List of Figures


1.1 Weekly Average Quantity of Sales: *For All Beingmate Products* . . . . . . . 38
1.2 Weekly Average Quantity of Sales: *For Only Milk Powder Products by Beingmate* 39
1.3 Weekly Average Quantity of Sales: *For All Other Products by Beingmate* . . 40

2.1 Distribution of Average Number of Patents by Sector . . . . . . . . . . . . . 75
2.2 Distribution of Average Diversity Measures by Sector . . . . . . . . . . . . . 76

3.1 Importance of Various Sources of Financing at Startup Stage . . . . . . . . . 102
3.2 Ease of Obtaining Financing during Growth Stage . . . . . . . . . . . . . . . 103
3.3 Bidding ROSCAs and Banks in Andhra Pradesh, India . . . . . . . . . . . . 118
3.4 Bidding ROSCAs and Banks with Buffer Circles (10 km) . . . . . . . . . . . 119
3.5 Average Number of Neighboring Banks by Month . . . . . . . . . . . . . . . 120
3.6 Average Number of ROSCA Participants by Month . . . . . . . . . . . . . . 124
3.7 Distribution of ROSCA Winning Bids . . . . . . . . . . . . . . . . . . . . . 125
3.8 Distribution of ROSCA Default . . . . . . . . . . . . . . . . . . . . . . . . 126

A.1 Number of Firms with ISO9001 Certification in China by Industry . . . . . . 158
A.2 Card Designs for Treatment Groups . . . . . . . . . . . . . . . . . . . . . . 159

# Acknowledgments

I would like to express my sincere gratitude to my advisors Oliver Hart, Michael Kremer, and Nathan Nunn for their guidance, motivation, and support throughout my graduate studies at Harvard University. Their insightful suggestions and thought-provoking questions were essential to the completion of this dissertation. Having worked with them at various times as a graduate student, research assistant, teaching fellow, and thesis writer, I have learned a tremendous amount from these distinguished and dedicated scholars.

Besides my advisors, I have also gained inspiration from interactions with many of my colleagues and other professors. In particular, I would also like to thank Mitra Akhtari, Shawn Cole, Rafael Di Tella, Ray Fisman, Yasheng Huang, Simon Johnson, Tarun Khanna, Divya Kirti, Stefan Klonner, Sara Lowes, Juan Ma, Jeffrey Picel, Ashok Rai, Gautam Rao, Martin Rotemberg, Andrei Shleifer, Daniel Trefler, Andrew Weiss, Jack Willis, Wentao Xiong, Dennis Yao, and Juanjuan Zhang, as well as seminar participants of the Development Lunch, the Development Tea, the Economic Development Workshop, the Contracts & Organizations Lunch, and the Strategy & Entrepreneurship Seminar at INSEAD Singapore for their helpful comments and suggestions.

In addition, I thank the Department of Economics, the Graduate School of Arts and Sciences, and Harvard University for their generous financial support through the Cheung Yan Family Fund, the Danielian/Kuznets Travel & Research Grant, the Douglas Dillon Fellowship, the Dissertation Completion Fellowship, the Hirtle Callaghan Research Grant, and the Weiss Family Program Fund.

Last but not least, this journey would not have been possible without the unconditional love and support from my parents and grandparents, who always encourage me to follow my dreams and work hard in all pursuits of life.

I welcome any comment on this dissertation. All errors are my own.

To my parents

# Introduction

Consumers, firms, and financial institutions are important components in modern economies. In recent decades, there has been an increasing amount of interest in these topics in the developing world, especially in the fastest growing countries such as China and India, where the socioeconomic structures are distinctively different from those in the developed world. In particular, there are existing observations or conclusions derived mostly from developed countries that may not apply to those in the developing world due to special characteristics of their populations, enterprises, and government functions. There are also important industry- or region-specific questions that have yet to be answered using data from various developing countries. Moreover, there are hypotheses or common beliefs about the developing world that lack solid empirical evidence. In this regard, this dissertation presents three chapters, each of which addresses a question in development economics using data from China or India. The findings also have policy implications for other similarly situated developing countries or even beyond the developing world.

The first chapter is joint work with Juan Ma and Tarun Khanna, where we study the negative effects of producers' voluntary information disclosure on consumer choices and market behaviors. We provide evidence from two field experiments in China's infant milk powder industry, which faces tremendous challenge in regaining consumer trust after several safety scandals over the past few years. Our results suggest that advertising high quality may backfire and that voluntary information disclosure may lead to inefficient market outcomes. In both experiments, we find that providing certain quality-related information, especially on the existence of a product traceability system and international quality certification, has a significantly *negative* impact on consumers' purchase decisions. A survey among individual consumers further indicates that voluntary information disclosure has a negative effect on consumers' impression of the entire industry. One explanation is that consumers become suspicious when provided with quality information and start questioning the intention of the producer, especially in light of the ongoing consumer trust crisis. We refer to this as a "reminder effect," where the disclosure of quality-related information reminds consumers of past scandals and diverts their attention to potential health and safety risks. Our findings also shed light on why many countries regulate the safety and quality of health-related products despite the fact that companies could potentially advertise simply based on observable and verifiable quality information.

In the second chapter, I study the impact of firms' industrial diversification structure on their innovation outcomes in China, which has been an important topic in recent years due to various government initiatives. Since early 2011, the central government of China has been encouraging domestic firms to engage in industrial and geographical diversification, which has been widely enforced by local governments and authorities. Using this new initiative as an exogenous variation, I examine the impact of diversification on corporate innovations measured by the number of new patents received after 2012. Specifically, I define two different diversity measures to indicate the relative levels of related and unrelated diversification, which are two different diversification strategies pursued by different firms. I also classify the firms as either production-oriented or service-oriented and conduct separate analyses for comparison purposes. In line with most existing work, I find that related diversification, or at least a combination of related and unrelated diversification, is key to generating new innovations among production-oriented firms. However, for service-oriented firms, my empirical results suggest that unrelated diversification may be more effective than related diversification in terms of inducing corporate innovations, which has rarely been mentioned or addressed in the literature. One explanation, which is supported by my subsequent analysis, is related to the transferability of the technology and knowledge required for innovations. Specifically, my results for the production sector are mainly driven by those high-technology firms for which related diversification is more realistic and efficient in using existing resources. In contrast, my results for the service sector are primarily driven by those less knowledge-intensive firms for which unrelated diversification is not only feasible but also broadens the scope of firms' innovation activities.

In the third chapter, I study the relationship between formal and informal financial institutions represented respectively by banks and bidding ROSCAs (Rotating Savings and Credit Associations) in India, a critical source of credit for many impoverished individuals, households, and small businesses. Some traditional views treat formal and informal finance as imperfect substitutes, where previously credit-constrained entities replace their informal credit channels by formal ones as soon as they gain access to the formal credit market. However, using monthly auction data from 219 ROSCA branches and operational data from over 5,000 bank branches in Andhra Pradesh, India from 1998 to 2000, I find that formal and informal financial institutions can coexist and complement each other. My results suggest that the emergence of formal finance, measured by nearby openings of all types of credit-issuing banks, increases ROSCA participation, reduces the cost of capital or winning bids for ROSCA participants, and lowers the amount of ROSCA default. Moreover, after breaking down the banks by type, I find that different types of banks have different impacts on ROSCA participation, winning bids, and the amount of default. In particular, I find that rural banks, represented by Grameen banks in my sample, have the strongest significant impact on ROSCA participation and on the reduction of ROSCA default among all types of banks, possibly due to the strong ties between ROSCAs and rural banks in India.

# Chapter 1

# Negative Effects of Voluntary Information Disclosure: Evidence from Two Field Experiments in China's Infant Milk Powder Industry[1]

## 1.1 Introduction

The quality of health-related products is an important issue in many countries. One way to ensure high quality is for the government to establish regulations from firms' production process to the sale of their final products. However, what remains unclear is whether the market will produce similar incentives for firms to produce high-quality products in the absence of regulation, especially in underdeveloped institutions where the lack of trust is a major social problem. For instance, a possible alternative to government regulation is

[1]This chapter is co-authored with Juan Ma (Strategy Unit, Harvard Business School) and Tarun Khanna (Strategy Unit, Harvard Business School).

to allow firms to advertise based on the quality of their products. If quality information cannot be conveyed directly, firms can also choose to provide signals that consumers believe are correlated with their product quality, thus indirectly informing potential consumers of their "true type." However, as we will show in this study, signaling might not provide the right incentives for firms to deliver high quality when consumers have already developed an extremely negative prior belief about a particular market before any information is provided. In this case, it becomes important for the government to intervene and to contemplate the most appropriate public policies to incentivize high quality. Our study is conducted in the context of the infant milk powder industry in mainland China, which is undergoing a serious consumer trust crisis due to previous scandals related to toxic chemicals added to the milk powder products. Specifically, the problem is twofold: first, no entrepreneurs or incumbents themselves have devised an effective system that credibly reveals their product quality; second, the Chinese government has not taken enough regulatory measures to combat milk safety issues.

In this study, we conduct two field experiments to evaluate possible ways in which producers of high-quality products in a certain industry can differentiate themselves from others when consumers hold a strong, negative belief about the collective reputation of the industry. There is a widely received idea that in markets with asymmetric information, consumer behaviors often change when quality information is revealed through various channels such as consumer reports, online forums, quality certification programs, etc. (Grossman, 1981; Nelson, 1970; Spence, 1973). For the rest of this chapter, we will focus on possible informational strategies for producers and examine the effectiveness of these strategies. We find that providing quality-related information may actually contribute negatively to the perceived quality because it may remind consumers of the negative images associated with a specific industry, which may affect every product that belongs to the industry. In fact, our results cast doubt on the hypothesis that providing information that signals higher product quality raises consumers' willingness to buy among similar goods. In other words, high-quality producers may be unable to differentiate themselves from low-quality producers by merely providing information about their costly, verifiable, and quality-improving efforts. If voluntary information revelation fails, government interventions such as mandatory information provision or forced shutdowns of some firms may become necessary in order to resolve the market failure.

There are two types of information strategies that we will examine in this study. One type is information about firm-specific investments. In the case of the milk powder industry, an important investment is the traceability system, which allows consumers to track every step along the supply chain, from the dairy farmers who supply raw milk to the supermarkets where milk powder products are purchased. Having such a system increases the accountability of each party in the food supply chain and thus their incentive to maintain high quality standards. Given the high cost of the traceability system, it is conceivable that those who have made this expensive investment might be in a better signaling position, but there is a possibility that consumers' prior belief is so negative that they choose not to trust such systems at all. Moreover, although the traceability system is expected to contribute to high quality on a macro level, it is less clear whether it actually adds value in terms of individual consumers' well-being. In other words, the traceability system does not necessarily guarantee that every product purchased by every consumer is absolutely safe and of high quality.

Another type is information about the quality certification program offered by a third party, such as the government or an international agency. In this study, we focus on the ISO9001:2008 quality certification. ISO (International Organization for Standardization) is an independent, non-governmental membership organization and "the world's largest developer of voluntary International Standards."[2] It is designed to give world-class specifications for products, services, and systems, to ensure quality, safety, and efficiency. However, there are several issues with the certification program: some people are simply not familiar with it,

[2]"The ISO Story." ISO. Retrieved January 26, 2015, from http://www.iso.org/iso/home/about/the_iso_story.htm.

while those who have the knowledge might not trust the program because producers may still be able to circumvent the standards by adopting inappropriate procedures that are not currently part of the test. Hence, despite its worldwide recognition, it is debatable whether the ISO9001 quality certification is actually effective in terms of conveying the idea of high quality to consumers and thus stimulating sales for producers.

For both field experiments, we partnered with Beingmate Baby & Child Food Co., Ltd. (hereafter "Beingmate"), the largest domestic infant formula manufacture headquartered in Hangzhou, Zhejiang Province.[3] Its products were tested clean during the melamine scandal in 2008. Beingmate runs five infant formula manufacturers that supply products to supermarkets across China. It has a traceability system based on QR codes and it has obtained the ISO9001:2008 quality certification.

In the first experiment ("the online study"), we investigate the effect of voluntary disclosure of quality-related information on consumers' purchase decisions, measured by the likelihood of purchase and quantity of purchase. Specifically, we use Beingmate's online store as a platform to randomize information provision based on the last digit of each visitor's IP address. In particular, we remind potential consumers of Beingmate's brand name, its traceability system, and its ISO quality certification, to observe whether providing such information changes consumers' purchasing decisions. Moreover, we conduct a survey at the end of each purchase to establish a direct, causal linkage between voluntary disclosure of quality information and consumers' reported impression of the infant milk powder industry in China.

Our main findings are summarized as follows: providing information about Beingmate's product traceability system or its ISO9001 quality certification has a significantly *negative* impact on both the quantity of sales and consumers' likelihood of purchase. Moreover, we find that consumers who are provided with quality-related information are more likely to

[3]Hangzhou, with a population of 21 million (according to the 2010 census), is the capital of Zhejiang Province and one of the largest cities in Eastern China.

report a more negative impression of the domestic infant milk powder industry. Hence, our results cast doubt on the hypothesis that providing information that signals higher quality has an unambiguously *positive* impact on consumers' willingness to buy. In contrast, we find that even positive information on product quality can lead to a *negative* impact on consumers' willingness to buy, which is consistent with some behavioral theories on memory and attention but cannot be explained by the hypothesis that assumes perfectly rational consumers.

In our second experiment ("the in-store study"), we provide the same types of product information about Beingmate products in around 90 supermarkets in Hangzhou, each of which sells an average of 30 different Beingmate products. Using difference-in-difference regressions, we estimate the treatment effect of each treatment by comparing the quantity of sales before, during, and after the experiment period. We also run separate regressions for stores located in high-income districts and those in low-income district to see if the treatment effects are the same. Overall, the results confirm our findings in the online study. In addition to the negative treatment effect of quality-related information, we find that the treatment effect on high-income districts is significantly more negative than that on low-income districts during the experiment period, possibly because consumers in high-income districts have better understanding of both the previous safety scandals and the limitations of the quality-related investments.

Our results suggest that in the context of a market failure, providing more information may actually be a worse decision than providing less information due to a "reminder effect," in which information triggers consumers' recall of and diverts their attention to health and safety risks, especially in light of the past safety scandals related to milk powder products. Consumers can also become suspicious and start questioning the intention of the producer. As a result, high-quality producers are unable to differentiate themselves from low-quality producers by merely providing information about their costly, verifiable, and quality-improving efforts. This implies that at least in some markets with asymmetric information, the separating equilibrium in the classic signaling model may not be achieved. Our results also shed light on the broader notion that government regulation may be necessary when the market-based information channel alone fails. Moreover, it is worth noting that since advertising consumes real resources, even if there is a separating equilibrium, it may not be efficient in terms of the cost-effectiveness for high-quality producers.

The remainder of this chapter is organized as follows. Section 1.2 describes the background and some previous literature. Section 1.3 sets up the theoretical framework to show why a separating equilibrium may fail to exist with the introduction of the "reminder effect." Section 1.4 describes the experiment design, lays out the empirical strategy, and presents the results of our online experiment. Section 1.5 discusses the results of our in-store study. This chapter concludes with Section 1.6.

## 1.2 Background and Previous Literature

### 1.2.1 The Melamine Scandal

Domestic companies are the key players of China's dairy market. Sanlu, which used to be the fourth largest dairy company, went bankrupt after the "melamine scandal" in 2008 that claimed the lives of six infants and sickened 300,000 others. Investigations by national and local authorities found that a low-end infant milk powder product made by Sanlu contained an elevated level of melamine, a toxic chemical that could cause kidney failure in infants. Banned in food processing, melamine is widely used in production of plastics, adhesives, countertops, and fertilizers. Adding melamine into diluted milk could artificially inflate the protein levels when the milk powder is subjected to standard tests for protein content. The reason is that standard protein tests used to measure only the level of nitrogen. As melamine is extremely high in nitrogen, blending raw milk with melamine could make the milk appear to have a higher protein level. The chairwoman of Sanlu's board of directors was sentenced to life imprisonment because of the melamine scandal. Dozens of milk brands including several industry giants were later found tainted by melamine. The court later found that Sanlu provided tacit support to such misconduct, and melamine addition had been a "known secret" of the dairy industry. The dispersed nature of milk supply chain also made tracing the source of contamination and enforcement of quality regulation a daunting challenge.

After the melamine scandal, consumers became more cautious. Some started to make soybean milk at home instead of buying milk. Others opted to buy imported products only. The milk scandal boosted the market share of foreign milk brands from about 30% in 2008 to over 50% and as high as 70% in high-end infant formula segment in 2012.[4] Some consumers went online or overseas to buy infant milk powder. Hong Kong implemented an "infant milk powder quota order," a policy that banned people from taking more than two tins of infant milk powder outside Hong Kong. However, a later issue involving New Zealand's dairy giant Fonterra in 2013 raised public concern over foreign brands as well. Fonterra-admitted clostridium botulinum, another kind of toxin, was found in its whey protein that other companies purchased to produce infant formula products.

### 1.2.2 Key Investments in Milk Safety

#### 1.2.2.1 Traceability System

A generic definition for traceability is provided by the International Organization of Standards: traceability is the "ability to trace the history, application or location of an entity, by means of recorded identifications" (ISO, 1995). The Europe Union later narrowed down the definition to food industry, defining traceability as "the ability to trace and follow food or any ingredient intended to be incorporated into food through all stages of production, processing and distribution" (EU, 2002). Traceability systems have been gradually adopted by large food producers in China. This is regarded as an innovative way to secure food quality by sourcing products directly from qualified farmers and thus achieving better control of the supply

---

[4] "Restore Faith in Dairy." (2013, August 8). *China Daily*. Retrieved January 26, 2015, from `http://usa.chinadaily.com.cn/opinion/2013-08/08/content_16879410.htm`.

chain. Although many producers claim that they have a traceability system of some sort or other similar systems, the quality of such systems varies. Moreover, it is not difficult for consumers to see the differences: most producers require their consumers to go through a lengthy and sometimes broken registration process in order to trace the products, while only a few producers are able to print a unique QR (Quick Response) code on every product so that consumers can simply use their smartphones to trace the products.[5] A good traceability system can be costly. In particular, Beingmate spent more than 20 million Chinese Yuan (approximately 3.17 million USD) in building its traceability system.

#### 1.2.2.2 ISO9001 Quality Certification

ISO is made up of 163 member countries that are the national standards bodies around the world, with a Central Secretariat based in Geneva, Switzerland. By 2013, there are more than 1.13 million enterprises worldwide certified by the ISO9001 quality management system, and ISO is better known and much more widely adopted in Europe and East Asian than in North America.[6] Unlike the traceability system, ISO certifications are not specific to the food industry. In fact, ISO has published more than 19,500 International Standards covering almost every industry, from technology to food safety, agriculture, and healthcare. Within China, the top five industries for ISO are metal, electronics, machinery, trade and repairs of motor vehicles, and rubber and plastic products.[7] Obtaining an ISO9001 certification requires creating and documenting a production process in accordance with the ISO9001 standards, training the entire organization for effective implementation of the standards, and getting an ISO-accredited certification body for initial assessment, multiple rounds of audits,

---

[5] QR code, also known as two-dimensional (or 2D) code, is a special type of machine-readable barcode invented by Denso Wave Inc. See the official website http://www.qrcode.com/en/index.html for a more detailed introduction.

[6] See Table A.1 for the number of firms with ISO9001 certification by region.

[7] See Figure A.1 for the number of firms with ISO9001 certification in China by industry.

and the final registration. The cost of obtaining an ISO9001 certification varies depending on the size and complexity of the organization and on whether the enterprise already has some elements of a quality management system in place. For an enterprise with 501 to 1000 employees and a basic quality system in place, it is estimated to cost $42,750 (registration and consultancy fees) and 432 employee hours to obtain an ISO9001 certification.[8]

### 1.2.3 Empirical Setting

In implementing the field experiments, we partnered with Beingmate, the largest domestic infant milk powder producer in China. Beingmate, founded in 1992, is an infant and child foods company headquartered in Hangzhou, Zhejiang Province, China, which specializes in providing products and services to mothers and their children aged from 0 to 12. In 2010, Beingmate entered the Hong Kong market after gaining accreditation from the Hong Kong Standards and Testing Centre, and it has been listed on the Shenzhen Stock Exchange since 2011.

Beingmate operates six plants and logistics centers, 80,000 retail terminals, 30 regional subsidiaries, and more than 1,000 Beingmate-authorized stores across China. It employs 7,000 full-time staff and 20,000 maternal service consultants in authorized stores.

Beingmate's products were tested clean during and after the melamine scandal. The most rapid sales growth took place after the melamine scandal. From 2008 to 2009, Beingmate's revenue grew by more than 65%. In the infant formula segment, Beingmate is the largest domestic player and third largest player among all producers, ranked only after Wyeth and Mead Johnson and accounting for 8.5% of the total market share in 2013.[9] Beingmate describes its competitive position as "International Quality, Chinese Formula." The company's

[8]See http://the9000store.com/ISO-9000-Tips-How-Much-Does-it-cost.aspx ("How Much Does ISO 9001 Certification Cost?") for the pricing of the certification.

[9]"China's Dairies Facing Fierce Competition." (2014, May 6). *Xinhua*. Retrieved January 26, 2015, from http://news.xinhuanet.com/english/china/2014-05/06/c_133314393.htm.

research and development department employs hundreds of professional R&D staff. Compared to international players such as Wyeth and Mead Johnson, Beingmate's competitive advantage centers on formula that mimics Chinese mothers' breast milk.

### 1.2.4 Review of Relevant Theories

#### 1.2.4.1 Rational Theory: Effect of Information on Market Behavior

There has been a series of papers that analyze the incentives for firms to reveal their private information, in which the idea of unraveling equilibria is developed in depth. Beginning with the pioneering study of Akerlof (1970), economists have theorized about the incentives for producers to voluntarily reveal their private information to consumers. If consumers believe that the high-quality non-disclosing sellers are no different from the low-quality non-disclosing sellers, there are incentives for the highest quality sellers to voluntarily reveal their quality. The next highest quality sellers then become the highest quality non-disclosing sellers, so these firms then have an incentive to disclose their quality as well. This unraveling process continues as long as the benefit of disclosure outweighs the cost. The notion of unraveling equilibria in settings with voluntary disclosure is initially put forward by Grossman and Hart (1980), Grossman (1981), and Milgrom (1981). Subsequent extensions to the theory include the effect of disclosure costs (Jovanovic, 1982), the effect of information acquisition costs (Farrell, 1986), the effect of some consumers being uninformed (Fishman and Hagerty, 2003), and the effect of competition (Jin, 2005).

The theoretical literature identifies several ways through which information may affect the efficiency of markets. The typical insight supports information disclosure, which has led economists to support policies that seek to increase the amount of information available to consumers. Meanwhile, existing empirical studies find small or negligible effects from increased information availability, casting doubt on the importance of such policies. For instance, health researchers have long debated the extent to which "report cards" — public disclosure of comparative information on the performance of doctors, hospitals, and insurers — affect the enrollment of health plans as well as the market share of highly rated hospitals (Beaulieu, 2002; Bundorf et al., 2009; Chernew et al., 2008; Cutler et al., 2004; Dafny and Dranove, 2008; Jin and Sorensen, 2005; Mukamel and Mushlin, 1998; Scanlon et al., 2002; Wedig and Tai-Seale, 2002). It has been contended that the failure is on the part of the empirical research, and is mainly due to the difficulty of observing exogenous variation in the amount of information available to consumers (Jin and Leslie, 2003, 2009). Estimating the causal effect of information on market outcomes therefore remains an open empirical issue.

#### 1.2.4.2 Behavioral Theory: Evidence and Consequence of Forgetfulness and Limited Attention on Information Processing

In contrast to models that have uniformly assumed full rationality, behavioral theories emphasize that the human brain has imperfect memory and limited information processing capacity, which respectively lead to two different yet closely related concepts — forgetfulness and limited attention. The literature on forgetfulness assumes that economic agents are unable to recall all the information they have once possessed when making decisions. Hence, reminders often play an important role in changing human behaviors, usually in the direction for which the reminders are intended (Calzolari and Nardotto, 2015; DellaVigna and Malmendier, 2006; Ericson, 2014; Finkelstein, 2009; Goldin and Homonoff, 2013; Karlan et al., 2010; Patrick et al., 2009; Taubinsky, 2014; Vervloet et al., 2012). The typical insight is that economic agents are forgetful, but will react rationally to a forgotten attribute (e.g., going to the gym) once it is explicitly reminded (Chetty et al., 2009; Gabaix and Laibson, 2006).

As for limited attention, it is ubiquitous in an information-rich economic environment, where the relevant information to be analyzed exceeds the information processing capacity of the human brain (Simon, 1957). In particular, Luca and Smith (2011), Pope (2009), and Scanlon et al. (2002) document situations where consumers rely on very coarse information while ignoring finer details, suggesting that consumers are indeed subject to limited attention in these settings. There is a growing amount of literature suggesting that limited attention can lead to suboptimal choices in both real-world and experimental settings for various reasons (Akerlof, 1991; Ater and Landsman, 2013; Bordalo et al., 2012, 2013; Bushong et al., 2014; Gabaix and Laibson, 2006; Gilchrist et al., 2015; Hastings and Shapiro, 2013; Huysentruyt and Read, 2010; Koszegi and Szeidl, 2013). However, according to some psychological models, limited attention can also be an effective strategy that allows for economical search for objects and thus enhances performance (Clark and Dukas, 2003; Simon, 1955). In this case, attention is limited to the information being processed in conscious thought, which enables the human brain to focus on the task at hand.

The actual allocation of attention depends on one's ability to retrieve relevant examples from memory, and the interaction between memory and attention is key to understanding how economic agents react to information. One heuristic the human brain uses is to make decisions based on the most salient or surprising factors in light of what is being reminded and what comes to mind first (Bordalo et al., 2012, 2013, 2015; Gennaioli and Shleifer, 2010; Tversky and Kahneman, 1973, 1974). Although this cognitive shortcut enables the human brain to make judgments easily, it can result in inefficient outcomes. For example, a reminder that a TV can break down causes consumers to overestimate the likelihood of a breakdown and thus buy overpriced warranties for newly purchased TVs (Huysentruyt and Read, 2010). Moreover, behavioral finance studies suggest that limited attention affects capital market outcomes: investors fail to make use of all available information and they tend to respond to salient, easy-to-process information while neglecting relevant aspects of the economic environment (Amir, 1993; Hirshleifer and Teoh, 2003; Lipe, 1998). There is also evidence that firms present accounting information in a way that is consistent with exploiting the perception of partially attentive investors (Kothari, 2001). In nearly every experimental study, agents demonstrate limited information processing power — even professional analysts resort to simplifying tools such as placement, categorization, and labeling in evaluating accounting information and do not fully adjust for differences in accounting alternatives (Libby et al., 2002; Maines, 1995).

Recent work by Bordalo et al. (2015) presents a theoretical model that combines the elements of selective memory and limited attention, and explains how consumers under- or over-react to information when their allocation of attention is distorted by selective recall. According to the model, disclosure of information does not necessarily change consumers' behaviors in the expected direction, because how attention is directed can be triggered by the information being provided and influenced by consumers' past experiences. Specifically, in the context of quality information disclosure, the effect of information may combine two effects: (1) a rational response that corresponds exactly to the informational content (and thus the belief update from the market average to the "true" quality of a particular product), and (2) a behavioral response that corresponds to the change in salience and allocation of attention triggered by quality information itself (e.g., a reminder of consumers' past experiences with quality of a particular product). The net effect of information is therefore *ex ante* unclear, which is different from what a conventional economic model would predict. The net effect could be positive or negative, depending on the direction and relative magnitude of the two responses. Moreover, any observed effect of information provision simultaneously contains two responses, which creates a potential challenge for empirically decomposing the rational and the behavioral effects, as well as for directly providing empirical evidence for the existence of distorted allocation of attention as put forward by the behavioral models.

## 1.3 Theoretical Framework

### 1.3.1 Basic Setup

#### 1.3.1.1 Assumptions

There are two types of products in the market, high-quality products with true value $\overline{v}$ and low-quality products with true value $\underline{v}$, where $\overline{v} > \underline{v} \geq 0$.

There are $n \geq 2$ sellers of two types: good sellers and bad sellers. Let $j \in \{g, b\}$ denote the type of seller $j$. Good sellers ($g$) sell high-quality products with true probability $p_g$ and low-quality products with probability $1 - p_g$, and bad sellers ($b$) sell high-quality products with probability $p_b$ and low-quality products with probability $1 - p_b$. We assume that $p_g > p_b$ so that good sellers have a higher probability of selling high-quality products than bad sellers do.

A seller can voluntarily make costly quality-improving investments and provide information to buyers about their investments. All information provided must be observable and verifiable, which means that a seller cannot provide information about an investment unless they have actually made the investment in the first place. Let $y \in [0, \infty)$ be an index that measures the value of the information provided by the seller, which is a continuous function that strictly increases with the amount of investment $I$. Specifically, we define $c(j, y)$, which is continuous and differentiable in $y$, as the cost of investment by seller $j$, which is incurred before any information about the investment can be provided.

We assume that

$$\frac{\partial c(j, y)}{\partial y} > 0$$

so that the value of information provided by the seller is positively associated with the cost of the investment. By construction, for any given seller of type $j \in \{g, b\}$, the inverse function $y = c^{-1}(j, \cdot)$ exists.

We further assume that $c(j, 0) = 0$ and that for all $y > 0$,

$$0 < c(g, y) < c(b, y)$$

which means that it is more costly for bad sellers than for good sellers to provide any information that is valuable. In the milk powder industry, the intuition behind this assumption is that good sellers are often equipped with better infrastructure or resources that allow them to make the same level of investment in a less costly way.

Then, suppose there are $m \geq 1$ buyers whose utility functions are all given by

$$u(x, w) = x - w$$

where $x$ is their expected value of the product and $w$ is the amount they are willing to pay for the product.

There are two types of buyers in the market: the "rational" buyers and the "naïve" buyers. Let $\alpha$ be the fraction of rational buyers and $1 - \alpha$ be the fraction of naïve buyers. Rational buyers always have correct beliefs of the probability of high-quality products sold by each type of seller, regardless of whether any information is provided to them. Naïve buyers, on the contrary, are too "optimistic" in the absence of information provision and thus believe that such probabilities are higher than they actually are. In particular, let $p_g'$ and $p_b'$ be their perceived probabilities of good sellers selling high-quality products and bad sellers selling high-quality products, respectively, when no information is provided to them. However, once the naïve buyers are provided with any type of information related to the milk-powder industry, their perceived beliefs of such probabilities will revert to the correct ones, i.e., $p_g$ for good sellers and $p_b$ for bad sellers. We assume that

$$0 < p_b < p_g < p_g' < 1$$
$$0 < p_b < p_b' < p_g' < 1.$$

Let $x_g$ and $x_b$ be the "rational" expected value of a product produced by a good seller and that by a bad seller, respectively. Similarly, let $x_g'$ and $x_b'$ be the "naïve" expected value of a product produced by a good seller and that by a bad seller, respectively. By construction,

$$\begin{aligned}
x_g &= p_g\overline{v} + (1 - p_g)\underline{v} \\
x_b &= p_b\overline{v} + (1 - p_b)\underline{v} \\
x_g' &= p_g'\overline{v} + (1 - p_g')\underline{v} \\
x_b' &= p_b'\overline{v} + (1 - p_b')\underline{v}.
\end{aligned}$$

Finally, we assume that the sellers have all the bargaining power and that all the sellers and buyers are risk neutral.

#### 1.3.1.2 Timing

1. The sellers privately observe their type $j \in \{g, b\}$.

2. Each seller chooses the amount of information $y \in [0, \infty)$ they would like to provide explicitly to the buyers.

3. The buyers observe the sellers' provision of information and make offers with price $w(y)$ to the seller.

4. Sellers decide whether to accept the offer. Once an offer is accepted, products are supplied and payoffs are realized.

### 1.3.2 Separating Equilibrium

#### 1.3.2.1 Benchmark

We first assume that $\alpha = 1$, which means that all buyers in the market are rational. We claim that there exists an infinite number of separating equilibria where the good sellers will choose to provide information and bad sellers will choose not to provide information. In return, buyers are willing to pay more for products produced by those who choose to provide information.

**Proposition 1.1.** *Suppose that buyers make the following decisions:*

$$w(y) = \begin{cases} x_g & \text{if } y \geq y^* \\ x_b & \text{if } y < y^* \end{cases}$$

*where $y^* > 0$.*

*Then, a separating equilibrium exists if the signal threshold $y^*$ satisfies the following:*

$$c^{-1}(b, x_g - x_b) \leq y^* \leq c^{-1}(g, x_g - x_b).$$

*Proof.* Given the buyers' belief, we know that a seller will choose either $y = y^*$ to get $x_g$ or $y = 0$ to get $x_b$. We need to check whether it is optimal for all the good sellers to choose $y = y^*$ and for all the bad sellers to choose $y = 0$.

For the good sellers, their incentive compatibility constraint is

$$x_g - c(g, y^*) \geq x_b \implies c(g, y^*) \leq x_g - x_b. \tag{1.1}$$

For the bad sellers, their incentive compatibility constraint is

$$x_b \geq x_g - c(b, y^*) \implies c(b, y^*) \geq x_g - x_b. \tag{1.2}$$

Equation (1.1) and Equation (1.2) jointly imply

$$c(g, y^*) \leq x_g - x_b \leq c(b, y^*) \implies c^{-1}(b, x_g - x_b) \leq y^* \leq c^{-1}(g, x_g - x_b). \tag{1.3}$$

By our assumption, since $x_g - x_b > 0$

$$\left[c^{-1}(b, x_g - x_b), c^{-1}(g, x_g - x_b)\right] \neq \varnothing.$$

Therefore, for any $y^*$ satisfying Equation (1.3), there is a separating equilibrium where good sellers provide information of value $y^*$ to get $x_g$ from the buyers and bad sellers do not provide any information so as to get $x_b$ from the buyers. □

As long as the separating equilibrium exists with some appropriate $y^*$, the optimal amount of investment $I$ by good sellers (denoted as $I_g$) and that by bad sellers (denoted as $I_b$) satisfy the following:

$$\begin{aligned} I_g^* &= y^{-1}(y^*) \\ I_b^* &= 0 \end{aligned}$$

which means that good sellers have the incentive to make a non-zero investment while bad sellers do not.

#### 1.3.2.2 Information Provision with a "Reminder Effect"

In the benchmark case, we only considered the case where all buyers are rational. However, it is conceivable that in real life, buyers do not readily recall all the information they have when choosing products. Their information retrieval might depend on what they see and what is being provided to them. For instance, when the reputation of a particular industry is at issue, disclosure of quality-related information by sellers in this industry may have two effects: the "direct effect" and the "reminder effect." While the direct effect only allows the buyers to learn what is being disclosed to them, the reminder effect might trigger their recall of some negative aspects associated with the industry. In other words, even if a piece of information involves only a quality-improving investment, the mentioning of the investment might bring up a negative prior that buyers have about the industry, thus lowering their willingness to pay.

In this case, we assume that $\alpha \in [0, 1)$, which means that there are at least some naïve buyers in the market. We claim that a separating equilibrium is no longer guaranteed.

**Proposition 1.2.** *There does not exist any separating equilibrium if*

$$\frac{p_b' - p_b}{p_g - p_b} \geq \frac{1}{1-\alpha}.$$

*Proof.* We use proof by contradiction in this case. There are two possible types of separating equilibrium we need to check:

(1) Good sellers provide a positive amount of information while bad sellers do not provide any information.

(2) Good sellers do not provide any information while bad sellers provide a positive amount of information.

For the first equilibrium candidate, suppose that there exists a separating equilibrium with some $y^* > 0$ in which the buyers classify whoever provides $y \geq y^*$ as a good seller and whoever provides $y < y^*$ as a bad seller.

As in the benchmark case, a seller will choose either $y = y^*$ to be perceived as good or $y = 0$ to be perceived as bad. Hence, rational buyers make the following decision

$$w(y) = \begin{cases} x_g & \text{if } y = y^* \\ x_b & \text{if } y = 0 \end{cases}$$

and naïve buyers make the following decision

$$w(y) = \begin{cases} x_g & \text{if } y = y^* \\ x'_b & \text{if } y = 0. \end{cases}$$

However, we will show that the good sellers have an incentive to deviate by not providing information at all.

For the good sellers, their incentive compatibility constraint is

$$x_g - c(g, y^*) \geq \alpha x_b + (1-\alpha)x'_b \implies c(g, y^*) \leq x_g - (\alpha x_b + (1-\alpha)x'_b). \tag{1.4}$$

For the bad sellers, their incentive compatibility constraint is

$$\alpha x_b + (1-\alpha)x'_b \geq x_g - c(b, y^*) \implies c(b, y^*) \geq x_g - (\alpha x_b + (1-\alpha)x'_b). \tag{1.5}$$

Note that

$$\begin{aligned} & x_g - (\alpha x_b + (1-\alpha)x'_b) \\ = \; & p_g\overline{v} + (1-p_g)\underline{v} - \alpha(p_b\overline{v} + (1-p_b)\underline{v}) - (1-\alpha)(p'_b\overline{v} + (1-p'_b)\underline{v}) \\ = \; & \alpha(p_g - p_b)(\overline{v} - \underline{v}) + (1-\alpha)(p_g - p'_b)(\overline{v} - \underline{v}) \\ = \; & (\overline{v} - \underline{v})(\alpha(p_g - p_b) + (1-\alpha)(p_g - p_b + p_b - p'_b)) \\ = \; & (\overline{v} - \underline{v})(p_g - p_b - (1-\alpha)(p'_b - p_b)). \end{aligned} \tag{1.6}$$

However, since

$$\frac{p_b' - p_b}{p_g - p_b} \geq \frac{1}{1-\alpha} \implies p_g - p_b - (1-\alpha)(p_b' - p_b) \leq 0$$
$$\implies x_g - (\alpha x_b + (1-\alpha)x_b' \leq 0$$

from Equation (1.6). Then, Equation (1.4) implies that

$$c(g, y^*) \leq 0.$$

However, by definition, $c(g, y^*) > 0$ for any $y^* > 0$, which is a contradiction. This means that good sellers do not have the incentive to provide any information. Hence, there is no separating equilibrium in which good sellers provide information while bad sellers do not.

For the second equilibrium candidate, suppose that there exists a separating equilibrium with some $y^* > 0$ in which buyers perceive whoever provides $y \geq y^*$ as a bad seller and whoever provides $y < y^*$ as a good seller. With a similar argument established for the first equilibrium candidate, the rational buyers make the following decision:

$$w(y) = \begin{cases} x_b & \text{if } y = y^* \\ x_g & \text{if } y = 0 \end{cases}$$

and the naïve buyers make the following decision

$$w(y) = \begin{cases} x_b & \text{if } y = y^* \\ x_g' & \text{if } y = 0. \end{cases}$$

For the bad sellers, their incentive compatibility constraint is

$$x_b - c(b, y^*) \geq \alpha x_g + (1-\alpha)x_g' \implies c(b, y^*) \leq x_b - (\alpha x_g + (1-\alpha)x_g'). \tag{1.7}$$

However, note that

$$x_b - (\alpha x_g + (1-\alpha)x_g') \leq x_g - (\alpha x_g + (1-\alpha)x_g') = (1-\alpha)(x_g - x_g') < 0$$

which obviously violates the assumption that $c(j, y) \geq 0$. Hence, there is no separating equilibrium in which good sellers provide no information while bad sellers provide a positive amount of information.

Now that we have checked both candidates, we can conclude that there is no separating equilibrium in this case. $\square$

When the separating equilibrium does not exist, the optimal amount of investment by each type becomes the following:

$$I_g^* = I_b^* = 0$$

which is inefficient compared to the benchmark case without the "reminder effect."

Intuitively, this result shows that good sellers are unable to distinguish themselves from bad sellers by merely providing positive quality-related information on their products if any of the following is true:

- The industry has a sufficiently negative image such that once the naïve buyers are reminded of quality-related issues, they experience a significant reduction in their perceived probability of bad sellers selling high-quality products. This corresponds to a large $p_b' - p_b$.

- The probability of good sellers selling high-quality products is not too much higher than the probability of bad sellers selling high-quality products, which corresponds to a small $p_g - p_b$. This applies to the infant milk powder industry when many producers, including several previous industry giants, turned out to add toxic chemicals to their products, though to varying degrees.

- The proportion of rational buyers is small relative to the proportion of naïve buyers, which means that for most buyers, their information retrieval process is subject to a "reminder effect" as described before. This corresponds to a small $\alpha$; the closer it is to zero, the less likely a separating equilibrium exists.

## 1.4 Online Study

### 1.4.1 Experiment Design

One key implication of the theoretical model is that some consumers become less likely to purchase from a producer who provides positive quality-related information because the information reminds them of other negative aspects of the industry. To provide some empirical evidence, we conduct an online experiment with a survey to examine the effect of voluntary disclosure of quality-related information when consumers have a strong negative prior belief about the industry. The survey experiment is done at Beingmate's online store "Motherbuy" (www.motherbuy.com) that offers the full range of the company's products.

The online survey experiment is designed to have one control group and three treatment groups as follows. First, we instruct site administrators to randomly assign site visitors to different experiment groups based on the last digit of each visitor's IP address.[10] The treatment assignment process is as follows:

- Control Group: IP addresses with the last digit of 0, 5, 6 or 9;
- Treatment 1 (Brand): IP addresses with the last digit of 1 or 4;
- Treatment 2 (Traceability): IP addresses with the last digit of 2 or 7;
- Treatment 3 (Certification): IP addresses with the last digit of 3 or 8.

Then, quality-related information is shown in a pop-up window at the center of the computer/smartphone screen when a visitor enters the online store, and we allow visitors to close the pop-up window as they wish. All the information is provided in Chinese and is translated as follows.

[10] The IP address is related to the visitor's physical location, but the last digit is randomly generated without any identifying information.

- Control Group: No information is provided.
- Treatment 1 (Brand)

  Visitors are provided with the message "Beingmate is the go-to brand for your babies" via a pop-up window.[11] (This is a standard phrase used on many of Beingmate's advertisements.)
- Treatment 2 (Traceability)

  Visitors are provided with the message "Beingmate provides a product traceability system that achieves traceability from milk powder's raw material to production, marketing, and sales" via a pop-up window.[12]
- Treatment 3 (Certification)

  Visitors are provided with the message "Beingmate has obtained the ISO9001 quality certification. ISO9001 is the world's leading product quality certification system" via a pop-up window.[13]

Next, for visitors who choose to make a purchase, we conduct a survey after the checkout page to empirically test if quality information leads to a negative impression of the industry. Upon completion of the purchase, visitors are provided with an option to fill out a four-question survey. For each customer with valid survey responses, we offer a 10 CNY (approximately 1.5 USD) coupon that can be redeemed towards future purchases. The survey questions are given in Chinese, and the translated version is as follows:

**Question 1.** What is your impression of the domestic infant milk powder industry?

0: Highly untrustworthy

[11]See the first panel of Figure A.2 for the original, Chinese-version of the message.

[12]See the second panel of Figure A.2 for the original, Chinese-version of the message.

[13]See the third panel of Figure A.2 for the original, Chinese-version of the message.

1: Untrustworthy

2: Somewhat trustworthy

3: Trustworthy

4: Highly trustworthy

**Question 2.** When choosing infant formula products, how often do you take into account traceability system as an important factor?

0: Never

1: Rarely

2: Sometimes

3: Often

4: Always

**Question 3.** When choosing infant formula products, how often do you take into account ISO9001 certification as an important factor?

0: Never

1: Rarely

2: Sometimes

3: Often

4: Always

**Question 4.** To ensure that you are not a robot, please choose the last option for this question. (This question is given in order to ensure that all questions are read carefully by the survey respondents.)

We conduct the online experiment from the week of April 13 to the week of June 1, 2015 (i.e., a total of eight weeks). For each online visit, we collected the following data:

- The visitor's IP address and the province in which the IP address is located;
- The purchase decision of the visitor, i.e., whether the visitor made a purchase (binary);

- For each visitor who made a purchase online: date of purchase, name(s) and unit price(s) of product(s) purchased, number of units purchased of each product, and whether the visitor is part of Beingmate's loyalty programs (binary);

- For visitors who filled out the survey after their purchase, we code their numerical responses to the survey questions.

### 1.4.2 Empirical Methodology

We examine two research questions with the data collected from the online experiment. First, for those who visited the online store, does the disclosure of quality information have a negative effect on purchase behavior? Second, for those who made a purchase and filled out the survey at the end of the purchase, does the disclosure of quality information lead to a negative impression of the industry?

For the first question, our baseline regression models are as follows:

$$\text{OLS}: \quad quantity_i = \beta_0 + \sum_{k=1}^{3} \beta_{1k} \cdot treatment_{ik} + \beta_2 \cdot affiliated_i + \sum_{l \in L} D_{il} + \varepsilon_i \tag{1.8}$$

$$\text{Logit}: \quad purchase_i = \beta_0 + \sum_{k=1}^{3} \beta_{1k} \cdot treatment_{ik} + \beta_2 \cdot affiliated_i + \sum_{l \in L} D_{il} + \varepsilon_i \tag{1.9}$$

where $i$ indicates each visitor to the online store. We define $quantity_i$ as the total number of units purchased by visitor $i$, which can be any positive integer greater than or equal to 0. We define a similar variable, $purchase_i$, as an indicator variable that takes the value of 1 if $quantity_i > 0$ and 0 otherwise to indicate visitor $i$'s purchase decision. The variable $treatment_{ik}$ is an indicator variable that takes the value of 1 if a particular visitor $i$ receives Treatment $k$, where integer $k$ ranges from 1 to 3. The variable $affiliated_i$ is an indicator variable that takes the value of 1 if visitor $i$ is a member of at least one of the loyalty programs offered by Beingmate. Finally, the $D_{il}$'s indicate location fixed effects at the province level and $\varepsilon_i$ is the error term.

For the second question, our baseline regression model is as follows:

$$\begin{aligned} impression_i \quad = \quad & \beta_0 + \sum_{k=1}^{3} \beta_{1k} \cdot treatment_{ik} + \sum_{k=2}^{3} \beta_{2k} \cdot Q_{ik} + \\ & \sum_{k=2}^{3} \beta_{3k} \cdot (treatment_{ik} \cdot Q_{ik}) + \sum_{l \in L} D_{il} + \varepsilon_i. \end{aligned} \tag{1.10}$$

In particular, $impression_i$ is defined as visitor $i$'s impression of the milk powder industry in China. This is the same as visitor $i$'s response to Question 1 of the survey, which is a numerical answer ranging from 0 to 4. The variables $Q_{i2}$ and $Q_{i3}$ are defined as visitor $i$'s response to Question 2 (on the importance of the traceability system) and Question 3 (on the importance of the ISO9001 quality certification) of the survey, respectively, both of which are numerical answers ranging from 0 to 4. All the other variables have exactly the same definition as before. Since the survey responses are all coded as ordinal variables, we use both OLS regressions and ordered logit regressions to estimate the coefficients on the interaction variables, which represent the treatment effect of the three types of information provided on consumers' impression of the milk powder industry in China.

### 1.4.3 Results and Discussions

Table 1.1 and Table 1.2 provide the summary statistics for all the relevant variables as described in Section 1.4.1, broken down by experiment group (one control group plus three treatment groups), with the former for all visitors and the latter for valid survey respondents only. There are a total of 569,272 potential consumers visiting the online store during the eight-week treatment period, of which 1,184 made a purchase. Among those who made a purchase, 737 consumers answered the survey. After eliminating the survey respondents who answered Question 4 incorrectly, we received a total of 493 valid survey responses.

Table 1.3 documents the effect of information disclosure on purchase decisions. Columns (1)–(4) present the results of the logit regression as detailed in Equation (1.8); and columns (5)–(8) present the results of the OLS regression as detailed in Equation (1.9). Overall,

**Table 1.1: Summary Statistics for Online Study**
*All Visitors*

| | $N$ | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Control Group* | | | | | | |
| Purchase Decision | 229,095 | .00226 | .0475 | 0 | 0 | 1 |
| Total Quantity | 229,095 | .00966 | .273 | 0 | 0 | 28 |
| Total Revenue | 229,095 | 1.17 | 38.9 | 0 | 0 | 5,820 |
| Affiliated | 229,095 | .125 | .33 | 0 | 0 | 1 |
| Impression of Industry | 323 | 2.65 | .873 | 0 | 3 | 4 |
| Importance of Traceability | 323 | 3.08 | 1.06 | 0 | 3 | 4 |
| Importance of Certification | 323 | 3.09 | 1.14 | 0 | 4 | 4 |
| *Treatment 1 (Brand) Group* | | | | | | |
| Purchase Decision | 115,192 | .00214 | .0463 | 0 | 0 | 1 |
| Total Quantity | 115,192 | .00911 | .26 | 0 | 0 | 22 |
| Total Revenue | 115,192 | 1.11 | 32.4 | 0 | 0 | 1,885 |
| Affiliated | 115,192 | .123 | .328 | 0 | 0 | 1 |
| Impression of Industry | 151 | 2.58 | .836 | 0 | 3 | 4 |
| Importance of Traceability | 151 | 3.22 | .965 | 0 | 4 | 4 |
| Importance of Certification | 151 | 3.17 | 1.07 | 0 | 4 | 4 |
| *Treatment 2 (Traceability) Group* | | | | | | |
| Purchase Decision | 113,526 | .00183 | .0428 | 0 | 0 | 1 |
| Total Quantity | 113,526 | .00735 | .231 | 0 | 0 | 24 |
| Total Revenue | 113,526 | .973 | 31.9 | 0 | 0 | 2,910 |
| Affiliated | 113,526 | .121 | .326 | 0 | 0 | 1 |
| Impression of Industry | 129 | 2.7 | .844 | 0 | 3 | 4 |
| Importance of Traceability | 129 | 3.11 | 1.01 | 0 | 3 | 4 |
| Importance of Certification | 129 | 3.26 | 1.04 | 0 | 4 | 4 |
| *Treatment 3 (Certification) Group* | | | | | | |
| Purchase Decision | 111,459 | .00188 | .0434 | 0 | 0 | 1 |
| Total Quantity | 111,459 | .00773 | .25 | 0 | 0 | 25 |
| Total Revenue | 111,459 | .949 | 29.9 | 0 | 0 | 2,328 |
| Affiliated | 111,459 | .119 | .324 | 0 | 0 | 1 |
| Impression of Industry | 134 | 2.51 | .907 | 0 | 3 | 4 |
| Importance of Traceability | 134 | 3.16 | 1.03 | 0 | 4 | 4 |
| Importance of Certification | 134 | 3.08 | 1.26 | 0 | 4 | 4 |

**Table 1.2: Summary Statistics for Online Study**

*Valid Survey Respondents Only*

| | $N$ | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Control Group* | | | | | | |
| Purchase Decision | 211 | .37 | .484 | 0 | 0 | 1 |
| Total Quantity | 211 | 1.85 | 3.53 | 0 | 0 | 22 |
| Total Revenue | 211 | 236 | 561 | 0 | 0 | 3,856 |
| Affiliated | 211 | 1 | 0 | 1 | 1 | 1 |
| Impression of Industry | 211 | 2.73 | .872 | 0 | 3 | 4 |
| Importance of Traceability | 211 | 3.27 | .965 | 0 | 4 | 4 |
| Importance of Certification | 211 | 3.2 | 1.12 | 0 | 4 | 4 |
| *Treatment 1 (Brand) Group* | | | | | | |
| Purchase Decision | 102 | .402 | .493 | 0 | 0 | 1 |
| Total Quantity | 102 | 1.83 | 3.76 | 0 | 0 | 22 |
| Total Revenue | 102 | 205 | 377 | 0 | 0 | 1,561 |
| Affiliated | 102 | 1 | 0 | 1 | 1 | 1 |
| Impression of Industry | 102 | 2.7 | .793 | 0 | 3 | 4 |
| Importance of Traceability | 102 | 3.42 | .849 | 0 | 4 | 4 |
| Importance of Certification | 102 | 3.35 | .992 | 0 | 4 | 4 |
| *Treatment 2 (Traceability) Group* | | | | | | |
| Purchase Decision | 89 | .315 | .467 | 0 | 0 | 1 |
| Total Quantity | 89 | 1.31 | 2.56 | 0 | 0 | 12 |
| Total Revenue | 89 | 167 | 364 | 0 | 0 | 1,508 |
| Affiliated | 89 | 1 | 0 | 1 | 1 | 1 |
| Impression of Industry | 89 | 2.73 | .822 | 0 | 3 | 4 |
| Importance of Traceability | 89 | 3.34 | .891 | 1 | 4 | 4 |
| Importance of Certification | 89 | 3.35 | .931 | 0 | 4 | 4 |
| *Treatment 3 (Certification) Group* | | | | | | |
| Purchase Decision | 91 | .429 | .498 | 0 | 0 | 1 |
| Total Quantity | 91 | 2.59 | 4.51 | 0 | 0 | 24 |
| Total Revenue | 91 | 247 | 396 | 0 | 0 | 1,463 |
| Affiliated | 91 | 1 | 0 | 1 | 1 | 1 |
| Impression of Industry | 91 | 2.46 | .911 | 0 | 3 | 4 |
| Importance of Traceability | 91 | 3.27 | .955 | 0 | 4 | 4 |
| Importance of Certification | 91 | 3.13 | 1.27 | 0 | 4 | 4 |

**Table 1.3: Treatment Effect on Online Purchase Decisions**

| | OLS | | | | Logit | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) Quantity | (2) Quantity | (3) Quantity | (4) Quantity | (5) Purchase | (6) Purchase | (7) Purchase | (8) Purchase |
| Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Brand | −0.001 | −0.000 | −0.001 | −0.000 | −0.053 | −0.040 | −0.047 | −0.055 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.077) | (0.078) | (0.077) | (0.079) |
| Traceability | −0.002*** | −0.002** | −0.002** | −0.002** | −0.211** | −0.186** | −0.199** | −0.161* |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.082) | (0.083) | (0.082) | (0.083) |
| Certification | −0.002** | −0.002 | −0.002* | −0.001 | −0.183** | −0.139* | −0.172** | −0.106 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.082) | (0.082) | (0.082) | (0.083) |
| Affiliated | | 0.071*** | | 0.072*** | | 9.059*** | | 9.145*** |
| | | (0.003) | | (0.003) | | (1.000) | | (1.002) |
| Location Dummies | No | No | Yes | Yes | No | No | Yes | Yes |
| Observations | 569272 | 569272 | 569272 | 569272 | 569272 | 569272 | 569272 | 569272 |
| $R^2$ | 0.000 | 0.008 | 0.001 | 0.009 | | | | |
| Pseudo $R^2$ | | | | | 0.001 | 0.293 | 0.033 | 0.311 |

Standard errors in parentheses

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

the results are surprising from a rational perspective but confirm our hypothesis about the "reminder effect," that is, disclosure of quality-related information, namely Treatment 2 (Traceability) or Treatment 3 (Certification), has a significantly negative effect on purchase behaviors. Translating to odds ratios, the coefficients of logit regressions indicate that Treatment 2 (Traceability) leads to a 16-21% reduction in the likelihood of purchase, and Treatment 3 (Certification) leads to a 14-18% reduction in the likelihood of purchase.

One possible interpretation of these results is that seeing the information about the traceability system or the quality certification makes consumers suspicious because it reminds them of the negative images associated with milk powder products produced in China (the "reminder effect"), thus making them less likely to make the purchase. This also leads to the broader research question of why sometimes providing less information is actually better than providing more information.

Table 1.4 and Table 1.5 test the central hypothesis of this study: disclosure of quality information makes consumers nervous and start questioning, "Why does the company need to assure me that their products are safe?" As detailed in Equation (1.10), the OLS regression results are reported in Table 1.4 and the ordered logit regression results are reported in Table 1.5. Overall, the results are consistent with the predictions of our model: disclosure of quality-related information leads to a more negative impression of the industry. In particular, this effect is significant for Treatment 3 (Certification) in all regression models: the coefficients of logit regressions indicate that Treatment 3 (Certification) leads to a 60-72% reduction in the likelihood of having a better impression of the milk powder industry. Moreover, in some regression models, we find that the negative effects of Treatment 2 (Traceability) and Treatment 3 (Certification) are mitigated by consumers who attach more importance to the traceability system and the ISO9001 quality certification. Specifically, the effect of Treatment 2 (Traceability) on the likelihood of having a better impression of the milk powder industry is 23-28 percentage points higher for people who attach one more degree of importance to product traceability systems. Similarly, the effect of Treatment 3 (Certification) is 41-45

**Table 1.4: Treatment Effect on Consumers' Reported Impression of the Milk Powder Industry (OLS)**

| | Dependent Variable: Impression of Industry (OLS) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Brand | −0.034 | −0.079 | −0.976** | −0.333 | 0.005 | −0.046 | −1.037** | −0.482 |
| | (0.099) | (0.093) | (0.448) | (0.412) | (0.105) | (0.097) | (0.464) | (0.415) |
| Traceability | 0.000 | −0.035 | −0.455* | 0.237 | 0.065 | −0.004 | −0.415 | 0.239 |
| | (0.106) | (0.103) | (0.247) | (0.288) | (0.113) | (0.110) | (0.254) | (0.292) |
| Certification | −0.268** | −0.256** | −0.268** | −0.256** | −0.267** | −0.266** | −0.267** | −0.265** |
| | (0.113) | (0.109) | (0.113) | (0.109) | (0.113) | (0.109) | (0.113) | (0.110) |
| Importance of | | 0.110** | | 0.097** | | 0.104** | | 0.083* |
| Traceability | | (0.043) | | (0.045) | | (0.044) | | (0.045) |
| Importance of | | 0.186*** | | 0.195*** | | 0.194*** | | 0.201*** |
| Certification | | (0.040) | | (0.045) | | (0.042) | | (0.047) |
| Traceability × | | | 0.275** | 0.074 | | | 0.304** | 0.129 |
| Importance of Traceability | | | (0.125) | (0.118) | | | (0.130) | (0.120) |
| Certification × | | | 0.136* | −0.081 | | | 0.142* | −0.071 |
| Importance of Certification | | | (0.077) | (0.089) | | | (0.080) | (0.092) |
| Location Dummies | No | No | No | No | Yes | Yes | Yes | Yes |
| Observations | 493 | 493 | 493 | 493 | 493 | 493 | 493 | 493 |
| $R^2$ | 0.014 | 0.101 | 0.033 | 0.103 | 0.089 | 0.169 | 0.110 | 0.173 |

Standard errors in parentheses

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

**Table 1.5: Treatment Effect on Consumers' Reported Impression of the Milk Powder Industry (Logit)**

| | Dependent Variable: Impression of Industry (Ordered Logit) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Brand | −0.108 | −0.236 | −2.307** | −0.929 | −0.031 | −0.157 | −2.500** | −1.192 |
| | (0.212) | (0.217) | (0.951) | (0.934) | (0.231) | (0.232) | (1.044) | (0.994) |
| Traceability | −0.100 | −0.149 | −1.245** | 0.208 | 0.058 | −0.063 | −1.209** | 0.272 |
| | (0.244) | (0.248) | (0.579) | (0.700) | (0.270) | (0.275) | (0.597) | (0.721) |
| Certification | −0.592** | −0.633** | −0.598** | −0.632** | −0.657** | −0.723*** | −0.668** | −0.723*** |
| | (0.244) | (0.256) | (0.247) | (0.257) | (0.262) | (0.277) | (0.264) | (0.278) |
| Importance of Traceability | | 0.276*** | | 0.244** | | 0.272*** | | 0.226** |
| | | (0.096) | | (0.100) | | (0.103) | | (0.106) |
| Importance of Certification | | 0.408*** | | 0.420*** | | 0.445*** | | 0.453*** |
| | | (0.094) | | (0.105) | | (0.103) | | (0.115) |
| Traceability × Importance of Traceability | | | 0.638** | 0.204 | | | 0.716** | 0.306 |
| | | | (0.267) | (0.269) | | | (0.295) | (0.288) |
| Certification × Importance of Certification | | | 0.349* | −0.109 | | | 0.381** | −0.101 |
| | | | (0.183) | (0.219) | | | (0.194) | (0.231) |
| Location Dummies | No | No | No | No | Yes | Yes | Yes | Yes |
| Observations | 493 | 493 | 493 | 493 | 493 | 493 | 493 | 493 |
| Pseudo $R^2$ | 0.005 | 0.041 | 0.014 | 0.042 | 0.042 | 0.076 | 0.051 | 0.078 |

Standard errors in parentheses

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

percentage points higher for people who attach more importance to quality certifications. These findings are intuitive to understand: despite a negative "reminder effect" due to the bad collective reputation of the industry, consumers are less likely to be affected when they are informed that a particular producer has made the exact type of quality-related investment that they value.

Our survey experiments are subject to several limitations. One potential concern with the survey responses is that they only come from visitors who made a purchase and that we have no information about visitors who decided not to make a purchase. However, it is reasonable to make the assumption that consumers who made a purchase should, on average, have a more positive impression of the milk powder industry than those who did not. In that case, our findings provide an upper bound for the negative "reminder effect." In other words, if we were able to conduct the same survey among all consumers who visited the website, we would expect an even more negative "reminder effect" than what we find among our selected sample in this experiment. Second, due to the cross-sectional nature of the online experiment, our results do not shed light on the extent to which the so-called "reminder effect" is transitory or persistent. Third, the experiment design does not rule out the possibility of a combination of multiple treatments, i.e., the situation in which one consumer visited the online store from two different IP addresses (e.g., one with the last digit of 2 from her office computer and the other with the last digit of 3 from her home computer), which makes the same person exposed to two different messages during the experiment period.

## 1.5 In-Store Study

### 1.5.1 Experiment Design

The in-store experiment is conducted in supermarkets in Hangzhou, Zhejiang Province, where Beingmate is headquartered. These stores are not owned or operated by Beingmate; instead, they sell a variety of products by different producers. We randomly assign stores to the same control and treatment groups as detailed in the online experiment. In order to minimize the effect of information spillover, randomization is done at the postal zone level so that stores located in the same postal zone are assigned to the same experiment group. There are 59 major postal zones in Hangzhou. Due to some practical constraints, only 10 postal zones are randomly selected for treatment and the other 49 zones are placed in the control group. Among the 10 selected postal zones, there are 3, 3, and 4 zones that receive Treatment 1 (Brand), Treatment 2 (Traceability), and Treatment 3 (Certification), respectively.

Then, we provide exactly the same information as in the online experiment to each corresponding treatment group. For supermarkets in the treated zones, there is one "information card" (a physical A4-sized board) per store attached to one of the product racks and placed at the most salient, "eye-level shelving" position.[14] All the information is provided in Chinese.

We trace what happens to dollar and volume sales before, during, and after the experiment period. The following data are collected in order to estimate the treatment effects and to test whether the information provided has a significant impact on consumer behavior:

- For each store, daily data (for 14 days) on dollar and volume sales for each product, *before* the experiment period ("period 0"), *during* the experiment period ("period 1"), and *after* the experiment period ("period 2");

- Store information: store ID, name, address, presence of sale representatives in the infant section (binary), existence of promotions on milk powder products produced by Beingmate (binary), and the average income of the district in which the store is located;

- Product information: product name, whether a product belongs to the milk category (binary), and daily unit price.[15]

[14]The design of the information cards is the same as shown in Figure A.2.

[15]Non-milk products include rice cereal, pork floss (protein supplement), biscuits, and other products for toddlers.

Figure 1.1, Figure 1.2, and Figure 1.3 illustrate the daily average quantity of sales for all Beingmate products, milk powder products, and all other products by Beingmate, respectively. As shown in the graphs, each period lasts for two weeks. Specifically, period 0 runs from July 28 to August 10, period 1 runs from August 11 to August 24, and period 2 runs from August 25 to September 7, 2014.

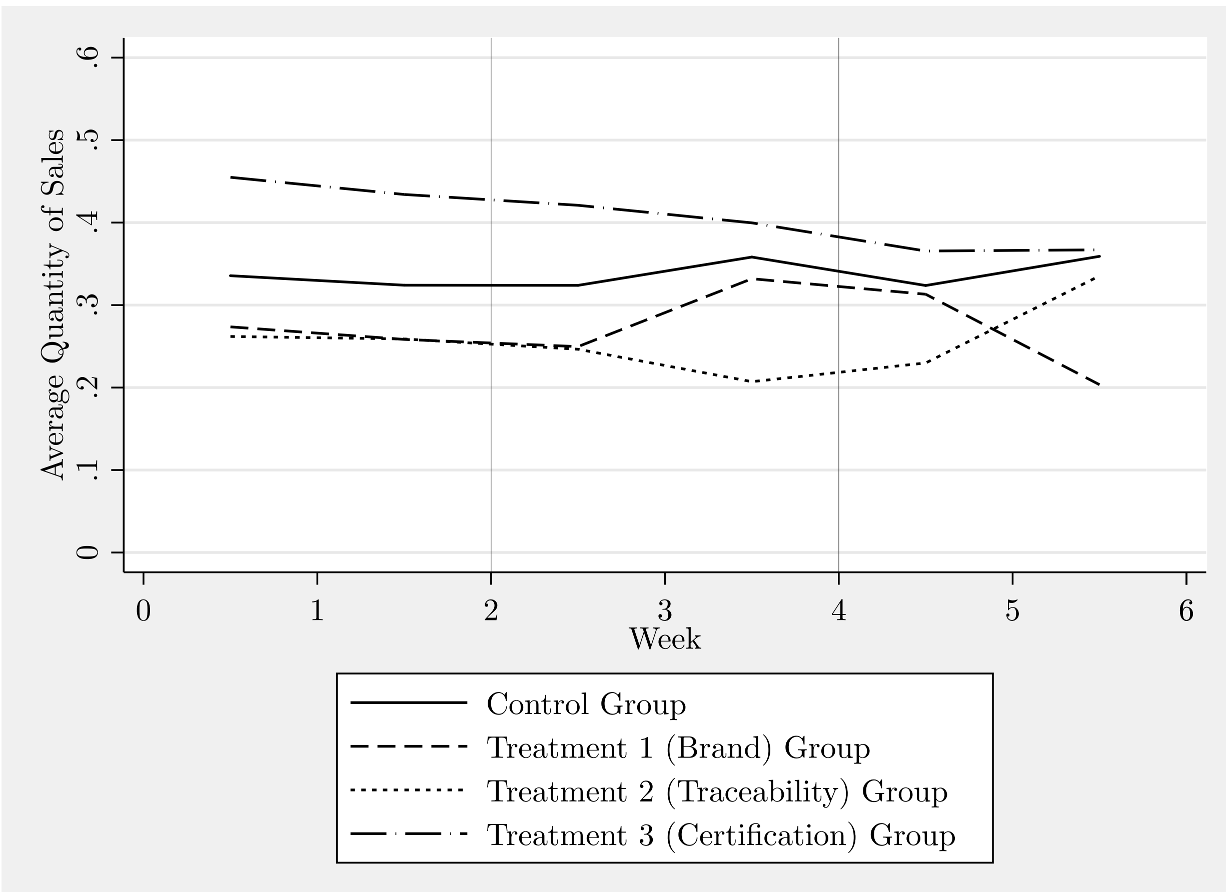



**Figure 1.1: Weekly Average Quantity of Sales**
*For All Beingmate Products*

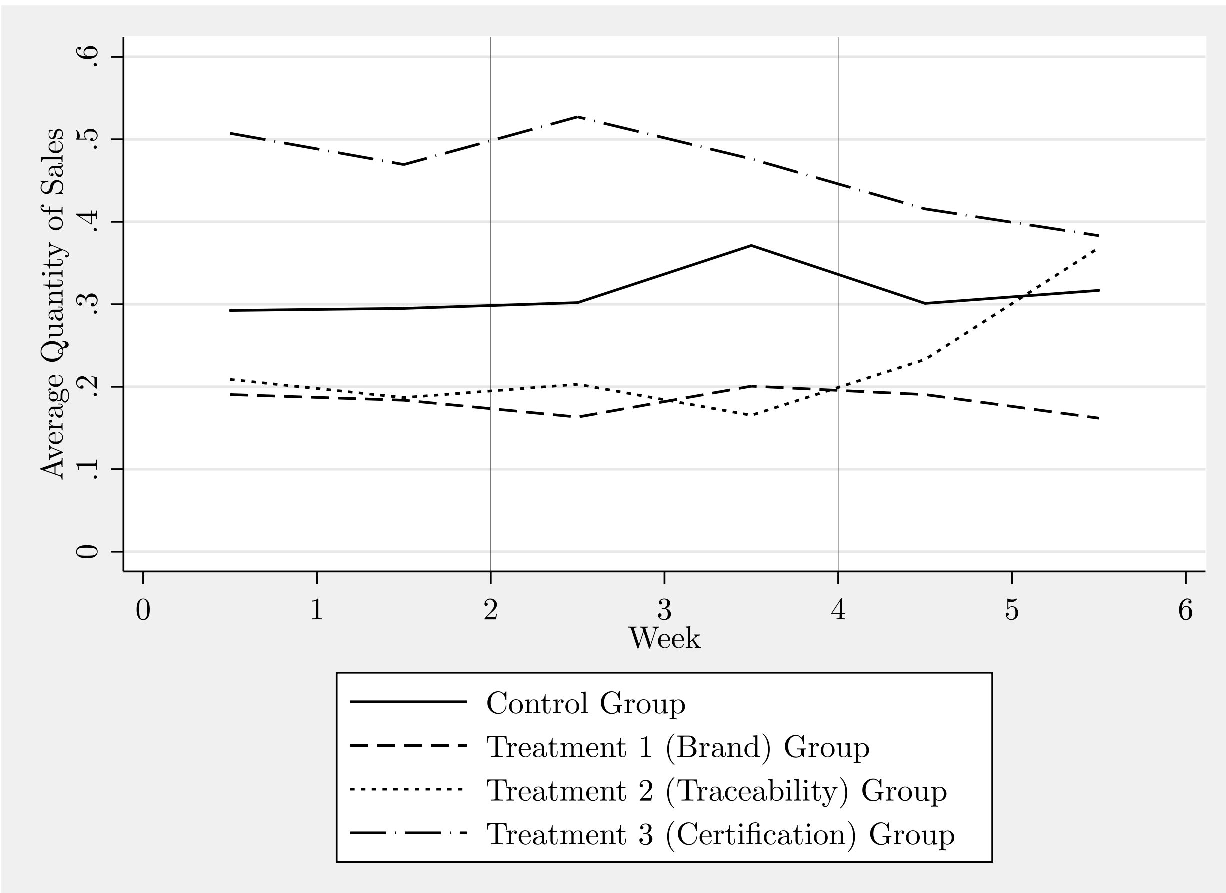



**Figure 1.2: Weekly Average Quantity of Sales**
*For Only Milk Powder Products by Beingmate*

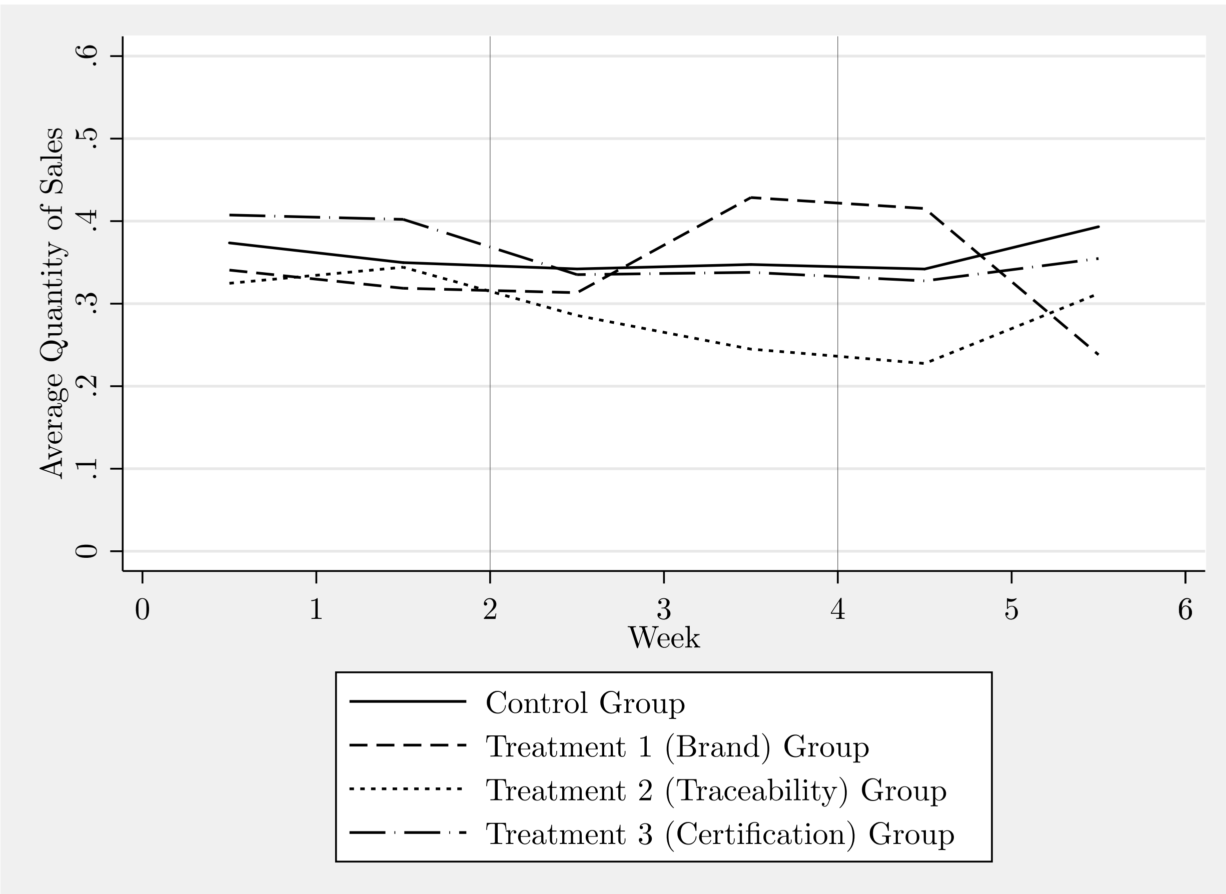



**Figure 1.3: Weekly Average Quantity of Sales**

*For All Other Products by Beingmate*

### 1.5.2 Empirical Methodology

We employ difference-in-difference regressions to estimate the effect of each treatment on the average daily sales. We compare two periods at a time (i.e., period 0 vs. period 1, period 1 vs. period 2, and period 0 vs. period 2) in order to capture different aspects of the effect of information that we are interested in.

For any two periods, our baseline regression model is as follows:

$$\begin{aligned} quantity_{ijt} \quad = \quad & \beta_0 + \beta_1 \cdot period_t + \sum_{k=1}^{3} \beta_{2k} \cdot (period_t \cdot treatment_{ik}) + \beta_3 \cdot price_{ijt} + \\ & \beta_4 \cdot salesperson_{it} + \beta_5 \cdot promotion_{it} + \gamma_i + \varepsilon_{ijt} \end{aligned} \tag{1.11}$$

where $i$, $j$, and $t$ indicate store, product, and period, respectively. We define $quantity_{ijt}$ as the average daily sales of product $j$ in store $i$ over period $t$. The variable $period_t$ equals 1 if $t$ is the latter period of the two that we are comparing and 0 otherwise. The variable $treatment_{ik}$ is an indicator variable that takes the value of 1 if a particular store $i$ receives Treatment $k$. The variable $price_{ijt}$ is defined as the average price of product $j$ in store $i$ over period $t$. The variables $salesperson_{it}$ and $promotion_{it}$ are indicator variables that take the value of 1 if in period $t$, a particular store $i$ has at least one salesperson in charge of milk powder products or has some form of store promotions on milk powder products, respectively. Finally, $\gamma_i$ captures store-level fixed-effects and $\varepsilon_{ijt}$ is the error term.

Since our randomization is done at the postal zone level, we do not have a large number of randomization units as desired. Therefore, we use randomization inference to construct our standard errors for the treatment effects. To do so, we generate 1,000 alternative randomization assignments. In each alternative assignment, we randomly reassign all the postal zones to different experiment groups, keeping all the outcome data the same as before. Then, we run the same regressions as specified for each comparison group and recalculate the coefficients on the interaction variables. The standard errors are calculated as the standard deviations of these new coefficients.

The $p$-values are generated from one-sided or two-sided hypothesis tests on the coefficients. In particular, we focus on the coefficient $\beta_{2k}$ of the interaction term between *period* and $treatment_k$ in Equation (1.11), where $k \in \{1, 2, 3\}$. We use $\beta'_{2k}$ to denote the new coefficient calculated from alternative random assignments.

For the pre-experiment period versus the experiment period, since we are interested in the negative "reminder effect," we conduct the following hypothesis test for each $k \in \{1, 2, 3\}$ in each regression model:

$$\begin{aligned} H_0 : &\quad \beta_{2k} = \beta'_{2k} \\ H_1 : &\quad \beta_{2k} < \beta'_{2k}. \end{aligned}$$

For the experiment period versus the post-experiment period, since we are interested in the positive effect due to the removal the "reminder effect," our hypothesis test is as follows for each $k \in \{1, 2, 3\}$ in each regression model:

$$\begin{aligned} H_0 : &\quad \beta_{2k} = \beta'_{2k} \\ H_1 : &\quad \beta_{2k} > \beta'_{2k}. \end{aligned}$$

For the pre-experiment period versus the post-experiment period, since we are interested in the zero effect of "pure" information, we have the following hypothesis test for each $k \in \{1, 2, 3\}$ in each regression model:

$$\begin{aligned} H_0 : &\quad \beta_{2k} = \beta'_{2k} \\ H_1 : &\quad \beta_{2k} \neq \beta'_{2k}. \end{aligned}$$

The $p$-values are calculated as the probability of rejecting the null hypothesis.

As a complement to our randomization inference method, we also run the same regressions using OLS to consolidate our results. Consistent with the level of randomization, we cluster the standard errors at the postal zone level in our OLS regressions. These regressions also allow us to estimate the effect of other variables than the treatments themselves. Besides the baseline regressions, we run separate regressions for stores located in high-income districts (districts with average per-capita income above or equal to the median income in Hangzhou) and in low-income districts (districts with average per-capita income below or equal to the median income in Hangzhou) in order to examine whether the effect of information is the same across different income groups.

### 1.5.3 Results and Discussions

Table 1.6, Table 1.7, and Table 1.8 provide summary statistics for all the relevant variables as described in Section 1.5.1 before, during, and after the experiment period, broken down by experiment group (one control group plus three treatment groups).[16] We present our main regression results using randomization inference in Table 1.9, Table 1.10, and Table 1.11. In addition, we also report corresponding OLS regression results in Appendix A.3.1 to complement our main results.[17] The comparisons between high-income districts and low-income districts are reported in Appendix A.3.2.[18]

#### 1.5.3.1 Period 0 vs. Period 1

In Table 1.9, when examining *all products* produced by Beingmate in columns (1)–(3), we find that all the three treatments have a negative impact on the average daily sales, which is surprising compared to the existing literature. This effect is significant for Treatment 2 (Traceability) in all regression models and for Treatment 3 (Certification) when we do not

[16]Note that the number of observations varies from one period to another. This is because some stores reported sales data in one period but failed to do so in another period. The collection of products may also be different from one period to another. However, we are able to verify that none of the stores failed to submit sales data in the middle of a period. We further verify that for each period, the number of observations is a multiple of 14, which is the length of each period.

[17]The corresponding tables are Table A.2, Table A.3, and Table A.4, respectively.

[18]Table A.5, Table A.6, and Table A.7 provide the regression results and Table A.8 provides the summary of the hypothesis tests between high-income and low-income districts.

**Table 1.6: Summary Statistics for In-Store Study**
*During the Two-Week Pre-Experiment Period*

| | $N$ | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Control Group* | | | | | | |
| Daily Quantity | 11,984 | .33 | .861 | 0 | 0 | 14 |
| Daily Revenue | 11,984 | 21.6 | 98.5 | 0 | 0 | 2,256 |
| Unit Price | 11,984 | 73.4 | 77.2 | 3.5 | 35.9 | 306 |
| Local Income | 11,564 | 32,045 | 9,253 | 12,875 | 37,511 | 40,000 |
| Milk Powder Product | 11,984 | .468 | .499 | 0 | 0 | 1 |
| Store Promotion | 11,984 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 11,984 | .113 | .316 | 0 | 0 | 1 |
| *Treatment 1 (Brand) Group* | | | | | | |
| Daily Quantity | 1,316 | .266 | .683 | 0 | 0 | 6 |
| Daily Revenue | 1,316 | 14.1 | 66.6 | 0 | 0 | 1,329 |
| Unit Price | 1,316 | 62.2 | 71 | 4.61 | 25.6 | 270 |
| Local Income | 1,316 | 38,305 | 1,161 | 37,511 | 37,511 | 40,000 |
| Milk Powder Product | 1,316 | .447 | .497 | 0 | 0 | 1 |
| Store Promotion | 1,316 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 1,316 | .0912 | .288 | 0 | 0 | 1 |
| *Treatment 2 (Traceability) Group* | | | | | | |
| Daily Quantity | 672 | .26 | .911 | 0 | 0 | 15 |
| Daily Revenue | 672 | 12.5 | 50.4 | 0 | 0 | 770 |
| Unit Price | 672 | 72.1 | 72.1 | 4.62 | 40.3 | 266 |
| Local Income | 672 | 37,511 | 0 | 37,511 | 37,511 | 37,511 |
| Milk Powder Product | 672 | .542 | .499 | 0 | 1 | 1 |
| Store Promotion | 672 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 672 | 0 | 0 | 0 | 0 | 0 |
| *Treatment 3 (Certification) Group* | | | | | | |
| Daily Quantity | 1,442 | .445 | 1.13 | 0 | 0 | 13 |
| Daily Revenue | 1,442 | 27.5 | 110 | 0 | 0 | 1,411 |
| Unit Price | 1,442 | 60.7 | 65 | 4.61 | 25.6 | 241 |
| Local Income | 1,442 | 34,240 | 9,404 | 13,956 | 37,511 | 40,000 |
| Milk Powder Product | 1,442 | .476 | .5 | 0 | 0 | 1 |
| Store Promotion | 1,442 | .169 | .374 | 0 | 0 | 1 |
| Salesperson | 1,442 | .169 | .374 | 0 | 0 | 1 |

**Table 1.7: Summary Statistics for In-Store Study**
*During the Two-Week Experiment Period*

| | $N$ | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Control Group* | | | | | | |
| Daily Quantity | 12,250 | .341 | 1.03 | 0 | 0 | 33 |
| Daily Revenue | 12,250 | 24.3 | 163 | 0 | 0 | 9,867 |
| Unit Price | 12,250 | 70.9 | 76.5 | 3.5 | 35 | 299 |
| Local Income | 12,250 | 31,415 | 9,638 | 12,875 | 37,511 | 40,000 |
| Milk Powder Product | 12,250 | .453 | .498 | 0 | 0 | 1 |
| Store Promotion | 12,250 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 12,250 | .117 | .321 | 0 | 0 | 1 |
| *Treatment 1 (Brand) Group* | | | | | | |
| Daily Quantity | 1,386 | .291 | 1 | 0 | 0 | 27 |
| Daily Revenue | 1,386 | 14.7 | 64.7 | 0 | 0 | 1,201 |
| Unit Price | 1,386 | 70.6 | 81.6 | 4.61 | 25.6 | 299 |
| Local Income | 1,386 | 38,265 | 1,144 | 37,511 | 37,511 | 40,000 |
| Milk Powder Product | 1,386 | .424 | .494 | 0 | 0 | 1 |
| Store Promotion | 1,386 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 1,386 | .105 | .306 | 0 | 0 | 1 |
| *Treatment 2 (Traceability) Group* | | | | | | |
| Daily Quantity | 560 | .227 | .607 | 0 | 0 | 6 |
| Daily Revenue | 560 | 10 | 46.3 | 0 | 0 | 770 |
| Unit Price | 560 | 57.4 | 61.3 | 4.62 | 25.6 | 230 |
| Local Income | 560 | 37,511 | 0 | 37,511 | 37,511 | 37,511 |
| Milk Powder Product | 560 | .475 | .5 | 0 | 0 | 1 |
| Store Promotion | 560 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 560 | 0 | 0 | 0 | 0 | 0 |
| *Treatment 3 (Certification) Group* | | | | | | |
| Daily Quantity | 1,316 | .41 | 1.02 | 0 | 0 | 11 |
| Daily Revenue | 1,316 | 32.2 | 148 | 0 | 0 | 1,887 |
| Unit Price | 1,316 | 63.5 | 69.5 | 4.61 | 25.6 | 282 |
| Local Income | 1,316 | 35,048 | 8,526 | 13,956 | 37,511 | 40,000 |
| Milk Powder Product | 1,316 | .447 | .497 | 0 | 0 | 1 |
| Store Promotion | 1,316 | .151 | .358 | 0 | 0 | 1 |
| Salesperson | 1,316 | .163 | .37 | 0 | 0 | 1 |

**Table 1.8: Summary Statistics for In-Store Study**
*During the Two-Week Post-Experiment Period*

| | $N$ | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Control Group* | | | | | | |
| Daily Quantity | 12,264 | .341 | .929 | 0 | 0 | 15 |
| Daily Revenue | 12,264 | 21 | 98.4 | 0 | 0 | 2,691 |
| Unit Price | 12,264 | 69.9 | 74.8 | 3.5 | 34.9 | 299 |
| Local Income | 12,264 | 30,990 | 9,614 | 12,875 | 37,511 | 40,000 |
| Milk Powder Product | 12,264 | .446 | .497 | 0 | 0 | 1 |
| Store Promotion | 12,264 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 12,264 | .115 | .319 | 0 | 0 | 1 |
| *Treatment 1 (Brand) Group* | | | | | | |
| Daily Quantity | 1,386 | .258 | .889 | 0 | 0 | 20 |
| Daily Revenue | 1,386 | 12.5 | 64.9 | 0 | 0 | 1,555 |
| Unit Price | 1,386 | 66.3 | 72.2 | 4.61 | 34.8 | 288 |
| Local Income | 1,386 | 38,165 | 1,096 | 37,511 | 37,511 | 40,000 |
| Milk Powder Product | 1,386 | .455 | .498 | 0 | 0 | 1 |
| Store Promotion | 1,386 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 1,386 | 0 | 0 | 0 | 0 | 0 |
| *Treatment 2 (Traceability) Group* | | | | | | |
| Daily Quantity | 644 | .283 | .91 | 0 | 0 | 12 |
| Daily Revenue | 644 | 14 | 65.3 | 0 | 0 | 898 |
| Unit Price | 644 | 59.6 | 68.7 | 4.61 | 18.1 | 222 |
| Local Income | 644 | 37,511 | 0 | 37,511 | 37,511 | 37,511 |
| Milk Powder Product | 644 | .413 | .493 | 0 | 0 | 1 |
| Store Promotion | 644 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 644 | 0 | 0 | 0 | 0 | 0 |
| *Treatment 3 (Certification) Group* | | | | | | |
| Daily Quantity | 1,428 | .366 | .954 | 0 | 0 | 12 |
| Daily Revenue | 1,428 | 23.8 | 114 | 0 | 0 | 1,539 |
| Unit Price | 1,428 | 57.4 | 65.1 | 4.61 | 18.1 | 270 |
| Local Income | 1,428 | 34,317 | 9,174 | 13,956 | 37,511 | 40,000 |
| Milk Powder Product | 1,428 | .431 | .495 | 0 | 0 | 1 |
| Store Promotion | 1,428 | 0 | 0 | 0 | 0 | 0 |
| Salesperson | 1,428 | .172 | .378 | 0 | 0 | 1 |

**Table 1.9: Treatment Effect on Beingmate Products**

*For Period 0 vs. Period 1*

| | All Products | | | Milk Products | | | Non-Milk Products | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) Quantity | (2) Quantity | (3) Quantity | (4) Quantity | (5) Quantity | (6) Quantity | (7) Quantity | (8) Quantity | (9) Quantity |
| Period 1 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 1 × | −0.000 | −0.001 | −0.016 | −0.056 | −0.057 | −0.056 | 0.057 | 0.056 | 0.037 |
| Brand | (0.056) | (0.058) | (0.121) | (0.082) | (0.080) | (0.149) | (0.058) | (0.058) | (0.114) |
| Period 1 × | −0.064** | −0.063** | −0.054* | −0.068* | −0.066* | −0.071* | −0.070** | −0.070** | −0.062* |
| Traceability | (0.052) | (0.050) | (0.120) | (0.083) | (0.078) | (0.152) | (0.047) | (0.044) | (0.108) |
| Period 1 × | −0.061** | −0.006 | −0.035 | −0.075* | 0.002 | −0.055 | −0.053* | −0.008 | −0.021 |
| Certification | (0.054) | (0.053) | (0.115) | (0.087) | (0.085) | (0.153) | (0.043) | (0.041) | (0.099) |
| Period 1 (vs. 0) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Unit Price | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Store Promotion | No | Yes | No | No | Yes | No | No | Yes | No |
| Salesperson | No | No | Yes | No | No | Yes | No | No | Yes |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2209 | 2209 | 2209 | 1017 | 1017 | 1017 | 1192 | 1192 | 1192 |
| $R^2$ | 0.164 | 0.215 | 0.513 | 0.166 | 0.249 | 0.479 | 0.247 | 0.283 | 0.609 |

Standard errors in parentheses (generated via randomization inference)

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

control for store promotions or sales representatives. On average, compared to the control group, Treatment 2 (Traceability) reduces the average daily sales (in terms of quantity) by about 0.064 unit. Given the pre-experiment average daily sales of 0.26 in the Treatment 2 (Traceability) group, this reduction is equivalent to about 24.6% of the average daily sales before the treatment. Similarly, Treatment 3 (Certification) reduces the average daily sales by about 0.061 unit, which is equivalent to about $0.061/0.445 \approx 13.7\%$ of its average daily sales before the treatment.

When examining *only milk powder products* or *all other products than milk powder* by Beingmate in columns (4)–(6) and columns (7)–(9), respectively, we observe a similar, significantly negative effect for Treatment 2 (Traceability) in all regression models and Treatment 3 (Certification) when we do not control for store promotions or sales representatives. Although there is no obvious reason to believe that the traceability system established specially for milk powder products should have a direct impact on products other than milk powder, it seems that the negative "reminder effect" has spread to other products under the larger umbrella of infant food.

In the appendix, we compare districts with higher average income to those with lower average income in terms of the treatment effects on average daily sales.[19] For either *all products* or *only milk powder products* produced by Beingmate, we find that all treatments have a negative impact on the average daily sales for both income groups, with the effect of Treatment 2 (Traceability) being significant at 1% or 5% in all regression models. The effect of Treatment 3 (Certification) is also significant for both types of districts when we do not control for store promotions. As for the difference between high-income and low-income districts, we find that for Treatment 2 (Traceability), the negative treatment effect is significantly stronger for high-income districts at 10% level when examine *only milk powder products* produced by Beingmate. However, we do not observe any differential impact of Treatment 1 (Brand) or

[19]See Table A.5 for detailed regression results.

Treatment 3 (Certification) in any regression models.[20]

The difference between the two income groups is in line with our hypothesis. Districts with higher average income tend to be urban areas where consumers are generally better educated and thus should understand the traceability system better than those consumers in districts with lower average income. Moreover, we expect consumers from districts with higher average income to have better access to information about the traceability system, possibly through the Internet, TV, newspaper, or other promotional events by milk powder producers. However, it is interesting to see that as consumers become more aware of the traceability system, they react more negatively towards the products that are advertised to be "traceable." One possible explanation is that the "reminder effect" is stronger among wealthier and better-educated consumers who tend to have a better understanding of the limitations of the traceability system as well as its merits. Going back to our theoretical framework, we believe that the driving force of this reduction is that consumers from high-income districts have a larger $p_b' - p_b$ than consumers from low-income districts do. High-income "naïve" consumers are more likely to be reminded of the past scandals because they are more aware of the seriousness of the scandals, which gives rise to a larger belief correction. For the low-income "naïve" consumers, information has a small "reminder effect," possibly due to their limited sources of information, which leads to a smaller belief correction for consumers in low-income districts.

Other than the treatment effects, we observe a very significant effect of store promotions, all at 1% significance level. On average, store promotions on milk powder products lead to an increase of about 2.2 units for all Beingmate products, about 2.9 for just milk powder products, and about 1.8 for products other than milk powder. As for product prices, the demand for milk powder products seems to be more inelastic than the demand for other products. On average, an increase in unit price does not have any significant effect on average

[20] All the corresponding $p$-values are reported in Table A.8.

daily sales of milk powder products, but it has a significantly negative impact on other products at 1% level.

#### 1.5.3.2 Period 1 vs. Period 2

In Table 1.10, when we examine *all products* by Beingmate in columns (1)–(3), we find that the removal of Treatment 2 (Traceability) has a positive impact on the average daily sales, which is significant in all regression models. The magnitude of this treatment effect is almost symmetric to what we observed when comparing period 0 to period 1. On average, compared to the control group, the removal of Treatment 2 (Traceability) increases the average daily sales (in terms of quantity) by about 0.062, which is equivalent to about $0.062/0.227 \approx 27.3\%$ of its average daily sales during the experiment period. When examining *only milk powder products* by Beingmate in columns (4)–(6), we observe a very similar pattern. These results suggest that by providing potential consumers with information about the traceability system but not reminding them of such a fact, Beingmate is able to achieve a significantly higher quantity of sales compared to the experiment period. In other words, the "reminder effect" is temporary for Treatment 2 (Traceability). As for the effect on *all other products than milk powder* in columns (7)–(9), we do not find any significant impact of the removal of Treatment 2 (Traceability), which is in line with our expectation. However, as for Treatment 1 (Brand) or Treatment 3 (Certification) in all three categories, the removal has a significantly positive impact on sales when we only control for sales representatives or store promotions, respectively.

Then, we run separate regressions for high-income districts and low-income districts to see if the removal of each treatment has the same impact.[21] Regardless of whether we examine *all products* or *only milk powder products* produced by Beingmate, we find that the removal of Treatment 2 (Traceability) has a significant, positive impact on both high-income and

[21] See Table A.6 for detailed regression results.

**Table 1.10: Treatment Effect on Beingmate Products**

*For Period 1 vs. Period 2*

| | All Products | | | Milk Products | | | Non-Milk Products | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) Quantity | (2) Quantity | (3) Quantity | (4) Quantity | (5) Quantity | (6) Quantity | (7) Quantity | (8) Quantity | (9) Quantity |
| Period 2 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 2 × | −0.035 | −0.035 | 0.186** | 0.019 | 0.019 | 0.188** | −0.047 | −0.047 | 0.198** |
| Brand | (0.069) | (0.095) | (0.102) | (0.116) | (0.154) | (0.130) | (0.064) | (0.084) | (0.105) |
| Period 2 × | 0.062* | 0.062* | 0.066** | 0.162** | 0.162** | 0.173** | −0.027 | −0.027 | −0.028 |
| Traceability | (0.059) | (0.074) | (0.078) | (0.092) | (0.114) | (0.103) | (0.051) | (0.064) | (0.077) |
| Period 2 × | −0.043 | 0.158** | −0.054 | −0.052 | 0.250** | −0.024 | −0.035 | 0.122** | −0.070 |
| Certification | (0.072) | (0.079) | (0.083) | (0.120) | (0.132) | (0.108) | (0.054) | (0.064) | (0.077) |
| Period 2 (vs. 1) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Unit Price | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Store Promotion | No | Yes | No | No | Yes | No | No | Yes | No |
| Salesperson | No | No | Yes | No | No | Yes | No | No | Yes |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2231 | 2231 | 2231 | 998 | 998 | 998 | 1233 | 1233 | 1233 |
| $R^2$ | 0.144 | 0.158 | 0.469 | 0.141 | 0.158 | 0.467 | 0.219 | 0.233 | 0.552 |

Standard errors in parentheses (generated via randomization inference)

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

low-income districts. The removal of Treatment 1 (Brand) and the removal of Treatment 3 (Certification) do not have any significant impact at 5% level.

As for the difference between high-income and low-income districts, we are able to reject the null hypothesis that the removal of Treatment 2 (Traceability) has the same impact on high-income districts and low-income districts at 10% for *all products* and 5% for *only milk powder products* produced by Beingmate. Since our earlier results indicate that the "reminder effect" induced by Treatment 2 (Traceability) disappears with the removal of the treatment, and since consumers in high-income districts experience a stronger negative "reminder effect" in the first place, we expect the reversal, positive effect that they experience with the removal of the treatment to be stronger as well. However, as for the removal of Treatment 1 (Brand) or the removal of Treatment 3 (Certification), we cannot reject the null hypothesis that it has the same impact on high- and low-income districts at 5%.[22]

#### 1.5.3.3 Period 0 vs. Period 2

Table 1.11 summarizes the results by comparing the pre-experiment period with the post-experiment period in order to make inferences about the "pure" effect of information provision. For these two periods, there is no "reminder effect" because no information is explicitly provided to consumers. The only difference is that consumers may not know about Beingmate's traceability system and ISO9001 quality certification in the pre-experiment period, while they should all have received the information in the post-experiment period.[23] Therefore, a comparison between the pre-experiment period and the post-experiment period should yield the "pure" value of information without any "reminder effect."

For Treatment 1 (Brand) and Treatment 2 (Traceability), we do not observe any significant effect on the average daily sales in any of the regression models. This result suggests that

[22] All the corresponding $p$-values are reported in Table A.8.

[23] Based on the recommended serving size, most Beingmate products can be consumed within two weeks or less. Hence, we expect each supermarket to have similar groups of consumers from one period to another.

**Table 1.11: Treatment Effect on Beingmate Products**

*For Period 0 vs. Period 2*

| | All Products | | | Milk Products | | | Non-Milk Products | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) Quantity | (2) Quantity | (3) Quantity | (4) Quantity | (5) Quantity | (6) Quantity | (7) Quantity | (8) Quantity | (9) Quantity |
| Period 2 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 2 × | −0.027 | −0.027 | 0.154 | −0.035 | −0.035 | 0.110 | 0.016 | 0.016 | 0.218 |
| Brand | (0.056) | (0.094) | (0.135) | (0.084) | (0.141) | (0.144) | (0.071) | (0.089) | (0.150) |
| Period 2 × | −0.000 | −0.000 | 0.015 | 0.095 | 0.095 | 0.102 | −0.102 | −0.102 | −0.090 |
| Traceability | (0.060) | (0.103) | (0.140) | (0.081) | (0.154) | (0.145) | (0.076) | (0.098) | (0.152) |
| Period 2 × | $-0.098^{**}$ | 0.149 | −0.088 | $-0.115^{**}$ | 0.261 | −0.085 | $-0.093^{**}$ | 0.069 | −0.090 |
| Certification | (0.034) | (0.084) | (0.106) | (0.061) | (0.130) | (0.111) | (0.046) | (0.067) | (0.121) |
| Period 2 (vs. 0) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Unit Price | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Store Promotion | No | Yes | No | No | Yes | No | No | Yes | No |
| Salesperson | No | No | Yes | No | No | Yes | No | No | Yes |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2224 | 2224 | 2224 | 1017 | 1017 | 1017 | 1207 | 1207 | 1207 |
| $R^2$ | 0.168 | 0.190 | 0.488 | 0.166 | 0.214 | 0.492 | 0.228 | 0.237 | 0.528 |

Standard errors in parentheses (generated via randomization inference)

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

providing information about product traceability system does not have any persistent impact once the "reminder effect" is removed. For Treatment 3 (Certification), however, we observe a significant, negative effect when we do not control for store promotions or sales representatives. These results suggest that the "pure" effect of information provision may be zero except for information about the ISO9001 quality certification, which seems to have a negative, lasting effect. In other words, the negative "reminder effect" seems to have "hysteresis," or persistence, when combined with some institutional imprimatur like ISO signals. This result further suggests that not only can "wrong" types of information provision backfire, but combining information provision with credible institutions can worsen the situation and thus backfire as well.

Finally, we examine the differential impact of information provision on high-income and low-income districts.[24] In particular, when we examine *only milk powder products* by Beingmate, we do not find any significant differential impact for any of the three treatments on the average daily sales in high-income and low-income districts. However, the "pure" value of information contained in Treatment 1 (Brand) and Treatment 2 (Traceability) is significantly higher in high-income districts than in low-income districts when we examine *all products* by Beingmate, which seems to be driven by other non-milk products given our results for milk powder products.[25]

## 1.6 Conclusion

In this study, we conduct two field experiments in order to evaluate the effect of voluntary information provision in China's infant milk powder industry, which is undergoing a serious collective reputation problem. We partner with one of the largest domestic infant milk

---

[24]See Table A.7 for detailed regression results.

[25]All the corresponding $p$-values are reported in Table A.8.

powder producers in China, Beingmate, to estimate what happens to sales when we provide different types of product information about Beingmate products, including its brand name, its traceability system, and its quality certification, to potential consumers in supermarkets. We identify a "reminder effect" of information provision, which is likely to be more evident in markets with a bad image among consumers and can lead to a negative response to positive quality-related information provided.

In order to establish a causal linkage between voluntary disclosure of quality information and the "reminder effect," we conduct an online study at Beingmate's online store that involves both random information provision and a survey at the end of each purchase. The survey involves questions directly on consumers' impression of the infant milk powder industry in China and the importance they attach to various quality-related factors when choosing milk powder products. One of the interesting observations we have is that when we provide information about Beingmate's product traceability system or its ISO9001 quality certification, there is a statistically significant reduction on both the quantity of sales and consumers' likelihood of purchase. Moreover, we find that being exposed to quality-related information has a negative impact on consumers' impression of the infant milk powder industry in China, which is in line with our hypothesis about the "reminder effect."

In the in-store study, we conduct a similar experiment in supermarkets in Hangzhou, Zhejiang Province in China, where Beingmate's headquarters is located. We employ a difference-in-difference strategy to estimate the effect of each treatment by comparing the quantity of Beingmate products sold before, during, and after the experiment period in supermarkets. As in the online study, we find a negative treatment effect on the average daily sales of milk powder products when we provide information about Beingmate's product traceability system or its ISO9001 quality certification. Moreover, this negative treatment effect is significantly stronger for high-income districts than for low-income districts. After the removal of information about Beingmate's product traceability system, the average daily sales increases by almost the same magnitude as the previous reduction, which brings it back to its original level before the experiment. However, the removal of information about Beingmate's quality certification does not have an immediate positive impact; the negative impact on the average daily sales still persists even after the experiment period.

One explanation for this negative treatment effect is that seeing information about product quality may remind potential consumers of various negative images associated with the infant milk powder industry in China, thus resulting in a worse impression of the industry as a whole. Hence, even with positive quality information, this "reminder effect" can lead to a reduction in consumers' willingness to buy when it is sufficiently strong. Moreover, this negative effect is more evident in high-income districts because consumers in these districts tend to have better access to knowledge so that they have a better understanding of the limitations of the traceability system as well as its merits.

One limitation of our study is that we cannot directly test the existence of the separating equilibrium because we have little information on what consumers opt for when they are deterred from buying Beingmate products as a result of the "reminder effect." However, there are four possibilities of consumers' alternative choices: (1) not buying milk powder at all, (2) buying milk powder from another domestic producer similar to or slightly better than Beingmate, (3) buying milk powder from another domestic producer worse than Beingmate, and (4) buying milk powder from an expensive foreign brand that is perceived to be more reliable than domestic producers including Beingmate. Due to the inelasticity of infant milk powder products, the first option is highly unlikely in this specific market. Similarly, due to increasingly tight restrictions on purchasing foreign-branded milk powder products, we do not believe that most consumers are able to choose the last option. If the second possibility is at play, it should not have any impact on the existence or nonexistence of the separating equilibrium because it can be regarded as a shift within the "good" producers. Finally, if consumers opt for the third possibility, it should be detrimental to the existence of the separating equilibrium or lead to the nonexistence of the separating equilibrium because the "bad" producers now receive more purchases than before.

Another limitation is that we do not have enough evidence to conclude on the external validity of our experiments. There are two key features of the infant milk powder industry in China: potentially high health-related risks from bad products and extremely poor collective reputation among consumers. While we expect to observe a similar effect in an industry that also possesses these two features, we are unable to determine whether these factors are necessary to observing the "reminder effect." In order to test the external validity, we need to conduct further experiments that randomize on both the information provision and those characteristics that are particularly pertinent to the infant milk powder industry.

In summary, our results suggest that at least in some markets, high-quality producers are unable to improve their sales by merely providing information about their quality-improving efforts. In other words, there does not exist any separating equilibrium where high-quality producers choose to provide a costly signal about their type while low-quality producers choose to remain silent. In light of our findings, it is conceivable that mandatory information disclosure of quality-related investments might circumvent the issue created by the "reminder effect." In other words, even if the information disclosure creates a negative reminder effect, it applies to every producer in the same industry, thus encouraging the "good" producers to make and advertise these quality-related investments. This hypothesis has found some empirical support when researchers examine the effect of an increase in product quality information to consumers on firms' choices of product quality. For instance, the Los Angeles County introduced and required hygiene quality grade cards to be displayed on restaurant windows in 1998. Jin et al. (2005) show that the grade cards lead restaurant health inspection scores to increase and the number of food-borne illness hospitalizations to decrease. Consumer demand also becomes more sensitive to changes in restaurants' hygiene quality. In such cases, some form of government intervention in the form of mandatory information disclosure might be necessary in order to improve the market efficiency.

# Chapter 2

# Industrial Diversification and Corporate Innovations: Evidence from Firms in China

## 2.1 Introduction

There has been a wide range of studies on the relationship between diversification and innovation in firms, especially in the developed world. In recent decades, these two concepts have become increasingly popular in many developing countries for several reasons. First, with the rapid growth of the economy, firms seek to explore new possibilities beyond their existing line of business to accommodate new market structures. Second, the protection of intellectual property, which was almost non-existent before, has been greatly improved as legal enforcement becomes more effective. Third, the government authorities have been more supportive of firms' innovation activities from a policy perspective as they are more aware of the importance of knowledge creation in economic development. One of the most prominent examples is the Chinese government, which has been actively promoting the idea of "cross-industrial" and "cross-regional" development since early 2011 in order to induce better firm performance and to stimulate more innovation activities among domestic firms. Therefore, by exploiting the introduction of the new policy, this study aims to contribute to the literature on the impact of diversification on innovation in developing countries like China, which has not been sufficiently studied before despite increasing interest from both scholars and policy-makers.

Theoretically, there is no clear prediction whether firms achieve better innovation outcomes by focusing on a few specific industries of operations or diversifying across many possible industries. On the one hand, by focusing its operations on a single or several similar areas, a firm is able to develop and strengthen its expertise in some highly specialized areas. Since expertise is essential to innovations, focusing allows the firm to be on the frontier of certain areas in which it possesses a comparative advantage to innovate. Hence, *ceteris paribus*, innovations should be weakly increasing in a firm's expertise in a certain industry. More importantly, a new innovation project is more likely to be approved by the management of a more focused firm because there is expected to be less conflict between different potential projects within the firm. In contrast, in a more diverse firm, a new project might be crowded out by projects in other areas of operations, especially if the management is being evaluated based on other projects that might not necessarily lead to more innovations. Finally, after a firm incurs some fixed cost in entering a new industry, it no longer needs to incur the same cost when it extends its business in the same area or operates in a very similar area. This way, new innovations become less costly to the firm.

On the other hand, however, diversifying across a wide array of areas may also promote innovations for different reasons. After a firm has been operating in a certain area for a long time, generating new innovations will become increasingly difficult. In other words, there is increasing marginal cost of new innovations in the same area. Hence, the firm may, at some point, find it optimal to operate in some areas that are relatively new to them and have not been exploited too heavily by other firms. Moreover, diversification allows the firm to benefit from knowledge and technological spillovers generated by other firms in those new areas. Finally, as a merit of diversification in general, diversifying across different industries is a good way to insure against adverse risk that may affect a certain market or industry in which the firm operates.

Within the concept of diversification, there are two subcategories, namely related diversification and unrelated diversification, whose pros and cons are currently under heated debate in the literature. Related diversification, also known as focused diversification, refers to the process where a firm explores new products, technologies, markets, or industries that are different from but still have meaningful commonalities with its current operations. Unrelated diversification, sometimes referred to as unfocused diversification, takes place where the firm expands its operations to areas completely beyond its existing expertise or structures. On the one hand, related diversification may spark new innovations by making use of the firm's core business and competitiveness, while unrelated diversification, if improperly conducted, may expose the firm's core business to risk by diverting attention and resources from it. On the other hand, unrelated diversification allows the firm to fully extend the breadth of its current business and to seek new areas of innovations, while related diversification naturally adds more limit to the firm's scope of business and thus the variety of innovations.

In my empirical analysis, I examine the impact of diversification on corporate innovations measured by the number of new patents a firm receives after 2012, which is the year immediately following the Chinese government's plan to strongly encourage cross-industrial development among domestic firms. To do so, I use firm-level data on both public and private companies that are located in mainland China, own at least one domestic subsidiary, and have at least one patent by the end of 2015. For each firm, I define two diversity measures — the "narrow diversity measure" and the "broad diversity measure," which can be thought of as indicators for the degrees of related diversification and unrelated diversification. The narrow diversity measure is constructed by calculating the number of subsidiaries with distinct 4-digit NACE industrial classification codes, while the broad diversity measure is defined as the number of subsidiaries with distinct 2-digit NACE codes. In other words, the narrow diversity measure treats any two subsidiaries in different industries, related or unrelated, as distinct; however, the broad diversity measure only distinguishes between subsidiaries in two different, unrelated industries but not those in related industries. Hence, the difference between the impacts of the two diversity measures on new patents indicates the relative effectiveness of the two diversification strategies in terms of generating innovations. Besides the two diversity measures, I also classify the firms as production-oriented or service-oriented to conduct separate analyses. Moreover, I identify those firms with at least one patent after 2012 as "actively innovating firms," on which I expect a stronger impact of diversification than the rest because the recency may imply more attention to innovation activities.

In general, the empirical results are different between diversity measures and between sectors in most cases, but not so much between innovation types (i.e., all firms versus actively innovating firms). For production-oriented firms, I find a significantly positive impact of the narrow diversity measure but an insignificant impact of the broad diversity measure on innovations, which suggests that related diversification (or at least a combination of related and unrelated diversification) positively contributes to innovations while unrelated diversification alone may not. For service-oriented firms, I observe interesting and opposite results, where the broad diversity measure has a positive impact on the number of new patents in some cases but the narrow diversity measure does not in any case. Therefore, the results indicate that unrelated diversification may be more effective than related diversification in terms of generating innovations among firms in the service sector, which is different from the production sector. As for actively innovating firms, I find similar results in terms of direction and significance level as those for all firms in each sector, but the magnitudes are all larger regardless of the diversity measure I use, which is exactly the same as expected due to the innovating nature of these firms.

Apart from the main analysis, I conduct additional hypothesis tests to quantify the different impact of diversification between different diversity measures, sectors, and innovation types. First, I compare the impact of the narrow diversity measure with that of the broad diversity measure on firms in each sector. Among production-oriented firms, despite the significant impact of the narrow diversity measure and the insignificant impact of the broad diversity measure, their difference is not significant. However, for service-oriented firms, I find a significantly stronger impact of the broad diversity measure than the narrow diversity measure in some cases. Moreover, these findings hold for both all firms and just actively innovating firms. Second, given each diversity measure, I compare its impact on firms from different sectors. Using the narrow diversity measure, I find that it has a stronger impact on production-oriented firms than on service-oriented firms regardless of innovation type. However, there does not seem to be any differential impact of the broad diversity measure on firms in the production sector and those in the service sector, with the exception of actively innovating firms in a few cases. Finally, I compare the impact of the two diversity measures on firms of different innovation types in each sector. For the production sector, I find a significantly stronger impact of diversification on actively innovating firms than on all firms regardless of the diversity measure I use. However, for the service sector, I only find such significant difference using the broad diversity measure in some cases but not the narrow diversity measure in any case.

Overall, the results for production-oriented firms in my sample are largely in line with the existing literature, which suggests that related diversification, or a combination of related and unrelated diversification, is key to generating new innovations. However, the results for those service-oriented firms seem to indicate that unrelated diversification may be more effective than related diversification in terms of corporate innovations, which has rarely been documented before in the literature. In order to provide an explanation for my findings, I examine subsamples of actively innovating firms characterized by the level of technology or the intensity of knowledge according to Eurostat (2014). Specifically, for production-oriented firms, I categorize them into either high-technology or low-technology firms. For service-oriented firms, I classify them as either more knowledge-intensive or less knowledge-intensive firms. After conducting separate analysis on each group within each sector, I find that my results for the production sector are mainly driven by those high-technology firms, for which the effect of diversification, especially using the narrow diversity measure, is much greater than that for low-technology firms. In contrast, my results for the service sector are primarily driven by those less knowledge-intensive firms, for which the effect of diversification, especially using the broad diversity measure, is stronger than that for their more knowledge-intensive counterparts.

Given my subsequent analysis, a plausible explanation for the difference between related and unrelated diversification for different sectors must relate to the key features of high-technology firms in the production sector and those of the less knowledge-intensive firms in the service sector. For the production sector, especially among high-technology firms, a high level of specialization is often required for new innovations that are usually in the shape of new production technologies or new products. The specialization may refer to either physical capital such as specialized machines or production facilities, or human capital such as expert skills or knowledge in a particular field. Regardless of the form, such specialization is likely to be very difficult to be applied to another unrelated industry immediately for new innovations, which may require a completely different set of specialization. Hence, for these high-technology firms in the production sector, related diversification should be more effective in generating new innovations because it allows at least some facilities or skills to be transferable.

The reverse of the story holds for service-oriented firms, especially those less knowledge-intensive ones, for which new innovations are often virtual products. These products either do not require specialized knowledge or allow for easier transferability of knowledge due to their similar underlying structures. As a result, firms are able to explore and innovate in a wide of variety of areas, which may be more effective than merely related diversification as long as the unrelated industries are all feasible within the scope of operations.

The remainder of this chapter is organized as follows. Section 2.2 describes the government initiatives on diversification in China and some previous literature on diversification and innovation. Section 2.3 describes the data and the construction of the diversity measures. Section 2.4 lays out the empirical methodology and Section 2.5 discusses the results of my analysis. This chapter concludes with Section 2.6.

## 2.2 Background and Previous Literature

### 2.2.1 Government Promotion of Industrial Diversification

Every five years since 1953, the Chinese government issues a "Five-Year Plan" (FYP) detailing a series of social and economic development initiatives to be carried out in the next five years. Through the plenary sessions of the Central Committee and national congresses, the Communist Party of China establishes growth targets, lays out new strategies for economic development, and discusses reforms in various sectors of the economy. The full texts of each Five-Year Plan are usually published online after the plenary sessions and promoted via television, broadcast, newspaper, and other social media. The Five-Year Plans are treated very seriously by local officials and firms (both state-owned and private), who adjust their goals and strategies to be in line with the plans.[1] After the issuance of each Five-Year Plan, individual industries also make detailed plans accordingly, many of which are subsequently published online and made available to the general public.

On March 14, 2011, the National People's Congress approved the 12$^{\text{th}}$ Five-Year Plan for the following five years from 2011 to 2015, which serves as the basis for provincial and industry-specific plans.[2] This Five-Year Plan reflects China's plan to focus on "cross-industrial" (or *kua hang ye* in Chinese) and "cross-regional" (or *kua di qu* in Chinese) development by

[1]"Why China's Five-Year Plans Are so Important." (2015, October 26). *The Economist*. Retrieved December 10, 2015, from http://www.economist.com/blogs/economist-explains/2015/10/economist-explains-24.

[2]The full translated texts of the 12$^{\text{th}}$ Five-Year Plan can be retrieved at http://www.cbichina.org.cn/cbichina/upload/fckeditor/Full%20Translation%20of%20the%2012th%20Five-Year%20Plan.pdf.

encouraging firms to engage in more industrial and geographical diversification in the next five years. Not only were these two phrases promoted heavily by the media, but they also became the keywords in many industry-specific plans published on the official website of the Chinese government (www.gov.cn). For instance, the State Administration for Industry and Commerce urged agricultural firms and culture & entertainment firms to diversify their operations across industries, regions, and ownership structures.[3] The Ministry of Industry and Information Technology delivered the same message to manufacturing firms, from those in the steel and metal industry to those in the pharmaceutical industry.[4] The General Administration of Press and Publication also promoted the idea of diversification among firms in the publishing industry.[5] In addition to industry-wide plans and reforms, successful stories of diversification by individual firms have also been circulated through both official and unofficial channels as an encouragement for other firms. Overall, in terms of industrial diversification, individual firms respond to this initiative in two ways. Some firms establish or acquire new subsidiaries that operate in new industries different from their main businesses; other firms with multiple existing subsidiaries in various industries tend to strengthen their operations in those subsidiaries and to reactivate the previously "dormant" ones for new business opportunities.

### 2.2.2 Previous Literature on Diversification and Innovation

There is a rich literature on how diversification affects innovation, and how firms choose to diversify and innovate differently depending on their capital (both external and internal) and

---

[3]The original articles in Chinese can be retrieved at http://www.gov.cn/jrzg/2011-10/31/content_1982592.htm (for agricultural firms) and http://www.gov.cn/jrzg/2011-12/05/content_2010712.htm (for culture & entertainment firms).

[4]The original article in Chinese can be retrieved at http://www.gov.cn/zwgk/2013-01/22/content_2317600.htm.

[5]The original article in Chinese can be retrieved at http://www.gov.cn/gongbao/content/2011/content_1987387.htm.

labor structures (Giroud and Mueller, 2015; Jefferson et al., 2003; Ramaswamy et al., 2004; Teece, 1996). In general, the studies in this area examine various forms of diversification, including product diversification, technological diversification, international diversification, and industrial diversification. They also employ different measures of innovation, such as R&D intensity and the number of patents.

In terms of product diversification, the majority of previous work has indicated a positive relationship between diversification and innovation, measured by the number of R&D personnel employed by the firm (Gort, 1963), R&D intensity (Grabowski, 1968; Teece, 1980), or the number of patents (Scherer, 1984). Then, in terms of international diversification, Hitt et al. (1997) also find a positive impact on R&D intensity when international diversification is examined alone but a negative impact when it is interacted with product diversification. In terms of technological diversification, Garcia-Vega (2006) uses data on European R&D active companies to show that both R&D intensity and the number of patents increase with the level of technological diversification of the firm. Finally, in terms of industrial diversification, Feldman and Audretsch (1999) provide evidence for the "diversity thesis" against the "specialization thesis," which suggests that diversity of economic activities conducted by the firm better promotes innovation and subsequent economic growth.

Apart from the impact on innovation, there are also studies that investigate the relationship between diversification and other aspects of firm performance, such as profitability and firm value. For instance, using data on multinational corporations (MNCs) from Hong Kong, Wan (1998) finds that international diversification has a positive impact on both profitability growth and sales growth, but not on profitability itself, and that industrial diversification leads to more stable but lower profitability. In addition, recent work by Rong and Xiao (2014) uses data on U.S. firms and indicates a significantly positive impact of innovation-related industrial diversification on firm value. In other words, when a firm diversifies into an industry where it has more existing, applicable innovations, the firm value increases as a result of such industrial diversification. However, there is no complete unanimity with regard to the effect of diversification in the literature. For instance, in a sample of forty firms in the industrial chemicals industry, Soni et al. (1993) find that less diversified firms actually perform better than highly diversified firms. Moreover, although they find support for the hypothesis that innovation leads to better firm performance, their empirical analysis suggests that diversification may not have any impact on innovation, either positive or negative.

### 2.2.3 Related Diversification and Unrelated Diversification

In general, for any type of diversification, there are two strategies firms can pursue: related diversification and unrelated diversification. When making related diversification, firms expand their operations beyond current products, technologies, markets, or industries, but are still operating within existing expertise or structures. In contrast, when making unrelated diversification, firms go beyond the familiar to explore new products, technologies, markets, or industries that are completely new given their existing capabilities.

In particular, some pervious literature has discussed the impact of related and unrelated diversification on innovation activities for various forms of diversification. For instance, using patent data from technology-based firms between 1990 and 1999, Sugheir et al. (2012) find a positive association between related product diversification and the quantity of created technological knowledge. Meanwhile, they also present evidence for a negative association between very high levels of unrelated product diversification and the amount of knowledge creation. In terms of technological diversification, Blundell et al. (1999) find a positive effect of market share but a negative effect of industry concentration on the number of patents, which suggests a positive relationship between related diversification and innovation. Breschi et al. (2003) further present robust evidence for knowledge-relatedness as an important factor in affecting firms' technological diversification and their innovation activities. Moreover, using a panel data set on patent portfolios pertaining to firms in major industrialized regions, Leten et al. (2007) find that technological diversification has an inverted U-shaped relationship with technological performance, which is positively moderated by the technological relatedness in firms' technology portfolios. Recent work by Chen et al. (2012) presents similar results using data from firms in Taiwan's semiconductor industry. Specifically, they suggest an inverted U-shaped impact of unrelated technological diversification and a monotonically positive impact of related technological diversification on both innovation performance and corporate growth.

### 2.2.4 Diversification and Performance of Chinese Firms

Due to the complexity of diversification, most of the empirical work in this area focuses on a particular region or industry. In recent years, an increasing amount of research has been conducted in developing countries. In particular, as firms become less diversified over time in most countries, Chinese firms do not exhibit a similar trend, making them the most diversified by 2005 (Fan et al., 2008). In terms of the direction of diversification, many Chinese firms have been expanding their operations to vastly different industries, either to seek better profitability or to cater to the general trend of diversification as encouraged by the government. For instance, some machinery firms in China have been diversifying by opening banks, because profitability has been declining in their machinery operations so that they need to "cast their net far and wide in an urgent search for new business."[6]

Much of the existing literature on diversification in mainland China tends to focus on firm performance or capital structure, but little has been done in terms of the impact of diversification on innovation among Chinese firms. Zhao and Luo (2002) conduct a comparison between subsidiaries pursuing a related product diversification strategy with parents and those pursuing an unrelated diversification strategy, where they find that the former perform better in terms of both sales growth and profitability than their counterparts. Then, instead of comparing the two, Li and Wong (2003) investigate the joint effect of related

[6]"China Machinery Firms Turn Bankers, Tractor Makers as Wheezing Economy Hits Profits." (2015, February 26). *Reuters*. Retrieved December 10, 2015, from http://reut.rs/1Fzqaws.

and unrelated diversification on firm performance using data from a large number of Chinese firms. Their empirical results support the notion that related diversification is important for both resource building and utilization, while unrelated diversification is essential for institutional environmental management. They also suggest that related and unrelated diversification should be considered together in improving firm performance in emerging economies. Finally, in regard to the effect on capital structure, recent work by Su (2010) suggests opposite effects of related or unrelated diversification on capital structure, which is consistent with the prediction of organizational economics. Specifically, using data on publicly listed Chinese firms, the empirical results imply that an increase in related diversification leads to a reduction in debt and that an increase in unrelated diversification results in an increase in debt.

## 2.3 Data

I use firm-level data obtained from Orbis. This database contains both public and private companies around the world. I first select all active companies located in mainland China with at least one patent and domestic subsidiary with 100% direct ownership by the end of 2015.[7] For each firm, I obtain data related to firm characteristics, such as size, location, industry classification, primary NACE industry code, total assets, operating revenue, number of employees, number of current directors, level of independence with regard to its shareholders, etc. I am also able to obtain the industry classification for most of the subsidiaries.

The parent firms can be largely divided into three categories according to their industry classification: the primary sector, the manufacturing sector, and the service sector. According to the definition of the NACE codes published by Eurostat (2008), the primary sector is

[7] According to Orbis, "the concept of subsidiary makes no reference to the percentage of ownership between the parent and the daughter." Since there is a large variation in the percentage of ownership, I focus only on those completely owned by the parent and believe that such subsidiaries better represent the firm's scope of operations.

composed of firms with 2-digit NACE codes ranging from 01 to 09, the manufacturing sector is composed of firms with 2-digit NACE codes ranging from 10 to 35, and the service sector is composed of firms with 2-digit NACE codes ranging from 36 to 99. Due to the small number of firms in the primary sector in my sample, I combine the primary sector and the manufacturing sector as the "production sector" and keep the rest unchanged. One empirical explanation for this treatment is that the relevant government policies are largely the same for these two sectors but may not work in the same way for those firms in the service sector.

Then, for each firm, I construct two indices indicating how diverse its subsidiaries are: the "narrow diversity measure" and the "broad diversity measure." This narrow diversity measure is defined as the number of distinct industries as defined by the 4-digit NACE codes (thus providing a narrower industry classification). The broad diversity measure is defined in a similar way from the first 2 digits of the NACE codes (thus providing a broader industry classification).

In order to capture the firms' recent innovation activities, I generate a new variable indicating the number of patents approved after 2012. These numbers are obtained from going through each firm's comprehensive report in the Orbis database. It is worth noting that this variable might be zero for some firms in my sample, which suggests that these firms may not have been "actively innovating" in recent years. Hence, I define those firms with at least one patent after 2012 as "actively innovating firms" and run separate regressions on those firms as their R&D incentives might differ from the other firms.

## 2.4 Empirical Methodology

In my empirical analysis, I examine the impact of a firm's industrial diversification on its innovation outcomes since 2012. For each firm, its industrial diversification is defined as the diversity measure generated from the number of its subsidiaries in distinct industries, and its level of innovation activities is measured by the number of patents it received from the beginning of 2012 to the end of 2015. I also control for firm-specific characteristic as detailed in Section 2.3.

My main regression model is as follows:

$$\begin{aligned} patents_i \quad = \quad & \beta_0 + \beta_1 \cdot diversity_i + \beta_2 \cdot subsidiaries_i + \beta_3 \cdot assets_i + \beta_4 \cdot revenue_i + \\ & \beta_5 \cdot employees_i + \beta_6 \cdot directors_i + \beta_7 \cdot city_i + \beta_8 \cdot size_i + \beta_9 \cdot independence_i + \\ & \sum_{s \in S} D_{is} + \varepsilon_i \end{aligned}$$

where $patents_i$ is the number of patents received by firm $i$ from 2012 to 2015 and $diversity_i$ is the diversity measure of firm $i$. Some other firm-specific characteristics are captured by $subsidiaries_i$, $assets_i$, $revenue_i$, $employees_i$, $directors_i$, which are defined as the number of direct subsidiaries, total assets (in billion dollars), operating revenue (in billion dollars), number of employees (in thousands), and number of current directors of firm $i$, respectively. In order to capture the difference of innovation capacity in top-tier cities (i.e., Beijing, Shanghai, Guangzhou, and Shenzhen, where the government initiatives are usually better promoted and enforced) versus other regions, I create a variable, $city_i$, which equals 1 if firm $i$ is located in any of the four top-tier cities. In addition, I include two indicator variables $size_i$ and $independence_i$ to indicate the firm $i$'s size and level of independence with regard to its shareholders. The firm size can be either small/medium or large/very large, and its level of independence can be low, medium, or high. Finally, for each firm $i$, I include sector fixed effects $D_{is}$, where the sectors are defined generically in the Orbis database. The classification is much coarser than the 2-digit NACE classification.[8] The error term is indicated by $\varepsilon_i$. For all the regressions I conduct, I cluster the standard errors by the 4-digit NACE code of the

[8] The sector classification in the Orbis database is adapted and abstracted from the ISIC Rev.2 codes, which are developed by the United Nations Statistics Division. My dataset spans 17 broadly defined sectors by the Orbis database, including "Primary Sector," "Food, Beverages, Tobacco," "Textiles, Wearing Apparel, Leather," "Wood, Cork, Paper," "Publishing, Printing," "Chemicals, Rubber, Plastics, Non-Metallic Products," "Metals & Metal Products," "Machinery, Equipment, Furniture, Recycling," "Gas, Water, Electricity," "Construction," "Wholesale & Retail Trade," "Transport," "Post & Telecommunications," "Banks," "Public Administration & Defense," "Education, Health," and "Other Services."

parent firms.

I first use the narrow diversity measure generated from 4-digit NACE codes and conduct separate regressions on the production sector and the service sector. Then, I run similar regressions by replacing the narrow diversity measure with the broad diversity measure. After the analysis on all the firms in my sample, I select only the "actively innovating firms" to examine the effect of diversification on their innovation activities. Finally, I conduct hypothesis testing on firms in the production sector and firms in the service sector to see if the effect of diversification is the same. Moreover, I compare the effect of the narrow diversity measure with that of the broad diversity measure in another set of hypothesis tests, which will shed light on the most impactful form of diversification in terms of generating corporate innovations.

## 2.5 Results and Discussions

Table 2.1 and Table 2.2 provide summary statistics for all firms and the actively innovating firms, respectively. In each table, the firms are divided into two large sectors: the production sector (a combination of the primary sector and the manufacturing sector) and the service sector. There are 2,929 firms in total, with 2,438 in the production sector and 491 in the service sector. Among all the firms, there are 1,363 actively innovating firms in total, with 1,107 in the production sector and 256 in the service sector. It is worth noting that not all firms have valid data for all the baseline variables employed in my regression model. Specifically, among those firms with valid data for all baseline variables, there are 1,432 in the production sector and 295 in the service sector. As for the actively innovating firms, 729 are in the production sector and 159 are in the service sector.

Figure 2.1 and Figure 2.2 illustrate the distribution of the average number of patents and the average diversity measures by sector (as defined generically in the Orbis database and explained in Section 2.4), respectively.

Table 2.1: **Summary Statistics for All Firms**

| | *N* | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Production Sector* | | | | | | |
| Number of Patents After 2012 | 2307 | 94 | 929 | 0 | 0 | 29839 |
| Diversity Measure (Narrow) | 2438 | 2.19 | 2.69 | 1 | 1 | 37 |
| Diversity Measure (Broad) | 2438 | 1.88 | 1.9 | 1 | 1 | 24 |
| Number of Subsidiaries | 2438 | 6.88 | 15.6 | 1 | 2 | 306 |
| Total Assets (in billion dollars) | 2401 | 1.91 | 16.4 | 0 | .261 | 616 |
| Operating Revenue (in billion dollars) | 2397 | 1.43 | 13 | 0 | .164 | 441 |
| Number of Employees (in thousands) | 1503 | 6.87 | 24.8 | .002 | 2.05 | 544 |
| Number of Current Directors | 2438 | 3.07 | 4.06 | 0 | 1 | 39 |
| Top City Indicator | 2438 | .191 | .393 | 0 | 0 | 1 |
| Small/Medium Company | 2438 | .0788 | .269 | 0 | 0 | 1 |
| Large/Very Large Company | 2438 | .921 | .269 | 0 | 1 | 1 |
| Low Independence | 1834 | .512 | .5 | 0 | 1 | 1 |
| Medium Independence | 1834 | .326 | .469 | 0 | 0 | 1 |
| High Independence | 1834 | .162 | .368 | 0 | 0 | 1 |
| *Service Sector* | | | | | | |
| Number of Patents After 2012 | 457 | 101 | 432 | 0 | 0 | 5699 |
| Diversity Measure (Narrow) | 491 | 2.97 | 4.5 | 1 | 2 | 63 |
| Diversity Measure (Broad) | 491 | 2.47 | 2.92 | 1 | 1 | 32 |
| Number of Subsidiaries | 491 | 8.96 | 18.5 | 1 | 3 | 168 |
| Total Assets (in billion dollars) | 368 | 29.9 | 197 | 0 | .736 | 2518 |
| Operating Revenue (in billion dollars) | 363 | 3.91 | 13.3 | 0 | .441 | 108 |
| Number of Employees (in thousands) | 311 | 13.3 | 44.1 | .002 | 1.5 | 368 |
| Number of Current Directors | 491 | 3.23 | 5.9 | 0 | 1 | 41 |
| Top City Indicator | 491 | .415 | .493 | 0 | 0 | 1 |
| Small/Medium Company | 491 | .285 | .452 | 0 | 0 | 1 |
| Large/Very Large Company | 491 | .715 | .452 | 0 | 1 | 1 |
| Low Independence | 410 | .649 | .478 | 0 | 1 | 1 |
| Medium Independence | 410 | .237 | .426 | 0 | 0 | 1 |
| High Independence | 410 | .115 | .319 | 0 | 0 | 1 |

Table 2.2: **Summary Statistics for Actively Innovating Firms Since 2012**

| | $N$ | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Production Sector* | | | | | | |
| Number of Patents After 2012 | 976 | 222 | 1419 | 1 | 20.5 | 29839 |
| Diversity Measure (Narrow) | 1107 | 2.71 | 3.39 | 1 | 1 | 37 |
| Diversity Measure (Broad) | 1107 | 2.25 | 2.34 | 1 | 1 | 24 |
| Number of Subsidiaries | 1107 | 9.72 | 20.8 | 1 | 3 | 306 |
| Total Assets (in billion dollars) | 1092 | 3.33 | 24.1 | 0 | .461 | 616 |
| Operating Revenue (in billion dollars) | 1091 | 2.49 | 19.1 | 0 | .287 | 441 |
| Number of Employees (in thousands) | 773 | 9.17 | 31.8 | .002 | 3 | 544 |
| Number of Current Directors | 1107 | 3.77 | 4.93 | 0 | 2 | 39 |
| Top City Indicator | 1107 | .233 | .423 | 0 | 0 | 1 |
| Small/Medium Company | 1107 | .0407 | .198 | 0 | 0 | 1 |
| Large/Very Large Company | 1107 | .959 | .198 | 0 | 1 | 1 |
| Low Independence | 947 | .504 | .5 | 0 | 1 | 1 |
| Medium Independence | 947 | .319 | .466 | 0 | 0 | 1 |
| High Independence | 947 | .177 | .382 | 0 | 0 | 1 |
| *Service Sector* | | | | | | |
| Number of Patents After 2012 | 222 | 208 | 602 | 1 | 26 | 5699 |
| Diversity Measure (Narrow) | 256 | 3.76 | 5.66 | 1 | 2 | 63 |
| Diversity Measure (Broad) | 256 | 2.98 | 3.55 | 1 | 2 | 32 |
| Number of Subsidiaries | 256 | 11.2 | 20.9 | 1 | 4 | 125 |
| Total Assets (in billion dollars) | 201 | 47.4 | 260 | 0 | 1.19 | 2518 |
| Operating Revenue (in billion dollars) | 199 | 6.33 | 17.6 | 0 | .745 | 108 |
| Number of Employees (in thousands) | 173 | 20.6 | 57.7 | .002 | 2.4 | 368 |
| Number of Current Directors | 256 | 3.66 | 6.68 | 0 | 1 | 41 |
| Top City Indicator | 256 | .477 | .5 | 0 | 0 | 1 |
| Small/Medium Company | 256 | .25 | .434 | 0 | 0 | 1 |
| Large/Very Large Company | 256 | .75 | .434 | 0 | 1 | 1 |
| Low Independence | 225 | .693 | .462 | 0 | 1 | 1 |
| Medium Independence | 225 | .222 | .417 | 0 | 0 | 1 |
| High Independence | 225 | .0844 | .279 | 0 | 0 | 1 |

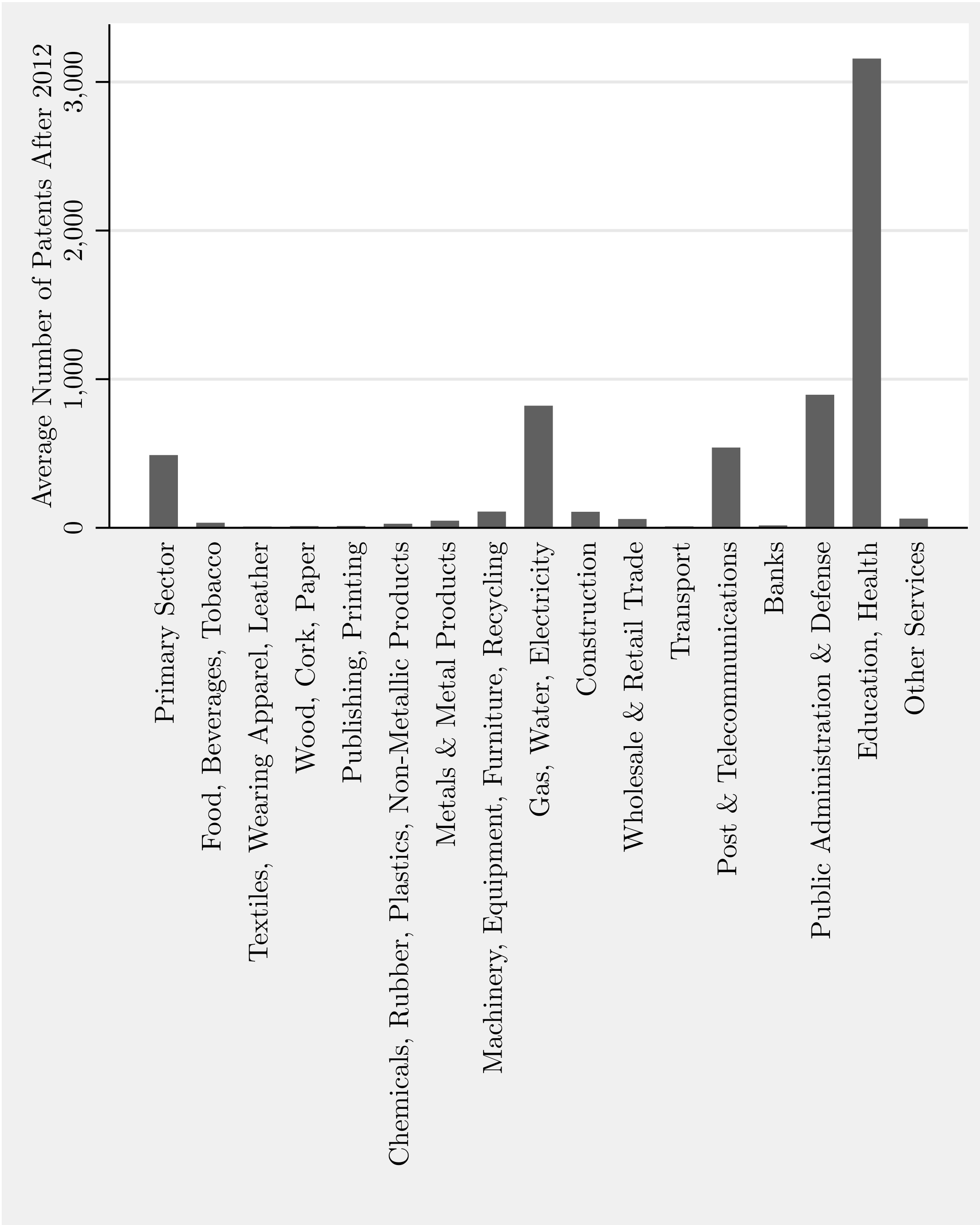



**Figure 2.1: Distribution of Average Number of Patents by Sector**

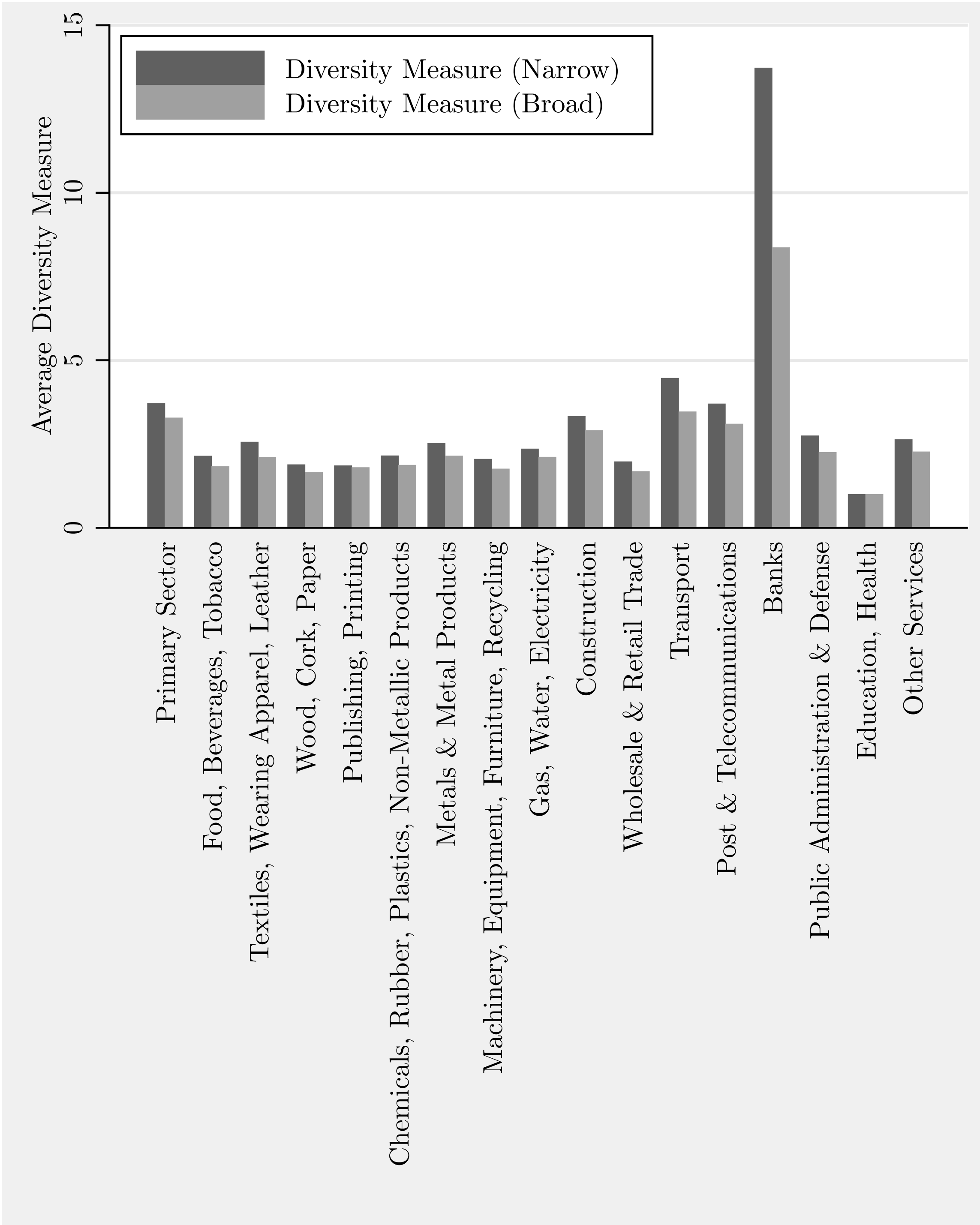



**Figure 2.2: Distribution of Average Diversity Measures by Sector**

### 2.5.1 All Firms

From Table 2.3 to Table 2.6, I present the regression results by examining the effect of diversification on innovation outcomes among *all firms* in different sectors with different diversity measures. Table 2.3 and Table 2.4 both examine firms in the production sector, with the former using the narrow diversity measure derived from 4-digit NACE industry codes and the latter using the broad diversity measure derived from 2-digit NACE codes. When using the narrow diversity measure, I find a positive impact of industrial diversification on corporate innovations among production-oriented firms after 2012. Moreover, this effect is significant in all regression models regardless of whether I control for sector dummies. In terms of the magnitude, a one-unit increase in the narrow diversity measure of a firm is expected to lead to an increase of about 43 (without sector dummies) to 54 (with sector dummies) new patents, *ceteris paribus*, which is equivalent to a 45.7% to 57.4% increase given the average number of patents of 94 in the production sector.[9] However, when I use the broad diversity measure instead of the narrow diversity measure, I no longer find any significant impact of diversification on corporate innovations, which implies that a firm is not expected to produce more innovations when it has a higher number of subsidiaries in distinct industries as defined by 2-digit NACE codes, *ceteris paribus*. Moreover, when I compare the effect of the two diversity measures, I do not find any differential impact of the narrow diversity measure from the broad diversity measure.[10]

Intuitively, my results signify the importance of related diversification versus unrelated diversification for production-oriented firms. In this case, related diversification, or at least a combination of related and unrelated diversification, is more relevant in generating corporate innovations. Firms benefit from diversifying their operations to other narrowly

---

[9]Empirically, this one-unit increase in the diversity measure could happen if a firm replaces one of its existing direct subsidiary with one in a new industry in which none of its other subsidiaries has operated before.

[10]See Table B.1 for the hypothesis test and the corresponding $p$-values.

**Table 2.3: Effect of Diversity on Corporate Innovations in the Production Sector**

*Diversity Measure Generated from 4-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 42.660* | 42.983* | 43.671* | 53.295* | 53.670* | 54.738* |
| (Narrow) | (24.740) | (24.944) | (25.640) | (27.048) | (27.288) | (28.099) |
| Direct Subsidiaries | −4.906* | −4.903* | −4.915* | −5.974** | −5.975** | −6.037** |
| | (2.792) | (2.797) | (2.805) | (2.826) | (2.833) | (2.856) |
| Total Assets | 9.973 | 9.974 | 9.751 | 9.455 | 9.454 | 9.186 |
| (in billion dollars) | (19.554) | (19.561) | (19.533) | (18.342) | (18.346) | (18.234) |
| Operating Revenue | 5.631 | 5.620 | 5.550 | 6.566 | 6.557 | 6.570 |
| (in billion dollars) | (17.047) | (17.050) | (17.099) | (16.057) | (16.059) | (16.032) |
| Number of Employees | 4.280 | 4.256 | 3.827 | 4.081 | 4.052 | 3.697 |
| (in thousands) | (5.966) | (5.969) | (6.050) | (5.931) | (5.935) | (5.979) |
| Number of Directors | 38.941 | 39.397 | 44.012 | 40.059 | 40.571 | 44.714 |
| | (27.594) | (28.031) | (30.132) | (27.629) | (28.035) | (30.413) |
| Top City Indicator | 198.729** | 200.010** | 207.217** | 164.458** | 166.250** | 168.183** |
| | (79.173) | (79.879) | (84.034) | (65.025) | (66.156) | (67.987) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | −85.835 | | | −94.765 | |
| Company | | (84.633) | | | (80.943) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −138.174 | | | −126.922 |
| | | | (93.997) | | | (92.876) |
| High Independence | | | −164.296* | | | −148.262 |
| | | | (97.653) | | | (99.493) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 1432 | 1432 | 1280 | 1420 | 1420 | 1270 |
| $R^2$ | 0.178 | 0.178 | 0.182 | 0.201 | 0.201 | 0.205 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

**Table 2.4: Effect of Diversity on Corporate Innovations in the Production Sector**

*Diversity Measure Generated from 2-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 42.408 | 42.987 | 45.065 | 53.473 | 54.115 | 56.530 |
| (Broad) | (35.450) | (35.568) | (36.659) | (39.701) | (39.913) | (40.988) |
| Direct Subsidiaries | −3.136 | −3.140 | −3.236 | −3.779 | −3.787 | −3.912 |
| | (2.414) | (2.418) | (2.499) | (2.699) | (2.705) | (2.797) |
| Total Assets | 9.765 | 9.760 | 9.505 | 9.130 | 9.123 | 8.822 |
| (in billion dollars) | (19.341) | (19.348) | (19.308) | (18.093) | (18.096) | (17.975) |
| Operating Revenue | 6.122 | 6.116 | 6.065 | 7.241 | 7.237 | 7.277 |
| (in billion dollars) | (16.791) | (16.794) | (16.835) | (15.723) | (15.724) | (15.687) |
| Number of Employees | 4.343 | 4.319 | 3.872 | 4.213 | 4.185 | 3.811 |
| (in thousands) | (6.002) | (6.006) | (6.083) | (5.938) | (5.943) | (5.985) |
| Number of Directors | 39.082 | 39.524 | 44.176 | 40.059 | 40.555 | 44.748 |
| | (28.358) | (28.771) | (30.837) | (28.402) | (28.786) | (31.129) |
| Top City Indicator | 203.903** | 205.153** | 212.281** | 170.309** | 172.057** | 173.867** |
| | (80.490) | (81.302) | (85.318) | (67.027) | (68.284) | (70.142) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | −84.823 | | | −93.354 | |
| Company | | (82.192) | | | (79.520) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −141.413 | | | −130.561 |
| | | | (94.603) | | | (92.934) |
| High Independence | | | −167.007* | | | −151.253 |
| | | | (98.861) | | | (100.088) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 1432 | 1432 | 1280 | 1420 | 1420 | 1270 |
| $R^2$ | 0.176 | 0.176 | 0.180 | 0.197 | 0.198 | 0.202 |

Standard errors in parentheses (clustered at the 4-digit industry level); $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

defined industries, but they may not gain additional benefits by having these industries in vastly different categories. In other words, it suffices for production-oriented firms to have subsidiaries in distinct 4-digit NACE industries for which some may be under the same umbrella of a more broadly defined industry, which is exactly what related diversification implies.

Other than the diversity measure, I find a significantly negative impact of direct subsidiaries when using the narrow diversity measure: one more direct subsidiary is, on average, associated with a decrease of about 5 to 6 new patents, *ceteris paribus*. Intuitively, given the same narrow diversity measure along with other variables, a higher number of direct subsidiaries implies a more narrowly focused firm because the firm must have more subsidiaries per industry on average than a firm with fewer direct subsidiaries in total. Hence, if a firm decides to develop or acquire a new direct subsidiary in a different industry from all its existing subsidiaries, the firm is expected to have $36 + (-5) = 31$ or $36 + (-6) = 30$ new patents because of the new subsidiary.[11] I do not find a similar effect of direct subsidiaries when using the broad diversity measure, which is expected given the insignificance of the broad diversity measure itself.

As for the other variables included in my regression model, I find most of them have positive yet insignificant impacts on corporate innovations regardless of the diversity measure I use. In particular, I find positive impacts from the amount of total assets, the amount of operating revenue, the number of employees, and the number of directors, but none of the effects is significant at 10% level. However, I find that being in a top-tier city is expected to significantly increase corporate innovations by about 170 (with sector dummies) to over 200 (without sector dummies). This makes sense intuitively for two reasons. One is that those top-tier cities are the center of technological progress, which is essential to the production sector. The other reason is that local governments in these top-tier cities usually put more emphasis on corporate innovations for socioeconomic and political reasons, so that firms may

[11] A new direct subsidiary in a new industry will increase both the diversity measure and the number of direct subsidiaries by 1.

feel a stronger urge to innovate. Finally, I do not find any significant impact of a firm's size or its level of independence on corporate innovations in most regressions. The only exception is when I leave out sector dummies, in which case I find that firms highly independent from shareholders are expected to have significantly fewer patents than others, possibly because new projects are less likely to be approved without a controlling shareholder (as mutual agreement across different parties might be more difficult to reach and usually takes longer due to the amount of communication required).[12]

However, when examining the service sector in Table 2.5 and Table 2.6, I find very different results. Not only are the coefficients smaller in magnitude compared to those in the production sector, but none of them is significant for the narrow diversity measure. In other words, an increase in the level of diversification across narrowly defined industries does not, on average, have any positive impact on new corporate innovation activities among service-oriented firms. For the broad diversity measure, it has a significant impact on innovations when I do not control for sector dummies and the firms' level of independence at the same time. In that case, a one-unit increase in the broad diversity measure is associated with about 12 new patents. Moreover, I find differential impact between the narrow diversity measure and the broad diversity measure.[13] Intuitively, this suggests that unlike the production sector, unrelated diversification may be better than related diversification for service-oriented firms in some cases in terms of stimulating new innovations.

When I compare the service sector with the production sector, I find that the impact of the narrow diversity measure is significantly stronger in the production sector than in the service sector in all regressions.[14] However, the broad diversity measure does not seem to have any differential impact on the two sectors. In other words, related diversification seems

[12]These "highly independent" firms are those for which none of the recorded shareholders has more than 25% of direct or total ownership.

[13]See Table B.1 for the hypothesis test and the corresponding $p$-values.

[14]See Table B.2 for the hypothesis test and the corresponding $p$-values.

**Table 2.5: Effect of Diversity on Corporate Innovations in the Service Sector**

*Diversity Measure Generated from 4-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 7.127 | 7.083 | 5.944 | 6.989 | 6.993 | 5.699 |
| (Narrow) | (4.940) | (4.920) | (4.732) | (5.545) | (5.545) | (5.337) |
| Direct Subsidiaries | −1.188 | −1.207 | −0.975 | −0.992 | −1.004 | −0.590 |
| | (0.858) | (0.864) | (0.985) | (0.967) | (0.972) | (1.068) |
| Total Assets | −0.275 | −0.272 | −0.271 | −0.204 | −0.206 | −0.225 |
| (in billion dollars) | (0.181) | (0.181) | (0.178) | (0.217) | (0.217) | (0.216) |
| Operating Revenue | −1.035 | −1.064 | −1.409 | −1.796 | −1.799 | −2.132 |
| (in billion dollars) | (4.484) | (4.509) | (4.395) | (4.853) | (4.867) | (4.712) |
| Number of Employees | 2.157 | 2.170 | 2.113 | 1.833 | 1.859 | 1.810 |
| (in thousands) | (1.843) | (1.849) | (1.833) | (2.137) | (2.146) | (2.109) |
| Number of Directors | −1.693 | −1.938 | −0.201 | 0.405 | 0.103 | 1.668 |
| | (2.540) | (2.565) | (2.486) | (3.626) | (3.661) | (3.713) |
| Top City Indicator | −49.003 | −49.851 | −73.820* | −57.402 | −57.531 | −85.960** |
| | (35.997) | (36.149) | (39.149) | (35.665) | (35.739) | (39.504) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 86.896*** | | | 69.090* | |
| Company | | (29.623) | | | (39.016) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −20.193 | | | −41.478 |
| | | | (49.165) | | | (54.922) |
| High Independence | | | −84.362** | | | −100.525** |
| | | | (35.973) | | | (43.072) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 295 | 295 | 278 | 285 | 285 | 270 |
| $R^2$ | 0.049 | 0.052 | 0.061 | 0.173 | 0.175 | 0.190 |

Standard errors in parentheses (clustered at the 4-digit industry level); $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Table 2.6: Effect of Diversity on Corporate Innovations in the Service Sector**

*Diversity Measure Generated from 2-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 12.068* | 11.766* | 9.973 | 11.909 | 11.762 | 9.879 |
| (Broad) | (7.010) | (6.953) | (6.686) | (8.013) | (7.976) | (7.777) |
| Direct Subsidiaries | −1.347 | −1.342 | −1.095 | −1.141 | −1.137 | −0.724 |
| | (0.844) | (0.845) | (0.957) | (0.969) | (0.968) | (1.053) |
| Total Assets | −0.268 | −0.265 | −0.265 | −0.201 | −0.203 | −0.223 |
| (in billion dollars) | (0.175) | (0.175) | (0.173) | (0.212) | (0.213) | (0.212) |
| Operating Revenue | −1.215 | −1.243 | −1.569 | −1.925 | −1.927 | −2.247 |
| (in billion dollars) | (4.450) | (4.475) | (4.368) | (4.812) | (4.826) | (4.674) |
| Number of Employees | 2.159 | 2.172 | 2.114 | 1.830 | 1.855 | 1.807 |
| (in thousands) | (1.838) | (1.844) | (1.829) | (2.128) | (2.138) | (2.101) |
| Number of Directors | −1.475 | −1.708 | 0.008 | 0.517 | 0.231 | 1.800 |
| | (2.555) | (2.581) | (2.523) | (3.633) | (3.676) | (3.730) |
| Top City Indicator | −48.779 | −49.627 | −73.474* | −57.671 | −57.826 | −86.031** |
| | (36.420) | (36.577) | (39.504) | (36.257) | (36.345) | (39.970) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 81.307*** | | | 64.200* | |
| Company | | (28.685) | | | (37.701) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −20.808 | | | −42.125 |
| | | | (48.915) | | | (54.705) |
| High Independence | | | −85.271** | | | −101.519** |
| | | | (35.992) | | | (43.068) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 295 | 295 | 278 | 285 | 285 | 270 |
| $R^2$ | 0.050 | 0.052 | 0.062 | 0.174 | 0.175 | 0.191 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to play a more important role among production-oriented firms than among service-oriented firms, while unrelated diversification may seem equally important or unimportant for firms in both sectors.

Apart from the diversity measure, I find that a large/very large firm, on average, produces significantly more new patents than a small/medium firm. Moreover, similar to those in the production sector, service-oriented firms that are highly independent from shareholders have significantly fewer patents than less independent ones. Most of the other variables do not have any significant impact on corporate innovations.

### 2.5.2 Actively Innovating Firms

From Table 2.7 to Table 2.10, I conduct a similar analysis as before but shift the focus to actively innovating firms only. In Table 2.7 and Table 2.8, I report my findings for those firms in the production sector using the narrow diversity measure and the broad diversity measure, respectively. With the narrow diversity measure, I again find a positive and significant impact of diversification on innovation outcomes after 2012. On average, a one-unit increase in the narrow diversity measure of a firm is associated with an increase of about 69 (without sector dummies) to 94 (with sector dummies) new patents, *ceteris paribus*. With the broad diversity measure, however, the positive coefficients remain but they are no longer significant. Moreover, the difference between the two diversity measures is not significant, which is similar to the case for all firms.[15]

Besides the diversity measure, I observe similar impacts of other variables as in the case for all firms. Using both the narrow and the broad diversity measure, I find a significantly negative impact of direct subsidiaries on corporate innovations. Moreover, the impacts of total assets, operating revenue, the number of employees, and the number of directors are all positive yet insignificant. Additionally, if an actively innovating firm in the production

[15] See Table B.1 for the hypothesis test and the corresponding $p$-values.

**Table 2.7: Effect of Diversity on Innovations among Actively Innovating Firms in the Production Sector**

*Diversity Measure Generated from 4-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 69.496* | 69.508* | 68.005* | 93.584** | 93.606** | 93.347** |
| (Narrow) | (37.985) | (37.955) | (38.169) | (43.672) | (43.541) | (43.970) |
| Direct Subsidiaries | −8.960** | −8.964** | −8.742** | −11.929*** | −11.934*** | −11.864*** |
| | (4.102) | (4.095) | (4.094) | (4.540) | (4.516) | (4.535) |
| Total Assets | 8.968 | 8.963 | 8.827 | 8.006 | 8.003 | 7.912 |
| (in billion dollars) | (19.461) | (19.489) | (19.417) | (17.524) | (17.552) | (17.531) |
| Operating Revenue | 4.889 | 4.889 | 4.728 | 6.061 | 6.060 | 5.947 |
| (in billion dollars) | (16.971) | (16.981) | (17.036) | (15.504) | (15.510) | (15.581) |
| Number of Employees | 4.722 | 4.735 | 4.154 | 4.591 | 4.599 | 4.095 |
| (in thousands) | (6.571) | (6.658) | (6.662) | (6.266) | (6.355) | (6.358) |
| Number of Directors | 54.507 | 54.460 | 59.218 | 59.032 | 59.003 | 63.203 |
| | (38.330) | (38.619) | (39.976) | (37.463) | (37.733) | (39.469) |
| Top City Indicator | 317.804** | 317.498** | 304.607** | 239.444** | 239.199** | 230.617** |
| | (129.322) | (130.261) | (126.433) | (99.114) | (100.087) | (97.526) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 32.588 | | | 20.438 | |
| Company | | (295.280) | | | (275.336) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −197.173 | | | −165.052 |
| | | | (144.120) | | | (139.880) |
| High Independence | | | −242.885* | | | −210.786 |
| | | | (136.493) | | | (138.528) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 729 | 729 | 709 | 721 | 721 | 702 |
| $R^2$ | 0.192 | 0.192 | 0.196 | 0.229 | 0.229 | 0.232 |

Standard errors in parentheses (clustered at the 4-digit industry level); $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

**Table 2.8: Effect of Diversity on Innovations among Actively Innovating Firms in the Production Sector**

*Diversity Measure Generated from 2-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 80.563 | 80.579 | 80.177 | 109.956 | 109.978 | 111.623 |
| (Broad) | (59.858) | (59.874) | (60.684) | (71.689) | (71.641) | (72.827) |
| Direct Subsidiaries | −6.983* | −6.987* | −6.908* | −9.340* | −9.344* | −9.428* |
| | (3.971) | (3.974) | (4.051) | (5.044) | (5.040) | (5.138) |
| Total Assets | 8.414 | 8.409 | 8.255 | 7.181 | 7.179 | 7.050 |
| (in billion dollars) | (19.086) | (19.114) | (19.024) | (17.067) | (17.096) | (17.052) |
| Operating Revenue | 5.762 | 5.762 | 5.582 | 7.333 | 7.332 | 7.220 |
| (in billion dollars) | (16.590) | (16.601) | (16.653) | (14.956) | (14.963) | (15.024) |
| Number of Employees | 4.736 | 4.749 | 4.148 | 4.711 | 4.716 | 4.188 |
| (in thousands) | (6.598) | (6.685) | (6.689) | (6.276) | (6.364) | (6.371) |
| Number of Directors | 54.712 | 54.668 | 59.487 | 58.932 | 58.911 | 63.192 |
| | (39.732) | (40.033) | (41.377) | (38.937) | (39.225) | (40.949) |
| Top City Indicator | 327.173** | 326.892** | 312.904** | 249.189** | 249.019** | 239.220** |
| | (131.691) | (132.747) | (128.560) | (102.715) | (103.956) | (100.882) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 30.031 | | | 14.163 | |
| Company | | (298.888) | | | (279.332) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −207.186 | | | −177.592 |
| | | | (147.312) | | | (141.738) |
| High Independence | | | −249.138* | | | −218.293 |
| | | | (141.258) | | | (143.656) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 729 | 729 | 709 | 721 | 721 | 702 |
| $R^2$ | 0.189 | 0.189 | 0.193 | 0.225 | 0.225 | 0.228 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

sector is located in a top-tier city, it is likely to have over 200-300 more new patents than its counterparts not located in a top-tier city. All these findings are in line with my expectation given the results of all firms in the production sector.

Then, I switch to service-oriented firms. In Table 2.9 and Table 2.10, similar as before, I only find significantly positive impact of diversification on corporate innovations when I use the broad diversity measure and leave out sector dummies, in which case a one-unit increase in the diversity measure is associated with an increase of 22-23 new patents. When comparing the two diversity measures, I find that the broad diversity measure has a significantly stronger impact when I do not include sector dummies, which suggests that unrelated diversification plays a more important role than related diversification in generating new patents for actively innovating firms in the service sector.[16]

When comparing the service sector with the production sector, I find slightly different results from before: the impact of the narrow diversity measure is significantly stronger in the production sector than in the service sector in all regressions, which also applies to the broad diversity measure at 10% significance level when sector dummies are included.[17]

Other than the diversity measure, most of the other variables have similar impacts on actively innovating firms as on all firms. The only exception is the "top city indicator," which has a significantly negative impact on all actively innovating firms in the service sector. In terms of the magnitude, this suggests that for a service-oriented firm, being in a top-tier city significantly reduces corporate innovations by about 124-190, which is exactly the opposite for production-oriented firms. One possible explanation is that for service-oriented firms, new innovations largely depend upon good entrepreneurial ideas, which tend to flourish in certain southeastern cities traditionally known for their entrepreneurial culture instead of traditional top-tier cities.

---

[16]See Table B.1 for the hypothesis test and the corresponding $p$-values.

[17]See Table B.2 for the hypothesis test and the corresponding $p$-values.

**Table 2.9: Effect of Diversity on Innovations among Actively Innovating Firms in the Service Sector**

*Diversity Measure Generated from 4-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 11.387 | 11.547 | 10.683 | 9.711 | 9.799 | 8.916 |
| (Narrow) | (8.690) | (8.750) | (8.766) | (9.225) | (9.291) | (9.332) |
| Direct Subsidiaries | −3.056 | −3.113* | −3.024 | −2.609 | −2.625 | −2.450 |
| | (1.829) | (1.844) | (2.192) | (2.013) | (2.023) | (2.335) |
| Total Assets | −0.307 | −0.305 | −0.315 | −0.087 | −0.094 | −0.128 |
| (in billion dollars) | (0.198) | (0.197) | (0.198) | (0.259) | (0.260) | (0.260) |
| Operating Revenue | −1.361 | −1.381 | −1.832 | −3.174 | −3.141 | −3.147 |
| (in billion dollars) | (4.521) | (4.571) | (4.324) | (4.774) | (4.800) | (4.568) |
| Number of Employees | 2.343 | 2.353 | 2.256 | 1.500 | 1.545 | 1.505 |
| (in thousands) | (1.866) | (1.875) | (1.862) | (2.291) | (2.302) | (2.280) |
| Number of Directors | −2.854 | −3.250 | −0.036 | 7.070 | 6.423 | 8.218 |
| | (4.373) | (4.372) | (4.410) | (9.134) | (9.161) | (9.610) |
| Top City Indicator | −151.593* | −154.170** | −190.470** | −124.249* | −125.967* | −163.796** |
| | (75.949) | (76.327) | (83.828) | (68.318) | (68.755) | (77.720) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 176.847*** | | | 126.462* | |
| Company | | (61.196) | | | (71.776) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −23.538 | | | −10.086 |
| | | | (92.515) | | | (86.313) |
| High Independence | | | −175.992** | | | −154.222* |
| | | | (72.130) | | | (80.725) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 159 | 159 | 153 | 156 | 156 | 150 |
| $R^2$ | 0.074 | 0.079 | 0.097 | 0.315 | 0.317 | 0.336 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

**Table 2.10: Effect of Diversity on Innovations among Actively Innovating Firms in the Service Sector**

*Diversity Measure Generated from 2-Digit Industry Codes*

| | Dependent Variable: Number of Patents After 2012 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Diversity Measure | 22.634* | 22.372* | 20.974 | 18.801 | 18.565 | 17.128 |
| (Broad) | (13.387) | (13.329) | (13.125) | (14.333) | (14.294) | (14.187) |
| Direct Subsidiaries | −3.739** | −3.731** | −3.628 | −3.099 | −3.068 | −2.876 |
| | (1.848) | (1.851) | (2.171) | (2.048) | (2.041) | (2.336) |
| Total Assets | −0.306 | −0.302 | −0.313 | −0.100 | −0.107 | −0.140 |
| (in billion dollars) | (0.189) | (0.189) | (0.190) | (0.259) | (0.261) | (0.262) |
| Operating Revenue | −1.618 | −1.650 | −2.087 | −3.318 | −3.292 | −3.294 |
| (in billion dollars) | (4.441) | (4.484) | (4.239) | (4.729) | (4.752) | (4.518) |
| Number of Employees | 2.337 | 2.347 | 2.249 | 1.522 | 1.564 | 1.523 |
| (in thousands) | (1.833) | (1.842) | (1.831) | (2.271) | (2.283) | (2.262) |
| Number of Directors | −2.594 | −2.928 | 0.258 | 6.713 | 6.124 | 7.941 |
| | (4.385) | (4.381) | (4.444) | (9.164) | (9.201) | (9.645) |
| Top City Indicator | −149.895* | −152.316* | −188.747** | −123.263* | −124.858* | −162.755** |
| | (75.886) | (76.273) | (83.458) | (68.280) | (68.743) | (77.428) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 165.280*** | | | 116.922* | |
| Company | | (58.186) | | | (69.409) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −24.441 | | | −11.467 |
| | | | (91.545) | | | (85.604) |
| High Independence | | | −177.360** | | | −155.330* |
| | | | (72.810) | | | (81.446) |
| Sector Dummies | No | No | No | Yes | Yes | Yes |
| Observations | 159 | 159 | 153 | 156 | 156 | 150 |
| $R^2$ | 0.077 | 0.082 | 0.100 | 0.317 | 0.319 | 0.338 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Finally, I formally compare the impact of industrial diversification for all firms with that for actively innovating firms.[18] My result suggests that in almost all regressions, the impact of diversification is significantly stronger for actively innovating firms in the production sector regardless of the type of diversity measures. In the service sector, the impact of the broad diversity measure is also significantly stronger for actively innovating firms when I leave out the sector dummies. The results are easy to understand: those actively innovating firms should, on average, place higher importance on innovations and be better positioned to innovate so that expanding to a new industry is more likely to spark new ideas and thus lead to more patents.

### 2.5.3 Further Discussions

From the empirical analysis, it seems that in terms of generating innovations among the Chinese firms in my sample, related diversification is more important for the production sector, while unrelated diversification plays a more prominent role for the service sector. One possible explanation pertains to the type of specialized technology and knowledge required in each sector. For production-oriented firms, new innovations are often in the form of a new production technology (e.g., a new process or a new method) or a new product (e.g., a new machine or a new drug), which may be very specific to a particular field and may require specialized knowledge or technical facilities in order to support the innovations. As a result, it may be very difficult for these production-oriented firms to explore a completely new industry and start innovating immediately given their specialized knowledge and facilities, which are not easily transferrable across vastly different industries. On the contrary, for the service-oriented firms, new innovations often come in the form of virtual products (e.g., a new financial product or some new educational software), for which the knowledge and facilities required may be similar or easily transferrable across subsidiaries even in seemingly different

[18]See Table B.3 for the hypothesis test and the corresponding $p$-values.

industries.

In order to explore this hypothesis in greater detail, I classify production-oriented firms into two categories according to their NACE codes: high-technology firms and low-technology firms. Similarly, I divide service-oriented firms into two groups: more knowledge-intensive firms and less knowledge-intensive firms. The specific categorizations are described in Eurostat (2014), where I combine "high-technology" and "medium-high-technology" to "high-technology," and "medium-low-technology" and "low-technology" to "low-technology" for the sake of simplicity. The rest remains the same. Since the regression results for actively innovating firms are similar as those for all firms in terms of direction and significance, I only focus on actively innovating firms for these comparisons. Moreover, to avoid repetitiveness, I include sector dummies in all the subsequent regressions because the conclusions will be similar with or without sector dummies.

Table 2.11 and Table 2.12 compare high-technology and low-technology firms in the production sector using the narrow diversity measure and the broad diversity measure, respectively. There are two interesting findings. First, it is easy to see that the strong effect of diversification on innovation is primarily driven by high-technology firms, where the effect of diversification is much stronger than that for low-technology firms regardless of the diversity measure I use. In other words, one should indeed focus on high-technology firms in explaining the effect of diversification in the production sector, which is in line with my assumption about production-oriented firms. Second, by shifting from the narrow diversity measure to the broad diversity measure, the effect of diversification for high-technology firms drops in both magnitude and significance level, which suggests that related diversification is especially important in terms of generating innovations for these high-technology firms in the production sector.

Table 2.13 and Table 2.14 compare more knowledge-intensive with less knowledge-intensive firms in the service sector using the narrow diversity measure and the broad diversity measure, respectively. I observe roughly the reverse as in the production sector. Specifically, the

**Table 2.11: Comparison between High-Technology and Low-Technology Firms in the Production Sector**

*Diversity Measure Generated from 4-Digit Industry Codes*

| | High-Technology | | | Low-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Patents | (2) Patents | (3) Patents | (4) Patents | (5) Patents | (6) Patents |
| Diversity Measure | 98.318*** | 98.268*** | 96.586*** | 16.595 | 16.557 | 15.719 |
| (Narrow) | (36.256) | (36.313) | (36.061) | (24.003) | (24.095) | (24.069) |
| Direct Subsidiaries | −13.448*** | −13.435*** | −12.925*** | −3.621 | −3.612 | −3.541 |
| | (3.789) | (3.788) | (3.541) | (2.660) | (2.674) | (2.696) |
| Total Assets | −10.604** | −10.620** | −10.849** | −10.270*** | −10.271*** | −10.292*** |
| (in billion dollars) | (5.125) | (5.114) | (5.129) | (0.895) | (0.897) | (0.914) |
| Operating Revenue | 20.028*** | 20.048*** | 20.156*** | 20.978*** | 20.978*** | 20.968*** |
| (in billion dollars) | (7.412) | (7.398) | (7.478) | (1.834) | (1.839) | (1.864) |
| Number of Employees | 7.009 | 6.990 | 6.400 | 14.038*** | 14.037*** | 13.941*** |
| (in thousands) | (5.136) | (5.133) | (5.238) | (0.816) | (0.818) | (0.831) |
| Number of Directors | 83.904* | 84.016* | 88.812* | −6.433 | −6.420 | −5.811 |
| | (42.514) | (42.735) | (45.498) | (6.451) | (6.477) | (6.387) |
| Top City Indicator | 296.736** | 298.030** | 284.329** | 36.000 | 36.308 | 33.421 |
| | (127.703) | (128.734) | (127.813) | (75.365) | (75.486) | (74.982) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | −123.594 | | | −29.568 | |
| Company | | (210.187) | | | (47.544) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −163.005 | | | −84.973* |
| | | | (193.177) | | | (44.792) |
| High Independence | | | −276.418 | | | −28.816 |
| | | | (195.483) | | | (32.414) |
| Sector Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 499 | 499 | 486 | 199 | 199 | 193 |
| $R^2$ | 0.339 | 0.339 | 0.345 | 0.905 | 0.905 | 0.906 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

**Table 2.12: Comparison between High-Technology and Low-Technology Firms in the Production Sector**

*Diversity Measure Generated from 2-Digit Industry Codes*

| | High-Technology | | | Low-Technology | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Patents | Patents | Patents | Patents | Patents | Patents |
| Diversity Measure | 91.807* | 91.741* | 92.534* | 29.623 | 29.587 | 28.425 |
| (Broad) | (53.738) | (53.746) | (51.686) | (35.424) | (35.580) | (35.798) |
| Direct Subsidiaries | −8.426* | −8.413* | −8.165* | −3.957 | −3.952 | −3.894 |
| | (4.778) | (4.779) | (4.544) | (2.662) | (2.678) | (2.718) |
| Total Assets | −10.731** | −10.747** | −11.072** | −10.568*** | −10.568*** | −10.579*** |
| (in billion dollars) | (4.992) | (4.981) | (4.868) | (1.082) | (1.085) | (1.105) |
| Operating Revenue | 20.857*** | 20.876*** | 21.052*** | 21.246*** | 21.246*** | 21.225*** |
| (in billion dollars) | (6.903) | (6.890) | (6.916) | (1.607) | (1.611) | (1.639) |
| Number of Employees | 7.377 | 7.357 | 6.687 | 14.133*** | 14.132*** | 14.032*** |
| (in thousands) | (4.841) | (4.837) | (4.900) | (0.865) | (0.868) | (0.877) |
| Number of Directors | 85.083* | 85.202* | 90.133* | −7.223 | −7.214 | −6.561 |
| | (44.307) | (44.531) | (47.295) | (6.863) | (6.894) | (6.755) |
| Top City Indicator | 307.891** | 309.257** | 293.703** | 27.670 | 27.870 | 24.922 |
| | (133.201) | (134.508) | (132.732) | (72.999) | (73.103) | (72.460) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | −130.891 | | | −18.520 | |
| Company | | (202.588) | | | (52.934) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −176.790 | | | −82.622* |
| | | | (188.036) | | | (42.963) |
| High Independence | | | −291.640 | | | −31.240 |
| | | | (203.439) | | | (32.428) |
| Sector Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 499 | 499 | 486 | 199 | 199 | 193 |
| $R^2$ | 0.330 | 0.330 | 0.336 | 0.906 | 0.906 | 0.907 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

**Table 2.13: Comparison between More and Less Knowledge-Intensive Firms in the Service Sector**

*Diversity Measure Generated from 4-Digit Industry Codes*

| | More Knowledge-Intensive | | | Less Knowledge-Intensive | | |
|---|---|---|---|---|---|---|
| | (1) Patents | (2) Patents | (3) Patents | (4) Patents | (5) Patents | (6) Patents |
| Diversity Measure | −3.389 | −3.221 | −3.608 | 15.556 | 14.543 | 2.413 |
| (Narrow) | (4.301) | (4.332) | (6.113) | (14.915) | (15.754) | (20.244) |
| Direct Subsidiaries | 1.578 | 1.517 | 1.398 | −4.451 | −4.237 | −2.876 |
| | (2.771) | (2.806) | (3.820) | (2.970) | (3.039) | (3.141) |
| Total Assets | 0.146 | 0.144 | 0.122 | 1.066 | 5.459 | 7.213 |
| (in billion dollars) | (0.210) | (0.211) | (0.212) | (27.951) | (26.627) | (26.931) |
| Operating Revenue | −1.009 | −0.977 | −2.120 | 5.042 | −0.935 | 6.567 |
| (in billion dollars) | (7.572) | (7.662) | (7.139) | (24.016) | (21.556) | (25.607) |
| Number of Employees | −0.380 | −0.370 | −0.601 | 21.654*** | 22.382*** | 21.391*** |
| (in thousands) | (2.180) | (2.198) | (2.230) | (3.635) | (3.384) | (3.473) |
| Number of Directors | 11.772 | 11.531 | 14.670 | −15.391 | −18.976 | −21.376 |
| | (10.708) | (10.955) | (11.033) | (13.381) | (12.921) | (15.852) |
| Top City Indicator | −209.306 | −209.254 | −284.546 | 41.467 | 40.256 | 20.149 |
| | (125.977) | (127.335) | (193.766) | (45.236) | (51.052) | (45.833) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 58.742 | | | 157.705** | |
| Company | | (111.183) | | | (57.953) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −190.161 | | | 86.822 |
| | | | (173.768) | | | (82.388) |
| High Independence | | | −263.582 | | | −9.320 |
| | | | (188.260) | | | (178.953) |
| Sector Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 65 | 65 | 62 | 42 | 42 | 41 |
| $R^2$ | 0.459 | 0.460 | 0.498 | 0.879 | 0.888 | 0.887 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

**Table 2.14: Comparison between More and Less Knowledge-Intensive Firms in the Service Sector**

*Diversity Measure Generated from 2-Digit Industry Codes*

| | More Knowledge-Intensive | | | Less Knowledge-Intensive | | |
|---|---|---|---|---|---|---|
| | (1) Patents | (2) Patents | (3) Patents | (4) Patents | (5) Patents | (6) Patents |
| Diversity Measure | 0.440 | 0.528 | −1.894 | 35.374* | 33.113 | 22.911 |
| (Broad) | (9.126) | (9.182) | (11.798) | (18.268) | (19.272) | (23.932) |
| Direct Subsidiaries | 0.319 | 0.298 | 0.468 | −5.292** | −5.031* | −4.285 |
| | (2.686) | (2.697) | (3.682) | (2.524) | (2.600) | (2.725) |
| Total Assets | 0.151 | 0.148 | 0.125 | 4.206 | 8.200 | 7.406 |
| (in billion dollars) | (0.212) | (0.213) | (0.215) | (27.819) | (26.428) | (28.707) |
| Operating Revenue | −1.214 | −1.172 | −2.234 | −0.848 | −6.174 | 1.276 |
| (in billion dollars) | (7.604) | (7.700) | (7.185) | (24.224) | (22.077) | (26.658) |
| Number of Employees | −0.371 | −0.360 | −0.588 | 22.182*** | 22.844*** | 21.787*** |
| (in thousands) | (2.174) | (2.191) | (2.229) | (3.191) | (2.965) | (3.275) |
| Number of Directors | 12.520 | 12.225 | 15.335 | −19.077 | −22.259 | −22.992 |
| | (11.002) | (11.253) | (11.557) | (13.832) | (13.448) | (17.131) |
| Top City Indicator | −202.806 | −202.963 | −280.923 | 44.833 | 43.450 | 28.753 |
| | (127.932) | (129.611) | (200.592) | (45.282) | (49.061) | (46.753) |
| Small/Medium Company | | ref. | | | ref. | |
| Large/Very Large | | 64.188 | | | 150.701*** | |
| Company | | (108.362) | | | (51.775) | |
| Low Independence | | | ref. | | | ref. |
| Medium Independence | | | −185.483 | | | 63.540 |
| | | | (176.970) | | | (80.339) |
| High Independence | | | −267.154 | | | −11.623 |
| | | | (186.373) | | | (158.323) |
| Sector Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 65 | 65 | 62 | 42 | 42 | 41 |
| $R^2$ | 0.459 | 0.459 | 0.497 | 0.885 | 0.893 | 0.890 |

Standard errors in parentheses (clustered at the 4-digit industry level); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

effect of diversification is stronger for less knowledge-intensive firms than for more knowledge-intensive firms (although most individual coefficients are insignificant), which means that my results for service-oriented firms are mainly driven by the less knowledge-intensive ones. This is also in line with my assumption about service-oriented firms. Moreover, by using the broad diversity measure instead of the narrow diversity measure, the effect of diversification increases in both magnitude and significance level for less knowledge-intensive firms. This signifies the importance of unrelated diversification for these less knowledge-intensive firms, for which diversifying broadly across many different industries that share transferrable skills is not only feasible but also more effective in generating new innovations than diversifying only across a few related industries.

## 2.6 Conclusion

In this chapter, I study whether firms are able to generate more innovations through focusing on a few specific areas of operations or diversifying across many possible areas. Since early 2011, the Chinese government has been encouraging industrial diversification for domestic firms in order to promote not only corporate profitability but also innovation activities. Specifically, with the government's new initiatives of "cross-industrial" and "cross-regional" development as an exogenous variation, I use firms' patent data after 2012 to examine the impact of the diversification of their subsidiaries on the firms' innovation outcomes.

There are two types of diversification that I explore in this study: related diversification and unrelated diversification. Through related diversification, a firm diversifies the operations of its subsidiaries mostly across related industries, which are defined as those with the same first two digits of the NACE codes. In contrast, through unrelated diversification, the subsidiaries of a firm can scatter across a variety of industries, including those vastly different ones, which are defined as those with different first two digits of the NACE codes. I then construct two diversity measures, the narrow diversity measure and the broad diversity measure, to indicate the degree of firm's related diversification and unrelated diversification, respectively.

In my empirical analysis, I examine the impact of the two diversity measures on the number of new patents a firm receives after 2012. The firms are divided into two groups by sector: the production-oriented firms and the service-oriented firms. In addition, I conduct separate analysis on "actively innovating firms," which are defined as firms with at least one patent after 2012.

Overall, for each diversity measure, I find similar results whether I examine all firms or only actively innovating firms, especially in terms of the direction and the significance level of the impact on corporate innovations. In particular, for production-oriented firms, the narrow diversity measure has a positive impact on innovations, while the broad diversity measure does not have any significant impact. This implies that related diversification, possibly combined with unrelated diversification, contributes significantly to innovations but that unrelated diversification alone may not. The reverse is true for service-oriented firms, where I find that the broad diversity measure has a positive impact on innovations in certain cases but that the narrow diversity measure may not lead to more innovations. Hence, it seems that unrelated diversification plays a more important role for the service sector than related diversification does in terms of generating innovations.

After the main analysis, I perform three sets of comparisons between diversity measures (narrow vs. broad), sectors (production vs. service), and innovation types (all vs. actively innovating). In terms of the narrow diversity measure versus the broad diversity measure, they do not seem to have any differential impact among product-oriented firms, but the broad diversity measure has a significantly stronger impact on service-oriented firms in some cases. As for the production-oriented firms versus service-oriented firms, I find that the narrow diversity measure has a stronger impact on firms in the production sector than those in the service sector, while the broad diversity measure does not have any differential impact between the two sectors. Finally, when comparing the impacts on all firms versus on actively innovating firms only, I find that regardless of the diversity measure, the impact of diversification is significantly stronger on actively innovating firms than on all firms in the production sector. In the service sector, the stronger impact remains only when I use the broad diversity measure in certain cases.

In general, the empirical study suggests that among the Chinese firms in my sample, the production sector relies more on related diversification while the service sector relies more on unrelated diversification in terms of generating new innovations. There are several possible explanations for this difference. One explanation, which is supported by my subsequent analysis, has to do with different natures of production-oriented firms and service-oriented firms. After comparing high-technology and low-technology firms in the production sector, I find that my results for production-oriented firms are mainly driven by those high-technology ones. For those high-technology firms in the production sector, typical innovations often come in the form of a new production technology or a new product, which may require very specific technical facilities, knowledge, or skills. Given the non-transferability of these tangible and intangible assets, related diversification would allow firms to explore another area and start innovating shortly afterwards by relying on their existing work and expertise. In contrast, when comparing more knowledge-intensive firms to less knowledge-intensive firms in the service sector, I find that my results for service-oriented firms are mainly driven by the less knowledge-intensive ones. For those firms, typical innovations usually take the form of virtual products, which may not require an extremely high level of specialization in a narrow field. In other words, the knowledge or facilities required for generating new innovations may be easier to transfer across subsidiaries, which makes unrelated diversification more appealing for service-oriented firms.

One limitation of my study is that there is no perfect mapping between the NACE codes and the industry information of the patents. Hence, it is hard to determine the channel through which diversification contributes to innovations. In other words, with the existing data, one cannot discern whether the new innovations result directly from new subsidiaries in new industries or from their spillovers to existing subsidiaries due to diversification. Further research may be useful in understanding the mechanism itself, especially in mainland China and some other developing countries where the concept of industrial diversification is still a relatively new topic to be fully studied.

# Chapter 3

# Complementarity of Formal and Informal Finance: Evidence from Bidding ROSCAs in India[1]

## 3.1 Introduction

Traditionally, informal financial institutions, especially those in developing countries, are often pictured as the "last resort" for credit-constrained entities to obtain credit (Aryeetey, 2002; Germidis et al., 1991; Masuyama et al., 1999). Indeed, for either impoverished individuals or startup firms, informal finance is a crucial channel for satisfying their consumption needs or providing financial support for their expansion of production. For instance, many small and medium enterprises (SMEs) in India are characterized as "credit-constrained" due to their lack of collateral and financial credibility (Banerjee and Duflo, 2012; Beck, 2007; Madestam, 2010). As a result, they rely primarily on informal financing options from private moneylenders, trusts, or credit agencies in order to survive and expand.

However, it should not be taken for granted that informal financing channels are simply

[1] This chapter is built upon my previous work on bidding ROSCAs (Wang, 2011).

imperfect substitutes for formal financing channels. On the contrary, both survey results and anecdotal evidence seem to suggest that in some areas where formal finance is available, informal finance is still very common and sometimes preferred by many individuals and small firms. According to a substantial survey conducted by Allen et al. (2012a) in 2004 among 212 SME executives, the importance of formal financing channels may have been overestimated. Their survey results suggest that many firms continue to seek informal financing options even after they have been granted formal financing opportunities. In certain cases, informal finance accounts for an impressive 80% of the total financing sources for small businesses. Using the survey data of Allen et al. (2012a), I generate two figures to illustrate their findings. Figure 3.1 depicts the importance of different financing sources at the startup of a firm and suggests that SME executives often value financial support from family and friends much more than bank loans. Figure 3.2 describes the ease of obtaining financing during the growth stage of the SMEs. Unsurprisingly, bank loans are indeed more difficult to obtain than funds from informal sources. The responses from SME executives further suggest that sometimes, even so-called "credit-constrained" entities do have at least some choices and preferences over their financing sources, including the formal ones. In this regard, informal finance may have their own advantages that cannot be easily replaced by formal financing options.

These intriguing observations imply the reverse to some traditional views on informal finance: rather than being the last resort for all credit-constrained entities, informal financing channels are actually preferred by many individuals and business owners that do have some access to the formal financial market (Allen et al., 2005, 2012a; Jain, 1999). Moreover, there are other studies that aim to address whether and why formal financial institutions coexist with the informal or less formal ones. For instance, Madestam (2010) contends that weak legal institutions play an important role in explaining the coexistence of formal and informal financial institutions in developing economies, while Fang et al. (2011) suggest that informal financial markets serve as an insurance mechanism for participants in the formal financial markets and thus are not crowded out by their formal counterparts.

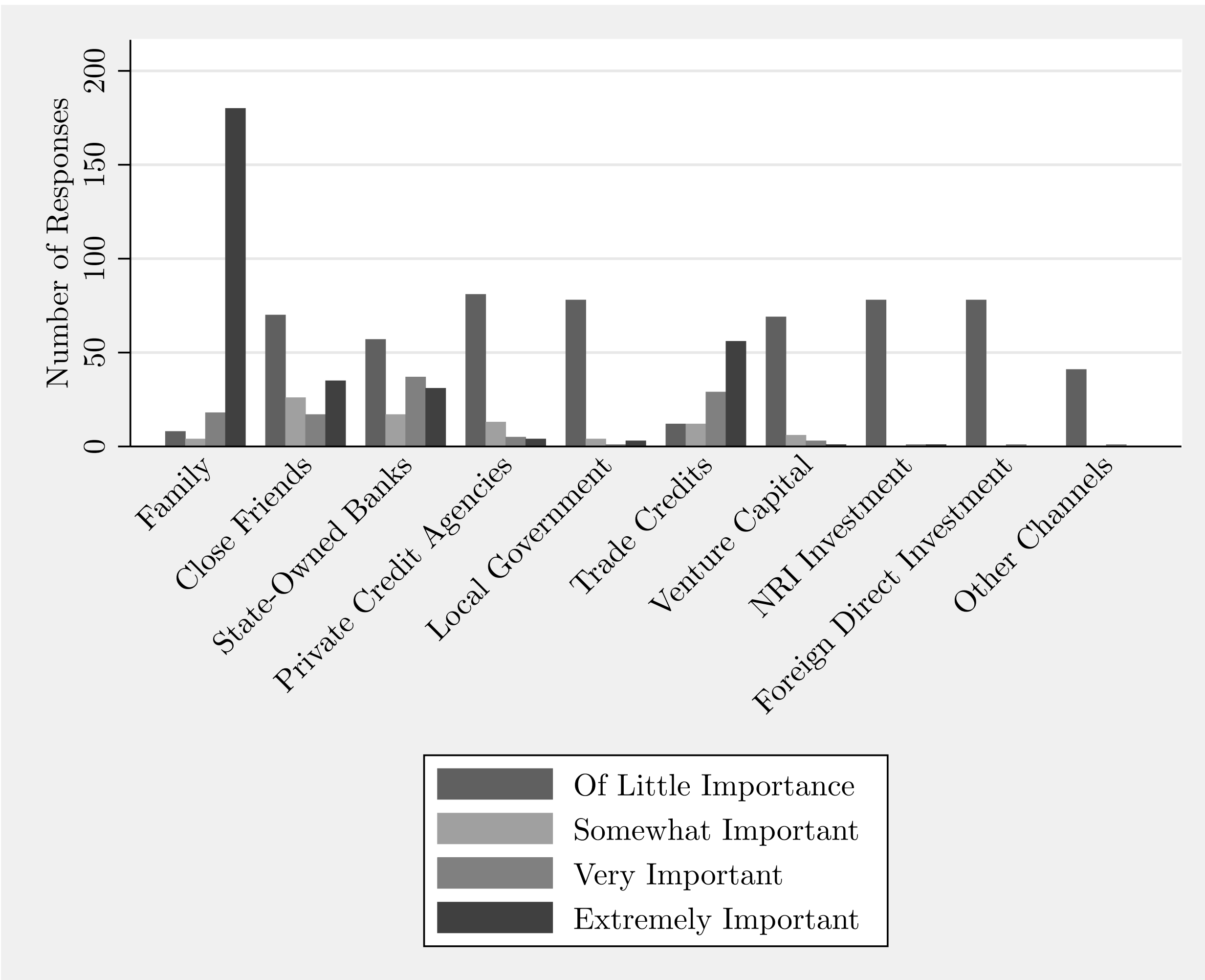



**Figure 3.1: Importance of Various Sources of Financing at Startup Stage**

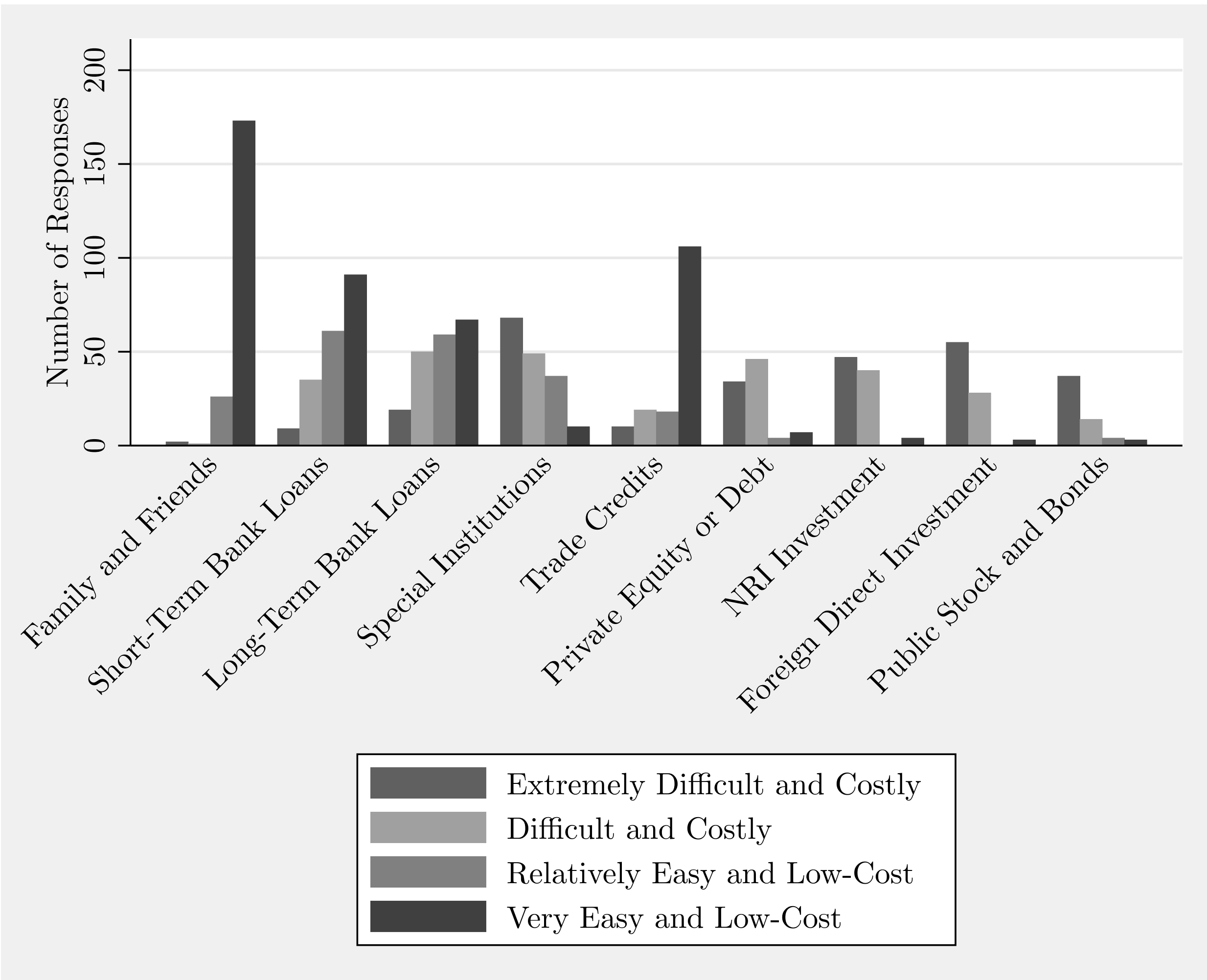



**Figure 3.2: Ease of Obtaining Financing during Growth Stage**

In light of the previous literature on informal finance, this study focuses on bidding ROSCAs, an important form of informal finance, and aims to investigate how informal financial institutions are influenced by newly emerged formal financial institutions. Specifically, I first develop a theoretical framework to capture the participation decisions and bidding behaviors of ROSCA participants with and without the existence of banks. Then, I test the predictions from the theoretical model using data from 219 ROSCA branches and over 5,000 banks in Andhra Pradesh, India. Since the financial liberalization in 1991, there has been a large number of reforms in India's formal financial sector, leading to a considerable increase in the number of banks in the previously unbanked or underbanked areas (United Nations: Economic and Social Commission for Asia and the Pacific, 2002). Using exogenous variation in the number of nearby banks from January 1998 to December 2000, the change in ROSCA participation, winning bids, and amount of default due to bank openings will partly reflect the influence of formal finance on informal finance.[2]

The theoretical model has several predictions. First, unlike in standard second-price sealed-bid auctions, almost all ROSCA participants have an incentive to overbid except for the one with the highest return to investment. Second, it predicts an ambiguous effect of nearby bank openings on ROSCA participation. On the one hand, a traditional "last resort" view on informal finance might expect the emergence of banks to *decrease* the attractiveness of bidding ROSCAs to their participants and thus the overall participation. On the other hand, however, one must also recognize that an increased availability of banks (and potentially increased amount of credit available from formal channels) enables those in severe poverty to engage in financial activities in general. Thus, ROSCA participation may actually *increase*. By looking at six different possibilities, I show that in most cases except for two extreme ones, the introduction of banks *does not affect* ROSCA participation. Finally, in the case where ROSCAs sustain, nearby bank openings weakly decrease ROSCA winning bids, which can be

[2]Potential caveats of the data are discussed in Section 3.4 and Appendix C.3.

explained by the increased competition between financial institutions that drives down the cost of capital.

In the empirical study, I test some of the model's predictions and also examine the amount of default in the ROSCAs. The regression results suggest that overall, bank openings have a positive yet insignificant impact on ROSCA participation. However, after breaking down the banks by type, I find that rural banks represented by Grameen banks have the strongest, significant, and positive impact on ROSCA participation among all types of banks. In terms of ROSCA winning bids, the empirical results confirm the theoretical prediction: bank emergence indeed has a significant, negative impact on ROSCA winning bids. When I look at different types of banks, the only exception seems to be the private banks. Finally, in terms of the ROSCA default, I find that instead of deteriorating the operations in the ROSCAs, nearby bank openings contribute to a reduction in the ROSCA default and that this effect is again significant. Using different regression specifications, I find that Grameen banks almost always have the strongest effect on ROSCA default among all types of banks, possibly due to the strong ties between ROSCAs and rural banks. Overall, the empirical results in this study seem to confirm the hypothesis that not only do ROSCAs and banks coexist, but that the formal financial institutions also foster the development of the informal ones in various aspects.

The difference between this project and the previous literature, such as Allen et al. (2012a) and Fang et al. (2011), is that not only do I show the coexistence of formal and informal financial institutions, but that I am also interested in how formal finance can be regarded as a complement to rather than a substitute for informal finance in some cases. In other words, previous literature puts a heavier emphasis on the channel through which people choose to participate in both formal and informal finance at the same time, but I would like to investigate whether formal finance is actually able to foster participation in informal finance and reduce the default rate as well as the cost of capital in informal financial institutions.

The remainder of this chapter is organized as follows. Section 3.2 describes the background and some previous literature. Section 3.3 sets up the theoretical framework to characterize the bidding behaviors of ROSCA participants with and without banks. Section 3.4 describes the data source used for the empirical analysis. Section 3.5 lays out the empirical strategy and the results are presented in Section 3.6. This chapter concludes with Section 3.7.

## 3.2 Background and Previous Literature

Rotating savings and credit associations (ROSCAs) constitute an important type of non-banking financial institutions worldwide (Besley and Levenson, 1996). Originated from India, ROSCAs are prevalent in many other countries, including China, Pakistan, South Africa, Japan, France, and even part of the United States. One major characteristic is that ROSCAs serve the impoverished and credit-constrained entities via a rotation scheme, where participants "take turns" to be lenders and borrowers. In general, a ROSCA can be formally defined as "a voluntary grouping of individuals who agree to contribute financially at each of a set of uniformly spaced dates towards the creation of a fund, which will then be allotted in accordance with some prearranged principle to each member of the group in turn" (Calomiris and Rajaraman, 1998).

Normally, a ROSCA can take two different forms: a *random* ROSCA or a *bidding* ROSCA, and they differ by the assignment of the ROSCA fund. In a random ROSCA, participants commit to putting a fixed sum of money into a "pot" for each period of the ROSCA. Then, in order to determine the recipient of the fund, a lottery will be conducted in each period after the collection of participants' mandatory contributions. The process will repeat in each period, and the winner will then be excluded from future lotteries, but is still obliged to make contributions until the end of the ROSCA. In contrast, in a bidding ROSCA, an auction will be conducted in place of the lottery so that the time at which a participant receives the fund might be associated with his or her own valuation of the fund and the expected return of future investment. Similarly, the winner of the auction will also be excluded from future auctions and continue to make the required dues for the rest of the ROSCA.

Previous literature, such as Besley et al. (1993) and Besley et al. (1994), has compared random ROSCAs with bidding ROSCAs in terms of their efficiency of allocation. Assuming the primary goal of homogeneous ROSCA participants is to purchase indivisible, durable goods, their papers suggest that random ROSCAs are weakly preferred to bidding ROSCAs. However, with the gradual integration of formal and informal credit markets over time, the value of ROSCA earnings is no longer limited to consumption needs, and for the analysis in this study, I will focus on bidding ROSCAs rather than random ROSCAs.

Bidding ROSCAs are growing rapidly in many regions of the world. For instance, in some slums in Kenya, the participation rates in ROSCAs can be as high as 57% per household (Anderson and Baland, 2002). Unlike formal financial institutions, the specific forms of bidding ROSCAs vary across countries, and the role of these informal financial intermediaries has been evolving over time as well. For example, in the 1980s when bidding ROSCAs started to be an interesting topic for economists, the main goal for credit-constrained ROSCA participants was to accumulate funds to purchase indivisible, durable goods (Besley et al., 1993; Bouman, 1995). However, in the late 1990s, bidding ROSCAs mushroomed in Southeast China and some states in South India, where the participants were mainly owners of small businesses with some access to formal financial markets. Instead of saving up for consumption purposes, these business owners, among others, utilize bidding ROSCAs as a form of insurance against unexpected productivity shocks (Allen et al., 2012b, 2007, 2012a; Calomiris and Rajaraman, 1998; Fang et al., 2011; Rao, 2007; Satkunasingam and Shanmugam, 2006; Zhang, 2008). As a result, ROSCAs become prevalent in many parts of the world even with the presence of formal credit markets fully or partially available to ROSCA participants.

As in the formal credit markets, default is also a serious problem among ROSCAs. Since ROSCA participants are usually entities without steady income or an established credit history, default is likely to take place from time to time despite relatively strict procedures in admitting new ROSCA members. In particular, in bidding ROSCAs, participants might overestimate their ability to repay or are too impatient to wait for the fund. Consequently, a ROSCA member may sometimes submit "irrationally" high bids in order to be the winner earlier in the life of the ROSCA, but fails to realize the risk involved until he or she defaults and thus gets prevented from future ROSCA participation (Eeckhout and Munshi, 2010). Moreover, defaulting from a bidding ROSCA not only makes the defaulter lose some of the future privileges, but it is also costly to other participants because everyone's payoff depends on whether other members in the same ROSCA pay their dues on time. For instance, in Hong Kong, a large fraction of bidding ROSCAs have been shut down by the government since 2009, partly because too many participants were unable to pay their ROSCA dues on time. Moreover, in recent years, stricter regulations on ROSCA operations have been passed and implemented in India (especially in Andhra Pradesh, a large and populated state in South India) in order to prevent a microcredit crisis like the ones in 2004 and 2010.

A more detailed description of the bidding ROSCAs, which is largely based on the work of Rao (2007), is provided in Appendix C.1, which includes the importance of bidding ROSCAs, their benefits for the poor and credit-constrained entities, the bidding mechanisms, the usage of ROSCA winnings, and the problems associated with bidding ROSCAs. It is worth noting that in many Asian countries, funds raised through ROSCAs are also called *chit funds*, and institutions conducting ROSCA activities are often referred to as *chit fund companies*.[3]

## 3.3 Theoretical Framework

In the theoretical framework, I characterize the bidding behaviors of ROSCA participants with and without banks. I first provide a simple example illustrating the auction mechanisms in Section 3.3.1, and then introduce an $n$-player model without and with access to the formal credit market in Section 3.3.2 and Section 3.3.3, respectively.

[3]This term is originated from India. See Klonner (2004) for the precise definition of chit funds.

Although in real life, ROSCAs are conducted in the form of open ascending-bid auctions, standard auction theory implies that one can model them as a variation of second-price sealed-bid auctions (Besley et al., 1993; Eeckhout and Munshi, 2010). Therefore, for the rest of the chapter, I will consider ROSCAs as series of second-price sealed-bid auctions instead of open ascending-bid auctions.

### 3.3.1 A Simple Example with Three Participants

Suppose that there are three participants (and hence three months) of a given bidding ROSCA. Each month, each participant makes a required contribution of \$100 so that the chit value or the "denomination" is \$300. Suppose that the winning bid is \$120 in the first month so that each losing participant receives a dividend of \$60. In the second month, there are only two eligible bidders. Suppose that the winning bid is \$100 so that each of the other two participants receives a dividend of \$50. In the last month, there is only one eligible bidder and thus the winning bid is zero.

The net gains (indicated by positive numbers) and contributions (indicated by negative numbers) are summarized in Table 3.1.

**Table 3.1: Payoffs for Each ROSCA Participant in the Example**

| | Month 1 | Month 2 | Month 3 |
|---|---|---|---|
| First Recipient | \$80 | −\$50 | −\$100 |
| Second Recipient | −\$40 | \$100 | −\$100 |
| Third Recipient | −\$40 | −\$50 | \$200 |

Note that the first recipient is a borrower, receiving \$80 in the first month and paying a net of \$50 and \$100 in subsequent months. The last recipient is a saver, saving \$40 and \$50 in the first two months and receiving \$200 in the last period. In this case, the second recipient is a partial saver and partial borrower (or a net borrower in this example by receiving \$100 and paying a total of \$140).

## 3.3.2 Model for Bidding ROSCAs without Banks

Suppose there are $n$ credit-constrained individuals having no access to the formal credit market. An important assumption in this model is that ROSCA participants have perfect information about each other. This assumption is plausible because ROSCA participants are often villager or neighbors who know each other's financial situation and ability to repay.

Due to publicly available information, any $n$-player bidding ROSCA can be modeled as a two-player auction, and the bidding equilibrium can be immediately generalized from a two-player model to an $n$-player model. Specifically, instead of considering $n$ participants with $n$ different returns to investment, I only consider the two participants with the two highest returns and explain how the results apply to an $n$-player model. With each participant's return observable to others, I show below that all the eligible bidders have an incentive to overbid except for the one with the highest return in each round of the auctions.

### 3.3.2.1 Model Setup

Consider a two-player bidding ROSCA that lasts for two periods. An auction is conducted only in period 1, because there is only one eligible bidder in the second period and thus no need for another auction. In each period, each player contributes 1 to the ROSCA fund (i.e., the "pot"), and the winner in the first period will receive the entire pot to make an investment. The winner has a net investment return of $y$, which can be treated as the productivity of the winner. The loser in the first period will not be able to make any investment in the first period due to the lack of funds. The productivity levels of the two players are drawn independently from the same continuous distribution $F(\cdot)$, and each player observes everyone's productivity. A repayment of $b$ is due from the winner of the first period.

In the second period, since there is only one eligible bidder left, the loser of the first period will automatically receive the pot after a contribution of 1 from each player. The winner in the first period will receive a net investment return of $2y$ in the second period. For simplicity, I assume that the players do not discount future incomes and that the commission is ignored because it is relatively small and can be treated as a sunk cost.

As explained before, open-ascending bid auctions in the ROSCAs can be modeled as second-price sealed-bid auctions. However, it is worth noting that such second-price sealed-bid auctions are slightly different from the standard ones, because in a classic second-price sealed-bid auction, the loser will get a payoff of zero. In contrast, the ROSCA participant who loses in the first auction will still make a positive profit of $b$, where $b$ is this participant's own bid in the first period. Therefore, for almost all participants (i.e., all potential losers of each round), it is in their best interest to overbid in order to gain a higher dividend payoff from the winner.

Recall that in a standard second-price sealed-bid auction, each bidder simply bids his or her true valuation. Now, it is important to calculate each ROSCA participant's valuation of the "pot" in order to characterize the Nash equilibrium, if any. Suppose that the productivity (i.e., investment return) of player $i$ is $y_i$ if this player wins the auction and thus invests the fund. Let $v_i$ be the amount that player $i$ is willing to pay for winning the pot. In the case of winning the auction, player $i$'s total payoff is $2y_i - v_i$ due to the investment return of $2y_i$ and a repayment of $v_i$ to the loser. Suppose, instead, that player $i$ loses the auction. The payoff in that case is merely $v_i$, since $v_i$ can be also regarded as the amount that player $i$ requires to get in order to compensate for losing the auction.

Hence, player $i$'s valuation $v_i$ for the "pot" in the first period can be calculated by equalizing the payoffs in both cases as described above, that is,

$$2y_i - v_i = v_i \implies v_i = y_i$$

so that each player's valuation $v_i$ is exactly the same as the player's productivity $y_i$.[4]

[4]Note that each player's valuation will be unaffected even if the commission is captured in the model, because in either case, the player is required to pay a fixed sunk cost regardless of the winning status.

The payoffs of the winner and the loser of the first period are summarized in Table 3.2.

**Table 3.2: Payoffs for Each ROSCA Participant with No Banks**

| | Payoff in the First Period | Payoff in the Second Period | Total Payoff |
|---|---|---|---|
| Winner | $-1+2-v_i$ | $-1+2y_i$ | $2y_i - v_i$ |
| Loser | $-1+v_i$ | $-1+2$ | $v_i$ |

#### 3.3.2.2 Bidding Equilibrium

Let $y_1$, $y_2$ be the productivity of player 1 and player 2, respectively. Since $F(\cdot)$ is a continuous distribution, $P(y_1 = y_2) = 0$. Without loss of generality, I assume that $y_1 > y_2$ so that player 1 has higher productivity and player 2 has lower productivity. I then solve for the bidding equilibrium and show that the player with lower productivity has an incentive to overbid.

**Proposition 3.1.** *Suppose player 1 has higher productivity than player 2, that is, $y_1 > y_2$. Then, there exists a Nash equilibrium where player 1 bids $y_1 = v_1$ and player 2 bids $y_1 - \varepsilon > y_2$, where $\varepsilon > 0$ is arbitrarily close to 0.*[5]

*Proof.* See Appendix C.2. □

From the proof, it has been shown that the optimal bidding strategies of ROSCA participants are determined by both their own productivity and that of their fellow participants. Specifically, those individuals who are more productive will end up being net borrowers by bidding a higher amount equal to their own productivity, while those less productive participants would rather bid as net savers and thus earn the dividend paid by the borrowers as their interest payment.

[5] There is no "closed form" of the Nash equilibrium (i.e., without using an arbitrarily small $\varepsilon$) due to the fact that there is no constraint on the bids in the model. In real life, however, one cannot bid arbitrarily close to the highest bid because there exists some smallest unit of currency $\Delta$, such as one rupee or one dollar, or even one paisa or one cent. In that case, the Nash equilibrium can be written as $(y_1, y_1 - \Delta)$, which becomes well defined.

In a ROSCA with $n$ participants, the result from a two-player auction can be immediately generalized. In each of the $n$ rounds of the ROSCA, the eligible bidder $i$ with the highest productivity will bid up to the true valuation $v_i$, while all the other eligible bidders will overbid by bidding slightly under $v_i$.

Consequently, bidding ROSCAs are able to provide financial support for different types of individuals in different ways. In this sense, bidding ROSCAs do have an advantage over formal financial institutions due to their flexibility, because in theory, they can automatically "screen" their participants through the bidding mechanism, while banks must conduct a rather complicated due diligence process in order to assess each applicant's ability to repay and determine the optimal loan package.

## 3.3.3 Model for Bidding ROSCAs with Banks

### 3.3.3.1 Model Setup

This section discusses how the market of bidding ROSCAs is affected by the formal credit markets that become available to ROSCA participants, that is, whether the auction mechanisms are effective enough for ROSCAs to sustain.

In the discussions below, assumptions from Section 3.3.2 still hold, and both players can now borrow from the bank at an interest rate of $r_b$ or save at an interest rate of $r_s$, where $r_s < r_b$. Suppose that apart from contributing 1 to the ROSCA fund in the first period, each player now has two other options that do not involve ROSCA participation: (a) saving 1 in the bank in order to collect the interest, and (b) borrowing an additional 1 from the bank and make an investment of 2.

Table 3.3 summarizes player $i$'s payoff in each stage for each option.

Moreover, I assume that player $i$ observes the productivity of both players with certainty before deciding whether to participate in the bidding ROSCA or to pursue banking instead. Player $i$ will be able to participate in the auction if and only if both players find it profitable to join the ROSCA.

**Table 3.3: Payoffs for Each ROSCA Participant with Banks**

| | First Period Payoff | Second Period Payoff | Total Payoff |
|---|---|---|---|
| Winner of the ROSCA | $-1+2-v_i$ | $-1+2y_i$ | $2y_i-v_i$ |
| Loser of the ROSCA | $-1+v_i$ | $-1+2$ | $v_i$ |
| Saving | $-1$ | $1+r_s$ | $r_s$ |
| Borrowing | $1$ | $2y_i-r_b-1$ | $2y_i-r_b$ |

#### 3.3.3.2 Bidding Equilibrium

Without loss of generality, I again assume that player 1 has higher productivity than player 2, that is, $y_1 > y_2$. It is a natural extension to check whether the ROSCA sustains with the introduction of banks. There are six possibilities as detailed below:

1. $r_s < r_b \leq y_2 < y_1$ (both players have high productivity);
2. $r_s \leq y_2 < r_b \leq y_1$ (one has high productivity and the other has mediocre productivity);
3. $y_2 \leq r_s < r_b \leq y_1$ (one has high productivity and the other has low productivity);
4. $r_s \leq y_2 < y_1 \leq r_b$ (both players have mediocre productivity);
5. $y_2 \leq r_s < y_1 \leq r_b$ (one has mediocre productivity and the other has low productivity);
6. $y_2 < y_1 \leq r_s < r_b$ (both players have low productivity).

**Proposition 3.2.** *There will be no bidding ROSCA if possibility 1 holds.*

*Proof.* See Appendix C.2. □

**Proposition 3.3.** *If possibility 2 holds, then the bidding ROSCA will sustain. In the Nash equilibrium, player 1 bids $r_b$ and player 2 bids $r_b - \varepsilon$, where $\varepsilon > 0$ is arbitrarily close to 0.*

*Proof.* See Appendix C.2. □

**Proposition 3.4.** *If possibility 3 holds, then the bidding ROSCA will sustain. In the Nash equilibrium, player 1 bids $r_b$ and player 2 bids $r_b - \varepsilon$, where $\varepsilon > 0$ is arbitrarily close to 0.*

*Proof.* See Appendix C.2. □

Note that the two possibilities above capture a negative effect of banks on ROSCA winning bids and thus the cost of capital faced by ROSCA participants. Originally, without banks, the equilibrium bids are $(y_1, y_1 - \varepsilon)$. After the introduction of banks, the new equilibrium bids are $(r_b, r_b - \varepsilon)$. Since $r_b \leq y_1$, both players now bid lower than before, which suggests that the emergence of banks leads to lower bids in the bidding ROSCA.

**Proposition 3.5.** *If possibility 4 holds, then the bidding ROSCA will sustain. In the Nash equilibrium, player 1 bids $y_1$ and player 2 bids $y_1 - \varepsilon$, where $\varepsilon > 0$ is arbitrarily close to 0.*

*Proof.* See Appendix C.2. □

**Proposition 3.6.** *If possibility 5 holds, then the bidding ROSCA will sustain. In the Nash equilibrium, player 1 bids $y_1$ and player 2 bids $y_1 - \varepsilon$, where $\varepsilon > 0$ is arbitrarily close to 0.*

*Proof.* See Appendix C.2. □

In the two possibilities above, the winning bids remain exactly the same as before.

**Proposition 3.7.** *There will be no bidding ROSCA if possibility 6 holds.*

*Proof.* See Appendix C.2. □

Table 3.4 summarizes the six possibilities discussed above.

It can be easily observed that bidding ROSCAs will be crowded out by banks only in two extreme cases, i.e., $r_s < r_b \leq y_2 < y_1$ or $y_2 < y_1 \leq r_s < r_b$. In the first case, both participants have excellent investment plans since both $y_1$ and $y_2$ are large enough compared to the interest rates offered at the bank. According to Proposition 3.1, the participant with

**Table 3.4: ROSCA Sustainability with the Introduction of Banks**

| Cases | Sustainability | Winning Bid | Winner's Payoff | Loser's Payoff |
|---|---|---|---|---|
| $r_s < r_b \leq y_2 < y_1$ | No | — | — | — |
| $r_s \leq y_2 < r_b \leq y_1$ | Yes | $r_b$ | $2y_1 - r_b + \varepsilon$ | $r_b - \varepsilon$ |
| $y_2 \leq r_s < r_b \leq y_1$ | Yes | $r_b$ | $2y_1 - r_b + \varepsilon$ | $r_b - \varepsilon$ |
| $r_s \leq y_2 < y_1 \leq r_b$ | Yes | $y_1$ | $y_1 + \varepsilon$ | $y_1 - \varepsilon$ |
| $y_2 \leq r_s < y_1 \leq r_b$ | Yes | $y_1$ | $y_1 + \varepsilon$ | $y_1 - \varepsilon$ |
| $y_2 < y_1 \leq r_s < r_b$ | No | — | — | — |

lower productivity has an incentive to overbid by bidding just below $y_1$. With a high $y_1$, the cost of borrowing through bidding ROSCA is high. Hence, borrowing money from the bank is relatively cheaper for the participant with higher productivity to obtain funds. Therefore, the participant with higher productivity is willing to abandon the bidding ROSCA and borrow from the bank at an interest rate of $r_b$.

On the other hand, the second case ($y_2 < y_1 \leq r_s < r_b$) suggests that if neither of the participants has a high-yield investment plan, they will then have an incentive to "play safe" by collecting interest payments on savings from the bank instead. In this case, even the more productive participant will be better off by saving money in the bank. Due to the low productivity of both players, it is no longer optimal for them to participate in the bidding ROSCA once there is an alternative financing channel.

For all other possibilities, however, the bidding ROSCA will sustain despite the introduction of banks. One explanation for this interesting result is that ROSCAs offer a relatively flexible and cheap way for their participants to obtain credit. If it is not the case that both participants happen to be extraordinarily productive or unproductive in terms of their investment returns at the same time, participating in the bidding ROSCA will make both parties better off by weakly increasing their payoffs. Therefore, in most cases, the bidding mechanism of ROSCAs are "competitive" enough for them to coexist with the formal credit market because the implied interest rates (on either savings or borrowings) offered by the ROSCAs are often more favorable than those offered by the banks.

## 3.4 Data Source

In order to test the coexistence and complementarity of banks and bidding ROSCAs, I use the following two main data sources to conduct my empirical analysis. The details of creating the datasets are provided in Appendix C.3.

### 3.4.1 Geographical Information of ROSCAs and Bank Branches

This dataset contains geographical information of 219 ROSCA branches and over 5,000 banks in Andhra Pradesh, India. Most of the geographical information is obtained using ArcGIS and Google Earth with the approximations based on banks' "pincodes" (similar to zip codes in the U.S.). Specifically, this dataset contains the following variables:

1. Type (state-owned, nationalized, private, or rural) of each bank;
2. Exact date when each bank was opened;
3. Exact date when each ROSCA branch was opened;
4. Number of banks of each type within a certain distance from each ROSCA branch. For the empirical study, I only report the results using 10 km as the distance. However, I have conducted the same analysis using other distances as a robustness check.

Figure 3.3 through Figure 3.5 illustrate some key information in this dataset. Figure 3.3 maps the banks and ROSCA branches in Andhra Pradesh, and Figure 3.4 adds a "buffer circle" with a radius of 10 kilometers from each of the 219 ROSCA branches. Figure 3.5 shows the average number of nearby banks of different types in each month.

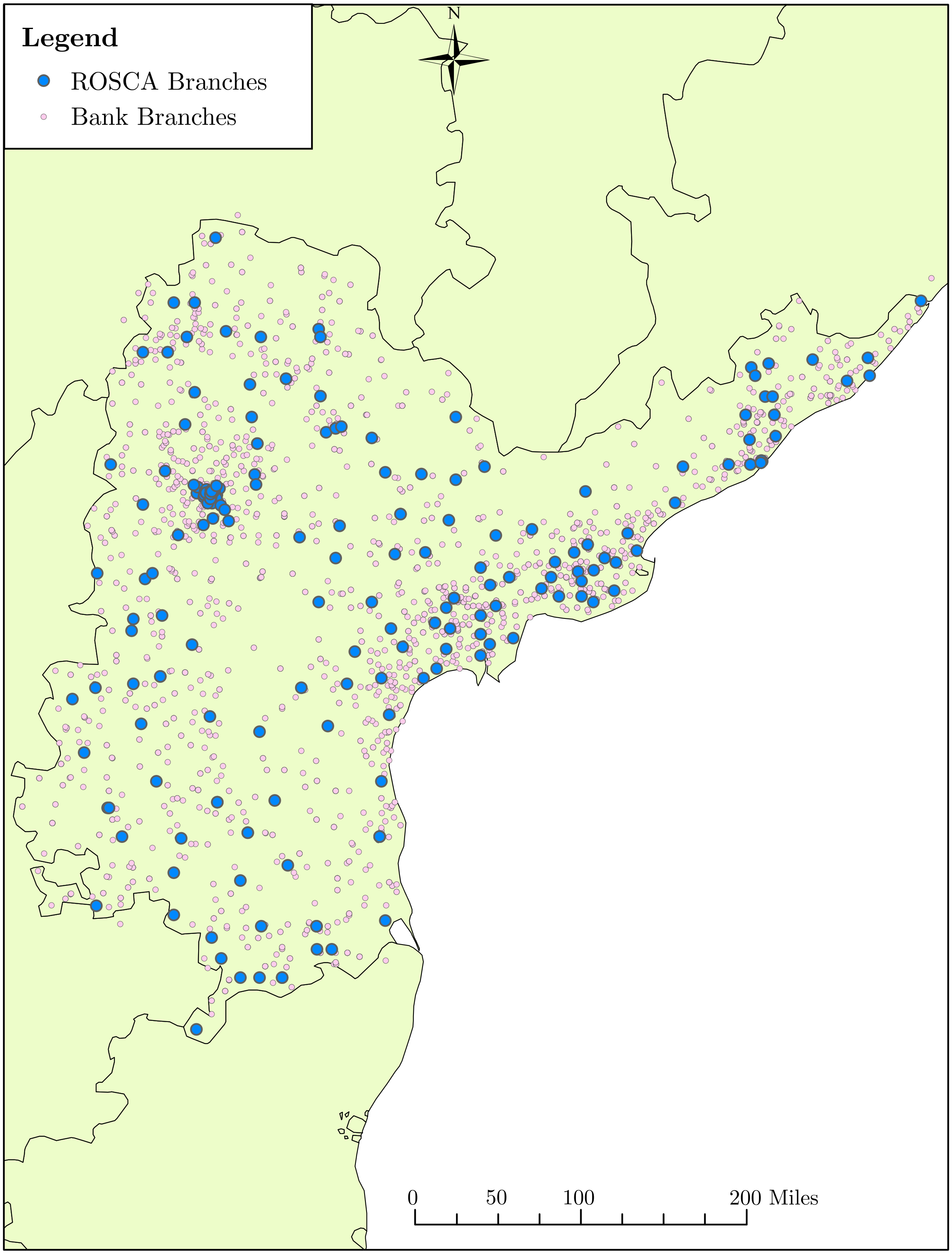



**Figure 3.3: Bidding ROSCAs and Banks in Andhra Pradesh, India**

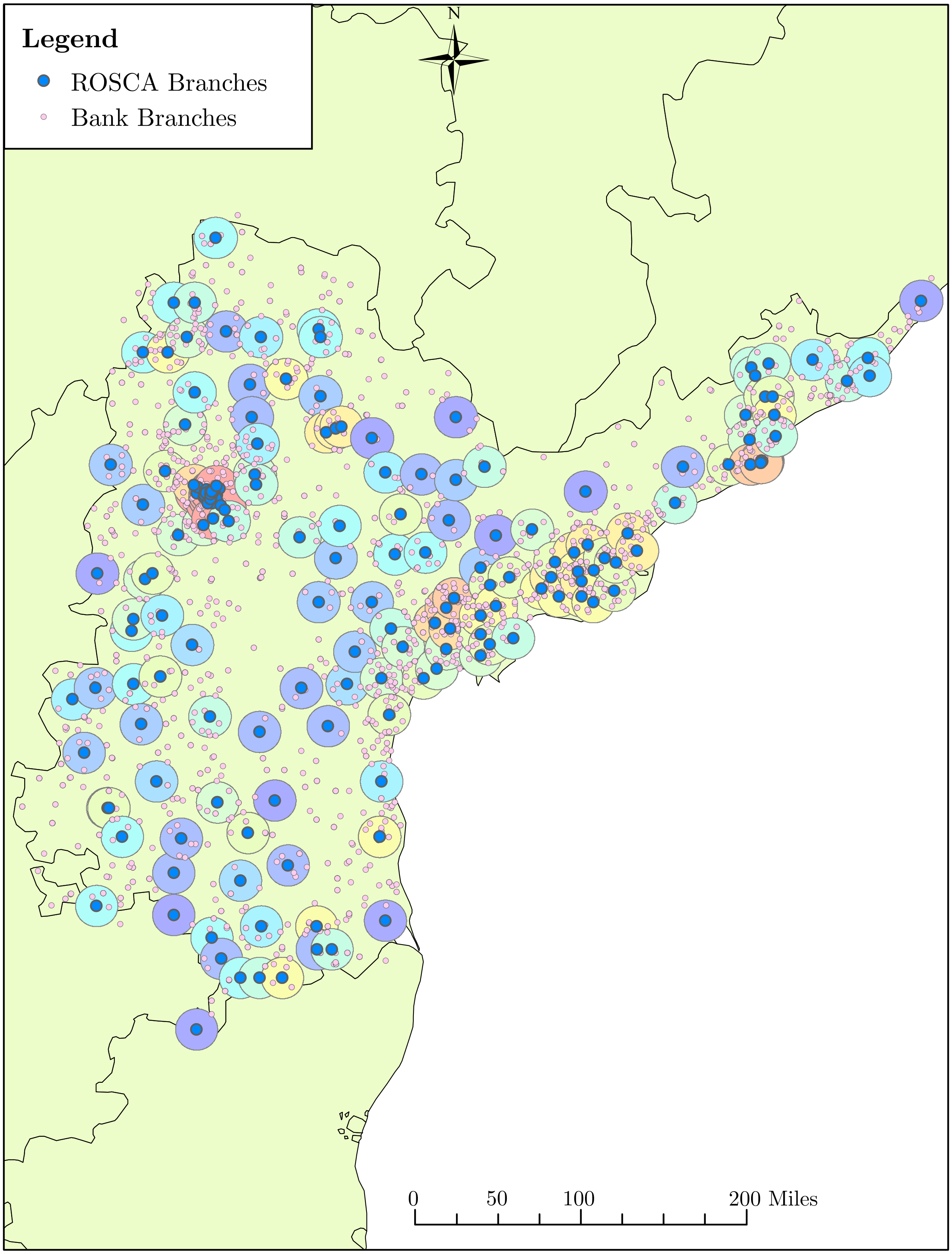



**Figure 3.4: Bidding ROSCAs and Banks with Buffer Circles (10 km)**

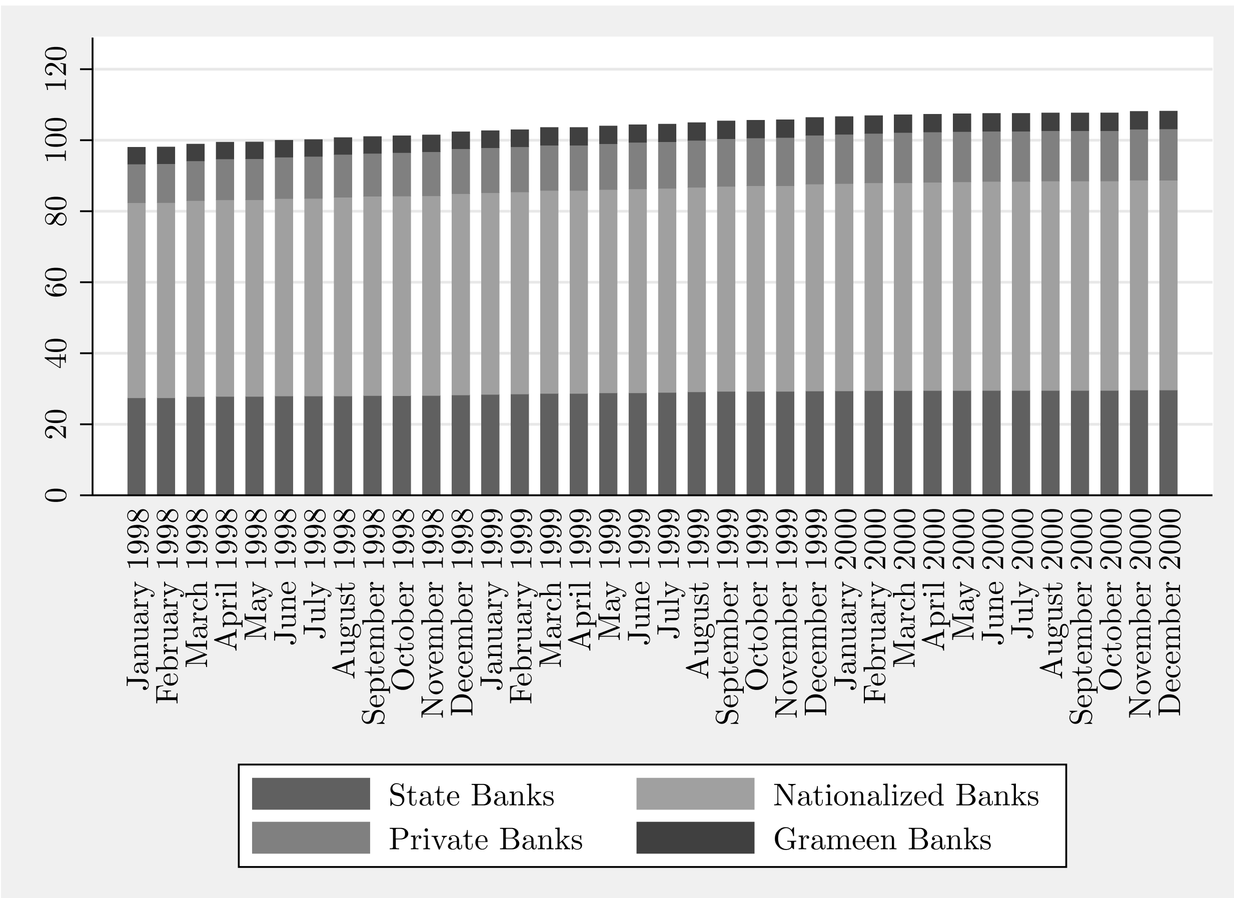



**Figure 3.5: Average Number of Neighboring Banks by Month**

### 3.4.2 Dataset on Bidding ROSCAs in Andhra Pradesh

This dataset contains the following information about the participation, winning bids, and the amount of default in each bidding ROSCA in each of the 219 branches located in Andhra Pradesh, India:

1. Number of participants in each bidding ROSCA branch in each month;
2. Winning bid in each round of auctions in each bidding ROSCA branch;
3. Amount of remaining default after each round of auctions in each bidding ROSCA branch;
4. Other auction-specific characteristics such as chit value and duration of each auction in each bidding ROSCA branch.

These data will be exploited in order to estimate the impact of nearby bank openings on the bidding behaviors of the ROSCA participants. For ROSCA participation, I use the aggregate number of participants in each ROSCA branch in each month since the empirical study will be conducted on the branch level instead of individual auction level. Note that hundreds of auctions usually take place in one ROSCA branch in a given month.

The summary statistics are provided in Table 3.5, and the numbers of auctions each month are provided in Table 3.6. Note that the number of auctions tends to increase in April or October each year, and remain mostly flat in other months of the year. Another interesting fact to note is that the starting dates of the ROSCAs (not shown in the tables) tend to peak around April and October as well, which is most likely due to two major Hindu festivals, Holi in March and Diwali in late October or early November.

Figure 3.6 depicts the average number of ROSCA participants by month. Figure 3.7 and Figure 3.8 are histograms of the distribution of ROSCA winning bids and amount of ROSCA default, respectively.

### 3.4.3 Limitations of the Existing Data

One major limitation is the short time span of the ROSCA data. In spite of a long history of ROSCAs in India, only in recent decades have their operations been digitalized so that most of the existing data were collected after 1995. Moreover, some ROSCAs can last for years and they had not yet reached the end when the data were collected. Therefore, I drop those unfinished bidding ROSCAs for my analysis of ROSCA default since default is recorded only after a given ROSCA ends.

As for the banks, their operational data after the year 2000 were unavailable at the time when I created the dataset. In this case, the only period during which I have a reasonably large amount of data on both the ROSCAs and the banks is from January 1998 to December 2000, which is hence selected as the time span to focus on for this study. Due to the relatively short time span, there is not a large amount of variation of the number of nearby banks as desired, which might undermine the efficiency of the fixed-effects regressions conducted in the study.

**Table 3.5: Summary Statistics**

| | $N$ | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| Chit Value | 263250 | 62635 | 96894 | 5000 | 50000 | 2500000 |
| Number of ROSCA Participants | 6156 | 108 | 219 | 0 | 0 | 2580 |
| Duration of Auctions (in months) | 14888 | 17.7 | 10.6 | 1 | 18 | 36 |
| Winning Bid Amount | 263250 | 18316 | 36402 | 0 | 8700 | 1560000 |
| Winning Bid (% of Chit Value) | 263250 | 25.4 | 16.3 | 0 | 23.3 | 76.8 |
| Default Amount | 204270 | 3119 | 13945 | 0 | 500 | 1651000 |
| Default (% of Chit Value) | 204270 | 4.13 | 8.57 | 0 | 2 | 91.6 |
| Total Number of Banks | 6156 | 105 | 200 | 0 | 21 | 710 |
| Number of State Banks | 6156 | 28.8 | 55.6 | 0 | 5 | 197 |
| Number of Nationalized Banks | 6156 | 57.3 | 114 | 0 | 11 | 403 |
| Number of Private Banks | 6156 | 13 | 26.1 | 0 | 2 | 94 |
| Number of Grameen Banks | 6156 | 4.9 | 4.24 | 0 | 4 | 17 |

Table 3.6: **Distribution of the Number of Auctions by Month**

| Year | Month | $N$ | Frequency (%) | Cumulative Frequency (%) |
|---|---|---|---|---|
| 1998 | January | 4828 | 1.8 | 1.8 |
| 1998 | February | 4839 | 1.8 | 3.6 |
| 1998 | March | 4864 | 1.8 | 5.5 |
| 1998 | April | 5397 | 2.0 | 7.5 |
| 1998 | May | 5480 | 2.1 | 9.5 |
| 1998 | June | 5553 | 2.1 | 11.6 |
| 1998 | July | 5537 | 2.1 | 13.7 |
| 1998 | August | 5510 | 2.1 | 15.8 |
| 1998 | September | 5532 | 2.1 | 17.9 |
| 1998 | October | 6315 | 2.4 | 20.2 |
| 1998 | November | 6479 | 2.4 | 22.7 |
| 1998 | December | 6533 | 2.5 | 25.1 |
| 1999 | January | 6522 | 2.4 | 27.6 |
| 1999 | February | 6491 | 2.4 | 30.0 |
| 1999 | March | 6558 | 2.5 | 32.5 |
| 1999 | April | 7219 | 2.7 | 35.2 |
| 1999 | May | 7465 | 2.8 | 38.0 |
| 1999 | June | 7468 | 2.8 | 40.8 |
| 1999 | July | 7503 | 2.8 | 43.6 |
| 1999 | August | 7438 | 2.8 | 46.4 |
| 1999 | September | 7457 | 2.8 | 49.2 |
| 1999 | October | 7879 | 3.0 | 52.2 |
| 1999 | November | 8578 | 3.2 | 55.4 |
| 1999 | December | 8637 | 3.2 | 58.6 |
| 2000 | January | 8610 | 3.2 | 61.9 |
| 2000 | February | 8542 | 3.2 | 65.1 |
| 2000 | March | 8526 | 3.2 | 68.3 |
| 2000 | April | 8730 | 3.3 | 71.5 |
| 2000 | May | 9104 | 3.4 | 75.0 |
| 2000 | June | 9090 | 3.4 | 78.4 |
| 2000 | July | 9018 | 3.4 | 81.8 |
| 2000 | August | 8937 | 3.4 | 85.1 |
| 2000 | September | 9055 | 3.4 | 88.5 |
| 2000 | October | 10089 | 3.8 | 92.3 |
| 2000 | November | 10265 | 3.9 | 96.2 |
| 2000 | December | 10220 | 3.8 | 100.0 |
| **Total** | | **266268** | **100.0** | — |

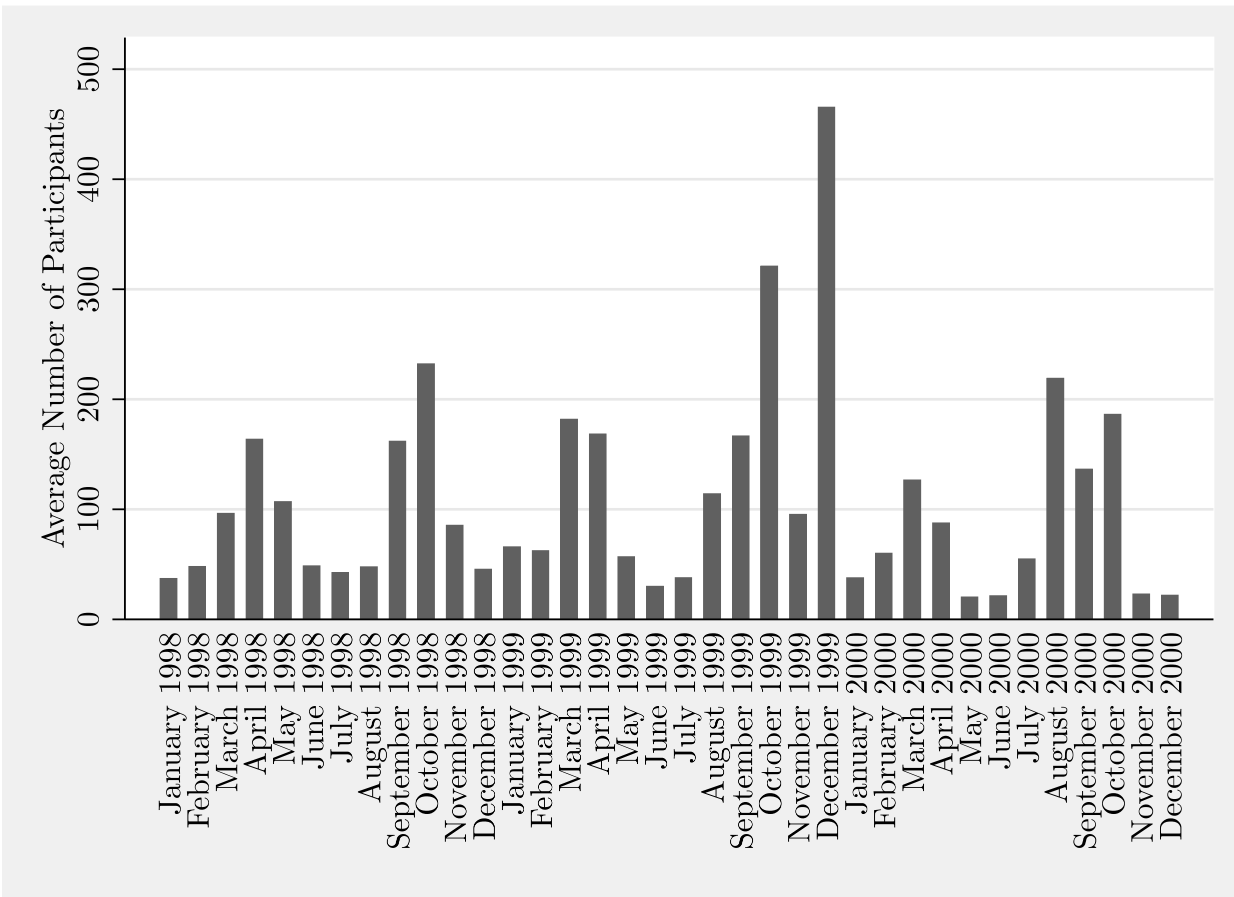



**Figure 3.6: Average Number of ROSCA Participants by Month**

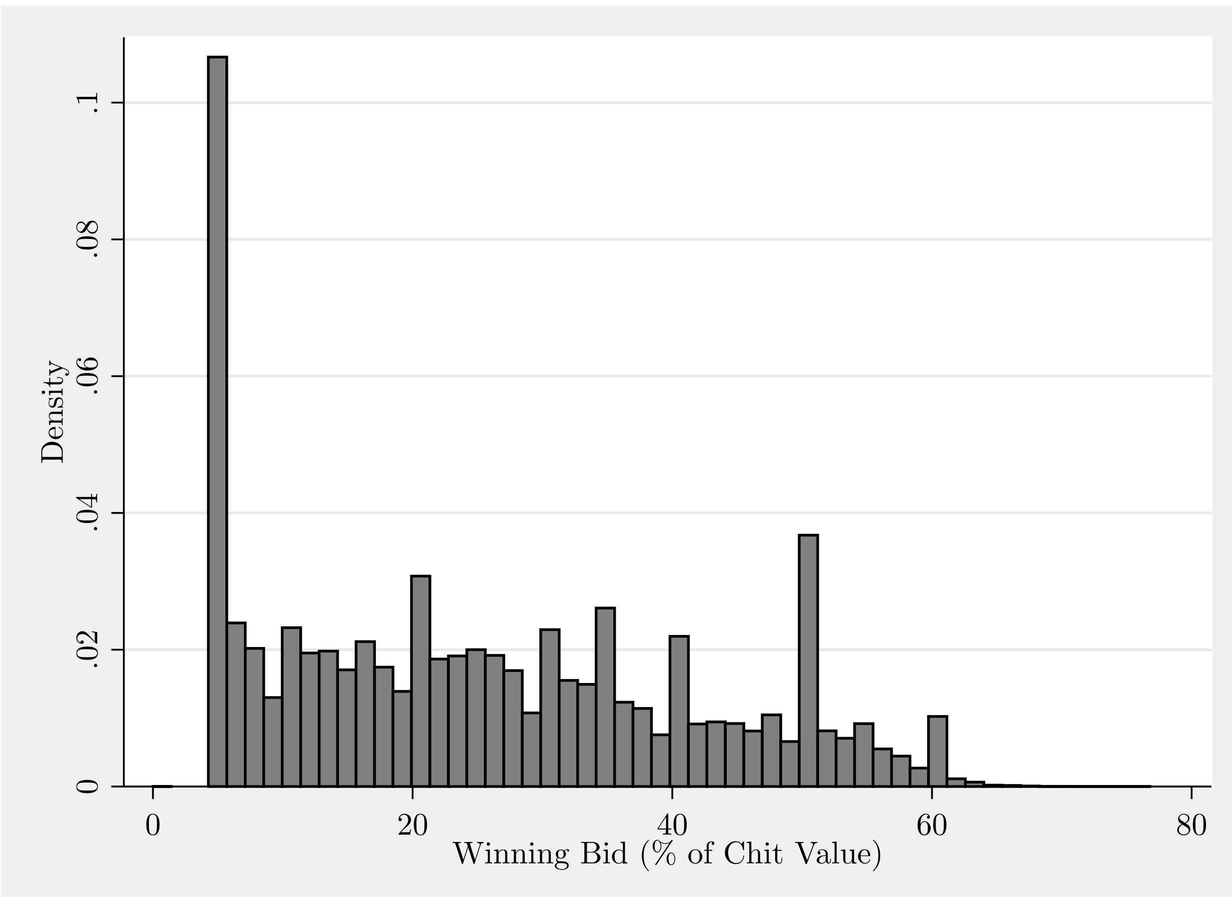



**Figure 3.7: Distribution of ROSCA Winning Bids**

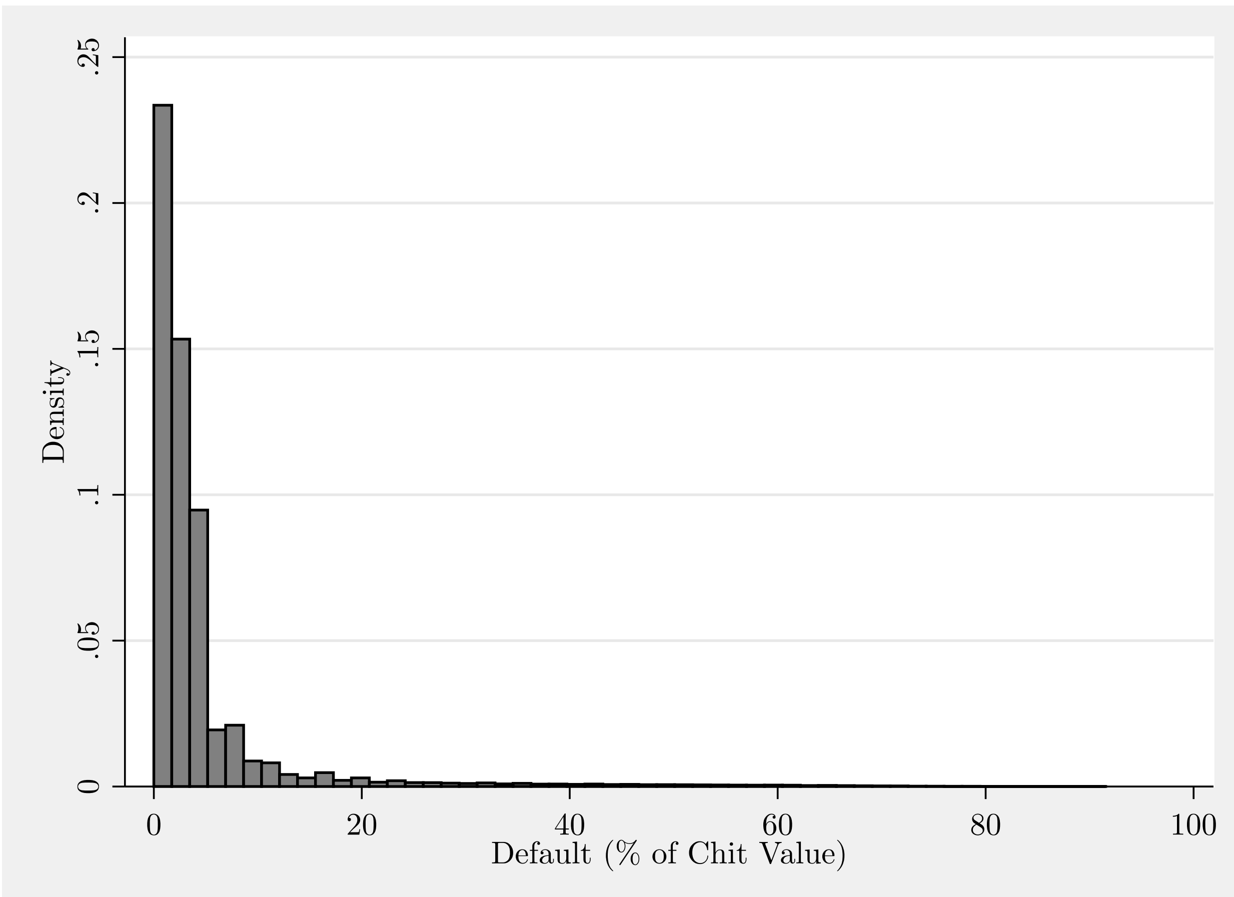



**Figure 3.8: Distribution of ROSCA Default**

## 3.5 Empirical Methodology

In order to study the impact of bank openings in the neighborhood of ROSCA branches on the participation, winning bids, and the amount of default of bidding ROSCAs, I employ fixed-effects models with slightly different explanatory variables for different outcome variables. In general, by including branch fixed effects, I am able to control for any time-invariant unobservable factors that may be different across ROSCA branches and would affect participants' decisions in joining the ROSCAs, such as locations and demographics. By including month fixed effects, my regression model also controls for any branch-invariant unobservable factors that would vary across time, such as seasonal effects on ROSCA activities. The regression models are outlined below.[6]

### 3.5.1 ROSCA Participation

The regression analysis for ROSCA participation is conducted on the branch level. The main regression equation is as follows:

$$participation_{it} = \beta_0 + \beta_1 \cdot banks_{it} + \beta_2 \cdot banks_{it-1} + \gamma_t + \delta_i + \varepsilon_{it}$$

where $participation_{it}$ is defined to be the number of new participants in ROSCA branch $i$ in a given month $t$, $banks_{it}$ is the total number of banks located in a circular neighborhood with a radius of 10 kilometers from ROSCA branch $i$ in month $t$, $bank_{it-1}$ is the one-month lagged variable, $\gamma_t$ is the month dummy, $\delta_i$ is the branch dummy, and $\varepsilon_{it}$ is the error term. To determine whether banks have a non-zero impact on ROSCA participation, I conduct a

[6] Given a relatively short time span, using the fixed-effects model may be inefficient due to a lack of within-unit variations. Therefore, for each regression specification, I conduct a Hausman specification test to check if a fixed-effects model and a random-effects model give essentially the same set of predictions. Standard econometric theory suggests that if these two models indeed give the same prediction, then the random-effects model should be employed since it is more efficient in terms of exploiting both within-unit and cross-sectional variations; otherwise, the fixed-effects model should be relied upon since the random-effects model is biased while the fixed-effects model gives unbiased estimates.

hypothesis test to see if the overall effect of bank openings is zero, that is, $\beta_1 + \beta_2 = 0$. Then, I break down $bank_{it}$ and $bank_{it-1}$ into different types and test if their effects on ROSCA participation are the same. These hypothesis tests will be done for all the other regression models as well.

### 3.5.2 ROSCA Winning Bids

The regression analysis for ROSCA winning bids is conducted on the individual ROSCA level (i.e., a complete set of auctions conducted among the same group of participants) instead of branch level because ROSCAs of different lengths and denominations often take place in the same ROSCA branch. Moreover, since the winning bid is associated with the ROSCA denomination (i.e., total amount of funds collected in a single round of a bidding ROSCA) and the bidding round, I need to control for these factors as well.

Specifically, I use the amount of winning bid as a percentage of the ROSCA denomination as the independent variable, but include ROSCA domination as an explanatory variable at the same time because not only the amount but also the percentage people are able to bid might depend on the denomination. Since ROSCA winning bids decline as the auctions progress (Besley et al., 1993; Klonner, 2004), I need to control for the "progress" of the ROSCAs as well. To do this, I define a new variable named "progress" as the percentage of auctions that have already taken place in a ROSCA. More formally,

$$progress_{ijt} = \frac{round_{ijt}}{length_{ij}} \times 100\%$$

so that the "progress" of the auction in ROSCA $j$ that takes place in month $t$ inside branch $i$ is the round of that auction as a percentage of the length of the ROSCA it belongs to.

The main regression model can be written as

$$\begin{aligned} bid_{ijt} \quad = \quad & \beta_0 + \beta_1 \cdot banks_{it} + \beta_2 \cdot banks_{it-1} + \beta_3 \cdot denomination_{ijt} + \\ & \beta_4 \cdot progress_{ijt} + \gamma_t + \delta_i + \varepsilon_{ijt} \end{aligned}$$

where $bid_{ijt}$ denotes the amount of winning bid (as a percentage of the total chit value) in ROSCA $j$ that takes place in month $t$ inside branch $i$, $denomination_{ijt}$ is the total chit value of ROSCA $j$ that takes place in month $t$ inside branch $i$, $\varepsilon_{ijt}$ is the error term, and all the other variables are the same as defined before. As the regression analysis for ROSCA participation, I then break down $bank_{it}$ and $bank_{it-1}$ into different types to test if their effects on ROSCA winning bids are the same.

### 3.5.3 ROSCA Default

Similar as the winning bids, the amount of default in the ROSCAs also depends on the denomination. Note that although the statistics for ROSCA default are collected after all the auctions end, the "progress" of the auctions is still likely to matter here. Moreover, the amount of default might also depend on the winning bids: the higher the winning bid in a particular round, the more likely for default to take place (Bouman, 1995).

The main regression equation is as follows:

$$\begin{aligned} default_{ijt} &= \beta_0 + \beta_1 \cdot banks_{it} + \beta_2 \cdot banks_{it-1} + \beta_3 \cdot denomination_{ijt} + \\ &\quad \beta_4 \cdot progress_{ijt} + \beta_5 \cdot bid_{ijt} + \gamma_t + \delta_i + \varepsilon_{ijt} \end{aligned}$$

where $default_{ijt}$ is the amount of outstanding, unpaid balance (as a percentage of the total chit value) that corresponds to ROSCA $j$ that takes place in month $t$ inside branch $i$. All the other variables are the same as defined before. Again, after running the main regression, I break down $bank_{it}$ and $bank_{it-1}$ into different types and test if their effects on ROSCA default are the same.

## 3.6 Results and Discussions

For all the regressions specified in Section 3.5, I conduct the Hausman specification test (test statistics not explicitly shown) to determine whether to use random-effects or fixed-effects models. Although random-effects models are believed to be more efficient, the Hausman specification test unfortunately rejects random-effects models for almost all the regression models at 5% significance level. Therefore, for all the regressions, I only report the estimates from fixed-effects models.

### 3.6.1 Effect of Bank Openings on ROSCA Participation

Table 3.7 and Table 3.8 provide the estimates of the effect of nearby bank openings on the number of participants in the bidding ROSCAs. The regression results suggest that overall, bank openings and thus an increased level of formal finance do not have a significant impact on ROSCA participation in either direction. However, after breaking down the banks into different types, rural banks (i.e., Grameen banks) do have a significant, positive impact on ROSCA participation. These findings are in line with some of the empirical observations from the previous literature: instead of being imperfect substitutes for formal financial institutions like banks, some informal financial institutions not only survive but sometimes also thrive even with the emergence of banks. In other words, an increased availability of formal finance does not, on average, cause a decline in ROSCA participation.

Specifically, Table 3.7 uses the total number of banks of all types as the explanatory variable and suggests that on average, an additional bank in the neighborhood of a bidding ROSCA has a positive yet insignificant impact on the ROSCA participation. In particular, column (2) is the regression with branch dummies but without month dummies. The estimates suggest that a new bank opened in the current month increases the number of ROSCA participants by about 14 on average, while a new bank opening in the previous month reduces the number of ROSCA participants by about 13 on average. Overall, for any two new banks opened in two consecutive months in the same neighborhood of a ROSCA branch, the ROSCA participation is expected to increase by about $14 + (-13) = 1$ person. However, after conducting the hypothesis test that $\beta_1 + \beta_2 = 0$, I cannot reject the null hypothesis that the number of nearby banks, in fact, has no impact on ROSCA participation.

**Table 3.7: Impact of Bank Emergence on ROSCA Participation**

| | Dependent Variable | | | | | |
|---|---|---|---|---|---|---|
| | (1) Participation | (2) Participation | (3) Participation | (4) Participation | (5) Participation | (6) Participation |
| Total Number of | 0.596 | 13.687 | 0.031* | 1.227 | 0.082 | 4.092 |
| Banks in Current Month | (0.381) | (8.127) | (0.015) | (6.068) | (0.391) | (7.540) |
| Total Number of | | −12.983 | | −1.199 | | −3.973 |
| Banks in Previous Month | | (7.743) | | (6.081) | | (7.164) |
| Constant | −197.658 | −294.731 | 34.329*** | 19.026** | 73.545 | 24.873 |
| | (259.363) | (366.917) | (6.486) | (6.018) | (261.939) | (365.903) |
| Branch Dummies | Yes | Yes | No | No | Yes | Yes |
| Month Dummies | No | No | Yes | Yes | Yes | Yes |
| Observations | 6156 | 5985 | 6156 | 5985 | 6156 | 5985 |
| $R^2$ | 0.070 | 0.072 | 0.180 | 0.178 | 0.248 | 0.248 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Column (4) reports the regression with month dummies but without branch dummies, and column (6) controls for both branch and month fixed effects. Compared with column (2), the estimates for both columns are much smaller, but the overall effects are both positive. As before, I cannot reject the null hypothesis that the overall effect of two nearby banks opened in two consecutive months have a zero impact on ROSCA participation.

In Table 3.8, I break down the total number of nearby banks by type. It is a natural question to ask whether different types of banks have differential impacts on ROSCA participation. Since different types of banks have different characteristics, certain types might have larger impact on ROSCA participation. For instance, rural banks, including Grameen banks and other community banks in rural areas, might have a stronger impact on ROSCA participation because they are traditionally connected to the informal or less formal financial sector, such as the bidding ROSCAs.

Specifically, columns (1), (3), and (5) of Table 3.8 provide the effects of state banks, nationalized banks, private banks, and Grameen banks on ROSCA participation without including one-month lagged variables. It is evident that rural banks, represented by Grameen banks, have the strongest positive impact on ROSCA participation among all types of banks. In particular, column (5) suggest that an additional state bank, nationalized bank, private bank, and Grameen bank is expected to increase ROSCA participation by about $-3$, 10, $-12$, and 20 persons, respectively. Moreover, I conduct the hypothesis test to see whether the effect of Grameen banks is different from the effect of other banks. I can reject the null hypothesis that the effect of Grameen banks is the same as that of private banks at 0.1% significance level, but I cannot reject the null hypothesis that Grameen banks have the same impact on ROSCA participation as state or nationalized banks even at 5% significance level. Finally, I conduct a joint test to see if these four types of banks all have the same effect. The test statistics suggest that different types of banks do have differential impacts on ROSCA participation at 1% significance level.

Columns (2), (4), and (6) of Table 3.8 provide the effects of state banks, nationalized

**Table 3.8: Impact of Different Types of Bank Emergence on ROSCA Participation**

| | Dependent Variable: Participation | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Number of State Banks in | 9.000 | 12.668 | −2.467*** | 27.001* | −2.691 | 21.687 |
| Current Month | (6.361) | (13.304) | (0.541) | (12.673) | (5.950) | (12.886) |
| Number of Nationalized | 10.419*** | 42.669** | 0.987** | 8.389 | 9.864*** | 17.521 |
| Banks in Current Month | (3.006) | (15.736) | (0.303) | (13.706) | (2.863) | (13.977) |
| Number of Private Banks | −18.747*** | −34.386** | 0.711 | −24.343* | −12.141** | −26.164** |
| in Current Month | (4.451) | (10.940) | (0.882) | (9.579) | (4.292) | (10.086) |
| Number of Grameen Banks | 29.234** | 96.125** | 4.089*** | 87.885* | 20.093* | 87.688* |
| in Current Month | (9.662) | (33.785) | (0.857) | (35.368) | (9.069) | (34.623) |
| Number of State Banks in | | −3.383 | | −29.591* | | −28.215* |
| Previous Month | | (13.772) | | (12.679) | | (13.052) |
| Number of Nationalized | | −33.632* | | −7.350 | | −10.800 |
| Banks in Previous Month | | (15.860) | | (13.767) | | (14.163) |
| Number of Private Banks | | 16.873 | | 25.081** | | 20.693* |
| in Previous Month | | (10.977) | | (9.597) | | (10.002) |
| Number of Grameen Banks | | −72.287* | | −83.820* | | −74.923* |
| in Previous Month | | (34.440) | | (35.359) | | (35.150) |
| Branch Dummies | Yes | Yes | No | No | Yes | Yes |
| Month Dummies | No | No | Yes | Yes | Yes | Yes |
| Observations | 6156 | 5985 | 6156 | 5985 | 6156 | 5985 |
| $R^2$ | 0.072 | 0.078 | 0.185 | 0.187 | 0.250 | 0.252 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

banks, private banks, and Grameen banks on ROSCA participation by including one-month lagged variables. As before, Grameen banks still have the strongest positive overall impact on ROSCA participation compared to other types.

From the strong impact of rural banks, it is worth noticing that banks such as Grameen banks and other banks designed especially for the poor not only do not drive ROSCA participants away, but that they also increase individuals' willingness to seek informal finance. This result makes sense intuitively: the more Grameen banks there are, the easier it is to spread the concept of group borrowing as a means of obtaining credit, and the more likely it is for these people to participate in informal finance because of the potential flexibility of ROSCAs over Grameen banks in terms of the "implied" interest rate. In this regard, an increased availability of formal financial institutions, especially those relevant to the poor, will contribute to an increased level of popularity of informal finance among previously credit-constrained individuals and small businesses.

These findings are also largely in line with the prediction from my theoretical model. Except for two special cases where the banks' interest rates are too high or too low, ROSCAs are able to survive despite the increased number of banks in the nearby area. One limitation to note is that in the heuristic two-player, two-period model, one cannot model the situation where the emergence of banks has a strictly positive impact on ROSCA participation. However, based on other empirical work in the literature, it is not uncommon to observe such situations in real life. With an increased availability of credit, these newly introduced formal financing channels will likely have a positive effect on people's accessible cash on hand, thus increasing their ability and willingness to participate in informal financial institutions like the bidding ROSCAs.

One potential concern about the results is the endogenous choice of locations by banks. Specifically, even with the regulations on the banking sector, some banks might still choose locations that are relatively more favorable to them in terms of demographics and financial development. However, this concern should not overturn the estimates above. Instead, it suggests that the estimates above might be a lower bound of the effect of bank openings on ROSCA participation. The intuition is simple: if the banks' locations are chosen endogenously, they would choose to open in places where they are more likely to attract a large body of customers and where people are more willing to switch from informal finance to formal finance. In this case, with banks' endogenous location choices, the effect of nearby bank openings on ROSCA participation should be less positive than the case with exogenous location choices. Therefore, the estimates in the study might be biased downwards and the actual effect of bank openings on ROSCA participation might be even larger.

### 3.6.2 Effect of Bank Openings on Winning Bids

Table 3.9 and Table 3.10 provide the estimates of the effect of nearby bank openings on the winning bids of the ROSCAs. The estimates suggest that the emergence of banks in the neighborhood of a ROSCA branch will reduce ROSCA winning bids, which are equal to the cost of capital in the bidding ROSCAs. Hence, the increased competition between banks and bidding ROSCAs is likely to benefit ROSCA participants in the sense that obtaining credit becomes more affordable than before. This result is in line with my theoretical model, which predicts a possible reduction in ROSCA winning bids as a result of new banks entering the area.

Specifically, Table 3.9 uses the total number of banks of all types as the explanatory variable, controlling for the effects of ROSCA denomination and progress. In particular, after including both month and branch fixed effects, column (5) suggests that an additional bank opened in the current month is expected to decrease ROSCA winning bids by about 0.006 percentage points, and this effect is significant at 5% level. After including the one-month lagged variable, column (6) still suggests a negative overall effect of bank openings on ROSCA winning bids, but the effect is no longer significant. However, after conducting the hypothesis test that $\beta_1 + \beta_2 = 0$, I am able to reject the null hypothesis that the number of nearby bank openings has zero overall impact on ROSCA winning bids at 1% significance level. As a side

**Table 3.9: Impact of Bank Emergence on ROSCA Winning Bids**

| | Dependent Variable | | | | | |
|---|---|---|---|---|---|---|
| | (1) Winning Bid | (2) Winning Bid | (3) Winning Bid | (4) Winning Bid | (5) Winning Bid | (6) Winning Bid |
| Total Number of | $0.170^{***}$ | $-0.320^{***}$ | $-0.005^{***}$ | 0.063 | $-0.006^{*}$ | 0.032 |
| Banks in Current Month | (0.003) | (0.032) | (0.000) | (0.033) | (0.003) | (0.035) |
| Total Number of | | $0.479^{***}$ | | $-0.068^{*}$ | | $-0.041$ |
| Banks in Previous Month | | (0.031) | | (0.033) | | (0.034) |
| Denomination | $0.027^{***}$ | $0.027^{***}$ | $0.023^{***}$ | $0.023^{***}$ | $0.024^{***}$ | $0.024^{***}$ |
| (in thousand rupees) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| ROSCA Progress | $-0.271^{***}$ | $-0.277^{***}$ | $-0.470^{***}$ | $-0.469^{***}$ | $-0.456^{***}$ | $-0.456^{***}$ |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant | $-78.848^{***}$ | $-69.942^{***}$ | $29.910^{***}$ | $68.089^{***}$ | $31.382^{***}$ | $68.825^{***}$ |
| | (1.831) | (1.955) | (0.210) | (0.206) | (1.887) | (2.074) |
| Branch Dummies | Yes | Yes | No | No | Yes | Yes |
| Month Dummies | No | No | Yes | Yes | Yes | Yes |
| Observations | 263250 | 258538 | 263250 | 258538 | 263250 | 258538 |
| $R^2$ | 0.294 | 0.299 | 0.377 | 0.380 | 0.391 | 0.393 |

Standard errors in parentheses

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

note, the coefficients on ROSCA "progress" are all negative and significant at 0.1% level in all regressions, which confirms the findings in previous literature that ROSCA winning bids decline as the ROSCA progresses. Based on my estimates from columns (5) and (6), from the beginning to the end of a ROSCA, the winning bid (as a percentage of the denomination) is expected to decline by 45.6 percentage points. Since there is no auction in the last period so that the winning bid in the last period is zero, this estimate suggests that the expected winning bid in the first period is about 45.6% of the denomination.

In Table 3.10, I break down the total number of nearby banks by type and test whether different types of banks have differential impacts on ROSCA winning bids. Since winning bids are effectively the cost of capital from a bidding ROSCA, a positive coefficient would suggest that nearby bank openings induce ROSCA participants to be willing to pay even more for the chit fund, while a negative coefficient, which is more likely, would suggest that bank emergence lowers the cost of capital in the bidding ROSCAs.

Specifically, columns (1), (3), and (5) of Table 3.10 provide the effects of state banks, nationalized banks, private banks, and Grameen banks on ROSCA winning bids without including one-month lagged variables. The estimates in columns (5) suggest that all types of banks except for the private ones are expected to reduce ROSCA winning bids. Specifically, an additional state bank, nationalized bank, and Grameen bank is expected to *reduce* ROSCA winning bids by 0.08, 0.18, and 0.05 percentage points, respectively, and an additional private bank is expected to *increase* the winning bids by 0.28 percentage points. Moreover, I conduct the hypothesis test of whether the effect of private banks is different from the effect of other banks. I can reject the null hypothesis that the effect of private banks is the same as that of state banks, nationalized banks, or Grameen banks at 5% significance level (with the first two types at 0.1% significance level). Then, I conduct the joint test to see if the effect is the same across the four types of banks. The test statistics suggest that the effects of different types of banks on ROSCA winning bids are indeed different at 0.1% significance level.

Columns (2), (4), and (6) of Table 3.10 are regressions with one-month lagged variables.

**Table 3.10: Impact of Different Types of Bank Emergence on ROSCA Winning Bids**

| | Dependent Variable: Winning Bid | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Number of State Banks in | 0.321*** | 0.129 | 0.080*** | 0.108 | −0.081 | 0.041 |
| Current Month | (0.048) | (0.094) | (0.006) | (0.096) | (0.050) | (0.102) |
| Number of Nationalized | 0.504*** | −0.411*** | −0.068*** | 0.022 | −0.177*** | −0.178* |
| Banks in Current Month | (0.033) | (0.075) | (0.004) | (0.071) | (0.034) | (0.079) |
| Number of Private Banks | −0.351*** | −0.262** | 0.093*** | 0.298*** | 0.283*** | 0.311*** |
| in Current Month | (0.048) | (0.080) | (0.010) | (0.076) | (0.050) | (0.085) |
| Number of Grameen Banks | 6.918*** | 3.572*** | −0.023** | −0.221 | −0.047 | −0.009 |
| in Current Month | (0.155) | (0.373) | (0.007) | (0.356) | (0.150) | (0.360) |
| Number of State Banks in | | 0.151 | | −0.032 | | −0.174 |
| Previous Month | | (0.097) | | (0.095) | | (0.105) |
| Number of Nationalized | | 0.993*** | | −0.089 | | −0.029 |
| Banks in Previous Month | | (0.065) | | (0.071) | | (0.071) |
| Number of Private Banks | | −0.196* | | −0.204** | | 0.044 |
| in Previous Month | | (0.079) | | (0.075) | | (0.083) |
| Number of Grameen Banks | | 3.710*** | | 0.197 | | 0.107 |
| in Previous Month | | (0.370) | | (0.356) | | (0.358) |
| Denomination | 0.027*** | 0.027*** | 0.023*** | 0.023*** | 0.024*** | 0.024*** |
| (in thousand rupees) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| ROSCA Progress | −0.281*** | −0.288*** | −0.469*** | −0.468*** | −0.457*** | −0.456*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Branch Dummies | Yes | Yes | No | No | Yes | Yes |
| Month Dummies | No | No | Yes | Yes | Yes | Yes |
| Observations | 263250 | 258538 | 263250 | 258538 | 263250 | 258538 |
| $R^2$ | 0.299 | 0.305 | 0.378 | 0.380 | 0.391 | 0.394 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Similar as the case without lagged variables, private banks still have the strongest positive overall impact on ROSCA winning bids. The channel through which private banks differ from other banks is not within the scope of this study, but it might be related to the different lending behaviors of private banks compared to other types of banks (Bhattacharyya et al., 1997; Bhaumik and Piesse, 2008). One possible explanation might be that it is too costly to borrow from private banks or that private banks have terms of lending that are unfavorable towards the poor so that ROSCA participants are willing to bid even more to obtain funds from ROSCAs rather than from private banks.

### 3.6.3 Effect of Bank Openings on ROSCA Default

Table 3.11 and Table 3.12 provide the estimates of the effect of nearby bank openings on the amount of default in the bidding ROSCAs. The regression results also suggest a significant, negative impact of the emergence of banks on ROSCA default, which might suggest a "monitoring" role played by the formal financial institutions on the operations of the informal sector. Moreover, the findings shed light on the complementarity between banks and the bidding ROSCAs. Instead of having a deteriorating impact on the bidding ROSCAs, the bank openings in the neighborhood of ROSCA branches seem to strengthen the operations in the ROSCAs by reducing the amount of default.

Specifically, Table 3.11 uses the total number of banks of all types as the explanatory variable, controlling for the effect of ROSCA denomination, auction progress, and winning bids. Except for the case with month dummies and without branch dummies, bank openings seem to have a negative effect on ROSCA default and this effect is significant at 0.1% level for all the regressions without the lagged variable. For instance, column (5) includes both month and branch fixed effects. The estimates suggest that an additional bank opened in the current month is expected to decrease ROSCA default by about 0.02 percentage points. After including the one-month lagged variable, the effect of bank openings in the current month is still negative yet insignificant, and the effect of bank openings in the previous month is also

**Table 3.11: Impact of Bank Emergence on ROSCA Default**

| | Dependent Variable | | | | | |
|---|---|---|---|---|---|---|
| | (1) Default | (2) Default | (3) Default | (4) Default | (5) Default | (6) Default |
| Total Number of | −0.009*** | −0.080*** | 0.001*** | 0.025 | −0.020*** | −0.047 |
| Banks in Current Month | (0.002) | (0.021) | (0.000) | (0.022) | (0.002) | (0.024) |
| Total Number of | | 0.069*** | | −0.024 | | 0.025 |
| Banks in Previous Month | | (0.021) | | (0.022) | | (0.024) |
| Denomination | 0.002*** | 0.002*** | 0.002*** | 0.002*** | 0.002*** | 0.002*** |
| (in thousand rupees) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| ROSCA Progress | −0.018*** | −0.019*** | −0.039*** | −0.040*** | −0.037*** | −0.038*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Winning Bid | 0.118*** | 0.116*** | 0.110*** | 0.108*** | 0.106*** | 0.105*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Constant | 9.068*** | 10.650*** | 1.517*** | 5.090*** | 15.982*** | 20.494*** |
| | (1.367) | (1.485) | (0.149) | (0.152) | (1.471) | (1.676) |
| Branch Dummies | Yes | Yes | No | No | Yes | Yes |
| Month Dummies | No | No | Yes | Yes | Yes | Yes |
| Observations | 204270 | 200642 | 204270 | 200642 | 204270 | 200642 |
| $R^2$ | 0.088 | 0.088 | 0.078 | 0.078 | 0.092 | 0.092 |

Standard errors in parentheses

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

insignificant. However, after conducting the hypothesis test that $\beta_1 + \beta_2 = 0$, I am able to reject the null hypothesis that the number of nearby bank openings has zero overall impact on ROSCA winning bids at 0.1% significance level.

In Table 3.12, I break down the total number of nearby banks by type and test whether different types of banks have differential impacts on ROSCA default. Overall, the effects of different types of banks are negative, but there are still a few exceptions. In column (1), I include branch fixed effects but not month fixed effects. All types of banks seem to have a negative impact on ROSCA default except for nationalized banks, which have a positive and significant impact on ROSCA default. Column (3) includes month fixed effects but not branch fixed effects, and the estimates suggest that all types except for state banks have a significant, negative impact on ROSCA default.

In column (5), after including both month and branch fixed effects, the results are somewhat reversed from columns (1) and (3). Both state and private banks have negative yet insignificant impacts on ROSCA default, while nationalized banks have a positive yet insignificant impact on ROSCA default. The estimate that clearly stands out is the coefficient for Grameen banks, which is significantly positive at 1% level and about an order of magnitude larger than all other coefficients. Hence, I conduct a hypothesis test to see whether the effect of Grameen banks is different from the effects of all other banks. Not surprisingly, I am able to reject the null hypothesis that the effect of Grameen banks is the same as that of state banks, nationalized banks, or private banks at 0.1% significance level. Moreover, after I conduct the joint test of whether the effect is the same across the four types of banks, the result suggests the effects of different types of banks on ROSCA default are indeed different at 0.1% significance level.

It seems that in all cases, rural banks represented by Grameen banks have the strongest impact (either positive or negative) on ROSCA default. This is not surprising because as discussed before, Grameen banks and ROSCAs are closely related so that Grameen banks should, in expectation, have a greater impact on ROSCAs than other types of banks. However,

**Table 3.12: Impact of Different Types of Bank Emergence on ROSCA Default**

| | Dependent Variable: Default | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Number of State Banks in | −0.030 | −0.074 | 0.042*** | 0.124 | −0.052 | −0.050 |
| Current Month | (0.032) | (0.069) | (0.005) | (0.075) | (0.036) | (0.079) |
| Number of Nationalized | 0.063** | −0.058 | −0.010** | −0.082 | 0.026 | −0.011 |
| Banks in Current Month | (0.022) | (0.049) | (0.004) | (0.048) | (0.025) | (0.056) |
| Number of Private Banks | −0.099** | −0.016 | −0.038*** | −0.041 | −0.067 | −0.022 |
| in Current Month | (0.032) | (0.056) | (0.008) | (0.058) | (0.037) | (0.064) |
| Number of Grameen Banks | 0.737*** | 0.284 | −0.031*** | 0.183 | 0.334** | 0.196 |
| in Current Month | (0.103) | (0.284) | (0.005) | (0.282) | (0.109) | (0.290) |
| Number of State Banks in | | 0.014 | | −0.084 | | −0.026 |
| Previous Month | | (0.072) | | (0.075) | | (0.081) |
| Number of Nationalized | | 0.175*** | | 0.074 | | 0.061 |
| Banks in Previous Month | | (0.042) | | (0.048) | | (0.048) |
| Number of Private Banks | | −0.146* | | 0.001 | | −0.068 |
| in Previous Month | | (0.058) | | (0.057) | | (0.065) |
| Number of Grameen Banks | | 0.560* | | −0.213 | | 0.209 |
| in Previous Month | | (0.279) | | (0.282) | | (0.285) |
| Denomination | 0.002*** | 0.002*** | 0.002*** | 0.002*** | 0.002*** | 0.002*** |
| (in thousand rupees) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| ROSCA Progress | −0.020*** | −0.021*** | −0.039*** | −0.039*** | −0.037*** | −0.038*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Winning Bid | 0.117*** | 0.116*** | 0.109*** | 0.108*** | 0.106*** | 0.105*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Branch Dummies | Yes | Yes | No | No | Yes | Yes |
| Month Dummies | No | No | Yes | Yes | Yes | Yes |
| Observations | 204270 | 200642 | 204270 | 200642 | 204270 | 200642 |
| $R^2$ | 0.089 | 0.089 | 0.079 | 0.079 | 0.092 | 0.092 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

what is interesting and somewhat surprising is that after including both month and branch fixed effects, Grameen banks turn out to have the strongest *positive* impact on ROSCA default. One potential explanation is that due to a strong connection between ROSCAs and rural banks, the emergence of nearby Grameen banks might contribute the most to the competition between formal and informal finance in terms of the cost of borrowing (i.e., explicit or implicit interest). With this increased level of competition, if ROSCA participants first submit a high bid and then observe the opening of a nearby bank with a lower borrowing rate, they might choose to default from the ROSCAs in order to pursue a less costly financing channel.

## 3.7 Conclusion

This chapter studies the role of informal finance and its relationship with the formal financial sector. Despite the traditional view that formal finance is unambiguously superior to informal finance and that informal financial institutions are merely a backup choice for credit-constrained entities, this study provides both theoretical support and empirical evidence for the new view that there is some complementary between formal and informal financial institutions. In particular, I examine the relationship between bidding ROSCAs and different types of banks in Andhra Pradesh, India. I find that the emergence of formal finance, rather than diverting people's interest away from pursuing informal finance, actually fosters the development of informal finance, especially in the case of bidding ROSCAs.

In the theoretical model, I find that the bidding behaviors of ROSCA participants are largely determined by how their productivity compares to the rest of the group, where productivity can be defined as the net return to their own investment using the ROSCA earnings. Specially, in each round of the ROSCA, all the eligible bidders except for the one with the highest productivity have an incentive to overbid because they get part of the premium as dividends. As banks start to open in the neighborhood, ROSCAs are not crowded out of the market unless all the ROSCA participants have extremely high or extremely low returns to their investment projects, which is highly unlikely. Moreover, the theoretical model predicts that increased competition due to the emergence of banks weakly reduces ROSCA winning bids — the cost of capital in the bidding ROSCAs.

The empirical study aims to test some of the predictions in the theoretical model. Using data from 219 ROSCA branches and over 5,000 banks in Andhra Pradesh, India, I find that overall, bank openings are expected to increase ROSCA participation. Although this effect is not significant for banks as a whole, after breaking down the banks by type, I find that this effect is very significant for rural banks, which also seem to have the strongest, positive impact on ROSCA participation among all types of banks. As for ROSCA winning bids, the empirical results are also in line with the theoretical predictions. I find that the emergence of banks in the nearby neighborhood of ROSCA branches has a significant, negative effect on ROSCA winning bids. Finally, my regression results also indicate a significant, negative impact of nearby bank openings on the amount of default in the ROSCAs. More broadly speaking, this suggests that informal financial institutions like the bidding ROSCAs not only coexist with banks but also benefit from the strengthening of the formal sector, which constitutes a complementary rather than substitutional effect.

An interesting extension of this study would be to examine how the "implied" interest rates in the ROSCAs are affected by an exogenous shock to the interest rates of nearby banks. According to the theoretical predictions, a reduction in the bank's interest rate might weakly reduce the ROSCA winning bids and thus the "implied" borrowing rate. Therefore, one can possibly compute the "implied" interest rate for each ROSCA participants using the bidding data, and use banks' interest rate data to test the theoretical hypothesis. One difficulty of this approach is that the calculation of the "implied" interest rate might be very messy, especially for the partial borrowers and partial savers (i.e., all the participants except for the one who wins first and the one who wins last). In particular, for each ROSCA participant, the calculation of the interest rate would involve solving a polynomial equation of degree $k-1$, where $k$ is the total number of participants in a given ROSCA. However, if such calculations were able to be done in an efficient manner, I would be able to test the theoretical predictions more directly and examine whether participating in the ROSCAs is indeed a rational and profitable choice for the current ROSCA participants.

# References

**Akerlof, George A.** 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(3): 488–500.

**Akerlof, George A.** 1991. "Procrastination and Obedience." *American Economic Review*, 81(2): 1–19.

**Allen, Franklin, Rajesh Chakrabarti, and Sankar De.** 2007. "India's Financial System." May, Working Paper.

**Allen, Franklin, Rajesh Chakrabarti, Sankar De, Jun Qian, and Meijun Qian.** 2012a. "Financing Firms in India." *Journal of Financial Intermediation*, 21(3): 409–445.

**Allen, Franklin, Jun Qian, and Meijun Qian.** 2005. "Law, Finance, and Economic Growth in China." *Journal of Financial Economics*, Forthcoming.

**Allen, Franklin, Jun Qian, Chenying Zhang, and Mengxin Zhao.** 2012b. "China's Financial System: Opportunities and Challenges." February, NBER Working Paper No. 17828.

**Amir, Eli.** 1993. "The Market Valuation of Accounting Information: The Case of Postretirement Benefits Other Than Pensions." *The Accounting Review*, 68(4): 703–724.

**Anderson, Siwan, and Jean-Marie Baland.** 2002. "The Economics of ROSCAs and Intra-household Resource Allocation." *The Quarterly Journal of Economics*, 117(3): 963–995.

**Aryeetey, Ernest.** 2002. "Informal Finance for Private Sector Development in Africa." February, African Development Bank, Working Paper 41.

**Ater, Itai, and Vardit Landsman.** 2013. "Do Customers Learn from Experience? Evidence from Retail Banking." *Management Science*, 59(5): 2019–2035.

**Banerjee, Abhijit, and Esther Duflo.** 2012. "Do Firms Want to Borrow? Testing Credit Constraints Using a Directed Lending Program." June, MIT Working Paper.

**Beaulieu, Nancy D.** 2002. "Quality Information and Consumer Health Plan Choices." *Journal of Health Economics*, 21(1): 43–63.

**Beck, Thorsten.** 2007. "Financing Constraints of SMEs in Developing Countries: Evidence, Determinants and Solutions." May, World Bank Working Paper.

**Besley, Timothy B., Stephen Coate, and Glenn C. Loury.** 1993. "The Economics of Rotating Savings and Credit Associations." *American Economic Review*, 83(4): 782–810.

**Besley, Timothy B., Stephen Coate, and Glenn C. Loury.** 1994. "Rotating Savings and Credit Associations, Credit Markets and Efficiency." *The Review of Economic Studies*, 61(4): 701–719.

**Besley, Timothy, and Alec R. Levenson.** 1996. "The Role of Informal Finance in Household Capital Accumulation: Evidence from Taiwan." *The Economic Journal*, 106(434): 39–59.

**Bhattacharyya, Arunava, C.A.K. Lovell, and Pankaj Sahay.** 1997. "The Impact of Liberalization on the Productive Efficiency of Indian Commercial Banks." *European Journal of Operational Research*, 98(2): 332–345.

**Bhaumik, Sumon Kumar, and Jenifer Piesse.** 2008. "Does Lending Behaviour of Banks in Emerging Economies Vary by Ownership? Evidence from the Indian Banking Sector." *Economic Systems*, 32(2): 177–196.

**Blundell, Richard, Rachel Griffith, and John Van Reenen.** 1999. "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms." *Review of Economic Studies*, 1999(66): 529–554.

**Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** 2012. "Salience Theory of Choice Under Risk." *Quarterly Journal of Economics*, 127(3): 1243–1285.

**Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** 2013. "Salience and Consumer Choice." *Journal of Political Economy*, 121(5): 803–843.

**Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** 2015. "Memory, Attention and Choice." March, Working Paper.

**Bouman, F.J.A.** 1995. "Rotating Savings and Credit Organizations: A Development Perspective." *World Development*, 23(3): 371–384.

**Breschi, Stefano, Francesco Lissoni, and Franco Malerba.** 2003. "Knowledge-Relatedness in Firm Technological Diversification." *Research Policy*, 32(1): 69–87.

**Bundorf, Kate, Natalie Chun, Gopi S. Goda, and Daniel P. Kessler.** 2009. "Do Markets Respond to Quality Information? The Case of Fertility Clinics." *Journal of Health Economics*, 28(3): 718–727.

**Bushong, Benjamin, Mathew Rabin, and Joshua Schwartzstein.** 2014. "A Model of Relative Thinking." November, Working Paper.

**Calomiris, Charles W., and Indira Rajaraman.** 1998. "The Role of ROSCAs: Lumpy Durables or Event Insurance?" *Journal of Development Economics*, 56(1): 207–216.

**Calzolari, Giacomo, and Mattia Nardotto.** 2015. "Nudging with Reminders: A Field Experiment." March, Working Paper.

**Chen, Yu-Shan, Chun-Yu Shih, and Ching-Hsun Chang.** 2012. "The Effects of Related and Unrelated Technological Diversification on Innovation Performance and Corporate Growth in the Taiwan's Semiconductor Industry." *Scientometrics*, 92(1): 117–134.

**Chernew, Michael, Gautam Gowrisankaran, and Dennis P. Scanlon.** 2008. "Learning and the Value of Information: Evidence from Health Plan Report Cards." *Journal of Econometrics*, 144(1): 156–174.

**Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–1177.

**Clark, Colin W., and Reuven Dukas.** 2003. "The Behavioral Ecology of a Cognitive Constraint: Limited Attention." *Behavioral Ecology*, 14(2): 151–156.

**Cutler, David M., Robert S. Ilckman, and Mary Beth Landrum.** 2004. "The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery." *American Economic Review*, 94(2): 342–346.

**Dafny, Leemore, and David Dranove.** 2008. "Do Report Cards Tell Consumers Anything They Don't Already Know? The Case of Medicare HMOs." *The RAND Journal of Economics*, 39(3): 790–821.

**DellaVigna, Stefano, and Ulrike Malmendier.** 2006. "Paying Not to Go to the Gym." *American Economic Review*, 96(3): 694–719.

**Eeckhout, Jan, and Kaivan Munshi.** 2010. "Matching in Informal Financial Institutions." *Journal of the European Economic Association*, 8(5): 947–988.

**Ericson, Keith M. Marzilli.** 2014. "On the Interaction of Memory and Procrastination: Implications for Reminders." August, NBER Working Paper No. 20381.

**EU.** 2002. "Regulation (EC) No. 178/2002 of the European Parliament and of the Council." *Official Journal of the European Communities.*

**Eurostat.** 2008. *NACE Rev.2: Statistical Classification of Economic Activities in the European Community.* Methodologies and Working Papers, Luxembourg: Office for Official Publications of the European Communities.

**Eurostat.** 2014. "High-Tech Aggregation of Manufacturing and Services (NACE Rev. 2)." January, Statistical Office of the European Union.

**Fan, Joseph P.H., Jun Huang, Felix Oberholzer-Gee, Troy D. Smith, and Mengxin Zhao.** 2008. "Diversification of Chinese Companies: An International Comparison." *Chinese Management Studies*, 2(1): 6–13.

**Fang, Hanming, Rongzhu Ke, and Li-An Zhou.** 2011. "The Insurance Role of ROSCAs in the Presence of the Credit Market: Theory and Evidence." May, Working Paper.

**Farrell, Joseph.** 1986. "Voluntary Disclosure: Robustness of the Unraveling Results and Comments on Its Importance." In *Antitrust and Regulation.* ed. by Ronald E. Grieson, Lexington, MA: Lexington Books, 91–103.

**Feldman, Maryann P., and David B. Audretsch.** 1999. "Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition." *European Economic Review*, 43(2): 409–429.

**Finkelstein, Amy.** 2009. "E-ZTax: Tax Salience and Tax Rates." *Quarterly Journal of Economics*, 124(3): 969–1010.

**Fishman, Michael J., and Kathleen M. Hagerty.** 2003. "Mandatory vs. Voluntary Disclosure in Markets with Informed and Uninformed Customers." *The Journal of Law, Economics, & Organization*, 19(1): 45–63.

**Gabaix, Xavier, and David Laibson.** 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics*, 121(2): 505–540.

**Garcia-Vega, Maria.** 2006. "Does Technological Diversification Promote Innovation?: An Empirical Analysis for European Firms." *Research Policy*, 35(2): 230–246.

**Gennaioli, Nicola, and Andrei Shleifer.** 2010. "What Comes to Mind." *Quarterly Journal of Economics*, 125(4): 1399–1433.

**Germidis, Dimitri, Denis Kessler, and Rachel Meghir.** 1991. *Financial Systems and Development: What Role for the Formal and Informal Financial Sectors*. Paris: OECD.

**Gilchrist, Duncan S., Michael Luca, and Deepak Malhotra.** 2015. "When $3 + 1 > 4$: Gift Structure and Reciprocity in the Field." June, Harvard Business School Working Paper.

**Giroud, Xavier, and Holger M. Mueller.** 2015. "Capital and Labor Reallocation within Firms." *The Journal of Finance*, 70(4): 1767–1804.

**Goldin, Jacob, and Tatiana Homonoff.** 2013. "Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity." *American Economic Journal: Economic Policy*, 5(1): 302–336.

**Gort, Michael.** 1963. *Diversification and Integration in American Industry*. Princeton, NJ: Princeton University Press.

**Grabowski, Henry G.** 1968. "The Determinants of Industrial Research and Development: A Study of the Chemical, Drug, and Petroleum Industries." *Journal of Political Economy*, 76(2): 292–306.

**Grossman, Sanford J.** 1981. "The Informational Role of Warranties and Private Disclosure of Product Quality." *Journal of Law and Economics*, 24(3): 461–484.

**Grossman, Sanford J., and Oliver D. Hart.** 1980. "Disclosure Laws and Takeover Bids." *The Journal of Finance*, 35(2): 323–334.

**Hastings, Justine S., and Jesse M. Shapiro.** 2013. "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks." *Quarterly Journal of Economics*, 128(4): 1449–1498.

**Hirshleifer, David, and Siew Hong Teoh.** 2003. "Limited Attention, Information Disclosure, and Financial Reporting." *Journal of Accounting and Economics*, 36(1-3): 337–386.

**Hitt, Michael A., Robert E. Hoskisson, and Hicheon Kim.** 1997. "International Diversification: Effects on Innovation and Firm Performance in Product-Diversified Firms." *The Academy of Management Journal*, 40(4): 767–798.

**Huysentruyt, Marieke, and Daniel Read.** 2010. "How Do People Value Extended Warranties? Evidence from Two Field Surveys." *Journal of Risk and Uncertainty*, 40(3): 197–218.

**ISO, E. S.** 1995. "EN ISO 8492.1995." European Committee for Standardization, Point 3.16.

**Izumida, Yoichi.** 1992. "The Kou in Japan: A Precursor of Modern Finance." In *Informal Finance in Low-Income Countries*. eds. by D.W. Adams, and D.A. Fitchett, Boulder, Colorado: Westview Press, 165–180.

**Jain, Sanjay.** 1999. "Symbiosis versus Crowding-out: The Interaction of Formal and Informal Credit Markets in Developing Countries." *Journal of Development Economics*, 59(2): 419–444.

**Jefferson, Gary, Albert G.Z. Hu, Xiaojing Guan, and Xiaoyun Yu.** 2003. "Ownership, Performance, and Innovation in China's Large- and Medium-Size Industrial Enterprise Sector." *China Economic Review*, 14(1): 89–113.

**Jin, Ginger Z.** 2005. "Competition and Disclosure Incentives: An Empirical Study of HMOs." *The RAND Journal of Economics*, 36(1): 93–112.

**Jin, Ginger Z. Jin, and Alan T. Sorensen.** 2005. "Information and Consumer Choice: The Value of Publicized Health Plan Ratings." *Journal of Health Economics*, 25(2): 248–275.

**Jin, Ginger Z., and Phillip Leslie.** 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *Quarterly Journal of Economics*, 118(2): 409– 451.

**Jin, Ginger Z., and Phillip Leslie.** 2009. "Reputation Incentives for Restaurant Hygiene." *American Economic Journal (Microeconomics)*, 1(1): 237–267.

**Jin, Ginger Z., Phillip Leslie, Paul Simon, Grace Run, Roshan Reporter, Arturo Aguirre, and Jonathan E. Fielding.** 2005. "Impact of Restaurant Hygiene Grade Cards on Foodborne Disease Hospitalizations in Los Angeles County." *Journal of Environmental Health*, 67(7): 32–36.

**Jovanovic, Boyan.** 1982. "Truthful Disclosure of Information." *Bell Journal of Economics*, 13(1): 36–44.

**Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman.** 2010. "Getting to the Top of Mind: How Reminders Increase Saving." July, NBER Working Paper No. 16205.

**Klonner, Stefan.** 2004. "Information Revelation and Bid Shading in Sequential Auctions with Price-Proportional Benefits to Bidders: Theory and Evidence from Rotating Savings and Credit Associations." January, Working Paper.

**Klonner, Stefan, and Ashok S. Rai.** 2010. "Financial Fragmentation And Insider Arbitrage." February, Working Paper.

**Koszegi, Botond, and Adam Szeidl.** 2013. "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics*, 128(1): 53–104.

**Kothari, S.P.** 2001. "Capital Markets Research in Accounting." *Journal of Accounting and Economics*, 31(1-3): 105–231.

**Leten, Bart, Rene Belderbos, and Bart Van Looy.** 2007. "Technological Diversification, Coherence, and Performance of Firms." *Journal of Product Innovation Management*, 24(6): 567–579.

**Li, Mingfang, and Yim-Yu Wong.** 2003. "Diversification and Economic Performance: An Empirical Assessment of Chinese Firms." *Asia Pacific Journal of Management*, 20(2): 243–265.

**Libby, Robert, Robert Bloomfield, and Mark W. Nelson.** 2002. "Empirical Research in Financial Accounting." *Accounting, Organizations and Society*, 27(8): 775–810.

**Lipe, M.G.** 1998. "Individual Investors' Risk Judgments and Investment Decisions: The Impact of Accounting and Market Data." *Accounting, Organizations and Society*, 23(7): 625–640.

**Luca, Michael, and Jonathan Smith.** 2011. "Salience in Quality Disclosure: Evidence from The US News College Rankings." September, HBS Working Paper 12-014.

**Madestam, Andreas.** 2010. "Informal Finance: A Theory of Moneylenders." November, IGIER Working Paper No. 347.

**Maines, Laureen A.** 1995. "Judgment and Decision-Making Research in Financial Accounting: A Review and Analysis." In *Judgment and Decision-Making Research in Accounting and Auditing*. ed. by Robert H. Ashton, New York: Cambridge University Press, 76–101.

**Masuyama, Seiichi, Donna Vandenbrink, and Siow Yue Chia.** 1999. *East Asia's Financial Systems: Evolution & Crisis.*: Institute of Southeast Asian Studies.

**Milgrom, Paul R.** 1981. "Good News and Bad News: Representation Theorems and Applications." *The Bell Journal of Economics*, 12(2): 380–391.

**Mukamel, Dana B., and Alvin I. Mushlin.** 1998. "Quality of Care Information Makes a Difference: An Analysis of Market Share and Price Changes after Publication of the New York State Cardiac Surgery Mortality Reports." *Medical Care*, 36(7): 945–954.

**Nelson, Phillip.** 1970. "Information and Consumer Behavior." *Journal of Political Economy*, 78(2): 311–329.

**Patrick, Kevin, Fred Raab, Marc A. Adams, Lindsay Dillon, Marian Zabinski, Cheryl L. Rock, William G. Griswold, and Gregory J. Norman.** 2009. "Text Message-Based Intervention for Weight Loss: Randomized Controlled Trial." *Journal of Medical Internet Research*, 11(1): , p. e1.

**Pope, Devin.** 2009. "Reacting to Rankings: Evidence from 'America's Best Hospitals'." *Journal of Health Economics*, 28(6): 1154–1165.

**Ramaswamy, Kannan, Mingfang Li, and Barbara S. Pécherot Petitt.** 2004. "Who Drives Unrelated Diversification? A Study of Indian Manufacturing Firms." *Asia Pacific Journal of Management*, 21(4): 403–423.

**Rao, Preethi.** 2007. "Chit Funds - A Boon to the Small Enterprises." January, Institute for Financial Management and Research Institute for Financial Management and Research, Small Enterprise Finance Centre, Working Paper Series.

**Rong, Zhao, and Sheng Xiao.** 2014. "Innovation-Related Diversification and Firm Value." In *Best Paper Proceedings of Seventy-Fourth Annual Meeting of the Academy of Management*. ed. by John Humphreys, January, No. 16765.

**Satkunasingam, Elsa Bawani, and Bala Shanmugam.** 2006. "Underground Banking in Malaysia: A Case Study of ROSCAs." *Journal of Money Laundering Control*, 9(1): 99–111.

**Scanlon, Dennis P., Michael Chernew, Catherine McLaughlin, and Gary Solon.** 2002. "The Impact of Health Plan Report Cards on Managed Care Enrollment." *Journal of Health Economics*, 21(1): 19–41.

**Scherer, F. M.** 1984. "Corporate Size, Diversification, and Innovative Activity." In *Innovation and Growth: Schumpeterian Perspectives*. ed. by F. M. Scherer, Cambridge, MA: The MIT Press, 222–238.

**Simon, Herbert A.** 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, 69(1): 99–118.

**Simon, Herbert A.** 1957. *Models of Man: Social and Rational*. New York: John Wiley and Sons, Inc..

**Soni, Praveen K., Gary L. Lilien, and David T. Wilson.** 1993. "Industrial Innovation and Firm Performance: A Re-Conceptualization and Exploratory Structural Equation Analysis." *International Journal of Research in Marketing*, 10(4): 365–380.

**Spence, Michael.** 1973. "Job Market Signaling." *The Quarterly Journal of Economics*, 87(3): 355–374.

**Su, Larry D.** 2010. "Ownership Structure, Corporate Diversification and Capital Structure: Evidence from China's Publicly Listed Firms." *Management Decision*, 48(2): 314–339.

**Sugheir, Jeff, Phillip H. Phan, and Iftekhar Hasan.** 2012. "Diversification and Innovation Revisited: An Absorptive Capacity View of Technological Knowledge Creation." *IEEE Transactions on Engineering Management*, 59(4): 530–539.

**Taubinsky, Dmitry.** 2014. "From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice." February, Working Paper.

**Teece, David J.** 1980. "Economies of Scope and the Scope of the Enterprise." *Journal of Economic Behavior & Organization*, 1(3): 223–247.

**Teece, David J.** 1996. "Firm Organization, Industrial Structure, and Technological Innovation." *Journal of Economic Behavior & Organization*, 31(2): 193–224.

**Tversky, Amos, and Daniel Kahneman.** 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology*, 5(2): 207–232.

**Tversky, Amos, and Daniel Kahneman.** 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185(4157): 1124–1131.

**United Nations: Economic and Social Commission for Asia and the Pacific.** 2002. *Rejuvenating Bank Finance For Development In Asia And The Pacific*. New York: United Nations: United Nations Publication.

**Vervloet, Marcia, Annemiek J. Linn, Julia C. M. van Weert, Dinny H. de Bakker, Marcel L. Bouvy, and Liset van Dijk.** 2012. "The Effectiveness of Interventions Using Electronic Reminders to Improve Adherence to Chronic Medication: A Systematic Review of the Literature." *Journal of the American Medical Informatics Association*, 19(5): 696–704.

**Wan, Chun-Cheong.** 1998. "International Diversification, Industrial Diversification and Firm Performance of Hong Kong MNCs." *Asia Pacific Journal of Management*, 15(2): 205–217.

**Wang, Zhaoning.** 2011. "Coexistence or Conflict in the Indian Financial Markets." Undergraduate Honors Thesis, Williams College, Williamstown, MA.

**Wedig, Gerard J., and Ming Tai-Seale.** 2002. "The Effect of Report Cards on Consumer Choice in the Health Insurance Market." *Journal of Health Economics*, 21(6): 1031–1048.

**Zhang, Guibin.** 2008. "The Choice of Formal or Informal Finance: Evidence from Chengdu, China." *China Economic Review*, 19(4): 659–678.

**Zhao, Hongxin, and Yadong Luo.** 2002. "Product Diversification, Ownership Structure, and Subsidiary Performance in China's Dynamic Market." *Management International Review*, 42(1): 27–48.

# Appendix A

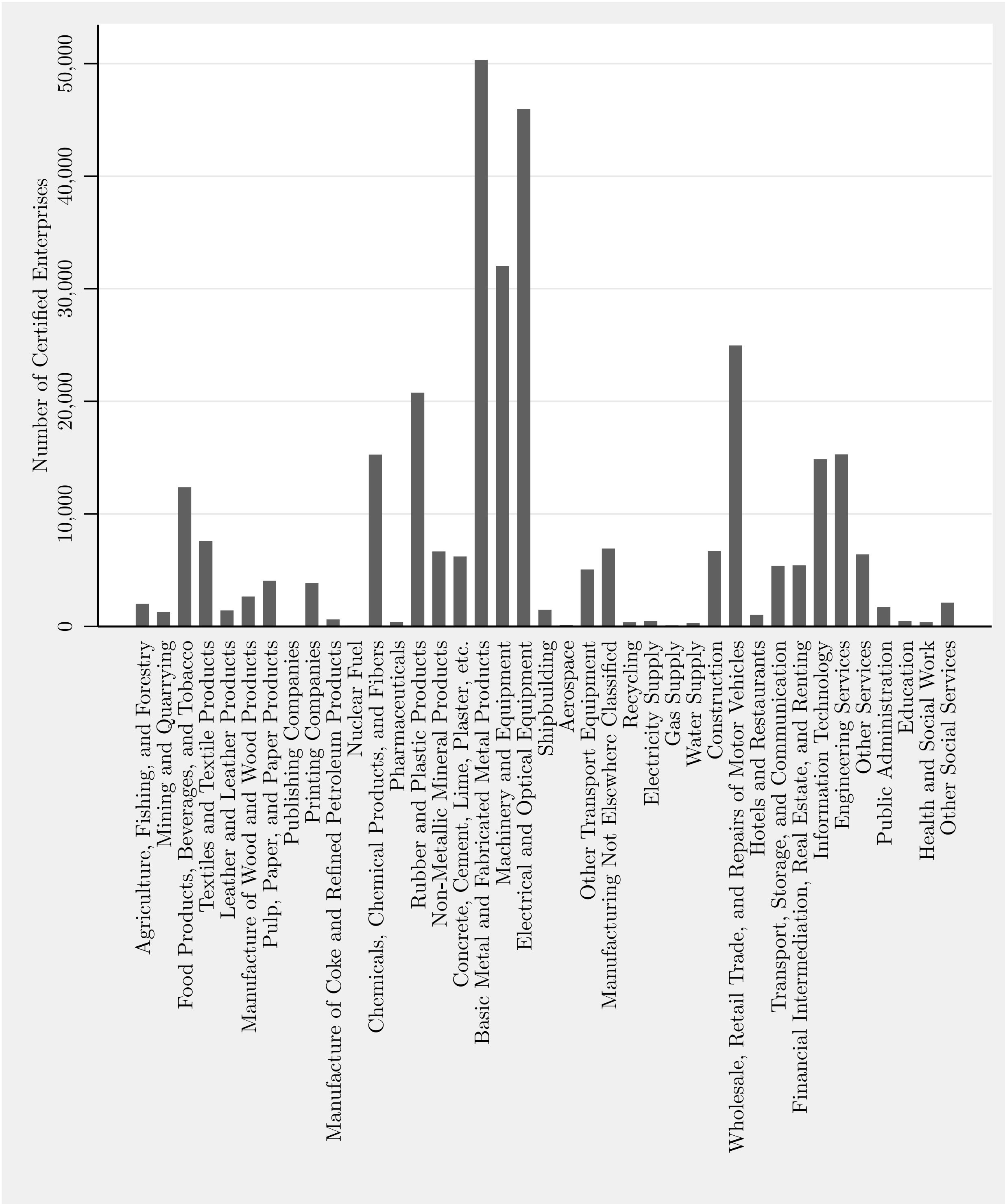
# Supplement to Chapter 1

## A.1 Overview of ISO9001 Certification

**Table A.1: Number of Firms with ISO9001 Certification by Region**

| | 1995 | 2000 | 2005 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|
| *Overview* | | | | | | | | |
| Africa | 1,563 | 4,769 | 6,763 | 7,667 | 8,164 | 9,674 | 9,816 | 10,308 |
| Central and South America | 1,220 | 10,805 | 22,498 | 49,260 | 51,685 | 51,459 | 52,466 | 50,256 |
| North America | 10,374 | 48,296 | 59,663 | 36,632 | 37,530 | 38,586 | 48,579 | 50,533 |
| Europe | 92,611 | 219,561 | 377,172 | 530,039 | 459,367 | 469,739 | 482,620 | 483,710 |
| East Asia and Pacific | 19,766 | 109,217 | 266,100 | 438,477 | 471,836 | 476,106 | 467,320 | 476,027 |
| Central and South Asia | 1,038 | 6,411 | 27,966 | 37,596 | 33,577 | 32,373 | 44,847 | 45,365 |
| Middle East | 776 | 9,003 | 13,681 | 18,839 | 17,069 | 19,050 | 20,812 | 21,956 |
| Total | 127,348 | 408,062 | 773,843 | 1,118,510 | 1,079,228 | 1,096,987 | 1,126,460 | 1,138,155 |
| | | | | | | | | |
| *Regional Share* | | | | | | | | |
| Africa | 1.2% | 1.2% | 0.9% | 0.7% | 0.8% | 0.9% | 0.9% | 0.9% |
| Central and South America | 1.0% | 2.6% | 2.9% | 4.4% | 4.8% | 4.7% | 4.7% | 4.4% |
| North America | 8.1% | 11.8% | 7.7% | 3.3% | 3.5% | 3.5% | 4.3% | 4.4% |
| Europe | 72.7% | 53.8% | 48.7% | 47.4% | 42.6% | 42.8% | 42.8% | 42.5% |
| East Asia and Pacific | 15.5% | 26.8% | 34.4% | 39.2% | 43.7% | 43.4% | 41.5% | 41.8% |
| Central and South Asia | 0.8% | 1.6% | 3.6% | 3.4% | 3.1% | 3.0% | 4.0% | 4.0% |
| Middle East | 0.6% | 2.2% | 1.8% | 1.7% | 1.6% | 1.7% | 1.8% | 1.9% |
| Total | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

Source: ISO Survey 2014, by International Organization for Standards




Source: ISO Survey 2014, by International Organization for Standards

**Figure A.1: Number of Firms with ISO9001 Certification in China by Industry**

## A.2 Card Designs

Treatment 1 (Brand):


Treatment 2 (Traceability):


Treatment 3 (Certification):


**Figure A.2: Card Designs for Treatment Groups**

## A.3 Supplementary Tables for In-Store Study

### A.3.1 Alternative Regression Results using OLS

For comparison purposes, Table A.2, Table A.3, and Table A.4 present alternative regression results using OLS, which is not used for the main regressions because of the small number of randomization units. It is worth noting that the magnitudes of the treatment effects are the same as in Table 1.9, Table 1.10, and Table 1.11, respectively, where randomization inference is employed.

### A.3.2 Comparison between High-Income Districts and Low-Income Districts

Table A.5, Table A.6, and Table A.7 provide the regression results comparing the treatment effects on high-income districts with those on low-income districts. Table A.8 provides the summary of the hypothesis tests between high-income and low-income districts.

**Table A.2: Treatment Effect on Beingmate Products**

*For Period 0 vs. Period 1*

| | All Products | | | Milk Products | | | Non-Milk Products | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) Quantity | (2) Quantity | (3) Quantity | (4) Quantity | (5) Quantity | (6) Quantity | (7) Quantity | (8) Quantity | (9) Quantity |
| Period 1 (vs. 0) | 0.022 | 0.022 | 0.021 | 0.054** | 0.054** | 0.061 | −0.006 | −0.006 | −0.012 |
| | (0.015) | (0.015) | (0.040) | (0.024) | (0.024) | (0.048) | (0.012) | (0.012) | (0.038) |
| Period 1 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 1 × | −0.000 | −0.001 | −0.016 | −0.056 | −0.057 | −0.056 | 0.057 | 0.056 | 0.037 |
| Brand | (0.020) | (0.020) | (0.040) | (0.035) | (0.035) | (0.051) | (0.034) | (0.034) | (0.042) |
| Period 1 × | −0.064*** | −0.063*** | −0.054 | −0.068*** | −0.066** | −0.071 | −0.070*** | −0.070*** | −0.062 |
| Traceability | (0.015) | (0.015) | (0.040) | (0.025) | (0.025) | (0.049) | (0.013) | (0.013) | (0.040) |
| Period 1 × | −0.061** | −0.006 | −0.035 | −0.075** | 0.002 | −0.055 | −0.053* | −0.008 | −0.021 |
| Certification | (0.026) | (0.030) | (0.042) | (0.036) | (0.028) | (0.050) | (0.029) | (0.038) | (0.047) |
| Unit Price | −0.000** | −0.000* | 0.000 | 0.000 | 0.000 | 0.000 | −0.005*** | −0.005*** | −0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.001) | (0.001) |
| Store Promotion | | 2.181*** | | | 2.856*** | | | 1.827*** | |
| | | (0.023) | | | (0.027) | | | (0.013) | |
| Salesperson | | | 1.985*** | | | 2.373*** | | | 1.830*** |
| | | | (0.166) | | | (0.326) | | | (0.159) |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2209 | 2209 | 2209 | 1017 | 1017 | 1017 | 1192 | 1192 | 1192 |
| $R^2$ | 0.164 | 0.215 | 0.513 | 0.166 | 0.249 | 0.479 | 0.247 | 0.283 | 0.609 |

Standard errors in parentheses (clustered at the postal zone level)

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

**Table A.3: Treatment Effect on Beingmate Products**

*For Period 1 vs. Period 2*

| | All Products | | | Milk Products | | | Non-Milk Products | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) Quantity | (2) Quantity | (3) Quantity | (4) Quantity | (5) Quantity | (6) Quantity | (7) Quantity | (8) Quantity | (9) Quantity |
| Period 2 (vs. 1) | 0.000 | −0.000 | −0.004 | −0.026 | −0.026 | −0.038 | 0.024 | 0.024 | 0.026* |
| | (0.021) | (0.021) | (0.017) | (0.038) | (0.038) | (0.032) | (0.018) | (0.018) | (0.015) |
| Period 2 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 2 × | −0.035 | −0.035 | 0.186 | 0.019 | 0.019 | 0.188 | −0.047 | −0.047 | 0.198 |
| Brand | (0.048) | (0.048) | (0.118) | (0.050) | (0.050) | (0.113) | (0.042) | (0.042) | (0.142) |
| Period 2 × | 0.062*** | 0.062*** | 0.066*** | 0.162*** | 0.162*** | 0.173*** | −0.027 | −0.027 | −0.028 |
| Traceability | (0.021) | (0.021) | (0.018) | (0.058) | (0.058) | (0.056) | (0.053) | (0.053) | (0.048) |
| Period 2 × | −0.043 | 0.158 | −0.054 | −0.052 | 0.250 | −0.024 | −0.035 | 0.122 | −0.070 |
| Certification | (0.030) | (0.146) | (0.060) | (0.040) | (0.238) | (0.071) | (0.033) | (0.119) | (0.068) |
| Unit Price | −0.000** | −0.000** | 0.000 | −0.000 | −0.000 | 0.000 | −0.004*** | −0.004*** | −0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.001) | (0.001) |
| Store Promotion | | 1.333*** | | | 1.735*** | | | 1.174*** | |
| | | (0.267) | | | (0.522) | | | (0.185) | |
| Salesperson | | | 2.053*** | | | 2.749*** | | | 1.834*** |
| | | | (0.133) | | | (0.271) | | | (0.136) |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2231 | 2231 | 2231 | 998 | 998 | 998 | 1233 | 1233 | 1233 |
| $R^2$ | 0.144 | 0.158 | 0.469 | 0.141 | 0.158 | 0.467 | 0.219 | 0.233 | 0.552 |

Standard errors in parentheses (clustered at the postal zone level)

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

**Table A.4: Treatment Effect on Beingmate Products**

*For Period 0 vs. Period 2*

| | All Products | | | Milk Products | | | Non-Milk Products | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) Quantity | (2) Quantity | (3) Quantity | (4) Quantity | (5) Quantity | (6) Quantity | (7) Quantity | (8) Quantity | (9) Quantity |
| Period 2 (vs. 0) | 0.017 | 0.017 | 0.008 | 0.020 | 0.020 | 0.012 | 0.017 | 0.017 | 0.008 |
| | (0.013) | (0.013) | (0.034) | (0.023) | (0.023) | (0.039) | (0.018) | (0.018) | (0.037) |
| Period 2 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 2 × | −0.027 | −0.027 | 0.154 | −0.035 | −0.035 | 0.110 | 0.016 | 0.016 | 0.218 |
| Brand | (0.032) | (0.032) | (0.109) | (0.048) | (0.048) | (0.081) | (0.017) | (0.017) | (0.154) |
| Period 2 × | −0.000 | −0.000 | 0.015 | 0.095 | 0.095 | 0.102 | $-0.102^{*}$ | $-0.102^{*}$ | −0.090 |
| Traceability | (0.013) | (0.013) | (0.036) | (0.061) | (0.062) | (0.072) | (0.056) | (0.056) | (0.064) |
| Period 2 × | $-0.098^{**}$ | 0.149 | −0.088 | $-0.115^{***}$ | 0.261 | −0.085 | $-0.093^{*}$ | 0.069 | −0.090 |
| Certification | (0.044) | (0.179) | (0.054) | (0.037) | (0.256) | (0.061) | (0.051) | (0.131) | (0.060) |
| Unit Price | $-0.001^{***}$ | $-0.001^{***}$ | −0.000 | $-0.000^{**}$ | $-0.000^{**}$ | −0.000 | $-0.006^{***}$ | $-0.006^{***}$ | $-0.004^{***}$ |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.001) | (0.001) |
| Store Promotion | | $1.423^{***}$ | | | $1.976^{***}$ | | | $0.997^{***}$ | |
| | | (0.315) | | | (0.449) | | | (0.230) | |
| Salesperson | | | $1.847^{***}$ | | | $2.129^{***}$ | | | $1.721^{***}$ |
| | | | (0.124) | | | (0.239) | | | (0.106) |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2224 | 2224 | 2224 | 1017 | 1017 | 1017 | 1207 | 1207 | 1207 |
| $R^2$ | 0.168 | 0.190 | 0.488 | 0.166 | 0.214 | 0.492 | 0.228 | 0.237 | 0.528 |

Standard errors in parentheses (clustered at the postal zone level)

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Table A.5: Comparison between High-Income Districts and Low-Income Districts**

*For Period 0 vs. Period 1*

| | All Products | | | | Milk Products | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) High | (2) Low | (3) High | (4) Low | (5) High | (6) Low | (7) High | (8) Low |
| Period 1 (vs. 0) | 0.026 | 0.028 | 0.026 | 0.028 | 0.077** | 0.053* | 0.076** | 0.053* |
| | (0.023) | (0.018) | (0.023) | (0.018) | (0.036) | (0.030) | (0.036) | (0.030) |
| Period 1 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 1 × | −0.007 | −0.004 | −0.007 | −0.005 | −0.079 | −0.081* | −0.081* | −0.083* |
| Brand | (0.027) | (0.026) | (0.027) | (0.026) | (0.047) | (0.041) | (0.047) | (0.042) |
| Period 1 × | −0.065*** | −0.072*** | −0.064*** | −0.071*** | −0.090** | −0.064** | −0.088** | −0.062** |
| Traceability | (0.021) | (0.017) | (0.021) | (0.017) | (0.034) | (0.030) | (0.035) | (0.030) |
| Period 1 × | −0.084*** | −0.065* | −0.018 | 0.018 | −0.106** | −0.094** | −0.018 | 0.010 |
| Certification | (0.026) | (0.037) | (0.037) | (0.019) | (0.047) | (0.038) | (0.040) | (0.033) |
| Unit Price | −0.000 | −0.001** | −0.000 | −0.000* | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Store Promotion | | | 2.199*** | 2.185*** | | | 2.861*** | 2.873*** |
| | | | (0.030) | (0.024) | | | (0.038) | (0.031) |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1452 | 1723 | 1452 | 1723 | 665 | 801 | 665 | 801 |
| $R^2$ | 0.151 | 0.170 | 0.224 | 0.231 | 0.166 | 0.178 | 0.264 | 0.286 |

Standard errors in parentheses (clustered at the postal zone level)

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

**Table A.6: Comparison between High-Income Districts and Low-Income Districts**

*For Period 1 vs. Period 2*

| | All Products | | | | Milk Products | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) High | (2) Low | (3) High | (4) Low | (5) High | (6) Low | (7) High | (8) Low |
| Period 2 (vs. 1) | −0.027 | 0.006 | −0.027 | 0.006 | −0.075 | −0.019 | −0.075 | −0.019 |
| | (0.033) | (0.024) | (0.033) | (0.024) | (0.063) | (0.045) | (0.063) | (0.045) |
| Period 2 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 2 × | −0.008 | −0.051 | −0.008 | −0.051 | 0.065 | 0.015 | 0.066 | 0.015 |
| Brand | (0.055) | (0.062) | (0.055) | (0.062) | (0.075) | (0.065) | (0.075) | (0.065) |
| Period 2 × | 0.089*** | 0.056** | 0.089*** | 0.056** | 0.213*** | 0.153** | 0.212*** | 0.152** |
| Traceability | (0.032) | (0.024) | (0.032) | (0.023) | (0.073) | (0.063) | (0.074) | (0.063) |
| Period 2 × | −0.024 | −0.059* | 0.226 | 0.281 | −0.010 | −0.049 | 0.364 | 0.438* |
| Certification | (0.040) | (0.035) | (0.187) | (0.173) | (0.066) | (0.048) | (0.290) | (0.245) |
| Unit Price | −0.000 | −0.000** | −0.000 | −0.000* | −0.000 | 0.000 | −0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Store Promotion | | | 1.412*** | 1.572*** | | | 1.879*** | 2.171*** |
| | | | (0.340) | (0.317) | | | (0.627) | (0.533) |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1404 | 1821 | 1404 | 1821 | 621 | 819 | 621 | 819 |
| $R^2$ | 0.142 | 0.143 | 0.164 | 0.162 | 0.145 | 0.142 | 0.169 | 0.170 |

Standard errors in parentheses (clustered at the postal zone level)

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

**Table A.7: Comparison between High-Income Districts and Low-Income Districts**

*For Period 0 vs. Period 2*

| | All Products | | | | Milk Products | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) High | (2) Low | (3) High | (4) Low | (5) High | (6) Low | (7) High | (8) Low |
| Period 2 (vs. 0) | −0.001 | 0.030** | −0.001 | 0.030** | 0.007 | 0.027 | 0.007 | 0.026 |
| | (0.019) | (0.014) | (0.019) | (0.014) | (0.037) | (0.028) | (0.037) | (0.028) |
| Period 2 × Control | ref. | ref. | ref. | ref. | ref. | ref. | ref. | ref. |
| Period 2 × | −0.010 | −0.046 | −0.010 | −0.047 | −0.022 | −0.063 | −0.022 | −0.062 |
| Brand | (0.035) | (0.042) | (0.035) | (0.042) | (0.056) | (0.067) | (0.056) | (0.068) |
| Period 2 × | 0.019 | −0.015 | 0.020 | −0.015 | 0.108 | 0.088 | 0.108 | 0.088 |
| Traceability | (0.020) | (0.014) | (0.020) | (0.014) | (0.066) | (0.064) | (0.067) | (0.065) |
| Period 2 × | −0.110*** | −0.119* | 0.203 | 0.332 | −0.123*** | −0.126*** | 0.372 | 0.449 |
| Certification | (0.035) | (0.061) | (0.244) | (0.211) | (0.040) | (0.045) | (0.324) | (0.283) |
| Unit Price | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.000** | −0.000* | −0.000* | −0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Store Promotion | | | 1.490*** | 1.768*** | | | 2.146*** | 2.320*** |
| | | | (0.430) | (0.373) | | | (0.567) | (0.495) |
| Store Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1432 | 1768 | 1432 | 1768 | 652 | 818 | 652 | 818 |
| $R^2$ | 0.166 | 0.169 | 0.201 | 0.201 | 0.181 | 0.167 | 0.250 | 0.239 |

Standard errors in parentheses (clustered at the postal zone level)

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

**Table A.8: Hypothesis Testing Results between High-Income Districts and Low-Income Districts**

*p-Values Corresponding to Each Hypothesis Test on the Interaction Terms*

| | Period × Brand | Period × Traceability | Period × Certification |
|---|---|---|---|
| *Period 0 vs. Period 1* | | | |
| Table A.5, (1) & (2) | 0.843 | 0.547 | 0.509 |
| Table A.5, (3) & (4) | 0.823 | 0.539 | 0.248 |
| Table A.5, (5) & (6) | 0.917 | 0.060 | 0.637 |
| Table A.5, (7) & (8) | 0.924 | 0.056 | 0.100 |
| *Period 1 vs. Period 2* | | | |
| Table A.6, (1) & (2) | 0.090 | 0.069 | 0.171 |
| Table A.6, (3) & (4) | 0.090 | 0.070 | 0.728 |
| Table A.6, (5) & (6) | 0.135 | 0.034 | 0.223 |
| Table A.6, (7) & (8) | 0.136 | 0.034 | 0.757 |
| *Period 0 vs. Period 2* | | | |
| Table A.7, (1) & (2) | 0.039 | 0.013 | 0.861 |
| Table A.7, (3) & (4) | 0.039 | 0.013 | 0.550 |
| Table A.7, (5) & (6) | 0.185 | 0.201 | 0.910 |
| Table A.7, (7) & (8) | 0.188 | 0.205 | 0.794 |

Note: In each interaction term, “Period” refers to the latter period within each comparison group.

# Appendix B

# Supplement to Chapter 2

## B.1 Supplementary Tables

**Table B.1: Hypothesis Testing Results between Different Diversity Measures**
*p-Values Corresponding to Each Hypothesis Test*

$H_0$: Diversity Measure (Narrow) = Diversity Measure (Broad)
$H_1$: Diversity Measure (Narrow) $\neq$ Diversity Measure (Broad)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *All Firms* | | | | | | |
| Production Sector | 0.992 | 1.000 | 0.953 | 0.994 | 0.986 | 0.943 |
| Service Sector | 0.076 | 0.087 | 0.125 | 0.107 | 0.114 | 0.158 |
| *Actively Innovating Firms* | | | | | | |
| Production Sector | 0.752 | 0.752 | 0.730 | 0.649 | 0.650 | 0.615 |
| Service Sector | 0.036 | 0.039 | 0.043 | 0.103 | 0.109 | 0.125 |

**Table B.2: Hypothesis Testing Results between the Production Sector and the Service Sector**

*p-Values Corresponding to Each Hypothesis Test*

$H_0$: Diversity Measure in Production Sector = Diversity Measure in Service Sector
$H_1$: Diversity Measure in Production Sector < Diversity Measure in Service Sector

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *All Firms* | | | | | | |
| Diversity Measure (Narrow) | 0.079 | 0.078 | 0.073 | 0.046 | 0.046 | 0.042 |
| Diversity Measure (Broad) | 0.200 | 0.193 | 0.172 | 0.151 | 0.147 | 0.130 |
| *Actively Innovating Firms* | | | | | | |
| Diversity Measure (Narrow) | 0.066 | 0.067 | 0.070 | 0.028 | 0.028 | 0.028 |
| Diversity Measure (Broad) | 0.166 | 0.165 | 0.164 | 0.099 | 0.099 | 0.094 |

**Table B.3: Hypothesis Testing Results between All Firms and Actively Innovating Firms**

*p-Values Corresponding to Each Hypothesis Test*

$H_0$: Diversity Measure for All Firms = Diversity Measure for Actively Innovating Firms
$H_1$: Diversity Measure for All Firms < Diversity Measure for Actively Innovating Firms

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *Production Sector* | | | | | | |
| Diversity Measure (Narrow) | 0.037 | 0.037 | 0.046 | 0.016 | 0.016 | 0.017 |
| Diversity Measure (Broad) | 0.067 | 0.069 | 0.081 | 0.045 | 0.046 | 0.048 |
| *Service Sector* | | | | | | |
| Diversity Measure (Narrow) | 0.159 | 0.152 | 0.149 | 0.258 | 0.253 | 0.234 |
| Diversity Measure (Broad) | 0.091 | 0.089 | 0.084 | 0.185 | 0.187 | 0.174 |

# Appendix C

# Supplement to Chapter 3

## C.1 A Detailed Description of ROSCAs[1]

### C.1.1 Importance and Prevalence of Chit Funds

In many developing countries, chit funds are often considered to be one of the most important instruments to cater to the financial needs of the poor (Besley et al., 1993; Bouman, 1995; Rao, 2007). On the one hand, due to the inconvenience of transportation or the lack of communication, many of the individuals or small businesses barely have any information on banks available in the neighborhood, nor are they acquainted with the process of taking loans. On the other hand, banks are highly unlikely to issue loans to the poor, who are often identified as the riskiest borrowers due to their lack of creditworthiness (Banerjee and Duflo, 2012). Hence, chit funds have become an easy and profitable source of funding for impoverished individuals and credit-constrained small and medium enterprises (SMEs).

Chit funds are most common in developing countries, but some immigrant groups in the United States also utilize them under certain circumstances as an insurance mechanism to their borrowings elsewhere. Such a phenomenon has also been observed among business

[1]Portions of this section are directly drawn from my previous work (Wang, 2011).

owners in Zhejiang Province in Southeast China. Moreover, some variations of ROSCAs are also beginning to develop in other developed countries like France and Japan.[2]

### C.1.2 Characteristics of ROSCA Participants

Similar as the beneficiaries of most microfinance institutions in developing countries, the majority of ROSCA participants are small traders and businesses that do not have an established credit history. In some cases, there is also extensive participation from housewives and salaried employees. In terms of the composition of participants, bidding ROSCAs resemble the Grameen banking system and other microfinance institutions.

Despite their role of providing credit for the credit-constrained entities, ROSCAs do have strict procedures in admitting new members. Due to the "relationship banking" nature of bidding ROSCAs, members are usually required to be connected to other existing members or to the chit fund manager directly. This admittance requirement is very different from formal financial institutions like banks, where borrowers are not usually known to the banks prior to applying for loans. Hence, compared with borrowers from banks, it is somewhat easier for each ROSCA participant to observe others' socioeconomic status and business skills, which can be indicators of their ability to repay.

Moreover, ROSCAs require rigorous verifications to ensure the credibility of the members in order to minimize the likelihood of default. For instance, even after the initial acceptance to the bidding ROSCAs, new members may be unable to participate in the auctions immediately. Instead, during the first few months of the ROSCA, they are required to make monthly contributions without being able to bid in the auctions.[3] Unlike in most banks, new members of

[2]However, according to the descriptions of Izumida (1992), the *Kou* system, which was first introduced in Japan around the thirteenth century, has many features that resemble those of the contemporary ROSCAs.

[3]Since each participant is guaranteed to be the winner once and only once in a given ROSCA, this practice of not allowing newcomers to bid in the beginning is simply an assessment of their financial capacity rather than an extra fee charged to them.

bidding ROSCAs must have a guarantor, sometimes referred to as the co-signer, whom the chit fund manager knows well and is able to trust. Finally, similar to most microfinance institutions, ROSCA participants are sometimes required to have a collateral before participating in the ROSCAs in case they default. However, once a member is formally admitted to the ROSCA and becomes eligible for bidding, little further documentation will be required. Nowadays, most of the chit fund companies require their members to have at least a deposit account since all of the transactions are made through check payments (Rao, 2007). It should also be noted that not all ROSCA participants are credit-constrained; in fact, some of them do have access to the formal credit market.

### C.1.3 Bidding Mechanisms

The specific mechanisms of the bidding ROSCAs are described as follows. Although these bidding mechanisms might vary slightly across regions, the basic "rotating" nature of the bidding ROSCAs is the same across different countries.

1. Each bidding ROSCA lasts for a fixed number of months, normally from a few months to over five years. In each ROSCA, the number of participants is equal to the number of months, both of which are denoted by $N$.

2. Participating members contribute a fixed amount of money $k$ to an imaginary "pot" each month. Here, the *chit value* $V$ is defined as the total amount of money in the "pot," i.e., $V = N \cdot k$.

3. An auction will then be conducted, and participants bid to receive the pot in an open ascending-bid auction, where previous winners must continue to make the required monthly contributions but are not eligible to bid until the end of a given ROSCA (with the exception of multiple-membership as discussed below).

4. The highest bidder in each month, also known as the *prized subscriber*, wins the chit

value and pays the bid amount called *discount*.[4] The discount is then distributed among the rest of the ROSCA participants as *dividend*, so that $dividend = discount/(N-1)$.

Multiple-membership is allowed for some bidding ROSCAs.[5] However, even with multiple-membership, participants can bid again in a given ROSCA only after half of all the auctions have been completed.

### C.1.4 Usage of ROSCA Winnings

There are two major uses of the chit funds obtained from ROSCAs: household expenditures and investment projects for small businesses.

For individual households, participation in bidding ROSCAs is mainly for consumption purposes, that is, for the purpose of purchasing large property, usually indivisible, durable goods in the traditional sense (Besley et al., 1993; Calomiris and Rajaraman, 1998). In recent decades, there have been considerable alterations in the role of chit funds, and investments on human capital have also been partly financed by chit funds (Eeckhout and Munshi, 2010; Rao, 2007). For instance, consumption purposes, in a broader sense, might also include buying land or paying for education, especially children's secondary education and beyond. Since a large portion of ROSCA participants are women, ROSCA winnings are also regarded as a way to save "free cash" by women of the households. In some cases, funds obtained through bidding ROSCAs may also be used to repay outstanding loans with other moneylenders (Rao, 2007). However, members using this practice may be subject to the risk of default as debt starts to accumulate (Besley et al., 1993; Bouman, 1995; Madestam, 2010; Rao, 2007). This situation is very similar to what has happened to many microfinance institutions. Moreover, like microfinance institutions, ROSCA members are not typically required to specify the

[4] In times of an equal bid, the decision of who is entitled to the loan will be made by means of a lottery.

[5] Although multiple-membership is allowed, I only consider the case of single-membership in the main study, because multiple-memberships can be approximated by two participants with exactly the same characteristics.

purpose of the funds to chit fund managers.

In terms of chit funds used towards small businesses, the funds are generally used as either working capital for business expansion or as emergency funds. Chit funds are also preferred by many such enterprises in order to overcome their financial constraints (Rao, 2007). Moreover, instead of receiving an interest rate quote from the banks, ROSCA participants can implicitly determine their own implied "interest rates" depending on their need.[6] More importantly, ROSCA participants are not required to file income tax forms, which potentially makes ROSCAs a more appealing financing channel than banks and other formal financial institutions, especially for small businesses.

### C.1.5 Regulation and Registration of Chit Funds

Most chit fund companies are loosely regulated by the government or unregulated at all. Although the regulation is far less strict than that for banks, regulation is generally considered to be beneficial in terms of ensuring proper operations of the bidding ROSCAs. Moreover, chit fund companies are required to register before their operations. The detailed terms of the regulation are outlined below based on the work of Rao (2007).

1. Registration of a chit scheme involves a number of fees and formalities. For instance, the chit fund manager is required to file returns, maintain minutes of meetings, audit accounts, and so on.

2. Prior sanction is required from the Registrar (an application plus a fee of 50 rupees, which is about one dollar). Then, the chit fund company must file a chit fund agreement with every member in a particular group, with a cost of around 20 rupees (about 44 cents) per member. Once the agreement is filed and approved, the Registrar will issue

[6]These interest rates are "implied" because, unlike banks, there is no publicly announced lending rate or borrowing rate in bidding ROSCAs. In other words, different participants might have different interest rates depending on their bidding strategies and abilities to repay (Klonner and Rai, 2010).

a certificate for the commencement of the ROSCA schemes.

3. The chit fund manager is required to deposit 100% of the chit value to the Registrar, which will be refunded upon successful completion of the chit cycle.

4. All registered chit fund companies are required to impose a 30% cap of the total chit value (increased to 40% in 2007) on the bidding amount to ensure that bidders do not bid beyond their ability to repay. At the same time, the minimum bid is restricted to be at least 5% of the total chit value, which is treated as the chit fund manager's commission.

5. The registration of chit funds improves the *transparency* and *accountability* of its operations. It also boosts the participants' confidence and reduces the associated risk.

### C.1.6 Possible Failures of Bidding ROSCAs

Like banks and other microfinance institutions, bidding ROSCAs do fail at times when their participants default. There are two possible forms of default in a bidding ROSCA. First, the winner in a particular round may fail to pay the winning bid to the other participants in the ROSCA. Second, participants may fail to pay their monthly contributions. As documented in Rao (2007), the default rate in the chit fund industry is usually as low as 1-2%. However, from the design of bidding ROSCAs, it is easy to see that even the default by a single person will cause serious issues and affect the payoffs of all other members. Moreover, late payments are common among ROSCA participants, although late payments do not count towards default as long as they are made before the end of a given ROSCA.

Three steps will be taken subsequently when members fail to make their monthly contributions: (1) oral correspondence, (2) a reminder sent by mail (if oral correspondence fails), and (3) a legal notice issued to take the person to court (if both fails).

In the event of default, the court has the right to seize the salary or the property of ROSCA members to repay the unpaid dues. In the most severe case, the guarantor will be asked to make the payments. Moreover, those who fail to make timely payments are charged explicit interest on their dues. If the payment is delayed for an excessively long period, the corresponding ROSCA member may be deprived of the right to receive future dividends.

## C.2 Proofs of Propositions

### C.2.1 Proof of Proposition 3.1

*Proof.* Let $b_1$, $b_2$ be the bids submitted by player 1 and player 2, respectively. In order to show the Nash equilibrium, the first step is to show that $b_2 = y_1 - \varepsilon$ is the best response to $b_1 = y_1$. If $b_2 < y_1 - \varepsilon$, player 2 will lose the auction (since $b_2 < y_1$) and get a payoff of $b_2$. Apparently, player 2 has an incentive to deviate by increasing the bid and getting $y_1 - \varepsilon$, which is greater than $b_2$. If $b_2 > y_1 - \varepsilon$ for any $\varepsilon > 0$, it must be true that $b_2 \geq y_1$ so that player 2 will win the auction and thus pay $y_1$ to player 1. Hence, player 2's payoff is

$$2y_2 - b_2 \leq 2y_2 - y_1 < y_2 < y_1 - \varepsilon$$

so that player 2 again has an incentive to deviate. Therefore, by overbidding $b_2 = y_1 - \varepsilon > y_2$, player 2 has the optimal payoff of $y_1 - \varepsilon$.

The next step is to show that $b_1 = y_1$ is the best response to $b_2 = y_1 - \varepsilon$. If $b_1 < y_1$, then there must exist some $\varepsilon > 0$ for which $b_1 < b_2$. Hence, player 1 loses and gets a payoff of $b_1 < y_1$. It is easy to see that player 2 has an incentive to deviate by bidding $y_1$ and winning the auction, in which case the payoff would be

$$2y_1 - (y_1 - \varepsilon) = y_1 + \varepsilon > y_1 > b_1.$$

If $b_1 > y_1$, player 1 will win the auction with a payoff of $2y_1 - b_1$, which is lower than the payoff of $y_1 + \varepsilon$ if player 1 bids exactly $y_1$, so that player 1 again has an incentive to deviate. Hence, by bidding $b_1 = y_1$, player 1 has the optimal payoff of $y_1 + \varepsilon$.

Therefore, the Nash equilibrium of this auction is $(b_1, b_2) = (y_1, y_1 - \varepsilon)$. □

### C.2.2 Proof of Proposition 3.2

*Proof.* From Proposition 3.1, it is clear that the player with lower productivity is always better off to be the loser by bidding $b_2 < b_1 - \varepsilon$. Note that for the bidding ROSCA to sustain, it must be true that $b_1 \leq r_b$, otherwise borrowing from the bank would give player 1 a payoff of $2y_1 - r_b$ in the second period, which is higher than the payoff of $2y_1 - b_2$ by joining the ROSCA. Knowing that the opponent has higher productivity, player 2 has an incentive to bid $b_2 = b_1 - \varepsilon \leq r_b - \varepsilon$. Thus, player 2's payoff is at most $r_b - \varepsilon$. However, player 2 will be better off by borrowing from the bank instead, in which case player 2 would get $2y_2 - r_b > r_b$ instead of $r_b - \varepsilon$. Therefore, the bidding ROSCA will not sustain in this case. $\square$

### C.2.3 Proof of Proposition 3.3

*Proof.* As shown in the previous proof, $b_1 \leq r_b$. Knowing that player 1 has higher productivity, player 2 will bid $b_2 = b_1 - \varepsilon \leq r_b - \varepsilon$, which yields a payoff of at most $r_b - \varepsilon$. By saving in the bank, player 2's payoff will be $r_s \leq r_b - \varepsilon$ since $r_s < r_b$. By borrowing from the bank, player 2's payoff will be $2y_2 - r_b \leq r_b - \varepsilon$. Therefore, player 2's best response is to join the bidding ROSCA. Therefore, the bidding ROSCA will sustain, where player 1 bids $r_b$ and player 2 bids $r_b - \varepsilon$. $\square$

### C.2.4 Proof of Proposition 3.4

*Proof.* The proof is exactly the same as the proof for Proposition 3.3. $\square$

### C.2.5 Proof of Proposition 3.5

*Proof.* Since $y_1 \leq r_b$, player 1's best response is to join the ROSCA and bid $b_1 = y_1$. Knowing that $y_1 > y_2$, player 2 will bid $b_2 = y_1 - \varepsilon$, which yields a payoff of $y_1 - \varepsilon$. By saving in the bank, player 2's payoff will be $r_s \leq y_1 - \varepsilon$ since $r_s < y_1$. By borrowing from the bank, player 2's payoff will be $2y_2 - r_b < 2y_1 - y_1 = y_1$ so that $2y_2 - r_b \leq y_1 - \varepsilon$. Hence, player 2's best response is to join the bidding ROSCA. Therefore, the ROSCA will sustain, where player 1 bids $y_1$ and player 2 bids $y_1 - \varepsilon$. □

### C.2.6 Proof of Proposition 3.6

*Proof.* The proof is exactly the same as the proof for Proposition 3.5. □

### C.2.7 Proof of Proposition 3.7

*Proof.* From Proposition 3.1, it is clear that the player with lower productivity is always better off to be the loser by bidding $b_2 < b_1 - \varepsilon$. Note that for the bidding ROSCA to sustain, it must be true that $b_1 \leq r_s$, otherwise saving in the bank would give player 1 a payoff of $r_s$ in the second period, which is higher than the payoff of $2y_1 - b_2 < 2r_s - r_s = r_s$ by joining the ROSCA. Knowing that the opponent has higher productivity, player 2 has an incentive to bid $b_2 = b_1 - \varepsilon \leq r_s - \varepsilon$. Thus, player 2's payoff is at most $r_s - \varepsilon$. However, player 2 will be better off by saving in the bank instead, in which case player 2 would get $r_s$ instead of $r_s - \varepsilon$. Therefore, the bidding ROSCA will not sustain in this case. □

## C.3 Detailed Dataset Creation Process[7]

### C.3.1 Creating the Dataset of Nearby Banks

To create the dataset of bank openings in Andhra Pradesh, I relied heavily on the ArcGIS Geoprocessing Tools. The bank location data were downloaded from the Database on Indian Economy in RBI's Data Warehouse (`http://dbie.rbi.org.in`). One crucial part of creating this dataset was to find the locations as accurately as possible for all the 5,292 banks in Andhra Pradesh that opened between January 1998 and December 2000. For most banks,

[7]Portions of this section are drawn from my previous work (Wang, 2011).

their addresses and "pincodes" (i.e., postal codes) were already available in the original dataset provided by RBI, which could then be used to find the locations. However, there were still around 1,000 banks for which I had to find the locations manually. Since assuring both the accuracy and the consistency of the location data was key to the successful completion of my work, I carefully chose Google Earth and the banks' official websites as major tools to determine the locations, occasionally with the aid of other unofficial yet legitimate sources. Using the addresses and "pincodes," I found the latitudes and longitudes of all the banks. For the 219 bidding ROSCA branches, the process was similar but much easier.

Then, I imported all the location data to ArcGIS to generate a map of Andhra Pradesh with all the ROSCA and bank branches. For each of the ROSCA branch, I created "buffer circles" of various radii (from 1 km to 30 km) before I decided the best distance to use. The way I chose the most suitable distance (10 km) to use for the regression analysis was to choose a radius such that almost all the banks are covered by the buffer circles and that the circles do not overlap too much. For each "buffer circle," I calculated the number of banks started in each month from January 1998 to December 2000 that fall into the "buffer circle." To do this, I created a set of new variables named $count_t$ and assigned 1 to all the banks that started before a particular month $t$ and 0 to all the banks that started after month $t$. Using ArcGIS, I was able to sum up all the $count_t$ variables in each month $t$ for each "buffer circle," and eventually I obtained a dataset containing the number of banks started in each month between January 1998 and December 2000 in the neighborhood of each ROSCA branch.

### C.3.2 Creating the Dataset of ROSCA Branches

To generate the dataset on the number of ROSCA participants, the winning bids, and the amount of default for each ROSCA in each branch, I utilized the original dataset, where the variables of interest were available for monthly auctions that took place in each ROSCA branch between 1984 and 2005. Since most ROSCAs started after 1998 and the bank opening data after 2000 were unavailable at the time when I created the dataset, I dropped all the observations before January 1998 or after December 2000. Moreover, although most ROSCAs do not last longer than three years, there do exist some ROSCAs that started after January 1998 but did not end before December 2000. I then dropped those observations that do not qualify for this study and created a new dataset containing only the ROSCAs that started after January 1998 and ended before December 2000. Finally, I aggregated the number of participants each month for each of the 219 ROSCA branches to create the participation data used in my regressions.