ESSAYS ON FIRM AND PLATFORM OPERATIONS: INFORMATION, MARKETPLACE DESIGN AND DATA-DRIVEN DECISION MAKING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Xiaoyan Liu

August 2021

ESSAYS ON FIRM AND PLATFORM OPERATIONS: INFORMATION, MARKETPLACE DESIGN AND DATA-DRIVEN DECISION MAKING

Xiaoyan Liu, Ph.D.
Cornell University 2021

The recent years have seen the emergence and proliferation of online platforms that serve as an intermediation to connect demand and supply, ranging from business-to-business (B2B) platforms connecting buyers and sellers to peer-to-peer labor platforms connecting customers and service providers in the gig economy. With the advancement of digital technologies, online platforms are superior in their capability to match demand and supply more accurately, economically, and timely. In this dissertation, I study the design of such platforms with a focus on data-driven decision making, operational efficiency, and economics in Chapter 1 and 2. The evolution of the financial market has placed a profound influence on firms. Information asymmetry exists between firms and the investors in the financial market, and managerial short-termism has reported become more phenomenal in firms today. In Chapter 3, I study the impact of managerial short-termism on firms' operational decisions and long-term value. The titles of the three chapters of this dissertation are:

– Chapter 1: Personalized Recommendation System Design for an Online B2B Platform.

– Chapter 2: Bonus Competition in the Gig Economy.

– Chapter 3: Operational Distortion: Compound Effects of Short-termism and Competition.

Chapter 1 is joint work with Professor Vishal Gaur. Chapter 2 is joint work with Professor Li Chen and Professor Yao Cui. Chapter 3 is joint work with Professor William Schmidt. This dissertation explores a suite of methodologies including data-driven predictive models, econometrics, and game theory. Chapter 1 develops a data-driven algorithm to design a recommendation system for an online B2B platform and tests its performance in a field experiment. Chapter 2 adopts a Hotelling framework to investigate the pricing strategies of competing labor platforms in the gig economy and their impact on market shares, platform profits, and social welfare. Chapter 3 constructs a signaling game model to examine the impact of managerial short-termism on firms' operational distortion and long-term profits in a competitive environment.

BIOGRAPHICAL SKETCH

Xiaoyan (Amber) Liu received a Bachelor's degree in Logistics Engineering from Tianjin University. She received her Master's degree and Doctoral degree in Operations, Technology, and Information Management from SC Johnson College of Business at Cornell University.

This document is dedicated to my beloved mother Aimin Liu, father Mingguang Liu, and my grandparents Yukun Liu, Heyuan Tang, Youyun Yang, and Zhengchun Liu.

ACKNOWLEDGEMENTS

On July 21 2015, I boarded a plane from Pudong, Shanghai heading to a town with a beautiful name called Ithaca. This marks the beginning of my PhD journey. That summer I was a college graduate not long before. Just as little did I expect I would be given the opportunity to study at Cornell, little did I envision that this journey would be so adventurous, bittersweet, and life-changing.

It would not be possible for me to navigate through this journey without the support and love of many wonderful people I met, whom I would like to extend my gratitude to. The first person I am deeply grateful to is my advisor Professor Vishal Gaur. He is a legendary scholar and his dedication to research has inspired me tremendously. Halfway through my doctoral study, I realized how much I was interested in empircal research. With a very humble start, I turned to him to embark on conducting data-driven research and he welcomed me warmly. He believed in me when I had a lot of self-doubt. He has given me so much invaluable advice and guidance on research, and academic and professional opportunities that broadened my horizon. He is a fabulous advisor who is very oepn-minded, patient and encouraging. Because of him, I had the opportunity to visit India with him for a research project. This research experience in the field was a highlight of my doctoral study. I had always wanted to visit India ever since I read the biography of my idol Steve Jobs, in which his trip to India before he founded Apple is depicted. Throughout my job search during Covid, he encouraged me and stood by me. His positive approach to research and life was a light to me in the darkness. Vishal is a life-long learner that I admire and he inspires me to keep learning and improving.

I am deeply indebted to my committee member and mentor Professor William Schmidt. He is an amazing scholar who has sharp eyes at spotting business problems and great intuition for bridging academic research and industry practice. He is the first professor that I started working on a research project with at Cornell, and he taught me so much. He is a wonderful mentor who always empowers me, uplifts me, and encourages me. He genuinely cares about his students, their career goals and mental health, and I am very lucky to be one of them. He is one of the most down-to-earth professors I have ever met. I treasure every minute of our discussion, including the minutes we were having different opinions. I am grateful to him for welcoming my opinions and always making me feel comfortable and valued for being myself and expressing myself. At a personal level, I get to know Bill as a warmhearted, fun, and caring person with a loving family. He invites our doctoral students to his family Thanksgiving meal every year, which is always a touching moment for me as an international student with no family members in the U.S. Bill has had a profound influence on me to become a better researcher, a better educator, and a better person.

I am sincerely thankful to Professor Yao Cui for giving me valuable advice on research and introducing me to conferences that helped me gain experience in communicating research outcomes. I extend my sincere gratitude to my committee member Professor Nagesh Gavirneni. He shared with me his wisdom on how to stay positive and navigate through the ups and downs of research trajectory.

Beyond my committee members, I am grateful to several other Professors. I am thankful to Professor Li Chen for giving me valuable advice on research and his effort in organizing our PhD forum that I benefited from. I am thankful to Professor Lawrence Robinson for all his help to my job market and doctoral studies. He brings great positive energy and always adds valuable input to our seminars. I am thankful to Professor Andrew Davis for all our interactions on research and teaching. He helped me get familiar with the course Operations Management and gave me valuable advice on teaching. I am thankful to Professor Shawn Mankad for giving me great advice on data science and machine learning. He was very generous in sharing with me his valuable insights into teaching machine learning.

This journey is made more colorful and unforgettable by the following people I met at Cornell. I want to thank my partner in crime Alesandro Arcuri for being there for me during highs and lows. I am thankful to the companionship of and friendship with my OTIM cohort Rakesh Allu, Kashish Arora, Junhyun Bae, Yuan Cheng, Xiaobo Ding, Shivank Goel, Rihuan Huang, Dayoung Kim, Wee Kiat Lee, Zhen Lian, Yu Nu, Lorraine Yuan, and Ziwei Zhu. I appreciate the opportunity to learn from them and progress together, the help they offered me, and the good times we spent together on hiking trails and at dinner tables.

I would like to thank my family. I am grateful to my parents for their unconditional love and support. I also thank them for respecting my decisions and letting me choose my path, even if it means me staying 8,000 miles away from them. I want to thank my grandfather. His words have had a profound influence on me and encouraged me to stay positive in adversary and be persistent.

My doctoral time is the most amazing, unforgettable, and crazy time of my life to date. I appreciate this period of encountering so many “unknowns” that my brain was stimulated to ask a lot of “whys”. It was one of the golden periods for me intellectually. Despite of all the excitement and cheerful moments, it has also been a journey with numerous difficulties and challenges. I thank all the people who nurtured my growth with love and support. I thank the people who showed me it is okay to fail and fall. I thank the people who inspired me

to keep going and overcome the fear of the unknown on the path of finding my purpose. Lastly, I would like to thank Ezra Cornell, the co-founder of Cornell University for his vision for Cornell to be “an institution where any person can find instruction in any study.” This motto has inspired me tremendously during my study at Cornell.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PERSONALIZED RECOMMENDATION SYSTEM DESIGN FOR AN ONLINE B2B PLATFORM

1.1 Introduction

The growth of business-to-business (B2B) E-commerce marketplaces has transformed sales and procurement for businesses all over the world. In 2019, the global B2B E-commerce market was valued at US$ 12.2 trillion, more than six times the size of the B2C market [124]. Leading online B2B platforms in different geographies include Alibaba in China, Mercateo in Europe, IndiaMart in India, and eWorldTrade in the U.S. These companies operate platforms that facilitate interaction between buyers and sellers across a variety of products and services. One dominant business model of such platforms works as follows: buyers share their business requirements or requests for quotation (RFQs) with the platform; sellers join the platform to discover RFQs that are relevant to their businesses; and a key task for a B2B platform is to generate personalized recommendations for sellers by accurately predicting which RFQs they are likely to accept. This is an important operational problem because a higher prediction accuracy leads to better quality matching between supply and demand and to a higher engagement of buyers and sellers on the platform, whereas a lower prediction accuracy can create goodwill costs and a loss of engagement from the platform. Thus, by improving the matching of RFQs to sellers, the platform can increase the value generated for participating firms, drive growth in engagement, and improve profitability.

Traditionally, platforms have relied on sellers' self-reported preferences to compute the matching of RFQs. For example, by administering a survey to sellers at the time of registration, a platform can directly measure the attributes of a seller's product and service offerings, such as preferred product categories, locations, order quantities, etc. Then, the platform can determine a new RFQ to be of interest to a seller if its attributes match the seller's stated preferences. However, this method has considerable shortcomings. For example, since it relies on a survey, its performance depends on the quality and comprehensiveness of the data captured and it suffers when the response rate is poor. Moreover, the behavior of sellers evolves over time in the dynamic marketplace as their business needs change, whereas this method does not utilize transaction data, and is unable to adapt to such changes.

The increasing volume of transactional data available to online B2B platforms presents a significant opportunity to learn sellers' latent preferences from their history and apply the resulting model to predict acceptance rates. Then, based on revealed preferences, RFQs that are more likely to be accepted can be presented to sellers. This problem of designing a recommendation system in B2B contexts has not been studied in the literature. In this paper, we present our collaborative research with IndiaMart, a B2B platform in India, to study this problem and develop a method to construct recommendations.

Our problem involves three types of challenges. First, how do we make predictions for new RFQs using historical data for other RFQs? The data in our context are high-dimensional because the B2B platform caters to a vast array of product categories from a wide range of industries, such as machine tools, construction materials, electronics, chemicals, pharmaceuticals, textiles, etc. Moreover, sellers are heterogeneous. Most sellers are active in only a narrow set of product categories as defined by their businesses, which results in a sparse structure of the data. To reduce the dimensionality of the problem, we propose a method that borrows ideas from choice estimation and demand forecasting in retailing and revenue management. Specifically, it has been shown in many contexts that future demand is correlated with past demand. This insight is used in models of Bayesian demand estimation, and is also the basis for choice models calibrated on historical data. Thus, we define new variables, as functions of attributes, that measure the *closeness* of an RFQ to previously transacted RFQs for a seller. This method enables us to construct a model that reduces the dimensionality of the data, adapts to changing seller preferences over time, and is able to generate personalized recommendations for heterogeneous sellers. We consider various characteristics of past transactions, such as the product categories and locations that the seller has transacted in, text descriptions, and RFQ attribute data, and show that this method yields an accurate prediction of the likelihood of acceptance.

Second, what are the implications of class imbalance in historical data and how can it be mitigated? Class imbalance occurs due to over-representation of accepted RFQs (viz., the majority class) and under-representation of declined RFQs (the minority class). In our data set, 97.04% of the records are accepts and only 2.96% are rejects. One reason for under-representation is that sellers may not decline all the RFQs that they chose not to accept, and instead, may just ignore them. Thus, declined RFQs cannot be identified accurately, and are vastly under-represented. A second reason is that the design of the recommendation system induces endogeneity because sellers can only accept RFQs that are displayed to them. Such imbalance can lead to a bias in learning from the past history of a seller's transactions. We compare three methods to address class imbalance: a naïve method based on a subset of the data, a method drawn from the computer science literature, which addresses under-representation due to the first reason, and a new panel data augmentation technique that we propose, which addresses both reasons. We empirically show that class imbalance results in adverse performance of the model in predicting the minority class, and that the proposed methods result in a significant improvement of recall of the minority class in out of sample tests.

Third, what is the value of implementing a data-driven recommendation method in an online B2B platform in practice? We estimate our model using data from IndiaMart including the full transaction history of 1,000 sellers for the time period from January-October 2018, and conduct a controlled field experiment at the company to evaluate performance. The experiment was conducted for more than six months beginning in January 2020 for 21 sellers in the treatment group and 63 sellers in the control group, and 5,686 RFQs were transacted using our model during this period. Results from a difference-in-differences analysis show that performance improved on multiple dimensions: the percentage of RFQs transacted from the top 5 positions on the recommendation page improved by 5.9%, from the top 10 positions improved by 9.0%, and from the top 25 positions improved by 8.2%. Moreover, the average position of accepted RFQs on the recommendation page improved from 37.95 pre-treatment to 16.21, showing that more attractive RFQs were displayed at the top, which led to an improvement in acceptance of RFQs and engagement of seller firms.

Our paper contributes to the literature on online marketplaces with respect to the problem of recommendation system design by applying attribute-based choice modeling to predict the probability of acceptance of new business enquiries. More importantly, we illustrate and solve the problem of class imbalance, which occurs commonly in such platforms and is severely detrimental to performance. Although we present our analysis using the terminology of buyers, sellers, and RFQs from B2B platforms, our method and insights should also be applicable to other types of online platforms that use recommendation systems and face similar challenges. Finally, we present practical evidence of the value of analytics in improving the performance quality of an online marketplace. The paper is organized as follows: we briefly review relevant literature in Section 2, formulate the problem and present the solution methodology in Section 3, describe our data in Section 4, show estimation results and out of sample performance statistics in Section 5, and present the field experiment in Section 6. Section 7 concludes the paper with a discussion of future research opportunities, and in the Appendix in Section 1.8, we present supporting descriptive statistics for our data set.

1.2 Literature Review

There are several streams of literature relevant to our paper: online marketplaces, recommendation systems, discrete choice modeling, and data-driven decision-making in operations management. The literature on online marketplaces is relatively nascent but is growing rapidly and is addressing a number of interesting topics. One set of papers analyzes empirical evidence regarding the existence of endogenous network effects in peer-to-peer online marketplaces. Li and Netessine [114] exploit an exogenous shock to the market size of a platform and show that search friction can reduce the effect of market thickness on matching rates. Cullen and Farronato [60] show that network effects can be limited by the local and time-sensitive nature of the exchanged services. A second set of papers studies marketplace design from different perspectives, such as market clearing in the sharing economy and gig economy. Taylor [160] shows the impact of delay sensitivity and agent independence on an on-demand service platform's optimal price and wage. Cachon et al. [39] show that surge pricing can benefit all stakeholders for platforms with self-scheduling capacity. Bimpikis et al. [29] study pricing for spatially dispersed demand in a ride-sharing network. Allon et al. [2] show the impact of operational efficiency and communication on the market equilibria for large-scale intermediated marketplaces. Dong and Ibrahim [64] study the optimal composition of flexible and fixed workers for on-demand platforms. Banerjee et al. [12] study the optimal search-based market segmentation for online platforms. Kabra et al. [99] use spatial discrete choice modeling to study the effect of bike station accessibility and bike availability on ridership in a bike-sharing system. A third area of work uses field experiments to assess the impact of operational levers on key performance metrics of online platforms. Besbes et al. [26] devise an algorithm to improve repeated interaction of consumers with an online service provider by introducing a representation of content along two dimensions and evaluate performance using a live pilot experiment. Zhang et al. [166] study the effect of price promotions on short-term sales and long-term consumer strategic behavior in Alibaba's B2C marketplace. Our paper contributes to this literature by studying recommendation system design, developing an algorithm to address the problems of high dimensionality and class imbalance in the data set used to match buyers and sellers, and providing evidence of practical impact.

Personalized recommendations are important for product recommendations on Amazon.com, movie recommendations to viewers on Netflix, apparel recommendations to shoppers in online retailing, and a variety of other applications. Different types of methods are used to generate personalized recommendations depending on the context and types of data (see [136], [147], [133] for an overview). Two methods that are most commonly used are collaborative filtering and content-based filtering [147]. Collaborative filtering clusters customers based on similarity of preferences and then utilizes these clusters to provide item recommendations to a customer based on data from past purchases and opinions of other customers in the same cluster ([142], [156]). This method is useful for online product recommendations to customers, but suffers from a cold-start problem, which makes it unsuitable for settings when products do not have previous purchase history with other customers. Content-based filtering uses item features to recommend other items similar to what a user likes, based on his previous explicit or implicit feedback [162]. This method is useful when items can be decomposed into features, such as in recommending movies based on year of make, actors, genre, etc. Our recommendation system shares similarity with content-based filtering by generating features of RFQs that measure the closeness of a target RFQ to the previous RFQs accepted or declined by a seller.

Our paper uses theory from discrete choice modeling to formulate the sellers' choice problem. The literature has extensively studied applications of different types of choice models—multinomial logit or MNL ([105], [127], [163], [113]), non-parametric rank ordering ([118], [83], and [91]), Markov chain-based choice, etc.—to learn how individuals substitute among products and to model their demand. With the growth of online retailing, the recent literature has focused on solving the dynamic assortment planning problem using approaches such as the multi-armed bandit ([1], [18], [141]) and inventory-balancing ([72])

to optimize assortment for a fixed set of products. Choice modeling is also beginning to be applied to personalized assortment and personalized pricing problems in the context of E-commerce, taking advantage of the vast amount of customer transaction and feature data available ([49]). Similarly, attribute-based choice modeling fits our problem extremely well because each product (i.e., RFQ) is unique and the business environment involves forecasting for new products using historical data for other products. However, adapting a model to a new problem context brings challenges. We encounter the problem of class imbalance which is analogous to the problem of unobservable stockouts in retailing. To address this problem in assortment planning, Musalem et al. [127] propose an approach using Gibbs sampling to generate potential sample paths of customers that lead to a given outcome of sales. With a similar motivation, we propose a panel data augmentation technique that utilizes transactions done by other firms during the same time interval to generate proxy data to address class imbalance. We also benchmark this method against alternatives and demonstrate that it performs well.

Methodologically, the literature on high dimensional estimation is also relevant to our paper. Common methods such as LASSO (see e.g., [161], [17], [11]), random projection (see e.g., [97], [125]), and principal component analysis (see [96] for a review) can effectively reduce the dimensionality of covariates when the coefficients are homogeneous. Our problem differs from those considered in this research because the coefficients of variables in our context are heterogeneous across sellers. Therefore, we propose a method that involves constructing new variables to capture the similarity of an RFQ to previously transacted RFQs of a seller and thereby homogenize the problem. We also borrow from the computer science literature to generate new synthetic data using the synthetic minority over-sampling technique (SMOTE, [45]) and propose a new panel data augmentation technique. Recently, Bastani [16] proposes a LASSO-based approach to combine proxy data with gold data to balance the competing objectives of prediction accuracy and bias reduction in recommendation system design. Such an approach may be applied to our problem after synthetic data has been generated.

Lastly, our paper contributes to the literature on practice-based research in operations management. Recently, other researchers have applied predictive and prescriptive analytics approaches to practical contexts [25]. For example, Ferreira et al. [68] develop a nonparametric demand prediction model and an efficient algorithm for subsequent multiproduct price optimization which was implemented at an online retailer for fashion sample sales, Glaeser et al. [71] develop a methodology for the retail location problem using machine learning, and Besbes et al. [26] optimize dynamic recommendations and conduct a live experiment at Outbrain, a leading content recommendation provider. Practice-based research has a long tradition in operations management, including topics such as inventory distribution at Zara [42], assortment planning in supermarket retailing [105], and choice estimation in textbook retailing [113]. We add to this literature by showing that matching in an online B2B marketplace can be improved using a recommendation system based on choice modeling.

1.3 Problem Statement

In this section, we first describe a seller's choice process for RFQs and formulate her decision problem using a binary logit model. We then discuss the source of high-dimensionality and sparsity of data and reformulate the problem to address this issue. Finally, we illustrate the origins of class imbalance in our setting and the potential bias in predictive outcomes resulting from class imbalance.

1.3.1 Research Context

Our research collaborator, IndiaMart (also abbreviated as IM in this paper), provides a platform, www.indiamart.com, for business buyers to discover products and contact their suppliers. IM is India's largest online B2B marketplace; in fiscal 2017, the company held approximately 60% market share of the online B2B E-commerce space in India with a revenue of INR 3.2 billion (1 USD is approximated equal to 70 INR), employing about 5,000 people and serving markets in more than 9,000 cities in India and in other countries. The company had 59.81 million registered buyers and 4.72 million supplier storefronts with 50.13 million products and services across 52 industries.

One method provided by the company for buyers and sellers to interact is through a proprietary recommendation system, which is the focus of our paper. A registered buyer wishing to enquire about a product or service can post a business enquiry or RFQ to the platform including a description of the product or service that he seeks to purchase, his intended use or application of the product, location, the estimated quantity for purchase, and an estimated order value. RFQs arrive continuously over time, and IM serves them to seller firms using an algorithmic matching approach. When a seller visits the platform, she is shown RFQs relevant to her in decreasing order of attractiveness. Sellers can select which of these RFQs to respond to—this is called *accepting* an RFQ—and can also optionally mark RFQs that are not relevant to their business as *not in-*

terested or *declined*. Sellers can also filter the recommendations of the platform and can alternatively use a search tool to discover RFQs using keywords. After an RFQ is accepted, the rest of the interaction between the buyer and seller, e.g., price negotiation, takes place outside the platform and is not tracked by IM. A total of 289.98 million business enquiries were delivered by the IM platform to sellers in fiscal 2018.

IM earns revenue through subscription packages for sellers; each package gives a seller a weekly allowance of credits that are consumed when RFQs are accepted. The more RFQs a seller accepts, the greater is her engagement with the platform, and the more likely she is to maintain or upgrade her subscription. Thus, IM seeks to improve the matching of RFQs to sellers to maximize the number of acceptances and minimize the number of RFQs shown that are not accepted (since those can result in a goodwill loss and decrease in engagement). Buyers post RFQs for free.

IM uses a proprietary method to rank RFQs for each seller based on the attributes of RFQs and the seller's preferences. This method works as follows: (1) A profile is constructed for each seller consisting of her preferred categories and locations elicited from the transactions in the preceding months as well as her declared preferred product categories and locations. (2) IM determines which attribute is more salient, and thus, converts seller profiles into a preference ordering. For instance, consider all paired combinations of primary/non-primary product category with primary/non-primary location, where primary and non-primary are determined from a seller's profile. If product category is the more salient attribute than location, then IndiaMart would rank an RFQ with primary product category, non-primary location as preferred to an RFQ with non-primary product category, primary location. (3) As RFQs arrive, they are given rankings using the preference ordering obtained in (2). (4) When a seller logs into the system, RFQs are sorted real-time by their rank and by their date and time stamp within each preference group, and then are displayed to the seller.

This business problem presents several opportunities for improvement. First, IM's ranking approach works by discretizing a small number of attributes, but will quickly become combinatorially complex as the number of attributes and their range of values grow. When products have rich attribute descriptions, it could be beneficial to use those attributes in a parametric choice model to estimate probabilities of acceptance. Second, the rankings approach yields an incomplete ordering on attributes. For example, if a seller accepts 10 RFQs from one category and 100 from another, both categories may be classified into the 'primary' group. Since there is sufficient data available, we expect that a parametric model with continuous variables will provide more granular estimates of acceptance probabilities that allow ranking of RFQs not only across but also within each discrete category. Finally, the existing method is based on data for accepted RFQs only. We expect that a data-driven approach that uses both accepted and rejected RFQs should yield more accurate predictions of the probability of acceptance of an RFQ.

The main challenge is that estimating a parametric model requires data for both accepted and declined RFQs, whereas IM's non-parametric preference ordering approach can be implemented using only data for accepted RFQs. Thus, the goal of our paper is to address the data limitations posed by this problem and evaluate whether the parametric model thus obtained doing so can provide an operational benefit compared to the existing method.

1.3.2 Sellers' Choice Process and Binary Logit Classifier

We formalize our model in this section. The B2B platform we consider is a two-sided marketplace consisting of buyer and seller firms who interact through the recommendation system of the platform. Buyer firms arrive randomly over time by posting new RFQs to the platform; we use the terms RFQ and *item* interchangeably throughout this paper. The population of buyers is practically infinite. Each RFQ has several attributes, such as product category, geographical location, product specifications, quantity, order value, and a time stamp. Moreover, each RFQ is on-demand and is available for only a short timespan of a few days because buyers have time-sensitive requirements. For example, if a buyer posted an RFQ a month ago, the buyer may very likely not have the same demand today (either the buyer's demand has been fulfilled or changed). There is a fixed population of seller firms who pay a periodic membership fee to the platform. Each seller arrives regularly on the platform in a repeated interaction, typically logging in several times a week. Each time when a seller logs in, the recommendation system selects RFQs from the available pool to display to the seller based on a current estimate of his likelihood of accepting them. Alternatively, the seller may search for RFQs using keywords. She *accepts* RFQs that are relevant to her business, otherwise she skips the RFQ or optionally marks it as *not interested* or *reject*.

The platform's objective is to maximize the value to sellers, which is the revenue side of the platform, by recommending RFQs that they are likely to accept and not recommending RFQs that they are likely to reject. An RFQ can be shown to several sellers, who independently decide whether to accept it. Moreover, a seller can accept any number of RFQs from the recommendations shown to her.

Thus, we make an important assumption that the seller evaluates each RFQ independently of the others. This assumption implies that there is no substitution between RFQs. Thus, we construct recommendations by predicting each seller's probability of accepting an RFQ based on its attributes and the seller's preferences, and then presenting those RFQs in decreasing order of acceptance probability. This assumption differentiates our problem from the traditional assortment optimization problem, in which a consumer selects at most one choice from the offered set.

In general, there can be two ways to estimate demand—using previous purchase history of the same product for long-lifecycle products or using purchase history of similar products for short-lifecycle products. Some recommendation systems, such as those in Amazon, Inc. or Netflix, Inc. deal with long lifecycle products where data on previous consumption and ratings of a product by other customers can be used to predict its attractiveness to a given customer. In contrast, an RFQ in our setting does not have sufficient previous purchase history that could be used to model its attractiveness. We address this problem by decomposing RFQs into their attributes and estimating sellers' preferences using an attribute-based utility model.

We adopt a binary logit model with heterogeneous coefficients as the basis of a classifier for each seller-RFQ pair into *accept* and *reject* groups. Binary logit is widely used as a discrete choice model to estimate and predict the choice probabilities of binary alternatives (see Luce 117). It is consistent with the classic assumption of utility maximizing decision makers. It can be utilized in our context instead of the multinomial logit model because each choice decision is independent of other alternatives. We use the following indexes throughout this paper: r denotes the index of an RFQ, s a seller, t a time epoch, c a product category, and g a geographical region or city. Let $\mathbf{x}_{rt}$ denote the attributes of RFQ r at time t. We make the time index explicit because the utility of an RFQ to a seller depends on its freshness, and moreover, we will use historical data before time t to train the model for prediction at time t. We subdivide the RFQ attributes into three components: $\mathbf{x}_{1rt} = [x_{1rtc}]'$ is a binary vector identifying the product category of the RFQ, $\mathbf{x}_{2rt} = [x_{2rtg}]'$ is a binary vector identifying the city or region that the RFQ originates from, and $\mathbf{x}_{3rt}$ consists of the remaining attributes such as product descriptors, quantity desired, and time elapsed since the arrival of the RFQ. We list these attributes in Section 1.4.2.

Let U_{srt} denote the utility obtained by seller s from RFQ r at time t. We assume that sellers accept RFQs that yield a positive utility and decline them otherwise. Let A_{st} denote the set of RFQs accepted and D_{st} denote the set of RFQs rejected by seller s up to time t; thus, $r \in A_{st}$ if $U_{srt} > 0$ and $r \in D_{st}$ otherwise. The seller's utility is latent, and is influenced by the attributes associated with the seller and the RFQ as follows:

$$U_{srt} = \beta'_{1s}\mathbf{x}_{1rt} + \beta'_{2s}\mathbf{x}_{2rt} + \beta'_{3}\mathbf{x}_{3rt} + \xi_{srt}, \tag{1.1}$$

where $\beta_{1s} = [\beta_{1sc}]'$ and $\beta_{2s} = [\beta_{2sg}]'$ denote the seller's preference vectors for product categories and geographies, β_3 are homogenous preferences for other attributes, $\mathbf{x}_{rt}$ provides the observable component and ξ_{srt} is the unobservable component. Let $\beta'_s \equiv (\beta_{1s}, \beta_{2s}, \beta_3)$ for brevity.

We assume here that sellers are homogeneous in their preferences for RFQ attributes in $\mathbf{x}_{3rt}$, i.e., all sellers prefer RFQs with larger quantities demanded, more recent time stamp, and richer information content about product descriptors. We explain the rationale for this assumption in Section 3.3. In practice, this assumption can be relaxed through various methods such as clustering the sellers, normalizing data, or estimating heterogeneous coefficients across the history of transactions of each seller. We also assume for ease of presentation that there is no interaction between the covariates. In other words, the seller's preferences for product categories, geographic regions, and remaining attributes are independent of each other, i.e., if a seller sells products in category P_1 in geography G_1 and category P_2 in geography G_2, then she is also likely to sell products in category P_2 in geography G_1 and category P_1 in geography G_2. Thus, the utility function is additive in the three components of $\mathbf{x}_{rt}$. Our model can be extended to allow interactions across these components.

We make the standard assumption that the unobserved component ξ_{srt} follows a logistic(0, 1) distribution. It follows that the choice probability is given by:

$$Pr(r \in A_{st}) = \frac{\exp(\beta_s' \mathbf{x}_{rt})}{1 + \exp(\beta_s' \mathbf{x}_{rt})}, \tag{1.2}$$

$$Pr(r \in D_{st}) = \frac{1}{1 + \exp(\beta_s' \mathbf{x}_{rt})}, \tag{1.3}$$

and the log likelihood function is given by:

$$\log L(\beta_s) = \sum_{r,t:r \in A_{st}} \beta_s' \mathbf{x}_{rt} - \sum_{r,t} \log(1 + \exp(\beta_s' \mathbf{x}_{rt})). \tag{1.4}$$

The log-likelihood function of a logit model of linear-in-parameters latent utility is globally concave in the parameters β_s [123]. Thus, we can obtain the log-likelihood estimators $\hat{\beta}_s$ of β_s given a training sample. The corresponding predicted probability of acceptance under $\hat{\beta}_s$ is $\hat{p} = Pr(r \in A_{st} | \beta_s = \hat{\beta}_s) = \frac{\exp(\hat{\beta}_s' \mathbf{x}_{rt})}{1+\exp(\hat{\beta}_s' \mathbf{x}_{rt})}$. We assign an accept outcome if $\hat{p} > 0.5$ and a reject outcome if $\hat{p} < 0.5$. The 0.5 threshold corresponds to the seller's latent utility specification of $\hat{U}_{srt} > 0$ for accept outcomes, and is widely used in predicting outcomes from a binary classifier (see Jiménez-Valverde and Lobo 93).

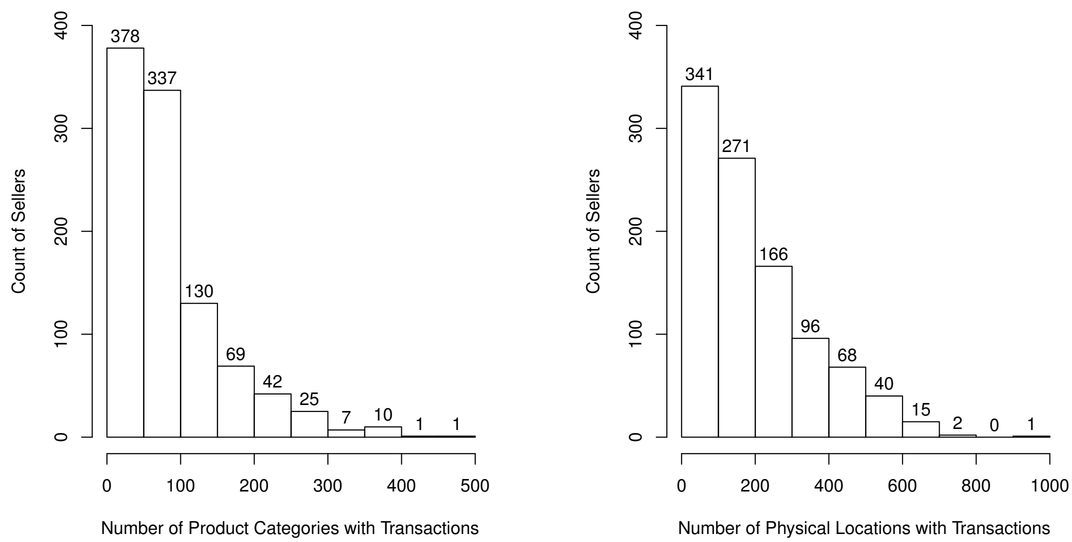
1.3.3 High Dimensionality and Sparsity

In online B2B platforms, the transactional data of RFQs is high-dimensional in product category, spatial engagement, and sellers. For example, since a B2B platform typically offers a vast number of product categories (e.g., Amazon Business, the B2B platform of Amazon.com Inc. offers millions of products), the dimensionality of transactional data regarding product engagement will inevitably be vast. Similarly, since an online B2B platform connects sellers and buyers from a wide range of geographical locations within a country or even globally, the vector of spatial attributes of transactional data is high-dimensional. As noted above, our research collaborator serves more than 4.72 million sellers and 50.13 million products across several thousand cities and product categories.

Besides being high dimensional, transactional data is also sparse in both product category and spatial attributes because a seller is actively engaged in only a small number of product categories and locations defined by the seller's businesses (i.e., a large number of columns of the high-dimensional binary vector are filled with 0). Sellers are also heterogeneous in their engagement with product categories and locations. For example, some sellers only sell chemical compounds whereas others sell only cotton fabrics. In Figure 1.1(a), we show the distribution of sellers against the number of product categories they have engaged with in our data set. We observe that 71.5% of the sellers have engaged with fewer than 100 product categories from a total of 13,395 product categories in our transactional history data. Similarly, Figure 1.1(b) shows the distribution of sellers against the number of cities that sellers have engaged in. We observe that 61.2% of the sellers have transacted in fewer than 200 buyer cities from a total of 8,712 buyer cities.

Figure 1.1: Distribution of seller firms with respect to number of product categories and geographical locations

(a) Product categories (b) Geographical locations



Thus, the model specified by (1.1) will have a vast number of parameters to be estimated across sellers, product categories, and locations, which can be computationally onerous as well as error prone. It also leads to other problems: in particular, the model can only be used to make predictions for those combinations of sellers, product categories, and locations on which the model was estimated, and it cannot be applied to new columns.

These problems can be addressed by constructing new predictors to replace the terms $\beta'_{1s}\mathbf{x}_{1rt}$ and $\beta'_{2s}\mathbf{x}_{2rt}$ in (1.1). Let n^a_{sct} and n^d_{sct} denote, respectively, the number of RFQs in product category c accepted and declined by seller s up to time t. The values of these variables are computed as a function of $\mathbf{x}_{1rt}$ as follows:

$$n^a_{sct} = \sum_{r \in A_{st}, \tau < t} x_{1r\tau c}, \quad \text{and } n^d_{sct} = \sum_{r \in D_{st}, \tau < t} x_{1r\tau c}.$$

Note that the total number of transactions in the marketplace is very large but the vast number of n^a_{sct} and n^d_{sct} are zero.

The rationale for this approach is that if product category were the only attribute of each RFQ, i.e., if $\mathbf{x}_{2rt}$ and $\mathbf{x}_{3rt}$ were zero for all RFQs, then the maximum likelihood estimator of β_{1sc} for seller s and category c with non-zero accept and reject transactions would be given by $\log(n^a_{sct}/n^d_{sct})$. In that case, n^a_{sct}/n^d_{sct} is a sufficient statistic for the high dimensional data. Intuitively, this implies that if a seller has accepted RFQs from a product category frequently in the past, the seller is more likely to accept this product category in the future, and similarly for declines. With this motivation, we assume that a seller's preference for a product category is directly proportional to the number of RFQs from that category accepted by the seller in the past and inversely proportional to the number of RFQs declined. Thus, we define the *closeness* of an RFQ to previous RFQs accepted or rejected by a seller in terms of the corresponding values of n^a_{sct} and n^d_{sct} for that RFQ's product category. Thus, we replace $\beta'_{1s}\mathbf{x}_{1rt}$ by the linear term $\beta^a_1 \log(n^a_{sct} + 1) + \beta^d_1 \log(n^d_{sct} + 1)$ in the utility function, where 1 is added to avoid zero count values from getting omitted and β^a_1 and β^d_1 are identical across sellers. This modified predictor enables us to reduce the dimensionality of the data and allows a seller's future choices to depend on her historical transactions.

Following a similar approach for spatial engagement, we assume that a seller's preference for a location is directly proportional to the number of RFQs from that location previously accepted by the seller and inversely proportional to the number of RFQs declined. Let m^a_{sgt} and m^d_{sgt} denote the number of RFQs in geography g accepted and declined, respectively, by seller s up to time t. The values of these variables are computed analogous to n^a_{sct} and n^d_{sct}, and we replace

$\beta'_{2s}\mathbf{x}_{2rt}$ by the term $\beta_2^a \log(m_{sgt}^a + 1) + \beta_2^d \log(m_{sgt}^d + 1)$ in the utility function. Thus, we measure the closeness of an RFQ to previous RFQs transacted by a seller using historical frequencies with respect to product category and location.

We also extend this method to include text-based measures of closeness. Our data present multiple attribute-types. In particular, user-specified text data are a salient proportion of an RFQ. Compared to data in quantitative format, text data are unstructured and require significant preprocessing. We apply text mining and extract useful text features that may influence a seller's consumption of an RFQ. Based on the text features, we create a variable that measures the approximate matching between the text content of historical RFQs and new target RFQs.

Altogether, these assumptions enable us to eliminate the high-dimensionality and sparsity of the estimation problem. They have two additional benefits: (1) *Since the explanatory variables for product category and location are measures of similarity of RFQs to historical transactions of each seller, the model dynamically evolves with the changing behavior of seller companies.* For example, if a seller begins transacting in a new product category, the model prediction for that category adjusts without requiring a re-estimation of the model. (2) *Since all coefficients in the reformulated model are homogenous, they can be estimated on a subset of the data and then applied to all the sellers, locations, and product categories on the IM platform.* Thus, we use a subset of the data for model estimation and testing with the goal that the resulting model should be scalable to the entire firm.

Thus, the modified log likelihood function becomes:

$$\log L(\beta) = \sum_{r,t:r\in A_{st}} \beta'\mathbf{z}_{rt} - \sum_{r,t} \log(1 + \exp(\beta'\mathbf{z}_{rt})), \tag{1.5}$$

where $\beta' = [\beta_1^a, \beta_1^d, \beta_2^a, \beta_2^d, \beta_3']$ and $\mathbf{z}_{rt}' = [\log(n_{sct}^a+1), \log(n_{sct}^d+1), \log(m_{sgt}^a+1), \log(m_{sgt}^d+1), \mathbf{x}_{3rt}']$.

1.3.4 Class Imbalance

Class imbalance occurs when one class, the minority group, contains significantly fewer examples than the other class, the majority group [95]. We find that the data from the online marketplace is skewed in the class distribution such that the class of accepted RFQs is disproportionately more represented than the reject class. We have two sources of class imbalance. First, there is a genuine lack of records of declined RFQs marked by sellers due to the low frequency with which sellers opt to explicitly express their negative feedback to the platform. Although the platform provides a 'not interested' option to sellers to better learn their preferences, it is more convenient for sellers to just skip past an RFQ that they are not interested in. Such RFQs, that are left unmarked, are not included in the archival data set because it is not possible to assess accurately whether a given RFQ was actually seen by a seller—the seller may have just scrolled past it, clicked on another page, had an internet outage, or may have gotten distracted by other tasks. Hence, the marketplace data consists only of those RFQs that were marked accepted or rejected, resulting in undersampling of the minority class. Second, there is model endogeneity because RFQs seen by sellers are either those that were selected by the recommendation system of the platform or those that were discovered by the seller through a search. In either case, these RFQs are not randomly drawn. Both these types of class imbalance, due to undersampling and endogeneity, can result in biased estimates. As a result, a choice model may fit the training data well, but not perform well in field tests. The extent of imbalance in a data set can be measured by its *accept ratio*, defined as the ratio of the number of accept observations to the total number of observations in the data set. Our model calibration data set has an accept ratio of 97.04% as it contains 869,896 accept records and 26,446 reject records.

The problem of an undesirable misclassification of the minority class can be addressed by resampling. In the following, we propose three alternative resampling approaches with different advantages and drawbacks. In each case, to ease the presentation of the method, we denote the sample before resampling as S_0 and after resampling as S_1.

1.3.4.1 Curated Sampling.

This is a naïve and intuitive resampling technique in which we construct a curated training sample S_1 by retaining the records of only a subset of representative sellers whose accept ratio is among the lowest in the initially skewed sample S_0. In particular, our data set contains 1,000 sellers from which we retain records for only the bottom 20 sellers by accept ratio in the curated training set. Although this approach dramatically reduces the size of the training data set, it also mitigates class imbalance by reducing the accept ratio. We find in the test results that curated sampling results in a materially significant improvement in model performance compared to the original data set S_0. This is all the more remarkable since our curated data set contains less than 3% of the observations in the original data set. We also note that this approach can address class imbalance caused by undersampling, but not that caused by model endogeneity.

1.3.4.2 SMOTE.

Synthetic Minority Over-sampling Technique (SMOTE) is a very successful approach created by [45] in the computer science literature to over-sample the minority class by creating synthetic data points. It is commonly applied to situations such as medical diagnosis and image recognition, where the minority class is small in size but its accurate estimation is important. To describe this method, we use Z to denote the set of feature vectors of data points in our minority class, the reject class, and let $z_i \in Z$ be a feature vector containing continuous features z_i^c and categorical features z_i^n where $z_i = [z_i^c, z_i^n]$. Following [45], we describe how to use SMOTE to generate a new synthetic data point $z_{syn,i}$ based on the target feature vector z_i in the pseudo-code below:

Step 1: Calculate the Euclidean distance d_{ij} between the target feature vector z_i and other feature vectors $z_j \in Z$ for all $j \neq i$.

Step 2: Identify the k-nearest neighbors of z_i based on the Euclidean distances.

Step 3: Choose the continuous features of a synthetic data point to be $z_{syn,i}^c = z_i^c + \delta(z_j^c - z_i^c)$, where δ is a random number between 0 and 1, and z_j is a feature vector randomly chosen from the set of k-nearest neighbors.

Step 4: Choose the categorical features of the synthetic data point to be the mode of all the z_j^n in the set of k-nearest neighbors.

Thus, SMOTE is an over-sampling method based on random interpolation using the set of k-nearest neighbors. There are two adjustable parameters in SMOTE: the number of nearest neighbors k, and the size of the synthetic sample $|S_1| - |S_0|$. Unlike curated sampling, SMOTE allows us to achieve any targeted accept ratio. As $|S_1| - |S_0|$ increases, the accept ratio of the SMOTEd sample S_1 decreases.

Similar to curated sampling, SMOTE addresses class imbalance by relying on the existing minority group observations. Thus, the performance of both these methods would depend on the number and distribution of minority group observations in the data set. If this number is small or is unevenly distributed across sellers, the performance of these methods may suffer. To address these potential shortcomings, we propose a third method as below.

1.3.4.3 PDATE.

We propose a Panel Data Augmentation Technique (PDATE) as a different kind of resampling method that works on the feature space of the majority group in the original data set in contrast to SMOTE which works on the feature space of the minority group. PDATE is more suited to operational settings where a panel data is usually available than to the types of problems where SMOTE is applied. In PDATE, we assume that when an accept transaction occurs in the data set at time t, the seller has potentially also considered other RFQs that are available at that time and chosen to not accept them. Further, the platform has evaluated other RFQs in deciding which RFQs to show to the seller. Thus, we proxy the other available RFQs at that time using the RFQ records transacted by other sellers within a proximate time window around time t. We infer that the seller has no intent to purchase those records, and thus, in PDATE we randomly restore them as synthetic reject records of the seller. Thus, to implement PDATE, we begin with a target RFQ accepted by a target seller and randomly sample from RFQs that were transacted by other sellers but not the target seller and that arrived in the same time window as the target RFQ. We then augment the target seller's records with these sampled RFQs marked as 'synthetic' rejects.

PDATE has an advantage compared to SMOTE or curated sampling that it can be utilized when the number of declined RFQs in the original data set is very small or is unevenly distributed. Moreover, it can address class imbalance due to endogeneity because it unconditions on the decisions of the recommendation system of the platform. However, PDATE could potentially introduce an estimation error if the synthetic declines were candidates that could have been accepted by the target seller. We show in Section 1.5 that all three methods, PDATE, SMOTE, and curated sampling, dramatically improve out-of-sample performance compared to the original data set, but in slightly different ways.

There are two adjustable parameters when using PDATE: the length of the time window δt, and the size of the synthetic sample $|S_1| - |S_0|$. Like SMOTE, PDATE is able to achieve any targeted accept ratio. The pseudo-code of PDATE is as follows:

Step 1: Randomly sample $|S_1| - |S_0|$ records of accepted RFQs and denote them as set A.

Step 2: For record i in set A of seller s_i and transaction date t_i, identify all the RFQ transaction records within the proximate time window, $[t_i - \delta t, t_i]$, that are not transacted by seller s_i, which we denote as set M_i.

Step 3: Randomly choose a record x from M_i, set the seller of record x to s_i, and set the response of record x as a reject.

Step 4: Repeat step 2 and 3 until we have created a synthetic reject record for all records in set A.

1.4 Data

We obtained the full historical transaction data from IM for 1,000 sellers served by the company for the time period between Jan 2018 and Oct 2018 for this research. Subsequently, the model was deployed live in a controlled field experiment from January 2020 onwards. In this section, we present data description and model results using the Jan-Oct 2018 data set, which was used to train our model and conduct an out of sample evaluation.

The company selected a sample of 1,000 sellers for our analysis by first identifying the top 10 product categories by volume of consumption in the last three months when they fetched the data (late Oct 2018). These top 10 product categories included a wide range of industries, such as apparel, pharmaceuticals, machinery for industrial plants, electronics products such as CCTV cameras and biometric systems, and electrical products such as generators and compressors. Then the company picked the top 100 sellers from each product category by consumption volume, and obtained the full transaction history of these selected sellers. The resulting data set encompasses 896,342 records, 549,105 unique RFQs, 479,645 buyers, and 13,395 product categories. Each data record is a time-stamped accepted or declined RFQ transaction for a seller.

1.4.1 Data Preparation

We list the data files that we received from IM's data analytics team and explain our data preparation efforts. The data consists of the following files: (1) *Supplier accept/reject history joined with offer details.* This file provides all the accepted and rejected RFQ records for the sample period, including the identities of the seller and buyer firms, the time stamp when the RFQ is offered, and the time stamps when the RFQ is transacted. (2) *Features of RFQs* This data contains all the features of all the RFQs offered, including title, text description, product category, and specifications that vary from one product type to another, such as quantity, weight, volume, size, etc. For example, the buyer can specify the purchase quantity of a construction crane to be one unit, the usage to be for business use, etc., whereas specifications for a pharmaceutical drug would contain different fields. (3) *User city mapping file.* The data provides home cities of buyers and sellers in our data set. The 1,000 sellers given to us encompass 109 unique cities. Of the 479,645 buyers that are present in the transaction history, city information is available for 447,088 buyers in 8,712 unique cities. The remaining 32,557 buyers are missing city information for reasons including new buyers, buyers who opt not to share location information, etc. (4) *Indian cities latitude-longitude mapping file.* This data file provides the latitude and longitude information for cities in our data set. It provided us with data to map the locations of all 109 cities of the 1,000 sellers in the transaction history, and for 425,214 buyers out of the 447,088 buyers with city information. Thus, we use 772,205 records out of 896,342 in our analysis and remove 124,137 records due to missing buyer city or lat-long data.

1.4.2 Predictors Formulation

We next formulate the predictors to enter our binary logit model based on the given data. As in Section 3, we let s denote a seller, b a buyer, r an RFQ (i.e., order), c a product category, g a geographical location, and t a timestamp. First, following the discussion in Section 1.3.3, for each RFQ, we formulate the following predictors that are functions of a seller's transaction history in the 3 months preceding the time stamp t of that RFQ:

- $product\ category\ accept\ (or\ resp.\ reject)\ frequency_{srt}$ represents the total volume of RFQs accepted (or resp. rejected) by seller s from the product category associated with RFQ r.
- $city\ accept\ (or\ resp.\ reject)\ frequency_{srt}$ represents the total volume of RFQs accepted (or resp. rejected) by seller s from the buyer city associated with RFQ r.
- $seller\ volume_{st}$ represents the total volume of RFQs accepted by seller s.
- $product\ category\ volume_{rt}$ represents the total volume of RFQs accepted in the product category of RFQ r.
- $city\ volume_{rt}$ represents the total volume of RFQs accepted in the buyer city of RFQ r.
- $minimum\ buyer\ group\ distance_{srt}$ represents the minimum distance (in kilometers) between the buyer city of RFQ r and the top 5 cities of seller s by the seller's volume of accepted RFQs.
- $title\ match\ score_{srt}$ represents the approximate match between the title of RFQ r and the set of titles of RFQs in the seller's acceptance history. It is

computed as the maximum of the Levenshtein ratio of the title of RFQ r with respect to this set.

Here, the predictors *seller volume*$_{st}$, *product category volume*$_{rt}$, and *city volume*$_{rt}$ are control variables. We use the remaining predictors to measure the closeness of an RFQ to the RFQs transacted by the seller in the recent past in terms of their product categories, geographical locations, and text matching. Note that it is possible to add other variables to this list since closeness of an RFQ to previous transactions can be measured along several dimensions. An important characteristic of this approach is that the predictors can capture a sellers' time-varying tastes for product categories and locations. When a seller begins transacting in new product categories or locations, the model learns the changing preferences and adjusts the predicted probability dynamically.

Second, we formulate predictors in $\mathbf{x}_{3rt}$ that capture important attributes of RFQs as follows. These attributes are visible to sellers when they decide whether to accept or reject the RFQ.

- $specification\ count_r$ represents the number of specifications provided by the buyer for RFQ r. Specifications vary by product and include product description, technical requirements, color, size, etc. A buyer may choose not to provide any specifications. We hypothesize that an RFQ with a larger number of specifications is more informative, and hence, is more valuable to a seller.
- *quantity*$_r$ is a binary variable that indicates whether the buyer specifies an order quantity requirement.
- *order value*$_r$ is a binary variable that indicates whether the buyer provides

an estimated monetary value of the desired RFQ.

- $recency_{rt}$ is the difference between the time when RFQ r was posted and the time when it is accepted (or rejected). We hypothesize that the less recent the RFQ, i.e., the larger the value of $recency_{rt}$, the lesser is the value of the RFQ to the seller.

We find that the abovementioned predictors ending with "frequency" (or "volume" or "count") and variable $recency_{rt}$ present a log-normal distribution, and hence we apply a log transformation on them with the form $f(x) = \log(x+1)$ for our logit model.

1.4.3 Sample Preparation

In this subsection, we describe the preparation of training and test samples from the original data set that we obtained from IM. Training samples are prepared for each of the three resampling methods to tackle class imbalance described in Section 1.3.4. We denote our original data set as S, which contains 896,342 historical transaction records (including those with missing location data). We split S into S_1, containing all the records for the first three months (before April 1, 2018), and S_2, containing all the records thereafter. Thus, S_2 contains all the records in S that are eligible for calculating the time-varying predictors based on a 180-day history. Since our original dataset S is highly imbalanced, S_2 is also highly skewed in class representation with 96.99% of the records as accepts (647,710) and 3.01% as rejects (20,088). Table 1.1 shows a summary of all the data sets that we created in sample preparation; we describe the construction of these data sets in the following paragraph.

Table 1.1: Overview of Samples

	Total	Accepts	Rejects	Accept Ratio	Reject Ratio
S: Original data set S	896342	869,896	26,446	97.04%	2.96%
S_1: First 3 months' data set	228540	222,182	6,358	97.22 %	2.78%
S_2: Rest 7 months' data set	667802	647,714	20,088	96.99%	3.01%
S_2^e: After creating predictors and removing missing observations in S_2	504,724	496,657	8,067	98.40%	1.60%
Train_s: 75% of observations in S_2^e	378,543	372,493	6,050	98.40%	1.60%
Test_s: 25% of observations in S_2^e	126,181	124,164	2,017	98.40%	1.60%
Train_{curate}: Curated sample constructed from Train_s	9,702	6,190	3,512	63.80%	36.20%
Train_{smote}: SMOTEd training sample constructed from Train_s for accept ratio 0.5	744,986	372,493	372,493	50.0%	50.0%
Train_{pdate}: PDATEd training sample constructed for accept ratio 0.5	744,986	372,493	372,493	50.0%	50.0%

We first form a panel S_2^e by generating all the predictors as specified in Section 1.4.2 for S_2 and deleting records with missing information. Information can be missing when a buyer geographical location is not available, or in a small number of cases, when product category is not identified, specifications are not available, or time stamp is not available or misrecorded. Sample S_2^e has 504,724 records and is imbalanced with 98.40% accepts (496,657 records) and 1.60% rejects (8,067 records). We randomly split S_2^e into a training set Train_s containing 75% (378,543) of the records and a testing set Test_s containing the remaining 25% (126,181) of the records. The accept-reject ratio in the training set and the testing set are constrained in the random split to be the same as in S_2^e.

To tackle class imbalance, we augment the minority class using different resampling approaches. We construct a curated training sample Train_{curate} from Train_s by keeping all the records from 20 representative sellers who have the highest reject ratios in Train_s, and deleting all other records. The curated training sample obtained thus contains 9,702 records in which 63.80% are accepts and 36.20% are rejects. Note that the number of records in this sample is only a small fraction of the original data set, but class imbalance is considerably mitigated.

We next apply SMOTE on Train_s and the resulting oversampled training set is denoted as Train_{smote}. To implement SMOTE, we first define an accept ratio, i.e., the ratio of the number of records in the majority class to the number of records in the training set after resampling, to set the number of synthetic rejects to create. We choose values of accept ratio from 0.5 to 0.9 in steps of 0.1. The default ratio is set to be 0.5 such that the minority class will have the same number of records as the majority class after oversampling. We also set the number of nearest neighbors in applying SMOTE to $k = 5$, which is a widely used default value. Since Train_s has 372,493 accepts and 6,050 rejects, we create 366,443 synthetic rejects to form Train_{smote} when the accept ratio is 0.5. Finally, we distinguish between categorical and continuous predictors in the original training set—as described in section 1.3.4.2, categorical predictors will be augmented following a different procedure than continuous predictors.

We apply PDATE for the same values of accept ratio as SMOTE. Here, we describe its implementation for an accept ratio of 0.5. To begin with, note that unlike SMOTE, data augmentation in PDATE is conducted directly on S_2, not on Train_s, using the procedure described in section 1.3.4.3. Our parameter choice for the length of proximate time window in applying PDATE is set to $\delta t = 1$ day. The predictors for the synthetic rejects are formed based on the original data set S. In the last step, we form the training set Train_{pdate} by merging Train_s with the synthetic records. Since PDATE randomly infers synthetic rejects from a large pool of unmarked (neither accepted nor rejected) RFQs, the reject history regarding those pairs of sellers and product categories and pairs of sellers and physical locations is scant. To smooth the values of the covariates *product reject frequency*$_{srt}$ and *city reject frequency*$_{srt}$ in sample Train_{pdate}, we add value 1 for these two covariates to all the observations in Train_{pdate}. (This approach is analogous to the Add-one (Laplace) smoothing method in natural language processing models.)

The test panel Test_s is kept identical across all methods. We present descriptive statistics of the Train_s and Test_s panels for the majority and minority classes in the Appendix in Section 1.8.

1.5 Empirical Results

In this section, we present estimation results from our model specification and compare the performance of different resampling methods.

1.5.1 Estimates of Coefficients

Table 1.2 presents the coefficients' estimates from our model on the original training sample Train_s, the curated training sample Train_{curate}, the SMOTEd training sample Train_{smote}, and the PDATEd training sample Train_{pdate} in columns (1), (2), (3), and (4), respectively. We discuss these results in this section. First, note that the variables measuring the closeness of a target RFQ to the previous RFQs accepted or rejected by a seller are statistically significant. If a seller has accepted more RFQs from a product category in the past, then she is significantly more likely to accept a new RFQ from that product category again. The opposite is true for categories from which a seller has rejected RFQs in the past. Similarly, we see that accept and reject frequencies for different cities, the distance of an RFQ from the top five locations that a seller has transacted in, and the approximate match of the title of the RFQ to previously accepted RFQs of the seller are also statistically significant. This shows that the closeness of an

RFQ to the sets of accepted and rejected RFQs by a seller is a good predictor of the likelihood of the seller accepting that RFQ. Moreover, this result is consistent across models, even though the magnitudes of the coefficients vary. This validates an important component of our model.

Second, the coefficients of accept and reject frequencies in each case have similar orders of magnitude and opposite signs. This implies that the difference between accept and reject frequencies is a key determinant of the likelihood of acceptance of an RFQ. To see this, consider the coefficients of product category accept frequency and reject frequency from column (1), with values of 0.797 and -1.349, respectively. These values translate into odds ratios of 2.219 (i.e., exp(0.797)) and 3.854 (i.e., exp(1.349)), respectively. Now suppose there is a product category in which the seller has zero reject frequency. Then, we expect to see a rapid 121.9% increase in odds of acceptance of an RFQ in this category for a one-unit increase in $\log(product\ category\ accept\ frequency_{srt} + 1)$. However, suppose there is another product category in which the seller has equal values of accept and reject frequencies. From the differences in the values of odds ratio, we now expect to see a significant decline in the probability of acceptance from this category if both frequencies change in the same proportion. Thus, our model utilized information from rejected RFQs to improve the prediction accuracy of the probability of acceptance.

Third, the variables measuring the informativeness of a target RFQ show significant predicting power. The model fit results show that RFQs with a larger number of product specifications are significantly more likely to be accepted. RFQs that provide an order value are also more likely to be accepted than those that don't (odds ratio = 1.229 in column (1)). However, RFQs with quantity specification are significantly less likely to be accepted, which is counter to expectations. Upon further investigation, we find that the quantity variable interacts with product category; for some product categories, e.g., ‘river water treatment plant’, and for services, e.g., ‘franchise consulting services’, ‘third party manufacturing services’, quantity is not relevant. Moreover, sellers vary in their preferences for quantity, i.e., large versus small orders, depending on the scale of their operations. Sometimes quantity information may even be included in the text description of the RFQ rather than as a separate field. These problems can be addressed by interacting quantity with category fixed effect or seller fixed effect, or constructing new variables through text mining. Finally, our model fitting results show that sellers are significantly more likely to accept more recent RFQs than older RFQs.

We note that although the statistical significance of variables is largely consistent across models, the magnitudes of coefficients vary. This can affect out of sample predictive performance, which we discuss next.

1.5.2 Predictive Performance

Imbalanced data requires more sophisticated measures to evaluate model performance. The performance of a binary classification is typically assessed by the Confusion Matrix as shown in Table 1.3. The rows represent the observed case of negative response and positive response, and the columns represent the predicted case of negative response and postive response. In our setting, a positive response to an RFQ corresponds to an accept and a negative response corresponds to a reject. The values in the cells are the number of observations falling into each of the four categories shown. The predictive accuracy is defined as

Table 1.2: Coefficients' Estimates for the Original, Curated, SMOTEd, PDATEd Training Samples

	(1) Original	(2) Curated	(3) SMOTEd	(4) PDATEd
product category accept frequency	0.797***	0.666***	0.823***	1.050***
	(0.018)	(0.034)	(0.048)	(0.007)
product category reject frequency	−1.349***	−0.723***	−2.021***	−1.974***
	(0.015)	(0.023)	(0.008)	(0.014)
city accept frequency	0.431***	0.390***	0.579***	0.960***
	(0.021)	(0.039)	(0.005)	(0.009)
city reject frequency	−0.905***	−0.523***	−1.179***	−2.122***
	(0.019)	(0.032)	(0.008)	(0.014)
minimum buyer group distance	−0.131***	−0.172***	−0.046***	−0.106***
	(0.009)	(0.015)	(0.002)	(0.004)
title match score	3.366***	2.628***	4.825***	7.752***
	(0.115)	(0.211)	(0.034)	(0.050)
specification count	0.236***	0.424***	0.337***	0.073***
	(0.045)	(0.077)	(0.012)	(0.019)
quantity	−0.108***	−0.440***	−0.822***	−0.081***
	(0.040)	(0.064)	(0.011)	(0.017)
order value	0.206***	−0.101	0.917***	0.281***
	(0.040)	(0.071)	(0.011)	(0.017)
recency	−0.031***	−0.075***	−0.059***	−0.299***
	(0.007)	(0.012)	(0.002)	(0.004)
seller volume	−0.102**	−0.037	−0.249***	−1.129***
	(0.019)	(0.036)	(0.006)	(0.010)
product category volume	−0.119***	−0.094***	−0.165***	−0.614***
	(0.010)	(0.018)	(0.003)	(0.004)
city volume	−0.155***	−0.074***	−0.201***	−0.337***
	(0.011)	(0.019)	(0.003)	(0.004)
constant	2.483***	−0.199	−1.263***	7.251***
	(0.173)	(0.328)	(0.048)	(0.078)
Observations	378,543	9,702	744,986	744,986

Standard errors are in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

$(TP+TN)/(TP+FP+TN+FN)$. However, simply using the predictive accuracy is not an appropriate metric when the data is imbalanced. For example, If the data is highly imbalanced with 95% of the data points as the accept class and 5% as the reject class, the predictive accuracy would be as high as 95% when we simply predict all the data points as the accept class, even though all the reject class are misclassified. Thus, imbalanced data can pose a big challenge to the effective learning of classification models [95]. In our setting, the classification model is likely to achieve a better prediction for the accept class at the expense of the predictive performance of the decline class, but, in applications such as B2B recommendation systems, correctly recalling the accept class is not the sole objective. Instead, recommending a wrong RFQ can be costly for sellers' satisfaction and the platform's reputation. This point was echoed to us by the CEO of our research partner IndiaMart, Dinesh Agarwal: "The recommendation engine should be a matching tool, not a marketing tool. The company cannot afford the cost of overwhelming our sellers with irrelevant RFQs and losing their trust in our recommendations."

Thus, B2B platforms place more value into correctly predicting the reject class in order to avoid recommending those RFQs to sellers. Therefore, we compute the true positive rate %TP = TP/(TP + FN), also called the recall for accepts, and the true negative rate %TN = TN / (TN + FP), also called the recall for rejects. We use these as additional metrics to assess the predictive performance of our model for the accept class and reject class, respectively. In particular, achieving a high true positive rate at the expense of a low true negative rate is not desirable in our setting.

Table 1.3: Figure 1: Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

The predictive performance of the original, curated, SMOTEd, and PDATEd

Table 1.4: Predictive Performance of Original, SMOTEd, Curated, PDATEd Samples

	(1) Original		(2) Curated		(3) SMOTEd		(4) PDATEd	
	Train_s	Test_s	Train_{curate}	Test_s	Train_{smote}	Test_s	Train_{pdate}	Test_s
True Negative	1,776	584	2,231	1,020	325,870	1,694	359,798	1,393
False Positive	4,284	1,433	1,281	997	46,623	323	12,695	624
False Negative	936	328	844	2,901	36,650	12,429	12,861	4,313
True Positive	371,557	123,836	5,346	121,263	335,843	111,735	359,632	119,851
Prediction Accuracy	0.986	0.986	0.781	0.969	0.888	0.899	0.966	0.961
Recall for Accepts	0.997	0.997	0.864	0.977	0.902	0.900	0.965	0.965
Recall for Rejects	0.292	0.290	0.635	0.506	0.875	0.840	0.966	0.691
Precision for Accepts	0.989	0.989	0.807	0.992	0.878	0.997	0.966	0.995
Precision for Rejects	0.654	0.640	0.726	0.260	0.899	0.120	0.965	0.244
Observations	378,543	126,181	9,702	126,181	744,986	126,181	744,986	126,181

samples are summarized in Table 1.4. We make the following observations from our empirical results. First, the results for the original data sample (without resampling) show that class imbalance hinders the learning of the minority class. The recall for accepts is high (99.7%) but the recall for rejects is low (29.2%). Second, the three resampling methods all help to facilitate the learning of the minority class without costing too much predictive accuracy for the majority class. The predictive accuracy for the reject class increases to 63.5%, 87.5%, and 96.6% in the training samples Train_{curate}, Train_{smote}, and Train_{pdate}, respectively. We also achieve higher predictive accuracy for the minority class in the testing sample across all three resampling methods. The recall for rejects is the highest (84.0%) for SMOTE, followed by PDATE (69.1%) and curate (50.6%) methods. Compared to the recall for rejects when the model is trained on the original sample (29.0%), the improvement is substantial. Correspondingly, we find that the recall for accepts remains high for the three resampling methods: it is 97.7% for Curate, 90.0% for SMOTE, and 96.5% for PDATE. Thus, we observe that resampling makes a significant improvement to the performance of the binary classifier, with even the naïve curated resampling method performing well. However, SMOTE and PDATE yield much higher benefit, and SMOTE performs slightly better than PDATE.

1.5.3 Effect of Data Augmentation

We further assess the effect of data augmentation on predictive preformance by varying the accept ratio in the training set after resampling from 0.5 to 0.9 in steps of 0.1. Note that the number of synthetic observations is the highest when the accept ratio is 0.5, and the lowest for 0.9. We present the model preformance using SMOTE and PDATE in Table 1.5 and Table 1.6, respectively. To better assess the relationship, we present the Receiver Operating Characteristic (ROC) graph in Figure 1.2. In the ROC space, the x-axis represents the false positive rate %FP (%FP = FP / (TN + FP)) and the y-axis represents the true positive rate %TP (%TP = TP / (TP + FN)). Note that the %FP equals (1- Recall for Rejects) and %TP equals Recall for Accepts in our setting.

The ROC graph provides a convenient view to compare the classification performance for the accept and reject classes. In Figure 1.2, we plot ROC curves under SMOTE and PDATE with the varying accept ratios from 0.5 to 0.9 in steps of 0.1. We also plot the prediction outcome for the original sample for comparison. The best value on the graph is the value closest to the top left corner (100% true positive rate, 0% false positive rate). We find that PDATE performs better than SMOTE on the training data set, viz., it is able to decrease the number of false positives (i.e., increase the recall for rejects) much more substantially without much decrease in the true positive rate. On the test sample, both methods have similarly shaped ROC curves. The SMOTE curve is slightly above the PDATE curve and is able to achieve a larger reduction in the false positive rate for the same accept ratio. Thus, we provide model fit results obtained from SMOTE and PDATE to the company for field experiments.

Figure 1.2: Comparision of ROC Curves for SMOTE and PDATE (Note that each curve sweeps from the left to the right as the accept ratio for oversampling increases from 0.5 to 0.9, i.e., as the number of synthetic records declines)

(a) Training Samples

(b) Testing Samples

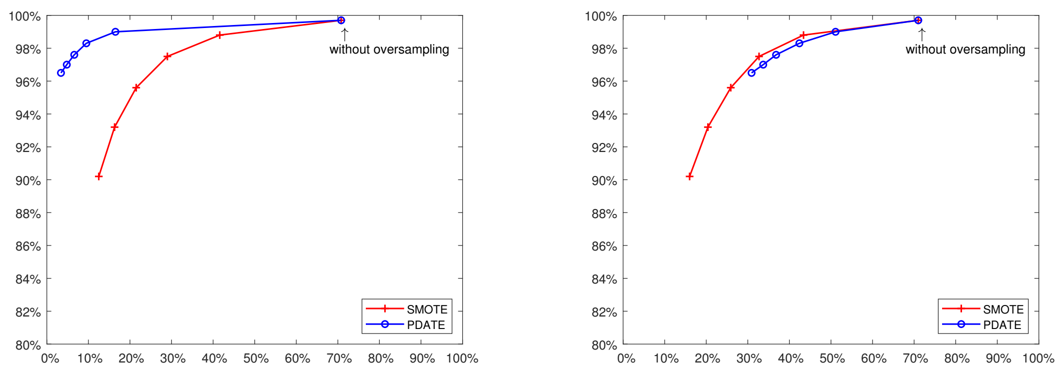


Table 1.5: Predictive Performance of SMOTEd Samples

	(1) Accept Ratio = 0.9		(2) Accept Ratio = 0.8		(3) Accept Ratio = 0.7		(4) Accept Ratio = 0.6	
	Train_{smote}	Test_{s}	Train_{smote}	Test_{s}	Train_{smote}	Test_{s}	Train_{smote}	Test_{s}
True Negative	24,180	1,142	66,094	1,357	125,336	1,494	207,861	1,605
False Positive	17,209	875	27,030	660	34,304	523	40,468	412
False Negative	4,354	1,469	9,359	3,116	16,521	5,435	25,503	8,585
True Positive	368,139	122,695	363,134	121,048	355,972	118,729	346,990	115,579
Prediction Accuracy	0.948	0.981	0.922	0.970	0.904	0.953	0.894	0.929
Recall for Accepts	0.988	0.988	0.975	0.975	0.956	0.956	0.932	0.931
Recall for Rejects	0.584	0.566	0.710	0.673	0.785	0.741	0.837	0.796
Precision for Accepts	0.955	0.993	0.931	0.995	0.912	0.996	0.896	0.996
Precision for Rejects	0.847	0.437	0.876	0.303	0.884	0.216	0.891	0.158
Observations	413,882	126,181	465,617	126,181	532,133	126,181	620,822	126,181

Table 1.6: Predictive Performance of PDATEd Samples

	(1) Accept Ratio = 0.9		(2) Accept Ratio = 0.8		(3) Accept Ratio = 0.7		(4) Accept Ratio = 0.6	
	Train_{pdate}	Test_{s}	Train_{pdate}	Test_{s}	Train_{pdate}	Test_{s}	Train_{pdate}	Test_{s}
True Negative	34,552	987	84,231	1,161	149,052	1,275	236,355	1,337
False Positive	6,837	1,030	8,893	856	10,588	742	11,974	680
False Negative	3,655	1,279	6,413	2,179	8,833	2,973	11,010	3,682
True Positive	368,838	122,885	366,080	121,985	363,660	121,191	361,483	120,482
Prediction Accuracy	0.975	0.982	0.967	0.976	0.964	0.971	0.963	0.965
Recall for Accepts	0.990	0.990	0.983	0.982	0.976	0.976	0.970	0.970
Recall for Rejects	0.835	0.489	0.905	0.576	0.934	0.632	0.952	0.663
Precision for Accepts	0.982	0.992	0.976	0.993	0.972	0.994	0.968	0.994
Precision for Rejects	0.904	0.436	0.929	0.348	0.944	0.300	0.955	0.266
Observations	413,882	126,181	465,617	126,181	532,133	126,181	620,822	126,181

1.6 Field Experiment

The purpose of this experiment is to estimate the impact of our recommendation algorithm, test it against the platform's existing method, and provide rationale for scaling the engine to a larger user base of the platform. Delivering positive impact of the engine was crucial in gaining buy-in from the executives of IndiaMart. We describe the implementation process of our method in section 1.6.1, present our controlled experimental design in section 1.6.2, and show the results from the field experiment in section 1.6.3.

1.6.1 Implementation

The platform provides each registered seller with alternative ways to select RFQs: a search page, a recency page, and a relevant page. The search page allows a seller to discover RFQs using search keywords. The recency page ranks RFQs based on their timestamps. The relevant page is subdivided into a default page, which presents the recommendation engine results (we refer to this page the "algorithm" page), and a filter page, which displays a subset of recommended RFQs according to filtering criteria such as geographical location. In the pre-treatment period, the algorithm page and filter page on both treatment and control groups were run using the company's proprietary model. In the treatment period, our recommendation algorithm was implemented on the algorithm page for a treatment group of sellers. The filter page for the treatment group and both filter and algorithm page for the control group continued to be based on the platform's proprietary model.

The implementation of our model involves constructing the predictors, computing estimated acceptance probabilities for all RFQs and sellers, and real-time handling of page requests from sellers. The algorithm page needs to respond within mere milliseconds at the front end of the platform upon request. Therefore, to ensure responsiveness, we divide the architecture of the recommendation engine into three modules: a batch processing module, an RFQ arrival module, and a real-time module. The batch processing module is executed every evening to construct an archive of the historical engagement frequencies of sellers from the transactions of the previous three months. The RFQ arrival module is executed upon the arrival of an RFQ, and includes mapping the RFQ to sellers and computing all the variables in the model except recency. Finally, when a seller logs in and requests the relevant page, the recency of available RFQs mapped to the seller is computed, acceptance probabilities are generated, and RFQs are sorted and displayed to the seller.

1.6.2 Experimental Design

In this subsection, we present our experimental design, the timeline of the experiment, and our treatment and control samples. The platform selected 21 sellers on which the model had not been trained as the treatment group and a set of 63 matched sellers as the control group. Sellers in both groups were selected based on common criteria: (i) they should be active users of the desktop system rather than the mobile app since our model was implemented for the desktop system, (ii) their weekly usage of credits should be around 70% of their subscribed allowances so that they have scope for benefiting from the new model, (iii) they should have a paid subscription package with IndiaMart. In addition, the sellers across the treatment and control groups were matched on product categories, volume of activity and locations. For the sellers in the treatment group, our recommendation engine was implemented in the backend to determine the ranking of the RFQs of the display pool on the algorithm page. For the sellers in the control group, the platform's existing engine was used to construct the algorithm page. The sellers were not aware of the change in the back end.

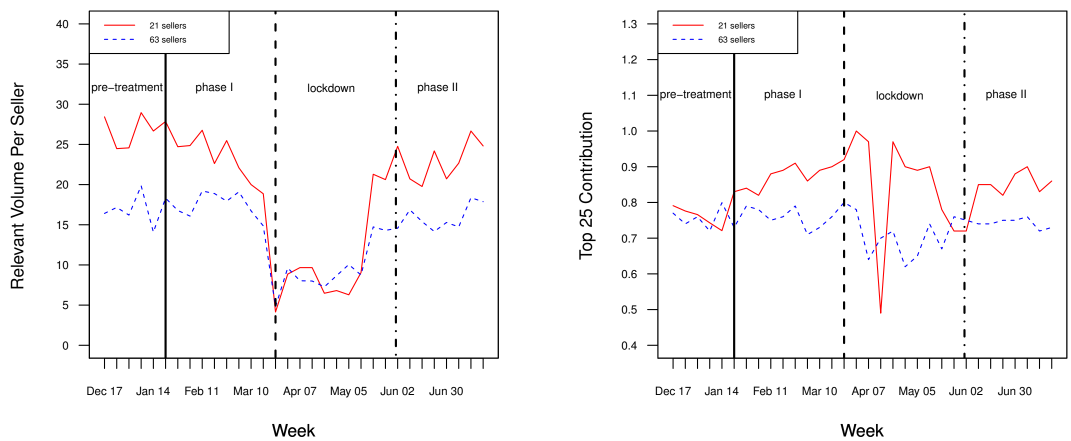
We implemented our recommendation engine on January 21, 2020. From January 21 to April 20, the platform used our recommendation engine excluding the title match predictor. From April 20 to July 27, the platform used our recommendation engine including the title match predictor. This is to purposefully evaluate the additional effect of the text variable.

We tracked all the transactions of the treatment and control groups starting 5 weeks before the implementation date on Dec 17, 2019 and throughout the experiment, and monitored several outcome metrics. The outcome metrics include the daily volume of transactions on the relevant page, the positions of accepted and rejected RFQs for each seller, the percentage of transactions in top 5, top 10, and top 25 positions, the recency of accepted RFQs, the computed probabilities of acceptance from the model, and so on in both the treatment group and the control group. Metrics were monitored for both the relevant and the algorithm page. These metrics indicate the performance of our recommendation engine, and a higher quality recommendation engine should yield a higher volume of transactions in the top positions. In particular, each page accommodates 25 RFQs, and thus, transactions occurring within the top 25 positions correspond to first-page transactions.

One unforeseeable event occurred during our study period. On March 24, 2020, the Indian government ordered a nationwide lockdown to contain the spread of COVID-19. Although the lockdown was initially planned for 21 days, it was eventually extended to May 31st. Business activity was severely affected during the lockdown, resulting in a decline of transactions on the platform. Thus, we exclude the lockdown period from the statistical analysis of performance, and comment on our learnings from the lockdown period separately towards the end of this section. For analysis in the remainder of this section, we refer to the time period from December 17, 2019 to January 20, 2020 as the pre-treatment period (35 days), the period from January 21 to July 27, 2020 as the treatment period (189 days), the sub-period from January 21 to March 23, 2020 (63 days) as phase I of the experiment, the sub-period from June 1 to July 27, 2020 (57 days) as phase II, and the sub-period from March 24 to May 31, 2020 (69 days) as lockdown period. We present statistics from all these periods, but use data from phase I and phase II to assess our methodology.

Figure 1.3: Weekly Transaction Volume and Top 25 Contribution of Treatment Group and Control Group [1]

(a) Transaction Volume (b) Top 25 Contribution



Our field experiment design provides us an opportunity to apply a difference-in-differences (DID) approach to evaluate the impact of our recommendation engine. Thus, we estimate the average treatment effect of our recommendation engine by comparing the changes before and after the treatment for both the treatment and the control groups. The DID approach relies on several assumptions that we address in the design of the experiment: sellers were selected for the treatment and control groups randomly by the data scientists at the platform, the dates of the experiment were unrelated to the intervention, and the treatment and control groups experienced parallel time trends in the pre-treatment period. One advantage of our method is that the company continued to offer the search, filter, and recency pages to sellers in the treatment group throughout the experiment. This enables us to rigorously assess performance with respect to volume of activity, share of activity, and quality metrics.

[1]There was a server outage in the third week of April 2020 resulting in a sharp dip in the Top

In Figure 1.3(a), we show the weekly total volume of transactions on the relevant page for the treatment and control groups over time. Note, with reference to the lockdown, that a significant drop in transactions occurred in the week of March 24, and the volume thereafter remained low for several weeks before gradually returning to the pre-lockdown levels. In Figure 1.3(b), we show the weekly values of top 25 contribution for treatment and control groups over time. Top 25 contribution is the fraction of relevant transactions that are conducted from the top 25 RFQs displayed to a seller on the algorithm page. This is one of the main performance metrics for our analysis. Figure 1.3 also shows that both the total volume of transactions for the relevant page (Figure 1.3(a)) and the proportion of these RFQs accepted within the top 25 positions (Figure 1.3(b)) exhibit a common trend for the treatment and control groups in the pre-treatment period. We expect the common trend assumption to hold throughout the treatment period.

1.6.3 Results and Implications

We first present a summary of performance metrics for the treatment group for each time period in Table 1.7. Columns (1)-(4) show combined statistics for the relevant pages (algorithm + filter) for the four time periods: pre-treatment, entire treatment, phase I, and phase II. Columns (5)-(7) show statistics for the algorithm page alone for the entire treatment, phase I, and phase II, respectively.

A total of 5,686 RFQs were transacted from the algorithm page based on our model in the treatment period, which amounts to 32% of the volume of trans-

25 contribution. The algorithms for the treatment and control group are run on different nodes of the server, so the outage affected only the treatment group.

Table 1.7: Key Performance Metrics for the 21 sellers in the Treatment Group

	Relevant Pages (Algorithm + Filter)				Algorithm Page		
Period	(1) Pre-treatment	(2) Treatment	(3) Phase I	(4) Phase II	(5) Treatment	(6) Phase I	(7) Phase II
Duration (days)	35	189	63	57	189	63	57
Desktop Volume	4,714	17,763	8,162	6,175	17,763	8,162	6,175
Page Volume	2,795	10,508	4,478	3,870	5,686	2,364	2,496
Page Contribution (% of Desktop Volume)	59.3%	59.2%	54.9%	62.7%	32%	28.9%	40.4%
Top 5 Volume	1,281	4,460	2,140	1,568	2,644	1,178	1,118
Top 5 Contribution (% of Page Volume)	45.8%	42.4%	47.8%	40.5%	46.5%	49.8%	44.8%
Top 10 Volume	1,648	5,787	2,817	1,991	3,642	1,590	1,579
Top 10 Contribution (% of Page Volume)	59.0%	55.1%	62.9%	51.4%	64.1%	67.3%	62.9%
Top 25 Volume	2,122	7,968	3,560	2,894	4,797	2,043	2,089
Top 25 Contribution (% of Page Volume)	75.9%	75.8%	79.5%	74.8%	84.4%	86.4%	83.7%
Average Position of Accept	37.95	31.65	24.31	43.19	16.21	12.57	17.24
Median Position of Accept	6.79	8.12	6.12	8.47	6.63	5.50	6.96
Average Recency of Accept (hrs)	10.71	14.01	14.02	12.61	15.22	14.73	14.06
Average Probability of Accept					0.83	0.84	0.87
Average Probability of Reject					0.35	0.32	0.44

actions by these sellers from all modes via desktop (including the search page and filter page). The number of transactions per day on the algorithm page was 37.5 in phase I and increased to 43.8 in phase II, even though overall relevant transactions per day declined from 71.1 in phase I to 67.9 in phase II. Thus, our method had a significant and sustained impact on these sellers during the 6-month long field experiment from Jan 21 to July 27, 2020 despite the challenge posed by the pandemic.

We now highlight observations from comparisons of the remaining performance metrics. First consider the top-5 contribution, which is defined as the number of accepted RFQs from the top-5 positions as a proportion of the total volume of accepted RFQs from the page. We observe that the top-5 contribution from the algorithm page is 49.8% in phase I, 44.8% in phase II, and 46.5% overall in the treatment period. Each of these numbers is higher than the corresponding numbers for the relevant pages. Moreover, the phase I and overall treatment numbers are higher than the pre-treatment statistic of 45.8%. We see a larger impact for the top-10 contribution as the algorithm page yields 64.1% of RFQs from top-10 treatment as opposed to 55.1% aggregate for the relevant pages and 59% for the pre-treatment period. This trend continues for the top-25 contribution: our model yields 84.4% contribution from top-25 in the treatment period compared to 75.8% aggregate for the relevant pages and 75.9% for the pre-treatment period. From these numbers, we conclude that our model ranks more high quality RFQs at the top of the page than the existing algorithm, resulting in a higher percentage contribution and increased share of volume. Consistent with this, note that the average and median positions of accepted RFQs are lower in our model than the pre-treatment period or the treatment aggregate for relevant pages. From the last three rows of the table, we note that the average recency of accepted RFQs remains about the same, and there is a large difference in the estimated acceptance probabilities of accepted and rejected RFQs. Thus, the results show the benefit of our recommendation system.

We now compare the treatment group with the control group with respect to the top-5, top-10, and top-25 contributions for phase I and phase II periods. (We present DID estimates for the lockdown period in the Appendix in Section 1.8). First, we conduct DID analysis at the seller-day level using the model:

$$y_{st} = \beta_0 + \beta_1 Tr_s + \beta_2 T_t + \beta_3 Tr_s \cdot T_t + \epsilon_{st}, \tag{1.6}$$

where s denotes the seller, t denotes the date of the transaction, y_{st} is a performance metric (top-5, top-10, or top-25 contributions), Tr_s indicates whether seller s belongs to the treatment group, $T_t = 1$ or 0 denote whether date t is after or before the implementation of our recommendation engine, respectively, and the β's denote the regression coefficients. The estimate of β_3 is the coefficient of interest for assessing the average treatment effect.

Table 1.8 presents the ordinary least squares regression results for (1.6) obtained for phase I and phase II. We observe that the performance improvement from our method is statistically significant for each performance metric for both phases. Specifically, using our recommendation engine has on average increased the likelihood for a seller to accept RFQs from the top-5, top-10, and top-25 positions in phase I by 5.9%, 9.0%, 8.2%, respectively, relative to the control group. For phase II, the positive impact is 5.8% for top-5, 9.9% for top-10, and 9.1% for top-25 compared to the control group.

This DID analysis can also be conducted at the RFQ level using the following regression. We extend the same notation as above for simplicity.

$$y_r = \beta_0 + \beta_1 Tr_r + \beta_2 T_r + \beta_3 Tr_r \cdot T_r + \epsilon_r, \tag{1.7}$$

where $y_r = 1$ if RFQ r was in the top-5 and 0 otherwise, Tr_r indicates whether RFQ r was accepted by a seller in the treatment group, T_r indicates whether RFQ r is in the post- or pre-treatment time period, and β's denote the regression coefficients as before. We again conduct the regression for top-5, top-10 and top-25 by changing the definition of y_r. The coefficient of interest is β_3. The bottom half of Table 1.8 shows the results from (1.7). This method also shows that implementing our recommendation engine has a significantly positive effect on the acceptance position of RFQs in both phases. Specifically, the likelihood for an accepted RFQ to be ranked within the top 5, top 10, and top 25 positions on average increased by 7.0%, 11.8%, and 14.2%, respectively, in phase I, and by 4.3%, 9.0%, and 12.8%, respectively, in phase II.

Beyond the positive impact of our recommendation engine indicated by the

Table 1.8: Difference-in-difference analysis of performance metrics for our recommendation engine compared to the control group

	(1) Top 5 Contribution	(2) Top 10 Contribution	(3) Top 25 Contribution
Seller-level Analysis			
Phase I Average Treatment Effect	0.059*	0.090***	0.082***
	(0.027)	(0.024)	(0.018)
Phase I Observations	3,642	3,642	3,642
Phase II Average Treatment Effect	0.058*	0.099***	0.091***
	(0.028)	(0.026)	(0.019)
Phase II Observations	3,212	3,212	3,212
RFQ-level Analysis			
Phase I Average Treatment Effect	0.070***	0.118***	0.142***
	(0.016)	(0.016)	(0.013)
Phase I Observations	20,380	20,380	20,380
Phase II Average Treatment Effect	0.043***	0.090***	0.128***
	(0.016)	(0.016)	(0.013)
Phase II Observations	18,581	18,581	18,581

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The observations exclude the seller-days with zero transactions.

DID analysis, we obtained additional insights from the field experiment. First, we noticed anecdotally that the behavior of sellers was sticky—if a seller accepted a high proportion of RFQs from a page on one day, then the seller was more likely to access that page on subsequent days. This led to cyclical deviations in the usage of the algorithm page. Unfortunately, such behavior could be triggered by uncertain quality of the pool of available RFQs; if available RFQs on a particular day were not well matched to a seller's preferences, the seller was more likely to direct efforts on the search and filter pages, and thus, less likely to use the algorithm page the next day. Thus, we did not see a jump in performance immediately upon implementing the system, but rather a gradual shift to the algorithm page. This can be observed from the upward sloping trend line in phase 1 in Figure 1.3(b). Second, we investigated the performance of our model during the lockdown period by conducting a DID analysis using the data from March 24 to May 31, 2020. Results from this analysis are shown in the Appendix in Section 1.8. We continue to find a significant difference between the performance of our recommendation engine and the control group during the lockdown period since the performance of the treatment group did not suffer as much as that of the control group.

1.7 Conclusion

In this paper, we formulate and solve the problem of designing a recommendation system for a B2B platform and demonstrate significant performance improvement in both out-of-sample and field experiment results. We attribute these improvements to the differences between the methods. In particular, by comparing prediction results across the original data set and the resampling methods, our out-of-sample tests show that correcting for class imbalance yields a substantially larger recall for the minority class. Then, the field experiment compares our approach to the company's existing non-parametric method, which was based on a fewer set of discretized attributes and used data for only accepted RFQs. This comparison shows that the acceptance rates from top-5, top-10, and top-25 recommendations increase significantly by expanding the set of attributes using a parametric method and employing data for rejected RFQs.

Our paper suggests several opportunities for future research. First, it would be valuable to incorporate attribute data and majority-minority classes in the design of non-parametric choice methods. This would be useful in improving both parametric and non-parametric methods by evaluating them against each other on standard problems. Second, the algorithm for generating recommendations could be enriched using the text data contained in RFQ items, which can be beneficial to parse the heterogeneity in business enquiries represented by RFQs. Although we have explored some aspects of the text data and adopted Levenshtein ratio as a predictor, it is possible to utilize more of the text data and examine different kinds of text-based metrics in future research. The algorithm could also incorporate heterogeneity in preferences across different types of sellers. For example, some sellers may be more sensitive to the recency of RFQs, whereas others may be more sensitive to geographical location. Third, using the data generated from the recommendation system, it would be valuable to study seller behavior with respect to display variety. As noted in Section 6, we found that seller behavior on the website takes time to adjust because randomness in the availability of RFQs makes it difficult for sellers to realize the improvement in recommendations. It would be useful to measure how fast users adopt the new page when shown an improved set of recommendations in a noisy environment. Finally, whereas our paper looked at the immediate impact of a recommendation page on the acceptance of RFQs by sellers, it would be valuable to assess the long-term impact of the quality of a recommendation engine on the engagement of sellers and buyers on the platform.

1.8 Appendix

Tables A.1. to A.4. show the summary statistics of the original training and testing samples by the class. Table A.5. shows the difference-in-differences estimates for the national lockdown period enforced in India during our field experiment.

Table A.1.: Train$_s$ Accept Class

	Mean	Std.	Median	Min	Max	Qu.1st	Qu.3rd
product category accept frequency	2.899	1.606	2.996	0.000	6.627	1.792	4.143
product category reject frequency	0.067	0.379	0.000	0.000	5.649	0.000	0.000
city accept frequency	2.053	1.588	1.946	0.000	6.184	0.693	3.219
city reject frequency	0.085	0.407	0.000	0.000	5.112	0.000	0.000
minimum buyer group distance	3.026	2.737	3.953	0.000	8.920	0.000	5.527
title match score	0.922	0.117	1.000	0.125	1.000	0.862	1.000
specification count	1.675	0.450	1.792	0.693	3.611	1.386	1.946
quantity	0.674	0.469	1.000	0.000	1.000	0.000	1.000
order value	0.285	0.452	0.000	0.000	1.000	0.000	1.000
recency	5.187	2.135	5.301	-2.485	9.303	3.823	7.011
seller volume	5.784	0.797	5.984	0.693	7.272	5.375	6.380
product category volume	6.024	2.098	6.328	0.000	9.010	4.663	7.779
city volume	6.988	2.101	7.442	0.000	9.717	5.565	8.928

Table A.2.: Train$_s$ Reject Class

	Mean	Std	Median	Min	Max	Qu.1st	Qu.3rd
product category accept frequency	1.683	1.538	1.386	0.000	5.892	0.000	2.996
product category reject frequency	1.696	1.665	1.386	0.000	5.690	0.000	2.944
city accept frequency	1.446	1.510	1.099	0.000	6.129	0.000	2.398
city reject frequency	1.107	1.328	0.693	0.000	5.106	0.000	1.946
minimum buyer group distance	3.841	2.810	5.064	0.000	7.704	0.000	6.217
title match score	0.803	0.182	0.828	0.255	1.000	0.667	1.000
specification count	1.651	0.434	1.609	0.693	2.944	1.386	1.946
quantity	0.700	0.458	1.000	0.000	1.000	0.000	1.000
order value	0.280	0.449	0.000	0.000	1.000	0.000	1.000
recency	5.126	2.322	5.091	-1.529	10.767	3.475	7.037
seller volume	5.330	1.168	5.517	0.693	6.878	4.522	6.286
product category volume	5.382	2.372	5.775	0.000	9.007	3.738	7.252
city volume	7.165	2.167	7.701	0.000	9.716	5.666	9.125

Table A.3.: Test$_s$ Accept Class

	Mean	Std	Median	Min	Max	Qu.1st	Qu.3rd
product category accept frequency	2.895	1.600	2.996	0.000	6.627	1.792	4.127
product category reject frequency	0.065	0.375	0.000	0.000	5.670	0.000	0.000
city accept frequency	2.050	1.584	1.946	0.000	6.184	0.693	3.219
city reject frequency	0.085	0.408	0.000	0.000	5.063	0.000	0.000
minimum buyer group distance	3.033	2.738	3.985	0.000	8.914	0.000	5.529
title match score	0.922	0.117	1.000	0.160	1.000	0.861	1.000
specification count	1.675	0.451	1.792	0.693	3.638	1.386	1.946
quantity	0.672	0.469	1.000	0.000	1.000	0.000	1.000
order value	0.284	0.451	0.000	0.000	1.000	0.000	1.000
recency	5.192	2.129	5.307	-2.485	9.302	3.829	7.010
seller volume	5.784	0.794	5.981	0.693	7.272	5.380	6.378
product category volume	6.018	2.099	6.319	0.000	9.010	4.673	7.774
city volume	6.989	2.098	7.436	0.000	9.717	5.583	8.922

Table A.4.: Test$_s$ Reject Class

	Mean	Std	Median	Min	Max	Qu.1st	Qu.3rd
product category accept frequency	1.638	1.528	1.386	0.000	6.351	0.000	2.890
product category reject frequency	1.637	1.633	1.099	0.000	5.690	0.000	2.833
city accept frequency	1.405	1.509	1.099	0.000	5.878	0.000	2.398
city reject frequency	1.102	1.359	0.693	0.000	5.063	0.000	1.946
minimum buyer group distance	3.862	2.812	5.008	0.000	7.704	0.000	6.228
title match score	0.804	0.183	0.830	0.214	1.000	0.667	1.000
specification count	1.646	0.434	1.609	0.693	2.996	1.386	1.946
quantity	0.690	0.463	1.000	0.000	1.000	0.000	1.000
order value	0.269	0.443	0.000	0.000	1.000	0.000	1.000
recency	5.133	2.333	5.125	-1.696	10.673	3.436	7.031
seller volume	5.298	1.163	5.468	0.693	6.963	4.522	6.275
product category volume	5.352	2.375	5.645	0.000	9.002	3.689	7.282
city volume	7.091	2.217	7.645	0.000	9.713	5.561	9.122

Table A.5.: Effect of Our Recommendation Engine on Performance Metrics during Lockdown Period

	(1) Top 5 Contribution	(2) Top 10 Contribution	(3) Top 25 Contribution
Seller-level Analysis			
Lockdown Average Treatment Effect	0.082^{*}	0.138^{***}	0.094^{***}
	(0.037)	(0.034)	(0.026)
Lockdown Observations	2,351	2,351	2,351
RFQ-level Analysis			
Lockdown Average Treatment Effect	0.095^{***}	0.118^{***}	0.100^{***}
	(0.022)	(0.022)	(0.018)
Lockdown Observations	27,142	27,142	27,142

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The observations exclude the seller-days with zero transactions.

CHAPTER 2

BONUS COMPETITION IN THE GIG ECONOMY

2.1 Introduction

In recent years, service platforms which rely on independent service providers have seen rapid growth with the rise of the gig economy. Several well-known platforms include Uber and Lyft for ride-hailing, Grubhub and DoorDash for food delivery, TaskRabbit and Handy for home service, and VIPKID and 51Talk for online English education. By 2018, 55 million people in the U.S. are "gig workers," which is more than 35% of the U.S. workforce. That number is projected to increase to 43% by 2020 [89]. Platforms in the gig economy utilize independent contractors (so called "gig workers") rather than traditional employees as a cost-effective means to provide on-demand service, and service providers benefit from greater flexibility in their work scheduling. However, this innovative business model poses a challenge to platforms in maintaining consistent participation of service providers, due to the platforms' lack of control over service providers' working schedules.

Moreover, because service providers in the gig economy are independent contractors, they can serve on more than one platform. Multihoming of service providers is ubiquitous in the gig economy. For instance, it is very common for drivers to register on both Uber and Lyft in the ride-hailing market [4]. The phenomenon of multihoming exacerbates the challenge that platforms face in maintaining consistent participation of service providers. This can lead to fierce competition between platforms because a steady supply base is pivotal to a platform's growth and seamless operations. In an NPR podcast titled "In the Battle

Between Lyft and Uber, the Focus is on Drivers," it is reported that "competition for drivers is so great that Uber even sent covert operatives into Lyft cars to recruit" [145].

Competition between gig economy platforms is two-sided. Unlike traditional businesses, gig economy platforms compete not only for customers (i.e., the demand side), but also for service providers (i.e., the supply side). It is believed in industry discussions that success of this business model is more crucially driven by the platform's ability to retain service providers [155, 145]. Gig services are essentially a commodity for customers. For example, a customer may not care about whether a home cleaner is hired from TaskRabbit or Handy as long as the work is done. Only with the consistent participation of service providers can a platform attract more customers and achieve sustainable profitability. However, as we discussed above, it is particularly challenging for gig economy platforms to maintain a steady supply base because of service providers' ability to switch between platforms. In practice, to tackle this problem, platforms have been commonly offering bonuses to service providers that are contingent on their consistent participation. The Quest program of Uber offers drivers a bonus once they complete a certain number of trips within a certain time period. Under this program, Uber drivers would obtain a $50 cash reward after completing 10 trips on a particular weekend. Similarly, Lyft has a competitive bonus scheme called Weekly Ride Challenge. Platforms in other markets have also been offering this type of contingent bonuses to service providers. For example, VIPKID has been offering registered teachers a bonus after they deliver a certain number of classes within a month.

The bonus strategy of a gig economy platforms is one way to subsidize service providers to drive their consistent participation. Moreover, the contingent feature of such bonuses is intended to lock in service providers and prevent them from switching to competing platforms. However, an interesting industry phenomenon is that platforms in certain markets end up paying too much in bonuses to service providers. The ride-hailing market is a case in point. As the major players in the U.S. ride-hailing market, both Uber and Lyft are known to commonly use bonus strategies to fight for drivers. However, the use of bonus strategies in this case has led both platforms to over-subsidize drivers and lose profits overall. In 2017, Uber's full-year net loss was $4.5 billion and Lyft's full-year net loss was $688 million [43, 165]. Similar bonus wars have also happened in China's ride-hailing market, where the two major players, Didi Dache and Kuaidi Dache, eventually merged into Didi Chuxing. Such phenomena have triggered many industry discussions and debates [149, 164], and also raise the questions of whether using a bonus strategy is actually profitable in competitive environments, and whether ending the bonus war through collaboration is the right path for competing platforms.

Motivated by the recent industry practice, our paper aims to provide insights into two key issues related to the contingent bonus strategy used by gig economy platforms. First, we aim to identify the driving forces for platforms to offer bonuses. Specifically, what market conditions will induce platforms to offer bonuses to service providers? What are the implications of platforms' bonuses on service providers' participation decisions, platforms' market coverage, and their profits? Will the bonus strategy be a competitive weapon or a corporate millstone for gig economy platforms? Second, motivated by the heated industry discussions of gig economy platforms losing huge money to subsidize service providers as well as the merger of platforms in certain markets, we aim to understand the implications of platform coordination on platforms' bonuses to offer providers. Specifically, would the implications vary in different market conditions? Would the platforms' profit improvement through bonus coordination come at the expense of the surplus of other stakeholders in the gig economy and social welfare?

To study these research questions, we develop a model that captures two-sided competition between platforms as well as their contingent bonus strategies in a stylized manner. In particular, we consider two platforms who compete in two periods. The platforms may choose to offer a bonus to the service providers. To reflect the industry practice that a bonus is contingent on the consistent participation of service providers, the criterion for a service provider to obtain a bonus from a platform is that he participates on that platform in both periods. Correspondingly, when the service providers choose which platform to participate on in each period, they take into consideration that they will only receive a bonus from a platform by participating on that same platform for both periods. Our model captures two-sided competition between the platforms in the following way. On the supply side, the service providers have heterogeneous preferences between the competing platforms, which we capture by using the Hotelling framework where the service providers are assumed to be located on a line segment connecting the two platforms. A service provider's proximity to the platforms indicates his platform preference. On the demand side, we assume that the platforms engage in a Cournot competition where the market price is influenced by the total number of service providers that participate in a period.

We first analyze a competitive bonus game where the platforms choose their bonuses to maximize their own profits as a response to their competing platforms' bonuses. We find that the platforms offer bonuses in equilibrium when the service providers' stickiness to their preferred platforms is either sufficiently low or sufficiently high. However, the driving forces for the platforms to offer bonuses are different in the two cases. First, in markets where the service providers' platform stickiness is low, the platforms need to fight for those service providers who do not exhibit a strong preference for either platform and could be easily swung by either platform. In this case, each platform can poach such service providers with a bonus. However, the other platform can respond by also offering a bonus to lock in the service providers. Therefore, bonuses in this case serve as a means to prevent the platforms' existing services providers from being poached by their competing platforms. However, we find that the service providers' low platform stickiness will force both platforms to over-subsidize the service providers with bonuses and end up in a prisoner's dilemma where they would lose profits. This finding can explain that in highly competitive markets such as ride-hailing, platforms can be easily trapped in a bonus war that leads to huge monetary losses.

Second, the platforms also offer bonuses in equilibrium when the service providers' platform stickiness is high and platforms' commission rate is high. This corresponds to markets where each platform covers a non-overlapping segment of the supply pool and it is too costly for each platform to poach. The platforms can thus use the bonus strategy to expand their market coverage by inducing more service providers, who would otherwise serve neither platform, to participate. Therefore, in markets where service providers have relatively higher platform stickiness (such as online education), the bonus strategy can be an effective lever for platforms to achieve market expansion and increase profitability.

To shed light on the implications of platform coordination, we next analyze a coordinated bonus game where the platforms choose their bonuses to maximize their total profit. We find that when the platforms coordinate, they will reduce their bonuses and achieve Pareto improvement. However, the implications on social welfare are different depending on the underlying mechanism for offering bonuses. First, in markets where the service providers' platform stickiness is low, coordination will eliminate bonuses completely, so that the platforms can steer clear of the prisoner's dilemma and improve profits significantly. Moreover, because the supply pool is fully utilized regardless of whether a bonus is offered or not, platform coordination will not hurt social welfare, and may in fact improve social welfare if the service providers' preferences between the platforms can change over time. This finding suggests that the merger of the two largest ride-hailing platforms in China not only helps save the platforms from an unsustainable subsidy war, but may also improve the overall efficiency for the society. Second, in markets where the service providers' platform stickiness is high, coordination may still reduce the platforms' bonuses. In this case, platform coordination will eliminate the platforms' incentive to over-supply. By offering a smaller bonus, the platforms will expand their market coverage less aggressively which improves their profits. However, social welfare will be compromised due to a reduced market coverage under platform coordination.

The rest of the paper is organized as follows. In Section 2.2, we review the relevant literature. In Section 3.3, we describe our model. We analyze the competitive and coordinated bonus games in Sections 2.4 and 2.5, respectively. Then, in Section 2.6, we consider a model extension where service providers' preferences between platforms can change over time. Finally, we conclude the paper in Section 2.7. All proofs are provided in the Appendix.

2.2 Literature Review

Our research is primarily related to three domains in the existing literature: pricing and operations management of peer-to-peer platforms, labor supply and social welfare in the gig economy, and platform competition in two-sided markets.

First, our paper belongs to the stream of research that studies pricing and other operational issues of peer-to-peer platforms (see 50 and 128 for reviews of the relevant literature). Several papers have studied the design of payment schemes for the gig economy. Gurvich et al. [77] investigate the platform's optimal decisions of the provider pool size, the wage for each provider, and the cap on the number of providers who are allowed to work. Taylor [159] employs a queueing model to study an on-demand platform's optimal price to charge delay-sensitive customers and optimal wage to pay independent providers. He characterizes how uncertainty in customer valuation of the service and the provider opportunity cost affect the platforms' optimal pricing decisions. Bai et al. [10] consider a platform using independent providers with heterogeneous reservation prices to serve waiting time sensitive customers with heterogeneous valuations for the service. They examine the impact of operating factors such as demand and supply pool size on the payout ratio (i.e., the ratio of wage over price). Hu and Zhou [86] study the performance of the fixed commission contract under which the platforms take a fixed cut of the price to customers and the remaining fraction is wage to providers. Besbes et al. [27] propose a framework where a ride-hailing platform chooses prices for different locations, and drivers respond by choosing where to relocate based on prices, travel costs and driver congestion levels. Cui et al. [59] consider a decentralized pricing model where the pricing decisions are made by individual service providers instead of the platform (e.g., Airbnb), and investigate the impact of social utility that arises from the social interaction between the guest and the host on host pricing, platform profit, and social welfare. Feldman et al. [66] study the contracting between delivery platforms and restaurants, and find that the "one-way" revenue sharing contract in which the platform keeps a fixed percentage of revenue from delivering orders often performs worse for the restaurant than a "no-contract" relationship in which the platform buys from the restaurant at the menu price and resells to customers with a mark-up.

Moreover, there are a number of papers studying surge pricing. Cachon et al. [39] investigate the role of surge pricing (when both prices and wages adjust in response to demand and the ratio of the two is fixed), and they find that surge pricing can substantially increase the platform's profit relative to contracts that have a fixed price or fixed wage (or both). Hu et al. [85] examine surge pricing from a temporal perspective and identify two regimes, which are skimming surge pricing and penetration surge pricing. They show that penetration surge pricing can mitigate controversies caused by sharp price surges. Guda and Subramanian [76] study the role of surge pricing in managing worker availability across market locations. They find that surge pricing can be useful even at locations with excess supply of workers. Besides pricing, there is also a stream of research that studies matching between providers and consumers when the participants' utilities vary with the matching outcomes [e.g., 2, 51, 67, 87, 100]. Our paper studies platforms' optimal bonuses to service providers contingent on consistent provider participation, which is a different and prevailing pricing strategy of gig economy platforms applicable to a wide range of markets. Kabra et al. [98] use data from a leading ride-hailing platform to estimate the effect of monetary incentives (i.e., bonuses) given to both demand and supply sides of the platform on the number of transacted trips. They find that passenger incentives are more effective than driver incentives in the short term and the opposite is true in the long term. In contrast, we use analytical modeling to characterize the market conditions under which platforms have incentives to offer bonuses and study the impact of platforms' bonus strategies on the gig economy.

Our paper is also related to the literature on labor supply and social welfare in the gig economy. Benjaafar et al. [22] study labor welfare in on-demand platforms. They show that labor pool size, delay cost, and variability in the providers' opportunity cost may have a non-monotonic effect on labor welfare. Benjaafar et al. [23] characterize equilibrium ownership and usage levels, consumer surplus, and social welfare in a peer-to-peer product sharing market. Jiang and Tian [92] study product sharing among consumers on a massive-scale platform, and find that transaction costs in the sharing market have a non-monotonic effect on consumer surplus and social welfare. Bimpikis et al. [28] explore spatial price discrimination in a ride-sharing platform and its impact on consumer welfare. Chen and Sheldon [47] and Sheldon [148] study the elasticity of labor supply in on-demand platforms using data from Uber, and find that drivers tend to drive more when earnings are higher. Jin et al. [94] study the impact of bilateral ratings on the platform's pricing policy as well as revenue and welfare implications of all stakeholders in the context of ride-sharing. They find that compared to unilateral rating system, bilateral rating system always improves drivers' welfare, and may also improve riders' welfare and social welfare. Our paper contributes to this literature by examining the implications of platforms' bonus strategies on provider welfare, consumer welfare, and social welfare.

Lastly, our paper is related to the literature on platform competition in a two-sided market. The operations management literature on the gig economy has so far largely focused on monopoly settings. Our paper belongs to the emerging stream of research that studies competition between gig economy platforms. A recent paper, Cohen and Zhang [57], uses a Multinomial Logit choice model to study competing platforms' price and wage decisions. Nikzad [130] studies the effects of thickness and competition on the equilibria of ride-sharing markets. Differently, our paper studies competing platforms' bonus offering decisions, and we use a Hotelling framework to capture service providers' heterogenous preferences for the platforms. The economics literature on platform competition in two-sided markets is fruitful (see the seminal work, 137 and 5, and the references therein). This literature places a major focus on the network externalities of two-sided markets, and investigates their impact on platform competition. While our two-sided competition model captures both cross-side and within-side network externalities, we focus on the bonus competition between platforms and its impacts on the gig economy.

2.3 Model

We consider a setting of two platforms ($i \in \{A, B\}$) who compete in two periods ($j \in \{1, 2\}$) . They compete for both the service providers (i.e., the supply side of the market) and the customers (i.e., the demand side of the market). We use “she” for a platform and “he” for a service provider. Figure 2.1 illustrates our two-sided competition framework.

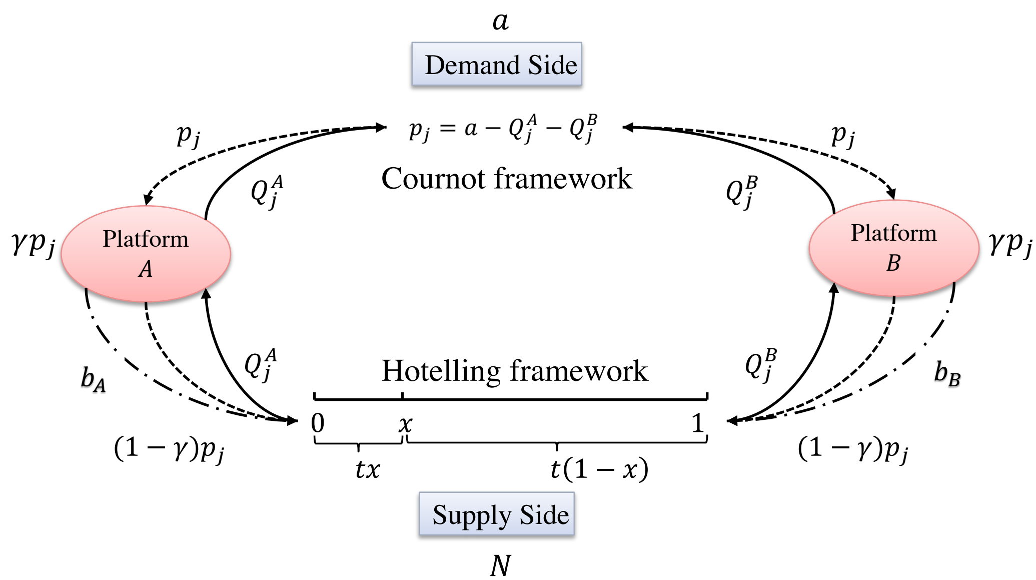


Figure 2.1: Illustration of Two-Sided Competition

On the supply side, there is a pool of N service providers (“providers” for short hereafter) who can participate on at most one platform in each period. We assume a provider offers one unit of service in each period if he chooses to participate. To capture the heterogeneity in the providers’ preferences between the two platforms, we adopt the Hotelling framework where the providers are uniformly located on the Hotelling line $x \in [0, 1]$, and platforms A and B are located at $x = 0$ and $x = 1$, respectively. A provider located at x incurs a participation cost tx if he participates on platform A and $t(1 - x)$ if he participates on platform B. Notice that a provider incurs a lower participation cost with the closer platform, and hence his location on the Hotelling line indicates his preference between the platforms. We refer to t as the marginal participation cost of the providers (it corresponds to the marginal transportation cost in the original Hotelling framework [84]). In our setting, the marginal participation cost measures the providers' stickiness to their preferred platforms. When the marginal participation cost is smaller, providers are less sticky to their preferred platforms, and they exhibit weaker preferences for one platform over the other.

On the demand side, platforms A and B engage in a Cournot competition. The market clearing price of the service in period j is $p_j = a - Q_j^A - Q_j^B$, where a indicates the market demand and Q_j^i is the number of providers that participate on platform i in period j. The market demand is assumed to be greater than the supply pool size (i.e., $a > N$) so that the market clearing price will always be positive. Note that the Cournot competition framework corresponds to the situation where providers are providing a commodity service. Furthermore, it is worth mentioning that unlike traditional firms who decide their own production quantities, gig economy platforms serve as intermediaries of the marketplace, where their supply quantities are an outcome of providers' participation decisions. Thus, our setting can alternatively be viewed as a N-player Cournot game played by the providers, where each provider decides his supply quantity of either one or zero, and they jointly determine the market price.

Platforms obtain revenues by charging commissions, and may incur bonus expenditure if they decide to offer bonuses to the providers. For each service completion in period j, the platform obtains revenue γp_j and the provider obtains wage $(1-\gamma)p_j$, where $\gamma > 0$ is the platform's commission rate.[1] The minimum market price that the platforms can charge a consumer is $p_{\min} = a - N$, and the minimum wage that the platforms can guarantee to a participating provider is $w_{\min} = (1-\gamma)(a-N)$. For analytical tractability, we assume that $\gamma \leq 2/3$. Note

[1]The commission rate is assumed to be exogenous in the model because in reality commission rates in gig economy marketplaces are usually subject to industry standards and the platforms have limited ability to set the commission rate freely.

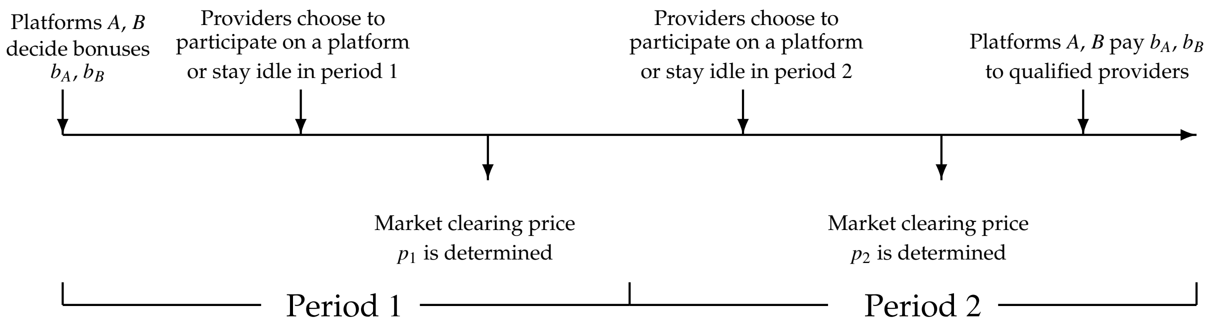
that this assumption is consistent with industry practice, as gig economy platforms typically do not charge very high commissions;[2] relaxing this assumption does not yield additional insights. To capture that the purpose of the bonuses is to drive consistent participation of providers, we assume that platform i fulfills a bonus in the amount of b_i to a provider if he participates on platform i in both periods.

The sequence of events is shown in Figure 2.2. At the beginning of period 1, platforms A and B decide their bonuses to offer the providers. A platform may choose not to offer a bonus by setting her bonus to zero. Then, providers choose a platform to participate on or not to participate on any platform in period 1, taking into account that they can only obtain a bonus by serving the same platform in period 2. Based on the number of providers that participate, the market price in period 1 is determined. Next, in period 2, providers make their participation decisions again. Providers who have participated in period 1 have three options: choose the same platform, switch to the other platform, or not participate in period 2. They will not obtain a bonus in the latter two cases. Providers who have not participated in period 1 decide to either participate on a platform or remain idle in period 2. Based on the number of providers that participate, the market price in period 2 is determined. Finally, platforms pay bonuses to those providers who qualify, and the surplus of all stakeholders is determined.

We assume that providers' platform preferences remain static between the two periods, i.e., in period 2 providers have the same locations on the Hotelling line as in period 1. This would correspond to situations where providers' pref-

[2]For example, the commission rate is 20% – 30% for ride-hailing platforms [88] and 15% – 20% for home service platforms (see `https://support.taskrabbit.com/hc/en-us/articles/204411610-What-is-the-TaskRabbit-Service-Fee-` and `https://www.fastcompany.com/3042248/the-gig-economy-wont-last-because-its-being-sued-to-death`).

Figure 2.2: Sequence of Events



erences between the platforms are mainly driven by long-term features of the platforms (e.g., brand image or service positioning of the platforms). For example, VIPKID and 51Talk are two major Chinese online English education platforms. VIPKID features a more helpful and friendly management team for the independent teachers regarding their inquiries and class cancelations, while 51Talk features a more diverse student demographics due to its expansion in Philippine. Another example is Uber and Lyft. Uber is better known for its efficient ride dispatching algorithms, while Lyft is better known for its emphasis on social value [4]. When providers' preferences are mainly driven by such solid features of the platforms, they are unlikely to change in the short term. However, one might also expect that in other situations, providers' preferences between the platforms may be also driven by factors that change over time. In our setting, this would correspond to the case where providers' locations on the Hotelling line are reassigned in period 2. We explore such an alternative model in Section 2.6.

A provider's utility is determined by the wage he receives from the platform, his participation cost, and the bonus offered by the platform if applicable in period 2. We normalize a provider's utility in a period to zero if he stays idle in that period. A provider located at x on the Hotelling line obtains utility $u^i_j(x)$ if he participates on platform i in period j. The provider utilities are as follows:

$$\begin{aligned} u_1^A(x) &= (1-\gamma)p_1 - tx, & u_2^A(x) &= (1-\gamma)p_2 - tx + b_A \cdot \mathbb{1}_{x\in[0,x_1^A]}, \\ u_1^B(x) &= (1-\gamma)p_1 - t(1-x), & u_2^B(x) &= (1-\gamma)p_2 - t(1-x) + b_B \cdot \mathbb{1}_{x\in[x_1^B,1]}, \end{aligned} \tag{2.1}$$

where x_1^A and x_1^B are the resulting segmentation points of the Hotelling line in period 1 based on the providers' participation decisions in period 1. The providers located in $[0, x_1^A]$, $[x_1^B, 1]$, and (x_1^A, x_1^B) choose to participate on platform A, platform B, and stay idle in period 1, respectively. As is common with Hotelling models, providers who are sufficiently close to a platform will participate on that platform in equilibrium, and this can be easily verified in our setting for both periods. In Eq. (2.1), the indicator functions $\mathbb{1}_{x\in[0,x_1^A]}$ and $\mathbb{1}_{x\in[x_1^B,1]}$ ensure that providers only obtain bonuses in period 2 if they participate on the same platform in both periods. Once the providers have made their participation decisions in period 2, the segmentation points of the Hotelling line in period 2 (i.e., x_2^A and x_2^B) is determined. In period j, the supply pool is fully utilized (i.e., all providers in the supply pool participate) if $x_j^A = x_j^B$ and partially utilized (i.e., some providers in the supply pool stay idle) if $x_j^A < x_j^B$. Providers are forward-looking utility maximizers. They make their participation decisions in period 1 to maximize their total utility from both periods. Their participation decisions in period 2 maximize their utility in period 2 given their decisions in period 1.

The profit functions of the two platforms are shown in Eq. (2.2). Each platform's profit function is given by the commission revenues in both periods minus the bonus expenditure:

$$\begin{aligned} \Pi_A(b_A, b_B) &= \gamma p_1 x_1^A N + \gamma p_2 x_2^A N - b_A \min(x_1^A, x_2^A) N, \\ \Pi_B(b_A, b_B) &= \gamma p_1 (1 - x_1^B) N + \gamma p_2 (1 - x_2^B) N - b_B \left(1 - \max(x_1^B, x_2^B)\right) N. \end{aligned} \tag{2.2}$$

Note that in the platforms' profit functions, the market prices, p_1 and p_2, as well as the provider segmentation points, x_1^A, x_1^B, x_2^A, and x_2^B, are all given by the equilibrium outcome of the game, so they are all functions of the platforms' bonuses b_A and b_B (For simplicity of notations, in Eq. (2.2) we suppressed b_A and b_B from the equilibrium market prices and segmentation points).

2.4 Competitive Bonus Equilibrium

In this section, we analyze the competitive bonus game between the platforms. In this case, each platform's objective is to choose a bonus to maximize her own profit as a response to her competing platform's bonus strategy, subject to the constraint that providers are willing to participate. Thus, in the competitive bonus game, platforms A and B will choose bonuses b_A^* and b_B^* such that

$$\begin{aligned} & b_A^* = \arg\max_{b_A \geq 0} \Pi_A(b_A, b_B^*),\ b_B^* = \arg\max_{b_B \geq 0} \Pi_B(b_A^*, b_B), \\ \text{s.t.}\quad & u_1^A(x) + u_2^A(x) \geq 0 \ \text{ for } x \leq x_1^A,\ u_2^A(x) \geq 0 \ \text{ for } x \leq x_2^A, \\ & u_1^B(x) + u_2^B(x) \geq 0 \ \text{ for } x \geq x_1^B,\ u_2^B(x) \geq 0 \ \text{ for } x \geq x_2^B, \end{aligned} \tag{2.3}$$

where $u_j^i(x)$ and Π_i are shown in Eq. (2.1) and (2.2), respectively. Note that the superscript $*$ represents the competitive equilibrium case. We derive the subgame perfect Nash equilibrium of the competitive bonus game. Using backward induction, we first characterize the market price and providers' participation decisions in period 2 and period 1 sequentially given the platforms' bonus decisions, and then derive the platforms' equilibrium bonuses.

Lemma 1 *When neither platform offers any bonus (i.e., $b_A = b_B = 0$), the equilibrium market price p_j^0 and the resulting supply pool utilization in period $j \in \{1, 2\}$ are given by the following:*

(i) If $t \leq \hat{t} = 2(1-\gamma)(a-N)$, the market price is $p_j^0 = a - N$ and the number of

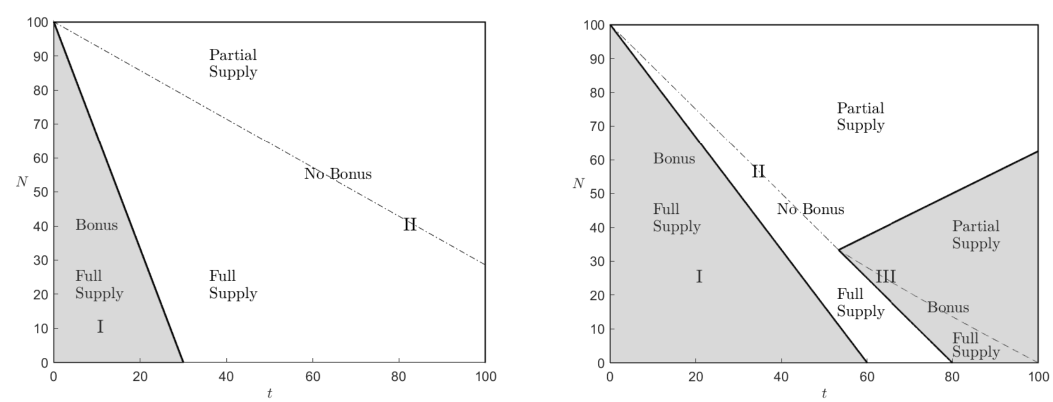
participating providers is $Q_j^{A0} = Q_j^{B0} = \frac{N}{2}$ *(i.e., the supply pool is fully utilized).*

(ii) If $t > \hat{t}$*, the market price is* $p_j^0 = \frac{at}{2(1-\gamma)N+t}$ *and the number of participating providers is* $Q_j^{A0} = Q_j^{B0} = \frac{(1-\gamma)aN}{2(1-\gamma)N+t} < \frac{N}{2}$ *(i.e., the supply pool is partially utilized).*

As a benchmark, we first present the equilibrium market price and supply pool utilization if neither platform offers any bonus (i.e., $b_A = b_B = 0$). Lemma 3 shows that the equilibrium depends on the providers' marginal participation cost t relative to the minimum wage $w_{\min}$ (recall that $w_{\min} = (1-\gamma)(a-N)$). When the marginal participation cost is small (i.e., $t \leq 2w_{\min}$), the equilibrium market price will induce all the providers to participate. In this case, the platforms may have an incentive to poach providers from their opponents. When the marginal participation cost is large (i.e., $t > 2w_{\min}$), the equilibrium market price will be higher so that only a proportion of providers will participate. In this case, the platforms serve as local monopolists on the Hotelling line. From Lemma 1, it is easy to see that as t increases, p_j^0 increases and Q_j^{i0} decreases. Recall that the marginal participation cost measures the providers' platform stickiness. Thus, Lemma 1 indicates that a lower platform stickiness corresponds to a higher supply pool utilization and a lower market price. We next derive the platforms' equilibrium bonuses.

Proposition 1 *The platforms offer bonuses in equilibrium when the service providers' marginal participation cost is sufficiently low, i.e.,* $t < \underline{t} = \gamma(a-N)$*, or when the marginal participation cost is sufficiently high, i.e.,* $t > \bar{t} = \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}\right)$*. Otherwise, the platforms do not offer any bonus in equilibrium.*

Proposition 1 shows that the platforms offer bonuses in equilibrium when the providers' marginal participation cost t is sufficiently low (i.e., providers' platform stickiness is sufficiently low) or sufficiently high (i.e., providers' platform stickiness is sufficiently high). On the other hand, when the marginal participation cost is moderate (i.e., providers' platform stickiness is moderate), neither platform offers any bonus. Figure 2.3 presents two examples to illustrate the competitive bonus equilibrium in the panel of marginal participation cost t and supply pool size N. In both examples, the market demand is $a = 100$ (which is also the upper bound for N). Figure 2.3(a) corresponds to a case where the commission rate is relatively low ($\gamma = 0.3$), and Figure 2.3(b) corresponds to a case where the commission rate is relatively high ($\gamma = 0.6$). We label region I, II, and III for the case of low, moderate, and high platform stickiness, respectively, which correspond to the three cases characterized in Proposition 1. We shade the regions where platforms offer bonuses in equilibrium. As Figure 2.3 shows, regions I and III correspond to the cases where platforms offer bonuses in equilibrium. By examining the market conditions in these two cases, we find that the incentive for platforms to offer bonuses is very different. We now discuss each case in detail.



(a) $\gamma = 0.3$ (b) $\gamma = 0.6$

Figure 2.3: Illustration of Competitive Bonus Equilibrium ($a = 100$)

- **Low Platform Stickiness**

Proposition 2 *When the marginal participation cost is sufficiently low, i.e., $t < \underline{t}$, the platforms offer bonuses $b_A^* = b_B^* = 2(\gamma(a - N) - t)$ in equilibrium, and obtain profits $\Pi_A^* = \Pi_B^* = Nt$ which decrease as t decreases and $\Pi_A^* = \Pi_B^* \to 0$ when $t \to 0$. Moreover, the number of participating providers in period $j \in \{1, 2\}$ is $Q_j^{A*} = Q_j^{B*} = \frac{N}{2}$.*

Proposition 2 characterizes the platforms' competitive bonus equilibrium as well as their equilibrium profits when providers' platform stickiness is low (i.e., when $t < \underline{t}$, which corresponds to region I in Figure 2.3). Notice that for $\gamma \leq 2/3$, $t \leq 2(1 - \gamma)(a - N)$ is implied by $t < \underline{t} = \gamma(a - N)$. Thus, according to Lemma 3, all providers would participate and the supply pool would be fully utilized had the platforms not offered any bonus. By offering bonuses in this case, the platforms do not expand their market coverage and the market price remains unchanged. However, since the providers' stickiness to their preferred platforms is low, the providers who are close to the middle point of the Hotelling line only weakly prefer one platform over the other, and they can be easily swung by a bonus offered by the less preferred platform. Thus, each platform has an incentive to offer a bonus to poach those providers from her competing platform. Anticipating the poaching incentive of the opponent, both platforms are forced to offer bonuses to lock in their existing providers. Thus, in equilibrium, both platforms will offer bonuses to retain labor supply and maintain market coverage. This means that neither platform poaches successfully in equilibrium. Since platforms obtain the same revenue but incur more bonus expenditure, their equilibrium profits will decline. Therefore, in this case, the platforms are trapped in a prisoner's dilemma where their profits will be improved if they both agree to offer smaller bonuses. Moreover, notice that $\underline{t} = \gamma(a - N)$ is the unit revenue that is earned by the platforms in this case. Thus, when $\gamma(a - N)$ is higher (e.g., the supply pool is smaller), the platforms' poaching incentive is more likely to be triggered.

Recall that a provider located at x on the Hotelling line incurs participation costs tx and $t(1-x)$ by participating on the two platforms, respectively. Thus, the provider incurs a disutility $t|1-2x|$ by switching to his less preferred platform. This indicates that when the providers' platform stickiness decreases (i.e., t decreases), the platforms' incentive to poach providers is more easily triggered. Proposition 2 shows that as the providers' platform stickiness decreases (i.e., t decreases), the platforms need to offer a greater amount of bonus in equilibrium to lock in their existing providers from being poached by their opponents. As a result of the increasingly fierce bonus competition between platforms, the platforms' equilibrium profits decrease. Further, as t approaches zero, the equilibrium bonus will be as high as $2\gamma(a-N)$, which is the same amount as the platform's share of revenue for a unit of completed service. This means that the platforms will transfer all their earnings to the providers through offering bonuses when the providers exhibit no platform stickiness. As formally shown in Proposition 2, the platforms' equilibrium profits will diminish to zero as t approaches zero.

Therefore, when providers' platform stickiness is low, platforms' bonuses serve as a means to lock in their existing providers. However, this comes with a significant cost for the platforms. Our results indicate that the bonus strategy is unsustainable to platform profits in this case. This finding can explain the heatedly discussed industry phenomenon that gig economy platforms in highly competitive markets are operating unprofitably. For example, as the two major platforms in the U.S. ride-hailing market, Uber and Lyft are facing extremely intense competition, and they are offering extensive bonuses to retain drivers. In

2017, Uber's full-year net loss was \$4.5 billion, and Lyft's full-year net loss was \$688 million [43, 165]. In such markets, by using the bonus strategy, platforms will very likely trap themselves in a prisoner's dilemma where they will end up paying too much in bonuses to the providers and making their business model unsustainable.

- **High Platform Stickiness**

Proposition 3 *When the marginal participation cost is sufficiently high, i.e., $t > \bar{t}$, the platforms offer bonuses* $b_A^* = b_B^* = \begin{cases} t - 2(1-\gamma)(a-N) & \text{if } t \leq \tilde{t} \\ \frac{2\left((2\gamma-1)t-2(1-\gamma)^2 N\right)at}{2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2} & \text{if } t > \tilde{t} \end{cases}$ *in equilibrium, and obtain profits* $\Pi_A^* = \Pi_B^* = \begin{cases} N((a-N) - \frac{t}{2}) & \text{if } t \leq \tilde{t} \\ \frac{2a^2Nt\left(t+(1-\gamma)N\right)\left(t+(2-\gamma)N\right)}{\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)^2} & \text{if } t > \tilde{t} \end{cases}$ *which are always strictly positive, where* $\tilde{t} = \frac{\left(2a-5N+2\gamma N+\sqrt{4\gamma^2N^2-4\gamma N(2a+N)+4a^2-4aN+9N^2}\right)}{4}$. *Moreover, the number of participating providers in period* $j \in \{1,2\}$ *is* $Q_j^{A*} = Q_j^{B*} = \begin{cases} \frac{N}{2} & \text{if } t \leq \tilde{t} \\ \frac{4(1-\gamma)aN+(b_A^*+b_B^*)N}{8(1-\gamma)N+4t} < \frac{N}{2} & \text{if } t > \tilde{t} \end{cases}$.

We have identified that the platforms offer bonuses in equilibrium when the providers' platform stickiness is sufficiently low. One might think that the platforms do not offer any bonus otherwise. However, we identify that there is another case for the platforms to offer bonuses in equilibrium. This case requires two conditions. First, this case occurs when the providers' platform stickiness is sufficiently high (i.e., when $t > \bar{t}$, which corresponds to region III in Figure 2.3(b)). By comparing Proposition 3 and Lemma 1, we can see that in this case, the supply pool would be only partially utilized had the platforms not offered any bonuses. The platforms serve as local monopolists on the Hotelling line, each covering a segment of providers whose locations on the Hotelling line are close enough to the end point, and the providers located closely to the center of the Hotelling line do not participate on either platform. Moreover, as shown in Lemma 1, the number of providers that would serve either platform without any bonus (i.e., Q_j^{i0}) decreases when t increases. Thus, when the marginal participation cost t is sufficiently high, both platforms have an incentive to increase their market coverage through offering a bonus. Second, the occurrence of this case also requires that the commission rate to be sufficiently high. Notice that if $\gamma \leq 1/2$, $\bar{t} \to \infty$ so $t < \bar{t}$ always holds. This is also shown in Figure 2.3, where region III only emerges in Figure 2.3(b) which corresponds to a high commission rate, but not in Figure 2.3(a). Intuitively, when the industry commission rate is low, the platforms need to forgo a large share of the market price to the providers. Consequently, it would not be profitable for either platform to cover more providers. Under a relatively high industry commission rate, platforms may have an incentive to increase market coverage by offering bonuses.

Therefore, when the bonus offering is not driven by the platforms' incentive to poach existing providers from their competing platforms but rather the incentive to expand market coverage, the bonus serves as a subsidy to increase the participation of providers who would otherwise stay idle. When the marginal participation cost t is smaller, it is easier to induce more providers to participate by offering a bonus. As a result, it is more likely that the bonus strategy can allow the platforms to cover the entire supply pool. Proposition 3 formally shows that the bonus strategy can expand the market coverage to the entire supply pool if the marginal participation cost is not too high (i.e., $t \leq \tilde{t}$).[3] The sup-

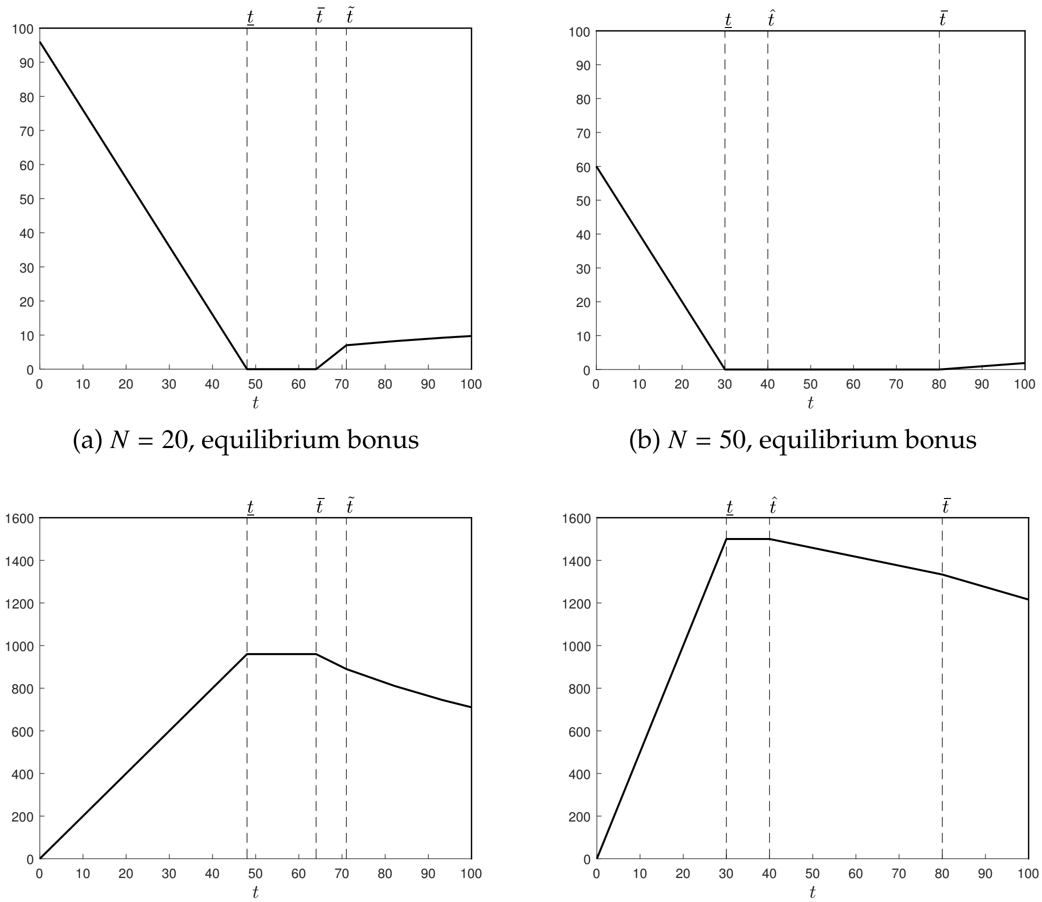
[3] In the Appendix, we show that $\tilde{t} > \bar{t}$ if $\frac{a}{N} > \frac{\gamma}{(2\gamma-1)^+}$. This indicates that in order to achieve full supply pool utilization with the bonus strategy, the supply pool cannot be too large compared to the market demand.

ply pool utilization under the bonus strategy is also depicted in Figure 2.3(b). Within region III, the bonus strategy enables the platforms to cover the entire supply pool below the dashed curve, whereas the supply pool is still partially utilized with the bonus strategy (so the platforms are still local monopolists) above the dashed curve.

In emerging marketplaces such as online education, a primary goal of platforms is to penetrate the market base. In this case, the bonus strategy can be an effective lever for market expansion. Because the bonus strategy is not mainly driven by the competition pressure, it is unlikely to lead the platforms to be unprofitable. Proposition 3 further shows that the equilibrium profit is guaranteed to be strictly positive, which is in clear contrast with Proposition 2.

As we have seen from the two bonus offering cases, the marginal participation cost t plays a dual role in our model. A small t corresponds to full utilization of the supply pool. In this case, t measures the providers' sensitivity of switching between platforms, and a smaller t indicates a stronger incentive of platforms to poach providers from their opponents. On the other hand, a large t corresponds to partial utilization of the supply pool. In this case, t measures the providers' sensitivity of participating on their preferred platforms, and a larger t indicates a stronger incentive of platforms to expand their market coverage. When t is moderate (i.e., $\underline{t} \leq t \leq \bar{t}$), the platforms do not have either incentive to offer bonuses, which corresponds to region II in Figure 2.3.

Figure 2.4 presents two examples to illustrate the equilibrium bonus and platform profit. The parameter values are consistent with Figure 2.3(b) (i.e., $a = 100$ and $\gamma = 0.6$). Figures 2.4(a) and 2.4(c) correspond to a relatively small supply pool ($N = 20$), and Figures 2.4(b) and 2.4(d) correspond to a relatively



(a) $N = 20$, equilibrium bonus

(b) $N = 50$, equilibrium bonus

(c) $N = 20$, equilibrium profit

(d) $N = 50$, equilibrium profit

Figure 2.4: Examples of Competitive Bonus Equilibrium and Platform Profit ($a = 100$, $\gamma = 0.6$)

large supply pool ($N = 50$). The annotated thresholds $\underline{t}$ and $\bar{t}$ correspond to the two bonus offering cases (i.e., low and high platform stickiness, respectively); $\hat{t}$ or $\tilde{t}$ is the threshold where the equilibrium supply pool utilization switches from full to partial. When the supply pool is relatively small, the equilibrium bonus strategy may enable the platforms to achieve full coverage of the supply pool in the high platform stickiness case. For example, in Figures 2.4(a) and 2.4(c), this occurs if $\bar{t} < t \leq \tilde{t}$. When the supply pool is relatively large, the platforms never achieve full coverage of the supply pool with the bonus strategy in the high platform stickiness case. For example, in Figures 2.4(b) and 2.4(d), $\tilde{t} < \bar{t}$

so the supply pool is always partially utilized for all $t > \bar{t}$. Moreover, a further examination of Figure 2.4 shows that the equilibrium bonus and platform profit exhibit different sensitivities with respect to the marginal participation cost t in the two bonus offering cases. In the low platform stickiness case (i.e., $t < \underline{t}$), the equilibrium bonus would increase substantially with a slight decrease in t, which would result in a considerable reduction in the equilibrium profit. This indicates that the bonus strategy can make the competitive environment extremely volatile in this case. On the other hand, in the high platform stickiness case (i.e., $t > \bar{t}$), the equilibrium bonus and platform profit are not sensitive in t, and the equilibrium bonus remains at a low level.

To summarize our main findings so far, when gig economy platforms do not coordinate on bonus offering, they will offer bonuses in equilibrium if the providers' platform stickiness is either sufficiently low or sufficiently high. We have also seen that in the former case, platforms' bonus strategy would trap the platforms in a prisoner's dilemma, and platform bonus competition intensifies and platform profits diminish as providers' platform stickiness decreases. Therefore, it is important to understand how the issue can be mitigated. A natural question that arises would be how coordination between platforms could affect their bonus offering decisions. Moreover, as platforms coordinate with each other, would social welfare be compromised? We analyze the bonus equilibrium under platform coordination and its implications on platform profit and social welfare in the next section.

2.5 Coordinated Bonus Equilibrium

In this section, we analyze the coordinated bonus game between the platforms. We first derive the bonus equilibrium when platforms coordinate in making bonus decisions. We then compare the coordinated bonus equilibrium to the competitive bonus equilibrium analyzed in Section 2.4 to derive insights regarding the impact of platform coordination on platforms' bonus strategies and profits. Finally, we investigate the impact of platform coordination on the surplus of other stakeholders of the gig economy (i.e., providers and consumers), as well as social welfare.

When platforms coordinate, they will decide their bonuses as if they were one platform whose objective is to maximize the total profit. The remaining stages of the game are the same as the competitive bonus game (see Figure 2.2), where the providers choose which platform to participate on or stay idle in each period, taking into consideration that they can obtain a bonus only if they participate on the same platforms in both periods. The platforms choose bonuses $b_A^\dagger$ and $b_B^\dagger$ such that

$$\begin{aligned} &(b_A^\dagger, b_B^\dagger) = \underset{b_A, b_B \geq 0}{\arg\max}\ \Pi_A(b_A, b_B) + \Pi_B(b_A, b_B), \\ \text{s.t.}\quad & u_1^A(x) + u_2^A(x) \geq 0 \ \text{ for } x \leq x_1^A,\ u_2^A(x) \geq 0 \ \text{ for } x \leq x_2^A, \\ & u_1^B(x) + u_2^B(x) \geq 0 \ \text{ for } x \geq x_1^B,\ u_2^B(x) \geq 0 \ \text{ for } x \geq x_2^B, \end{aligned} \tag{2.4}$$

where $u_j^i(x)$ and Π_i are shown in Eq. (2.1) and (2.2). Note that the superscript † represents the coordinated equilibrium case.

2.5.1 Implications on Bonus Strategy and Platform Profit

We use backward induction to first characterize the market price and providers' participation decisions in both periods given the platforms' bonus decisions, and then derive the coordinated bonus equilibrium. The coordinated bonus equilibrium is characterized in Proposition 4.

Proposition 4 *The platforms' coordinated bonus equilibrium is characterized by the following:*

(i) If $t \leq \max(\hat{t}, \frac{2(1-\gamma)N}{(2\gamma-1)^+})$*, the platforms do not offer any bonus, and obtain profits* $\Pi_A^\dagger = \Pi_B^\dagger = \begin{cases} \gamma N(a-N) & \text{if } t \leq \hat{t} \\ \frac{2\gamma(1-\gamma)a^2Nt}{\left(2(1-\gamma)N+t\right)^2} & \text{if } t > \hat{t} \end{cases}$*. Moreover, the number of participating providers in period* $j \in \{1,2\}$ *is* $Q_j^{A\dagger} = Q_j^{B\dagger} = \begin{cases} \frac{N}{2} & \text{if } t \leq \hat{t} \\ \frac{(1-\gamma)aN}{2(1-\gamma)N+t} < \frac{N}{2} & \text{if } t > \hat{t} \end{cases}$.

(ii) If $t > \max(\hat{t}, \frac{2(1-\gamma)N}{(2\gamma-1)^+})$*, the platforms offer bonuses* $b_A^\dagger = b_B^\dagger = \begin{cases} t - 2(1-\gamma)(a-N) & \text{if } t \leq a-2N \\ \frac{((2\gamma-1)t-2(1-\gamma)N)a}{2N+t} & \text{if } t > a-2N \end{cases}$*,and obtain profits* $\Pi_A^\dagger = \Pi_B^\dagger = \begin{cases} N((a-N)-\frac{t}{2}) & \text{if } t \leq a-2N \\ \frac{a^2N}{2(2N+t)} & \text{if } t > a-2N \end{cases}$*. Moreover, the number of participating providers in period* $j \in \{1,2\}$ *is* $Q_j^{A\dagger} = Q_j^{B\dagger} = \begin{cases} \frac{N}{2} & \text{if } t \leq a-2N \\ \frac{4(1-\gamma)aN+(b_A^\dagger+b_B^\dagger)N}{8(1-\gamma)N+4t} < \frac{N}{2} & \text{if } t > a-2N \end{cases}$.

Proposition 4 shows that when the two platforms coordinate, they will offer bonuses only when the providers' marginal participation cost is sufficiently high (i.e., $t > \max(\hat{t}, \frac{2(1-\gamma)N}{(2\gamma-1)^+})$). Figure 2.5 illustrates the coordinated bonus equilibrium using the same example of Figure 2.3(b). The shaded area corresponds to

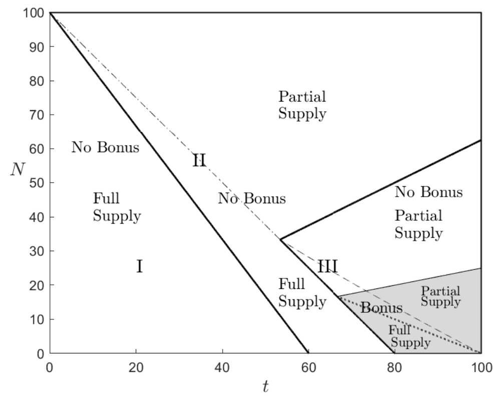


Figure 2.5: Comparison of Coordinated and Competitive Bonus Equilibria (a = 100, γ = 0.6)

the case where platforms offer bonuses in the coordinated equilibrium. To ease the comparison to the competitive bonus equilibrium, Figure 2.5 also shows the three regions corresponding to the three cases identified in the competitive bonus equilibrium. Different from the competitive equilibrium, platforms no longer choose to offer bonuses in the low platform stickiness case (i.e., region I) when they coordinate, and they only offer bonuses in the high platform stickiness case (i.e., region III). Therefore, when platforms coordinate, the bonus strategy is only used as a means to achieve market expansion. Moreover, as Figure 2.5 shows, the region where the platforms offer bonuses does not cover the entire region III, meaning that even in the high platform stickiness case, platform coordination would reduce bonus offering. This follows immediately from comparing the condition in Proposition 4(ii) to that in Proposition 2. Compared to the competitive equilibrium, it requires a higher marginal participation cost t (or a smaller supply pool size N) to trigger platforms' market expansion incentive under platform coordination. Proposition 4(ii) also reveals that by offering bonuses, the platforms can achieve full coverage of the supply pool if and only if the marginal participation cost is lower than a threshold (i.e., $t \leq a - 2N$).[4] In

[4]In the Appendix, we show that $a - 2N > \max(\hat{t}, \frac{2(1-\gamma)N}{(2\gamma-1)^+})$ if $\frac{a}{N} > \frac{2\gamma}{(2\gamma-1)^+}$. Thus, similar to the

Figure 2.5, the region below the dotted curve corresponds to the case where the bonus strategy enables the platforms to cover the entire supply pool, and the region above the dotted curve corresponds to the case where the supply pool is still partially utilized with the bonus strategy.

Proposition 5 *(i) The platforms' bonuses reduce under the coordinated equilibrium compared to the competitive equilibrium, i.e.,* $b_i^\dagger \leq b_i^*$. *Moreover,* $b_i^\dagger < b_i^*$ *when the marginal participation cost is sufficiently low, i.e.,* $t < \underline{t}$, *or when the marginal participation cost is sufficiently high, i.e.,* $t > \max(\bar{t}, a - 2N)$.

(ii) The platforms' profits improve under the coordinated equilibrium compared to the competitive equilibrium, i.e., $\Pi_i^\dagger \geq \Pi_i^*$, *where equality holds only when* $b_i^\dagger = b_i^*$. *Moreover,* $\Pi_i^\dagger > 0$ *as* $t \to 0$.

Proposition 5 examines the impact of platform coordination on the bonus amount and platform profit. As we have seen from Proposition 4, when the platforms coordinate, they will offer bonuses in fewer cases. Proposition 5(i) further shows that the platforms will never offer a larger bonus when they coordinate, and their profits will strictly increase if they offer a strictly smaller bonus under coordination.

Recall that when the providers' platform stickiness is sufficiently low (i.e., $t < \underline{t}$), the platforms use the bonus strategy to prevent existing providers from being poached by the competing platform. However, without coordination, the competition pressure will trap the platforms in a prisoner's dilemma where they are forced to offer too much bonus and lose profits overall. In this case, coordination will allow the platforms to overcome the competition pressure and

competitive equilibrium case, in order to achieve full supply pool utilization with the bonus strategy, the supply pool cannot be too large compared to the market demand.

avoid the prisoner's dilemma. Because the supply pool is already fully covered regardless of the bonus strategy, the coordinated equilibrium maintains zero bonus offering without hurting the platforms' profits. Moreover, as shown in Proposition 5(ii), under the coordinated bonus equilibrium, the platforms' profits no longer diminish to zero when providers exhibit no platform stickiness (i.e., $t \to 0$). This clear contrast with Proposition 2 underscores the important role that coordination plays in making the platforms' business sustainable in highly competitive environments.

On the other hand, when the providers' platform stickiness is sufficiently high (i.e., $t > \bar{t}$), the platforms use the bonus strategy to expand their market coverage. In this case, somewhat surprisingly, coordination may again reduce the bonus amount of platforms. This indicates that even when the platforms operate as local monopolists with each covering a non-overlapping segment of providers, they may still end up over-subsidizing the providers without coordination. This is because each platform faces the threat that her competing platform may want to expand market coverage more aggressively, which would decrease the market price and hurt the platform's profit if she does not respond by also expanding her market coverage with a larger bonus. Therefore, even though the bonus offering in this case is not driven by the poaching incentive, it could still result in a prisoner's dilemma for the platforms and lead to overproduction. By coordination, the platforms will offer smaller bonuses and expand their market coverage less aggressively.

2.5.2 Welfare Implications

So far, we have seen that coordination benefits the platforms by enabling them to offer smaller bonuses and improve profits. However, that may adversely impact other stakeholders of the gig economy, and could also raise the concern of whether social welfare would be compromised. Given the prosocial nature of the gig economy, policymakers should be particularly concerned about the welfare performance of the gig economy. We now examine the impact of platform coordination on provider surplus, consumer surplus, and social welfare of the gig economy.

In our setting, the provider surplus U_P is the total utility of all providers in the supply pool from both periods. Using the provider utility functions from Eq. (2.1), we have

$$\begin{aligned} U_P = {} & N\int_0^{x_1^A} ((1-\gamma)p_1 - tx)\,\mathrm{d}x + N\int_{x_1^B}^1 ((1-\gamma)p_1 - t(1-x))\,\mathrm{d}x \\ & + N\int_0^{x_2^A} \left((1-\gamma)p_2 - tx + b_A \cdot \mathbb{1}_{x\in[0,x_1^A]}\right)\mathrm{d}x + N\int_{x_2^B}^1 \left((1-\gamma)p_2 - t(1-x) + b_B \cdot \mathbb{1}_{x\in[x_1^B,1]}\right)\mathrm{d}x. \end{aligned} \tag{2.5}$$

The consumer surplus U_C is the accumulated difference between the consumers' willingness to pay for the service and the market price from both periods:

$$U_C = \int_0^{(x_1^A+1-x_1^B)N} (a - q_1 - p_1)\,\mathrm{d}q_1 + \int_0^{(x_2^A+1-x_2^B)N} (a - q_2 - p_2)\,\mathrm{d}q_2. \tag{2.6}$$

The social welfare W is the sum of provider surplus, consumer surplus, and platform profits, i.e., $W = \Pi_A + \Pi_B + U_P + U_C$. Proposition 6 summarizes the results comparing the provider surplus, consumer surplus, and social welfare between the competitive and coordinated equilibria.

Proposition 6 *When $b^{\dagger} < b^{*}$, the provider surplus reduces under the coordinated equilibrium compared to the competitive equilibrium, i.e., $U_P^{\dagger} < U_P^{*}$. The consumer surplus and the social welfare remain the same when the marginal participation cost is sufficiently low, and reduce under the coordinated equilibrium when the marginal participation cost is sufficiently high, i.e., $U_C^{\dagger} = U_C^{*}$ and $W^{\dagger} = W^{*}$ if $t < \underline{t}$, and $U_C^{\dagger} < U_C^{*}$ and $W^{\dagger} < W^{*}$ if $t > \max(\bar{t}, a - 2N)$.*

Proposition 6 shows that when platforms offer smaller bonuses under coordination, platform coordination hurts provider surplus. This is consistent with the notion that the bonus redistributes the surplus between the platforms and the providers. However, consumer surplus and social welfare may not decrease when platforms offer smaller bonuses under coordination. When the platform stickiness is sufficiently low (i.e., $t < \underline{t}$), the supply pool is fully covered regardless of whether a bonus is offered or not. In this case, a bonus only transfers surplus from the platforms to the providers, so platform coordination will not affect consumer surplus or social welfare. However, when the platform stickiness is sufficiently high (i.e., $t > \max(\bar{t}, a - 2N)$), a bonus will expand the platform's market coverage and hence improve consumer surplus and social welfare. In this case, since platforms offer smaller bonuses under coordination, platform coordination will hurt consumer surplus and social welfare. These findings speak directly to the fact that in a gig economy, the platforms serve as intermediaries rather than producers of the service/product. The social welfare in a gig economy crucially depends on what proportion of the market is covered.

The finding that platform coordination may not hurt social welfare when providers have low platform stickiness has important implications. When platforms are fighting a bonus war in a highly competitive environment, they may proactively seek collaboration with each other to reduce bonus offering and achieve Pareto improvement. Our results provide a theoretical ground for collaboration and even merger of intensely competing platforms. For example, in February 2015, ride-hailing companies Didi Dache (backed by Tencent Holdings Limited) and Kuaidi Dache (backed by Alibaba Group) merged into one company Didi Chuxing [164]. Before the merger, the market was highly competitive[5] and the two companies were suffering mounting losses in their effort to secure market share. Our findings predict that after the merger, the new platform will reduce bonus offering and the provider surplus will decline accordingly. However, the consumer surplus and social welfare may not necessarily suffer as a result of the merger. On the other hand, our findings for the high platform stickiness case indicate that policymakers should be more cautious about platform coordination/merger in emerging markets. In this case, platform coordination on the bonus strategy will increase platform profits at the expense of hurting provider surplus and consumer surplus, as well as social welfare.

2.6 Independent Provider Preferences Across Periods

In our main model, we have considered a setting where the providers' preferences between the platforms do not change over time (i.e. the providers' locations on the Hotelling line remain the same across different periods). This would correspond to situations where the providers' preferences are primarily driven by features that form the brand images of the platforms. Such brand-related factors are unlikely to change in the short term, leading to a fixed provider pref-

[5] A study in December 2013 by Analysis International estimated Didi Dache to hold approximately 55% of the smartphone-based taxi-hailing market, with Kuaidi Dache holding nearly all of the rest of the market share [149].

erence structure. In this section, we extend our main model to consider an alternative setting where the providers' preferences between the platforms are mainly driven by factors that change over time. In particular, a freelancer may prefer a platform in a certain period because the task fits his skill set better or aligns better with his personal tastes. For example, a freelance dog walker may prefer a golden retriever over a chihuahua, and a chihuahua over a bulldog. In one period, the task on platform A is a golden retriever and on platform B is a chihuahua, and hence he would prefer platform A. In another period, the task on platform A is a bulldog and on platform B is a chihuahua, and hence he would prefer platform B. Similarly, in the ride-hailing market, drivers may be inclined to accept rides with closer pickup locations even though such rides may not always be offered on the same platform. To capture the impact of such random factors on the providers' platform preferences, we follow the Hotelling literature (e.g., 104) and assume that the providers' locations on the Hotelling line will be randomly and independently reassigned in period 2 (their locations still follow a uniform distribution). We now explore this case.

We adopt Hotelling line $z \in [0, 1]$ to track the providers' locations in period 1. Due to the reshuffling of the providers' locations, we further adopt Hotelling lines α, β, $\eta \in [0, 1]$ to track the providers' locations in period 2 for those who have participated on platform A, platform B, and stayed idle in period 1, respectively. Our equilibrium analysis first derives the providers' participation decisions in each period, which are characterized by the segmentation points on all Hotelling lines, i.e., z_A and z_B, α_A and α_B, β_A and β_B, η_A and η_B. In period 1, the providers located at $z \in [0, z_A]$ and $z \in [1 - z_B, 1]$ will participate on platforms A and B, respectively. In period 2, the providers located at $\alpha \in [0, \alpha_A]$, $\beta \in [0, \beta_A]$, and $\eta \in [0, \eta_A]$ will participate on platform A, and the providers located at $\beta \in [\alpha, 1]$, $\beta \in [\beta_B, 1]$, and $\eta \in [\eta_B, 1]$ will participate on platform B. Thus, α_A and $1-\beta_B$ capture the retention rate of platforms A and B, respectively. For ease of exposition, we also denote $\bar{z}_i = 1 - z_i$, $\bar{\alpha}_i = 1 - \alpha_i$, $\bar{\beta}_i = 1 - \beta_i$, $\bar{\eta}_i = 1 - \eta_i$ for $i \in \{A, B\}$. The utilization of the supply pool in period 2 is $\rho = z_A(\alpha_A + \bar{\alpha}_B) + (z_B - z_A)(\eta_A + \bar{\eta}_B) + \bar{z}_B(\beta_A + \bar{\beta}_B)$.

Same as our main model, providers are forward-looking utility maximizers. In period 2, providers choose a platform to participate on or stay idle to maximize their second-period utility based on their new locations on the Hotelling line, taking into consideration that they will lose the bonus if they do not participate on the same platform as in period 1. In period 1, providers make their participation decisions to maximize their expected total utility from both periods. Provider z's expected total utility is given by Eq. (2.7) if he participates on platform A in period 1 (i.e., $z \leq z_A$), Eq. (2.8) if he participates on platform B in period 1 (i.e., $z \geq z_B$), and Eq. (2.9) if he stays idle in period 1 (i.e., $z_A < z < z_B$):

$$u^A(z) = (1-\gamma)p_1 - tz + \int_0^{\alpha_A} \left((1-\gamma)p_2 - t\alpha + b_A\right) \mathrm{d}\alpha + \int_{\alpha_B}^1 \left((1-\gamma)p_2 - t(1-\alpha)\right) \mathrm{d}\alpha, \tag{2.7}$$

$$u^B(z) = (1-\gamma)p_1 - t(1-z) + \int_0^{\beta_A} \left((1-\gamma)p_2 - t\beta\right) \mathrm{d}\beta + \int_{\beta_B}^1 \left((1-\gamma)p_2 - t(1-\beta) + b_B\right) \mathrm{d}\beta, \tag{2.8}$$

$$u^0(z) = \int_0^{\eta_A} \left((1-\gamma)p_2 - t\eta\right) \mathrm{d}\eta + \int_{\eta_B}^1 \left((1-\gamma)p_2 - t(1-\beta)\right) \mathrm{d}\eta. \tag{2.9}$$

We first characterize the providers' participation decisions as a function of platforms' bonus decisions. We then analyze the platforms' bonus decisions. The profit functions of the two platforms are shown in Eq. (2.10):

$$\begin{aligned} \Pi_A(b_A, b_B) &= \gamma p_1 z_A N + \gamma p_2(z_A\alpha_A + \bar{z}_B\beta_A + (z_B - z_A)\eta_A)N - b_A z_A \alpha_A N, \\ \Pi_B(b_A, b_B) &= \gamma p_1 \bar{z}_B N + \gamma p_2(z_A\bar{\alpha}_B + \bar{z}_B\bar{\beta}_B + (z_B - z_A)\bar{\eta}_B N - b_B \bar{z}_B \bar{\beta}_B N. \end{aligned} \tag{2.10}$$

Consistent with our main model, we consider both competitive and coordinated bonus equilibria. Moreover, we use consistent measures for provider surplus, consumer surplus, and social welfare as our main model. The provider surplus U_P and consumer surplus U_C are shown in Eq. (2.11):

$$\begin{aligned} U_P &= N\left(\int_0^{z_A} u^A(z)dz + \int_{z_A}^{z_B} u^0(z)dz + \int_{z_B}^1 u^B(z)dz\right), \\ U_C &= \int_0^{(z_A+1-z_B)N}(a - q_1 - p_1)dq_1 + \int_0^{\rho N}(a - q_2 - p_2)dq_2. \end{aligned} \tag{2.11}$$

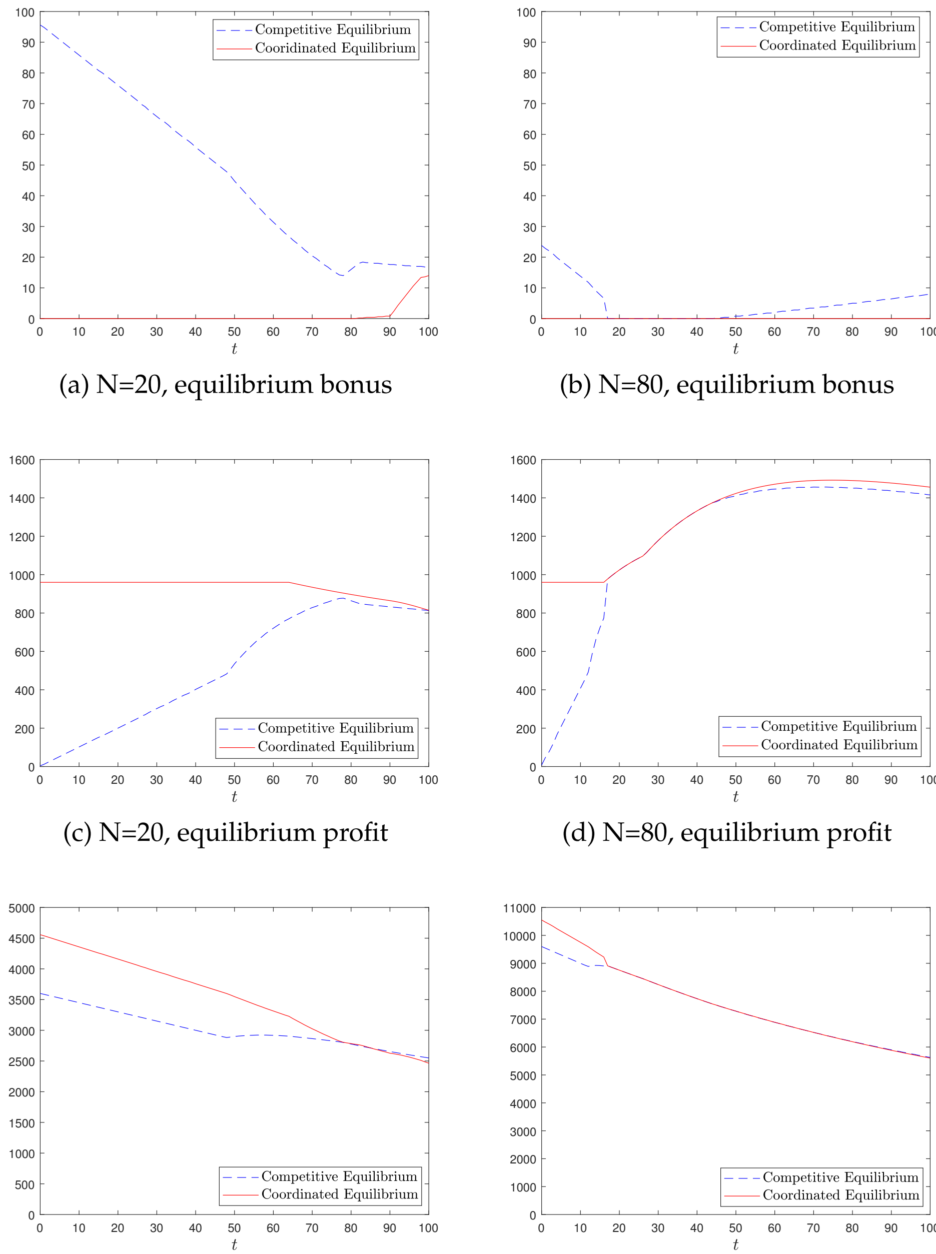
The social welfare is $W = \Pi_A + \Pi_B + U_P + U_C$. Under the setting of independent provider preferences, we are able to obtain analytical results for the case of low platform stickiness. These results are summarized in Proposition 7. The remaining case is not analytically tractable due to the complexity of the problem (the platforms' profit is a quartic function of the bonus). We thus rely on numerical analysis for the remaining case.

Proposition 7 *The following holds when the marginal participation cost is sufficiently low, i.e.,* $t \leq \underline{t}$*:*

(i) The platforms offer bonuses in the competitive equilibrium, i.e., $b_A^* = b_B^* > 0$*, and do not offer any bonus in the coordinated equilibrium, i.e.,* $b_A^\dagger = b_B^\dagger = 0$*. Under both equilibria, the supply pool is fully utilized in both periods with* $z_A = z_B = \frac{1}{2}$*,* $\alpha_A = \alpha_B = \min(\frac{1}{2} + \frac{b_A}{2t}, 1), \beta_A = \beta_B = \max(\frac{1}{2} - \frac{b_B}{2t}, 0)$*.*

(ii) The platforms' profits strictly improve under the coordinated equilibrium compared to the competitive equilibrium, i.e., $\Pi_i^\dagger > \Pi_i^*$*. Moreover, the provider surplus reduces, i.e.,* $U_P^\dagger < U_P^*$*, the consumer surplus remains the same, i.e.,* $U_C^\dagger = U_C^*$*, and the social welfare improves, i.e.,* $W^\dagger > W^*$*.*

Proposition 7(i) shows that when the providers have low platform stickiness (i.e., $t < \underline{t}$), the platforms are forced to offer bonuses in the competitive equilibrium, whereas if they coordinate, they should not offer any bonus. Figure 2.6



(a) N=20, equilibrium bonus (b) N=80, equilibrium bonus

(c) N=20, equilibrium profit (d) N=80, equilibrium profit

(e) N=20, social welfare (f) N=80, social welfare

Figure 2.6: Equilibrium Bonus, Platform Profit, and Social Welfare under Independent Preferences ($a = 100, \gamma = 0.6$)

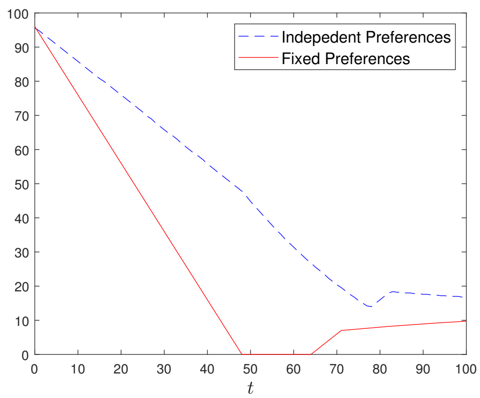
presents two sets of numerical examples under the independent preference setting. The left panel corresponds to a small supply pool ($N = 20$) and the right panel corresponds to a large supply pool ($N = 80$). Note that $\gamma(a - N) = 48$ when

$N = 20$ and $\gamma(a - N) = 12$ when $N = 80$, hence Proposition 7 corresponds to the region of $t \leq 48$ in the left panel and $t \leq 12$ in the right panel. As Figure 2.6 shows, bonus offering under the competitive equilibrium hurts the platforms' profits when providers' platform stickiness is low. Further, the platforms' profits in the competitive equilibrium will diminish to zero when providers exhibit no platform stickiness (i.e., $t \to 0$), which suggests that the bonus strategy is unsustainable for platform profit in this case. These results are consistent with those under the fixed preference setting.

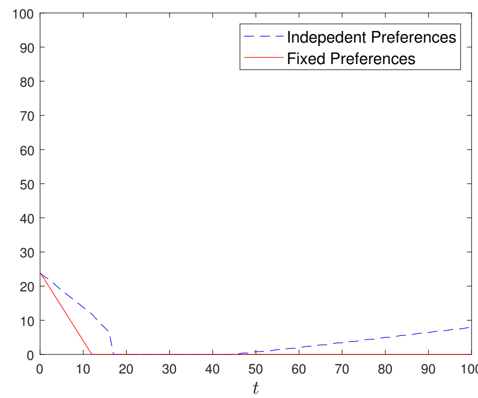
Proposition 7(ii) shows that the provider surplus will decrease while the consumer surplus will remain the same if platforms coordinate on bonus offering. Thus, the bonus strategy transfers surplus from the platforms to the providers, which is consistent with our previous findings with fixed provider preferences. However, somewhat surprisingly, we find that when providers' platform stickiness is low, the social welfare becomes higher if the platforms coordinate on bonus offering. This is also clearly shown in Figure 2.6. Thus, if the providers' platform preferences change over time, the bonus strategy can result in a friction to the social welfare when providers' platform stickiness is low.

Recall that if the providers' platform preferences remain fixed over time, the bonus strategy serves as a lever to lock in the platforms' existing providers when the providers' platform stickiness is low. If the providers' platform preferences change over time, the lock-in incentive of bonus offering becomes more nuanced. In this case, the platforms compete in period 2 to lock in not only the providers whose preferred platforms remain the same, but also the providers whose preferred platforms have changed. Without bonus offering, each platform can only retain half of her existing providers. Specifically, the providers who participate on platform A in period 1 (i.e., $z \leq 1/2$) will form the α-line in period 2, and only those at $\alpha \leq 1/2$ will continue to participate on platform A in period 2 without her bonus. For the providers whose preferred platforms remain the same (e.g., $\alpha \leq 1/2$ for platform A), the bonus will prevent them from being poached by the competing platform, which is similar to the fixed preference setting. Moreover, the bonus can also prevent a proportion of the providers whose preferred platforms change in period 2 from switching to a different platform. As indicated by Proposition 7(i), platform A can retain an additional $b_A^*/(2t)$ proportion of providers under the competitive bonus equilibrium, who would otherwise switch to platform B in period 2 because their preferences have changed significantly. The reduction in social welfare in the competitive equilibrium compared to the coordinated equilibrium is caused by the retention of this additional proportion of providers in period 2. These providers incur extra participation costs by serving their less preferred platforms in period 2, which results in a friction to the social welfare. When the platforms coordinate, such friction can be eliminated because platforms will not offer any bonus. This finding reinforces our previous insight that in highly competitive markets, platforms' coordination on the bonus strategy improves their profits without hurting the social welfare. Moreover, platform coordination may in fact improve the social welfare if the providers' preferences between platforms can change over time.

Next, consider the case where the providers' platform stickiness is not low. Figure 2.6 shows that when the platform stickiness is sufficiently high (i.e., t is sufficiently large), the platforms offer smaller bonuses in the coordinated equilibrium than in the competitive equilibrium. The platforms achieve higher profits under coordination but social welfare decreases. These results are consistent with the fixed preference setting. Additionally, Figure 2.6 indicates that when platform stickiness is high, the impact of platform coordination on platform profits and social welfare is not substantial. This could alleviate the policymakers' concerns about platform coordination in emerging markets.



(a) N=20, equilibrium bonus



(b) N=80, equilibrium bonus

Figure 2.7: Comparison of Competitive Equilibrium Bonuses under Independent and Fixed Preferences ($a = 100, \gamma = 0.6$)

Finally, note that the independent preference structure introduces another dynamic for providers to switch between platforms. Thus, the providers are less sticky to their preferred platforms compared to the case of fixed preference structure, and the platforms may have stronger incentives to offer bonuses, which exacerbates platforms' bonus competition. In Figure 2.7, we provide examples to compare the competitive bonus equilibrium between the two preference structures. As shown in Figure 2.7, the platforms offer higher bonuses if provider preferences are independent across time. Moreover, the impact of independent preferences on bonus can be more substantial when the supply pool is small (i.e., Figure 2.7(a)).

2.7 Conclusion

The self-scheduling feature of independent contractors creates a unique challenge for gig economy platforms in competing for labor supply. One strategy that is commonly used by gig economy platforms to drive labor supply is to offer a bonus contingent on the consistent participation of service providers. However, this bonus strategy has been viewed as controversial in industry discussions because platforms in certain industries (e.g., ride-hailing) have lost profit and been suffering steep losses by over-subsidizing providers. In this paper, we study the bonus competition between gig economy platforms. We find that platforms will offer bonuses in two cases: 1) when the service providers' platform stickiness is sufficiently low, or 2) when the service providers' platform stickiness and the commission rate of platforms are both sufficiently high. However, the driving forces of the platforms' bonus offering as well as its managerial and policy implications are different in these two cases.

In markets with low platform stickiness, platforms offer bonuses to prevent their opponents from poaching their existing service providers. We find that platforms will be trapped in a prisoner's dilemma by the bonus strategy and end up paying too much in bonuses to service providers. Moreover, as providers' platform stickiness keeps lowering, platforms will keep raising their bonuses which leads to a bonus war that is unsustainable to platforms' profits. Coordination can thus help the platforms to regain profitability. Moreover, we find that platform coordination on bonus offering in this case will not hurt social welfare, and may in fact improve social welfare if service providers' platform preferences can change over time. On the other hand, in markets with high platform stickiness, platforms offer bonuses to expand their own market coverage. In this case, platforms' bonus strategy can help improve their profits, but competition may still force the platforms to offer higher bonuses to service providers compared to if they coordinate on bonus offering. In this case, platform coordination will improve platforms' profits but hurt social welfare.

Therefore, when bonus offering is driven by intense competition pressure arising from the low platform stickiness of providers (e.g., ride-hailing), platform coordination could be a win-win solution to both platforms and policymakers. On the other hand, when service providers' platform stickiness is sufficiently high such that platforms do not compete directly with each other (as is likely the case with emerging marketplaces such as online education), although the bonus strategy can still be a profitable lever for the platforms, policymakers should be cautious about platform coordination on bonus offering.

Our work is a first step toward understanding the implications of bonus strategies in the gig economy. There are other important issues in this domain that future research could study. First, we have focused on a type of bonus that is offered to the service providers and contingent on their consistent participation. Gig economy platforms have also used other types of bonus strategies in practice, such as bonuses offered to the customers and bonuses that are offered on a dynamic basis. It would be interesting to understand when a certain type of bonus strategy is more profitable and how platforms should manage multiple bonus strategies at the same time. Second, platforms may be able to improve profit further by offering different bonuses to different participants. How should the platforms design such a personalized bonus scheme, and how would that impact social welfare? Finally, our findings can also generate hypotheses for empirical testing. For example, future empirical research can investigate the causal impact of bonus offering on platform profit and social welfare under different market conditions.

2.8 Appendix

2.8.1 Proofs of Results for Primary Model

2.8.1.1 Proof of Lemma 1

We first show that in equilibrium the segmentation points satisfy $x_1^A = x_2^A$ and $x_1^B = x_2^B$. Then we derive the equilibrium market segmentation points and market prices under an arbitrary bonus scheme (b_A, b_B), from which we can obtain the no-bonus equilibrium as a special case where $b_A = b_B = 0$.

We show $x_1^A = x_2^A$ and $x_1^B = x_2^B$ through ruling out six alternative cases: (1) $x_2^A \leq x_1^A$ and $x_2^B < x_1^B$; (2) $x_2^A > x_1^A$ and $x_2^B \geq x_1^B$; (3) $x_2^A > x_1^A$ and $x_2^B < x_1^B$; (4) $x_2^A < x_1^A$ and $x_2^B \leq x_1^B$; (5) $x_2^A \geq x_1^A$ and $x_2^B > x_1^B$; (6) $x_2^A < x_1^A$ and $x_2^B > x_1^B$. For Case (1)-(3), more providers participate in period 2 than period 1, and the market price is lower in period 2, i.e. $x_2^B - x_2^A < x_1^B - x_1^A$ and $p_2 < p_1$. For Case (4)-(6), more providers participate in period 1 than period 2, and the market price is higher in period 2, i.e. $x_2^B - x_2^A > x_1^B - x_1^A$ and $p_2 > p_1$.

In Case (1), for a provider at $x \in (\max(x_1^A, x_2^B), x_1^B)$, he stays idle in period 1 and participates on platform B in period 2, and thus he obtains utility $u_1 = 0$ in period 1 and $u_2 = (1-\gamma)p_2 - t(1-x)$ in period 2. Since he participates in period 2, $u_2 \geq 0$ holds. Suppose that this provider deviates to participate on platform B in both periods, he would obtain utility $\tilde{u_1} = (1-\gamma)p_1 - t(1-x)$ in period 1 and $\tilde{u_2} = (1-\gamma)p_2 - t(1-x) + b_B$ in period 2. Since $p_2 < p_1$ and $b_B \geq 0$, we obtain that $\tilde{u_1} > u_2 \geq 0$, $\tilde{u_2} \geq u_2$, and hence his total utility $\tilde{u} > u$. Thus, Case (1) cannot occur in equilibrium. Similarly, a provider located at $x \in (x_1^A, \min(x_2^A, x_1^B))$ in Case (2), $x \in (x_1^A, x_2^A)$ and $x \in (x_2^B, x_1^B)$ in Case (3), $x \in (x_2^A, \min(x_1^A, x_2^B))$ in Case (4), $x \in (\max(x_1^B, x_2^A), x_2^B)$ in Case (5), $x \in (x_2^A, x_1^A)$ and $x \in (x_1^B, x_2^B)$ in Case (6) have incentives to deviate. Thus, $x_1^A = x_2^A$, $x_1^B = x_2^B$, and $p_1 = p_2$ must hold in equilibrium.

Hereafter, we denote $x_A = x_1^A = x_2^A$ and $x_B = x_1^B = x_2^B$. In equilibrium, the supply pool can be fully utilized, i.e., $x_A = x_B = \bar{x}$, or can be partially utilized, i.e., $x_A < x_B$. We analyze the two cases as follows. (i) Suppose $x_A = x_B = \bar{x}$. Then $p_1 = p_2 = a - N$. For a provider at $\bar{x}$, his total utility is $u_A = 2(1-\gamma)(a-N) - 2t\bar{x} + b_A$ if he participates on platform A, or $u_B = 2(1-\gamma)(a-N) - 2t(1-\bar{x}) + b_B$ if he participates on platform B. First, he should obtain the same utility from either choice, i.e., $u_A = u_B$, and hence $\bar{x} = \frac{1}{2} + \frac{b_A - b_B}{4t}$. Second, the provider's utility at $\bar{x}$ should be higher than his utility when he stays idle, i.e., $u(\bar{x}) = 2(1-\gamma)(a-N) + \frac{b_A + b_B}{2} - t \geq 0$, and hence $b_A + b_B \geq 2(t - 2(1-\gamma)(a-N))$.

(ii) Suppose $x_A < x_B$. Then $p_1 = p_2 = p = a - N(x_A + 1 - x_B)$. For a provider at x_A, his total utility from participating on platform A should be zero, i.e., $u_A = 2((1-\gamma)p - tx_A) + b_A = 0$. For a provider at x_B, his total utility from participating on platform B should be zero, i.e., $u_B = 2((1-\gamma)p - t(1-x_B)) + b_B = 0$. Thus, we obtain that $x_A = \frac{(1-\gamma)(a+N\frac{b_A-b_B}{2t})+\frac{b_A}{2}}{2(1-\gamma)N+t}$, $1 - x_B = \frac{(1-\gamma)(a+N\frac{b_B-b_A}{2t})+\frac{b_B}{2}}{2(1-\gamma)N+t}$, and $p_1 = p_2 = \frac{2at-(b_A+b_B)N}{4(1-\gamma)N+2t}$. Also, partial utilization requires that $x_A + 1 - x_B < 1$, i.e., $b_A + b_B < 2\left(t - 2(1-\gamma)(a-N)\right)$.

Therefore, when $b_A = b_B = 0$, the supply pool is fully utilized if $t \leq 2(1-\gamma)(a-N)$, with $x_A^0 = x_B^0 = \frac{1}{2}$ and $p_1^0 = p_2^0 = a - N$. The supply pool is partially utilized if $t > 2(1-\gamma)(a-N)$, with $x_A^0 = 1 - x_B^0 = \frac{(1-\gamma)a}{2(1-\gamma)N+t}$ and $p_1^0 = p_2^0 = \frac{at}{2(1-\gamma)N+t}$. ■

2.8.1.2 Proof of Propositions 1-3

We combine the derivation of the competitive bonus equilibrium in this proof. We first derive the best responses of the two platforms, and then solve for the Nash equilibrium. Recall that in the proof of Lemma 1, we have derived the equilibrium supply pool utilization given any bonus scheme (b_A, b_B). By substituting the supply pool segmentation points into Eq. (2.2), the profit functions of the platforms simplify to Π_i^F if the supply pool is fully utilized and to Π_i^P if the supply pool is partially utilized, as shown in Eq. (2.A.1), where $i, j \in \{A, B\}$ and $i \neq j$:

$$\begin{aligned}\Pi_i^F(b_i, b_j) &= N\left(1 + \frac{b_i - b_j}{2t}\right)\left(\gamma(a-N) - \frac{b_i}{2}\right),\\ \Pi_i^P(b_i, b_j) &= N\frac{(1-\gamma)(a+N\frac{b_i-b_j}{2t})+\frac{b_i}{2}}{2(1-\gamma)N+t}\left(\frac{2at-(b_i+b_j)N}{2(1-\gamma)N+t}\gamma - b_i\right).\end{aligned} \tag{2.A.1}$$

- **Best Response**

We focus on platform A's best response $b_A(b_B)$. Platform B's best response $b_B(b_A)$ can be obtained similarly. Lemma 1 shows that the supply pool is fully utilized if $b_A \geq \tilde{b}_A$ and partially utilized if $b_A < \tilde{b}_A$ where $\tilde{b}_A = 2(t - 2(1-\gamma)(a-N)) - b_B$. Thus, Π_A is a piecewise function such that $\Pi_A = \Pi_A^P$ if $b_A < \tilde{b}_A$ and $\Pi_A = \Pi_A^F$ if $b_A \geq \tilde{b}_A$, and Π_A is continuous at the cutoff point $\tilde{b}_A$. From Eq. (2.A.1), the first-order derivative of Π_A w.r.t. b_A under full utilization $\frac{\partial \Pi_A^F}{\partial b_A}$ and under partial utilization $\frac{\partial \Pi_A^P}{\partial b_A}$ are shown in Eq. (2.A.2):

$$\begin{aligned}\frac{\partial \Pi_A^F}{\partial b_A} &= N\left(-\frac{1}{2t}b_A + \frac{1}{4t}b_B - \frac{1}{2} + \frac{\gamma(a-N)}{2t}\right),\\ \frac{\partial \Pi_A^P}{\partial b_A} &= -\frac{N\Big(2\big((1-\gamma)N+t\big)\big((2-\gamma)N+t\big)b_A - \big(2(1-\gamma)^2N+(1-2\gamma)t\big)(Nb_B-2at)\Big)}{2t(2(1-\gamma)N+t)^2}.\end{aligned} \tag{2.A.2}$$

From setting $\frac{\partial \Pi_A^F}{\partial b_A} = 0$, the unconstrained optimizer of Π_A^F is $b_A^F = \gamma(a-N) - t + \frac{b_B}{2}$. When evaluated at $b_A = \tilde{b}_A$, $\frac{\partial \Pi_A^F}{\partial b_A}|_{b_A=\tilde{b}_A} = N(\frac{3}{4t}b_B + \frac{3(1-\gamma)(a-N)}{2t} + \frac{a-N}{2t} - \frac{3}{2})$. From setting $\frac{\partial \Pi_A^P}{\partial b_A} = 0$, the unconstrained optimizer of Π_A^P is $b_A^P = \frac{(Nb_B - 2ta)\left(2N(1-\gamma)^2 + (1-2\gamma)t\right)}{2\left(N(1-\gamma)+t\right)\left((2-\gamma)N+t\right)}$. When evaluated at $b_A = \tilde{b}_A$, $\frac{\partial \Pi_A^P}{\partial b_A}|_{b_A=\tilde{b}_A} = \frac{\left((3-2\gamma)N+2t\right)N}{2t\left(2(1-\gamma)N+t\right)} b_B - \frac{N\left(2t^2 - (3-2\gamma)(a-2N)t - 2(2-\gamma)(1-\gamma)(a-N)N\right)}{\left(2(1-\gamma)N+t\right)t}$. We identify five thresholds of b_B that can affect the shape of Π_A, and they are $b_0 = 2(t - 2(1-\gamma)(a-N))$, $b_1 = 2t - \frac{2(a-N)}{3} - 2(1-\gamma)(a-N)$, $b_2 = \frac{2\left(2t^2 - (3-2\gamma)(a-2N)t - 2(2-\gamma)(1-\gamma)(a-N)N\right)}{(3-2\gamma)N+2t}$, $b_3 = 2(t - \gamma(a-N))$, $b_4 = \frac{2at}{N}$. First, threshold b_0 determines the number of pieces in the profit function Π_A for $b_A \geq 0$. If $b_B \geq b_0$, there is one piece, and $\Pi_A = \Pi_A^F$. If $b_B < b_0$, there are two pieces, and $\Pi_A = \Pi_A^P$ for $b_A < b_0 - b_B$ and $\Pi_A = \Pi_A^F$ for $b_A \geq b_0 - b_B$. Note that $b_0 \leq 0$ implies that $b_B \geq b_0$. Second, thresholds b_1 and b_2 determine the signs of the first-order derivatives of Π_A^F and Π_A^P at $\tilde{b}_A$, respectively. $\frac{\partial \Pi_A^F}{\partial b_A}|_{b_A=\tilde{b}_A} \geq 0$ if and only if $b_B \geq b_1$ and $\frac{\partial \Pi_A^P}{\partial b_A}|_{b_A=\tilde{b}_A} \geq 0$ if and only if $b_B \geq b_2$. Third, thresholds b_3 and b_4 determine the signs of the unconstrained optimizers of Π_A^F and Π_A^P, respectively. $b_A^F \geq 0$ if and only if $b_B \geq b_3$. If $2(1-\gamma)^2 N + (1-2\gamma)t \geq 0$, $b_A^P \geq 0$ if and only if $b_B \geq b_4$; if $2(1-\gamma)^2 N + (1-2\gamma)t < 0$, $b_A^P \geq 0$ if and only if $b_B \leq b_4$. Note that $2(1-\gamma)^2 N + (1-2\gamma)t \geq 0$ is equivalent to $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$.

Since $\gamma \leq \frac{2}{3}$, the comparison of the five thresholds shows that $b_4 = \max(b_0, b_1, b_2, b_3, b_4)$, and $b_3 \geq b_1 \geq b_0$. We also obtain that $b_0 \leq b_2$ if and only if $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$, and $b_1 \geq b_2$ if and only if $t \geq \gamma N$. Moreover, $\gamma N \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$ given that $\gamma \leq \frac{2}{3}$. We now derive $b_A(b_B)$ in two cases: (1) $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$; (2) $t > \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$.

Case (1): $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$. In this case, $b_0 \leq b_1 < b_2$ if $t < \gamma N$, and $b_0 \leq b_2 \leq b_1$ if $t \geq \gamma N$. (i) If $b_B < b_0$, $\Pi_A = \Pi_A^P$ and is decreasing in b_A for $0 \leq b_A < b_0 - b_B$; $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq b_0 - b_B$. Thus, $b_A(b_B) = 0$. (ii) If $b_0 \leq b_B \leq b_3$, $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq 0$. Thus, $b_A(b_B) = 0$. (iii) If $b_B > b_3$, $\Pi_A = \Pi_A^F$ and is first increasing in b_A for $0 \leq b_A \leq b_A^F$ then decreasing in b_A for $b_A \geq b_A^F$. Thus, $b_A(b_B) = b_A^F$.

Case (2): $t > \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$. In this case, $b_2 < b_0 \leq b_1$. (i) If $b_B < b_2$, $\Pi_A = \Pi_A^P$ and is first increasing in b_A for $0 \leq b_A \leq b_A^P$ then decreasing in b_A for $b_A^P \leq b_A < b_0 - b_B$; $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq b_0 - b_B$. Thus, $b_A(b_B) = b_A^P$. (ii) If $b_2 \leq b_B < b_0$, $\Pi_A = \Pi_A^P$ and is increasing in b_A for $0 \leq b_A < b_0 - b_B$; $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq b_0 - b_B$. Thus, $b_A(b_B) = b_0 - b_B$. (iii) If $b_0 \leq b_B \leq b_3$, $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq 0$. Thus, $b_A(b_B) = 0$. (iv) If $b_B > b_3$, $\Pi_A = \Pi_A^F$ and is first increasing in b_A for $0 \leq b_A \leq b_A^F$ then decreasing in b_A for $b_A \geq b_A^F$. Thus, $b_A(b_B) = b_A^F$.

Similarly, we can derive the best response of platform B. We summarize the best responses of the two platforms in Eq. (2.A.3) if $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$ and in Eq. (2.A.4) if $t > \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$:

$$b_i(b_j) = \begin{cases} 0 & \textit{if } \ b_j \leq 2(t - \gamma(a-N)), \\ \gamma(a-N) - t + \frac{b_j}{2} & \textit{if } \ b_j > 2(t - \gamma(a-N)). \end{cases} \tag{2.A.3}$$

$$b_i(b_j) = \begin{cases} \frac{(Nb_B - 2at)\left(2(1-\gamma)^2 N + (1-2\gamma)t\right)}{2\left((1-\gamma)N+t\right)\left((2-\gamma)N+t\right)} & \textit{if } \ b_j < \frac{2\left(2t^2-(3-2\gamma)(a-2N)t-2(2-\gamma)(1-\gamma)(a-N)N\right)}{(3-2\gamma)N+2t}, \\ 2(t - 2(1-\gamma)(a-N)) - b_j & \textit{if } \ \frac{2\left(2t^2-(3-2\gamma)(a-2N)t-2(2-\gamma)(1-\gamma)(a-N)N\right)}{(3-2\gamma)N+2t} \leq b_j \\ & \qquad < 2(t - 2(1-\gamma)(a-N)), \\ 0 & \textit{if } \ 2(t-2(1-\gamma)(a-N)) \leq b_j \leq 2(t - \gamma(a-N)), \\ \gamma(a-N) - t + \frac{b_j}{2} & \textit{if } \ b_j > 2(t-\gamma(a-N)). \end{cases} \tag{2.A.4}$$

- **Equilibrium**

From the best responses obtained in Eq. (2.A.3) and Eq. (2.A.4), we now derive the equilibrium bonuses and the conditions to support each equilibrium outcome.

Case (1): $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$. Two equilibrium outcomes can occur. (i) $b_A^* = b_B^* = 0$. Best responses $b_A(b_B) = 0$ and $b_B(b_A) = 0$ will induce this equilibrium. This equilibrium requires that $b_i^* \leq 2(t - \gamma(a-N))$, i.e., $t \geq \gamma(a-N)$. Thus, this equilibrium occurs if $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$ and $t \geq \gamma(a-N)$. (ii) $b_A^* = b_B^* = 2((a-N)\gamma - t)$. Best responses $b_A(b_B) = \gamma(a-N) - t + \frac{b_B}{2}$ and $b_B(b_A) = \gamma(a-N) - t + \frac{b_A}{2}$ will induce this equilibrium. This equilibrium requires that $b_i^* > 2(t - \gamma(a-N))$, i.e., $t < \gamma(a-N)$. Thus, this equilibrium occurs if $t \leq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$ and $t < \gamma(a-N)$.

Case (2): $t > \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$. Four equilibrium outcomes can occur. (iii) $b_A^* = b_B^* = \frac{2\left((2\gamma-1)t - 2(1-\gamma)^2 N\right)at}{2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2}$. Best responses $b_A(b_B) = \frac{(Nb_B-2at)\left(2(1-\gamma)^2N+(1-2\gamma)t\right)}{2\left((1-\gamma)N+t\right)\left((2-\gamma)N+t\right)}$ and $b_B(b_A) = \frac{(Nb_A-2at)\left(2(1-\gamma)^2N+(1-2\gamma)t\right)}{2\left((1-\gamma)N+t\right)\left((2-\gamma)N+t\right)}$ will induce this equilibrium. First, $b_A^* \geq 0$ and $b_B^* \geq 0$ since $t > \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$. Second, this equilibrium requires that $b_i^* < \frac{2\left(2t^2-(3-2\gamma)(a-2N)t-2(2-\gamma)(1-\gamma)(a-N)N\right)}{(3-2\gamma)N+2t}$, i.e., $f_1(t) = 2t^2 - (2a-(5-2\gamma)N)t - 2(1-\gamma)(a-N)N > 0$. Since $f_1(t)$ is an upward opening quadratic function of t and $f_1(0) = -2(1-\gamma)(a-$

$N)N < 0$, the smaller root of $f_1(t) = 0$ must be negative while the larger root must be positive. We denote the larger root as $\tilde{t}$, and hence $f_1(t) > 0$ if $t > \tilde{t}$ where

$$\tilde{t} = \frac{1}{4}\left(2a - 5N + 2N\gamma + \sqrt{4\gamma^2N^2 - 4\gamma N(2a + N) + 4a^2 - 4aN + 9N^2}\right) \tag{2.A.5}$$

Thus, equilibrium outcome (iii) occurs if $t > \max\left(\frac{2(1-\gamma)^2N}{(2\gamma-1)^+}, \tilde{t}\right)$. Note that $b_A^* = b_B^* = t - 2(1-\gamma)(a-N)$ if $t = \tilde{t}$.

(iv) $b_A^* = b_B^* = t - 2(1-\gamma)(a-N)$. Best responses $b_A(b_B) = 2(t - 2(1-\gamma)(a - N)) - b_B$ and $b_B(b_A) = 2(t - 2(1-\gamma)(a-N)) - b_A$ will induce this equilibrium. From the best responses, we can see that multiple equilibria can be obtained, and we employ a refinement criterion as follows. From the first case of (2.A.1), we obtain that $\Pi_A + \Pi_B = N\left(2\gamma(a-N) - \frac{b_A+b_B}{2} - \frac{(b_A-b_B)^2}{4t}\right)$. Since in equilibrium $b_A + b_B = 2(t - 2(1-\gamma)(a-N))$, $\Pi_A + \Pi_B$ is maximum when $b_A = b_B$. Thus, $b_A^* = b_B^* = t-2(1-\gamma)(a-N)$ is the unique equilibrium that generates the maximum total profit of the two platforms, and we select it as the equilibrium for this case. First, this equilibrium requires that $b_i^* < 2(t-2(1-\gamma)(a-N))$, i.e., $t > 2(1-\gamma)(a-N)$. Second, this equilibrium requires that $b_i^* \geq \frac{2\left(2t^2-(3-2\gamma)(a-2N)t-2(2-\gamma)(1-\gamma)(a-N)N\right)}{(3-2\gamma)N+2t}$, i.e., $f_1(t) = 2t^2 - (2a - (5-2\gamma)N)t - 2(1-\gamma)(a-N)N \leq 0$. Since we have obtained that $f_1(t) \leq 0$ if $t \leq \tilde{t}$, this equilibrium occurs if $\max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}\right) < t \leq \tilde{t}$. Note that $b_A^* = b_B^* = 0$ if $t = 2(1-\gamma)(a-N)$.

(v) $b_A^* = b_B^* = 0$. Best responses $b_A(b_B) = 0$ and $b_B(b_A) = 0$ will induce this equilibrium. First, this equilibrium requires that $b_i^* \leq 2(t - \gamma(a-N))$, i.e., $t \geq \gamma(a-N)$. Second, this equilibrium requires that $b_i^* \geq 2(t - 2(1-\gamma)(a-N))$, i.e., $t \leq 2(1-\gamma)(a-N)$. Thus, this equilibrium occurs if $t > \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}$ and $\gamma(a-N) \leq t \leq 2(1-\gamma)(a-N)$.

(vi) $b_A^* = b_B^* = 2(\gamma(a-N) - t)$. Best responses $b_A(b_B) = \gamma(a-N) - t + \frac{b_B}{2}$ and $b_B(b_A) = \gamma(a-N) - t + \frac{b_A}{2}$ will induce this equilibrium. This equilibrium requires that $b_i^* > 2(t-\gamma(a-N))$, i.e., $t < \gamma(a-N)$. Thus, this equilibrium occurs if $t > \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}$ and $t < \gamma(a-N)$. Note that $b_A^* = b_B^* = 0$ if $t = \gamma(a-N)$.

To summarize, the equilibrium outcomes are as follows: for $i \in \{A, B\}$,

① If $t < \gamma(a-N)$, $b_i^* = 2(\gamma(a-N) - t)$ which is obtained from (ii) and (vi), and $\Pi_i^* = Nt$;

② If $\gamma(a-N) \leq t \leq \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}\right)$, $b_i^* = 0$ which is obtained from (i) and (v), and $\Pi_i^* = \gamma N(a-N)$ if $t \leq 2(1-\gamma)(a-N)$ and $\Pi_i^* = \frac{2\gamma(1-\gamma)a^2Nt}{(2(1-\gamma)N+t)^2}$ if $t > 2(1-\gamma)(a-N)$;

③ If $\max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}\right) < t \leq \tilde{t}$, $b_i^* = t - 2(1-\gamma)(a-N)$ which is obtained from (iv), and $\Pi_i^* = N((a-N) - \frac{t}{2})$;

④ If $t > \max(\frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}, \tilde{t})$, $b_i^* = \frac{2\left((2\gamma-1)t - 2(1-\gamma)^2 N\right)at}{2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2}$ which is obtained from (iii), and $\Pi_i^* = \frac{2a^2 Nt\left(t+(1-\gamma)N\right)\left(t+(2-\gamma)N\right)}{\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)^2}$. Notice that case ① corresponds to Proposition 2, and case ③ and ④ correspond to Proposition 3.

Finally, we derive the number of providers who participate in equilibrium. From Lemma 1, we obtain that the supply pool is fully utilized in cases ① and ③. Moreover, a provider at $\bar{x} = \frac{1}{2} + \frac{b_A^* - b_B^*}{4t}$ obtains utility $u(\bar{x}) = 2(a-N)(1-\gamma) - 2t\bar{x} + b_i^* = b_i^* - (t - 2(1-\gamma)(a-N))$. Thus, $u(\bar{x}) > 0$ in case ① and $u(\bar{x}) = 0$ in case ③. Thus, even though the supply pool is fully utilized in both case ① and ③, the platforms are faced with poaching threat from their competing platforms only in case ① but not in case ③.

For case ②, if $2(1-\gamma)(a-N) \geq \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$, the supply pool is fully utilized; if $2(1-\gamma)(a-N) < \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$, the supply pool is fully utilized when $t \leq 2(1-\gamma)(a-N)$ and partially utilized if $t > 2(1-\gamma)(a-N)$. For case ④, the supply pool is partially utilized since the underlying best responses $b_A(b_B) = b_A^P$ and $b_B(b_A) = b_B^P$ are obtained under partial utilization. In this case, the number of providers who participate in equilibrium is $\frac{4(1-\gamma)aN + (b_A^* + b_B^*)N}{4(1-\gamma)N + 2t}$, which is higher than that in the case without bonuses, $\frac{2(1-\gamma)aN}{2(1-\gamma)N + t}$.

Additionally, we derive a necessary condition for case ③ to occur in equilibrium. Case ③ requires that $\tilde{t} > \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}\right)$. We first show that $\tilde{t} > 2(1-\gamma)(a-N)$ if and only if $\gamma > \frac{a}{2a-N}$. First, $\tilde{t} - 2(1-\gamma)(a-N) = \frac{f_2(\gamma) - f_3(\gamma)}{4}$ where $f_3(\gamma) = 3(2a-N) - 2\gamma(4a-3N)$ and $f_2(\gamma) = \sqrt{4(1-\gamma)^2 N^2 + 4(1-\gamma)N(2a-N) + 4a^2 - 12Na + 9N^2}$. $f_3(\gamma) > 0$ if $\gamma < \frac{3(2a-N)}{2(4a-3N)}$. Also, $f_3^2(\gamma) - f_2^2(\gamma) = 32(1-\gamma)(a-N)(a-\gamma(2a-N))$ and hence $f_3^2(\gamma) > f_2^2(\gamma)$ if and only if $\gamma < \frac{a}{2a-N}$. Since $\frac{3(2a-N)}{2(4a-3N)} - \frac{a}{2a-N} = \frac{4a^2 - 6aN + 3N^2}{2(4a-3N)(2a-N)} > 0$ when $\frac{a}{N} > 1$, we obtain that $f_2(\gamma) < f_3(\gamma)$ if and only if $\gamma < \frac{a}{2a-N}$. Thus, $\tilde{t} > 2(1-\gamma)(a-N)$ if and only if $\gamma > \frac{a}{2a-N}$.

If follows from our above analysis that $\tilde{t} > \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}\right)$ only if $\tilde{t} > 2(1-\gamma)(a-N)$, i.e., $\gamma < \frac{a}{2a-N}$. Next we show that $2(1-\gamma)(a-N) > \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$ if $\gamma > \frac{a}{2a-N}$. First, if $\gamma \leq \frac{1}{2}$, $\frac{2(1-\gamma)^2 N}{(2\gamma-1)^+} = \infty$, and thus $\frac{2(1-\gamma)^2 N}{(2\gamma-1)^+} > 2(1-\gamma)(a-N)$. If $\gamma > \frac{1}{2}$, $\frac{2(1-\gamma)^2 N}{(2\gamma-1)^+} - 2(1-\gamma)(a-N) = \frac{2(1-\gamma)\left(a-\gamma(2a-N)\right)}{2\gamma-1}$, and hence $\frac{2(1-\gamma)^2 N}{(2\gamma-1)^+} < 2(1-\gamma)(a-N)$ if $\gamma > \frac{a}{2a-N}$. Thus, $2(1-\gamma)(a-N) > \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}$ if $\gamma > \frac{a}{2a-N}$. From the above analysis, we obtain that $\tilde{t} > \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}\right)$ if and only if $\gamma > \frac{a}{2a-N}$. Note that $\gamma > \frac{a}{2a-N}$ is equivalent to $\frac{a}{N} > \frac{\gamma}{(2\gamma-1)^+}$. ■

2.8.1.3 Proof of Proposition 4

We first prove that platform A and platform B will offer the same bonus in the coordinated equilibrium, i.e., $b_A^\dagger = b_B^\dagger$. From Eq. (2.A.1), the total profit of the two platforms when the supply pool is fully utilized, i.e., Π^F, and when the supply pool is partially utilized, i.e., Π^P are shown in Eq. (2.A.6).

$$\begin{aligned}\Pi^F(b_A,b_B) &= N\left(2\gamma(a-N)-\tfrac{2(1-\gamma)N+t}{2}-\tfrac{(b_A-b_B)^2}{4t}\right)\\ \Pi^P(b_A,b_B) &= N\big(\tfrac{4(1-\gamma)a^2t}{(2(1-\gamma)N+t)^2}-\tfrac{a(2\gamma^2N-(6N+t)\gamma+4N)}{(2(1-\gamma)N+t)^2}(b_A+b_B)-\tfrac{N}{2(2(1-\gamma)N+t)^2}(b_A+b_B)^2\\ &\quad -\tfrac{1}{2(2(1-\gamma)N+t)}(b_A^2+b_B^2)-\tfrac{(1-\gamma)N}{2(2(1-\gamma)N+t)t}(b_A-b_B)^2\big)\end{aligned} \tag{2.A.6}$$

It is easy to see that Π^F achieves its maximmum when $b_A = b_B$. Now we focus on Π^P. Suppose $b_A < b_B$ and $b_A + b_B = K$, and denote $b_A = \frac{K}{2} - m$ and $b_B = \frac{K}{2} + m$, where $m \geq 0$. Thus, $b_A^2 + b_B^2 = (\frac{K}{2} - m)^2 + (\frac{K}{2} + m)^2 = \frac{K^2}{2} + 2m^2$, and hence $b_A^2 + b_B^2$ is increasing in m. Also, $(b_A - b_B)^2$ is increasing in m. Thus, for any K, when $b_A + b_B = K$, Π^P can always improve by equating $b_A = b_B$. Similar analysis can be shown when we suppose $b_A > b_B$. Thus, Π^P achieves its maximum only if $b_A = b_B$.

When $b_A = b_B = b$, according to the proof of Lemma 1, the supply pool is fully utilized if $b \geq t - 2(1-\gamma)(a-N)$, and the supply pool is partially utilized if $b < t - 2(1-\gamma)(a-N)$. The total profit of the two platforms simplify to $\Pi^F(b)$ and $\Pi^P(b)$ as shown in Eq. (2.A.6).

$$\begin{aligned}\Pi^F(b) &= 2N\left(2(a-N)\gamma - b\right),\\ \Pi^P(b) &= 2N\frac{-(2N+t)b^2-\big(2(1-2\gamma)at+4(1-\gamma)aN\big)b+4\gamma(1-\gamma)a^2t}{(2N(1-\gamma)+t)^2}.\end{aligned} \tag{2.A.7}$$

Under coordination, the platforms choose b to maximize their total profit $\Pi = \Pi^P \cdot \mathbb{1}_{b<t-2(1-\gamma)(a-N)} + \Pi^F \cdot \mathbb{1}_{b\geq t-2(1-\gamma)(a-N)}$. First, Π^F is decreasing in b. Second, the unconstrained maximizer of Π^P is $\tilde{b} = \frac{(2\gamma-1)at-2(1-\gamma)aN}{2N+t}$. Thus, $\tilde{b} > 0$ if $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$, and $\tilde{b} \geq t - 2(1-\gamma)(a-N)$ if $t \leq a - 2N$. We derive the optimal bonus $b^\dagger$ in the following two cases.

Case (1): $t \leq 2(1-\gamma)(a-N)$. In this case, $\Pi = \Pi^F$ for $b \geq 0$. Hence, $b^\dagger = 0$ and $\Pi^\dagger = 2\gamma N(a-N)$. The supply pool utilization follows from Lemma 1.

Case (2): $t > 2(1-\gamma)(a-N)$. In this case, $\Pi = \Pi^P$ for $b < t-2(1-\gamma)(a-N)$, and $\Pi = \Pi^F$ for $b \geq t-2(1-\gamma)(a-N)$. We consider three cases for the optimal bonus. (i) If $t \leq \frac{2(1-\gamma)N}{(2\gamma-1)^+}$, then $\tilde{b} \leq 0$. Thus, Π is decreasing in b for $b \geq 0$, and we obtain that $b^\dagger = 0$ and $\Pi^\dagger = \frac{4\gamma(1-\gamma)a^2Nt}{\left(2(1-\gamma)N+t\right)^2}$. The supply pool utilization follows from Lemma

1. (ii) If $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t \leq a-2N$, then $\tilde{b} \geq t-2(1-\gamma)(a-N)$. Thus, Π is increasing in b for $0 \leq b < t-2(1-\gamma)(a-N)$ and decreasing in b for $b \geq t-2(1-\gamma)(a-N)$, and we obtain that $b^\dagger = t-2(1-\gamma)(a-N)$ and $\Pi^\dagger = N(2(a-N)-t)$. In this case, the platforms achieve coverage of the supply pool through offering the optimal bonus. Moreover, a provider at $\bar{x} = \frac{1}{2}$ obtains utility $u(\bar{x}) = b^\dagger - (t-2(1-\gamma)(a-N)) = 0$. Thus, even though the supply pool is fully utilized, neither platform is faced with poaching threat from her competing platform. (iii) If $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t > a-2N$, then $0 < \tilde{b} < t-2(1-\gamma)(a-N)$. Thus, Π is increasing in b for $0 \leq b < \tilde{b}$ and decreasing in b for $b \geq \tilde{b}$, and we obtain that $b^\dagger = \tilde{b} = \frac{(2\gamma-1)at-2(1-\gamma)aN}{2N+t}$ and $\Pi^\dagger = \frac{a^2N}{2N+t}$. In this case, the platforms achieve partial coverage of the supply pool through offering the optimal bonus. The number of providers who participate under the optimal bonus is $\frac{4(1-\gamma)aN+(b_A^\dagger+b_B^\dagger)N}{4(1-\gamma)N+2t}$. In all the cases above, since $b_A^\dagger = b_B^\dagger = b^\dagger$, we obtain that $\Pi_A^\dagger = \Pi_B^\dagger = \frac{\Pi^\dagger}{2}$.

Additionally, using a similar approach as in the proof of Proposition 1-3, the comparison of the thresholds of t, i.e., $2(1-\gamma)(a-N)$, $\frac{2(1-\gamma)N}{(2\gamma-1)^+}$, $a-2N$, shows that $\frac{2(1-\gamma)N}{(2\gamma-1)^+} < 2(1-\gamma)(a-N) < a-2N$ if and only if $\frac{a}{N} > \frac{2\gamma}{(2\gamma-1)^+}$. ■

2.8.1.4 Proof of Proposition 5

We first prove part (i) of the Proposition. From Propositions 1-3, we know that if $t < \gamma(a-N)$, then $b^* > b^\dagger = 0$; if $\gamma(a-N) \leq t \leq \max(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+})$, then $b^* = b^\dagger = 0$. Moreover, the platforms offer bonuses in the competitive equilibrium if $t > \max(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+})$, and offer bonuses in the coordinated equilibrium if $t > \max(2(1-\gamma)(a-N), \frac{2(1-\gamma)N}{(2\gamma-1)^+})$. We divide the case of $t > \max(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+})$ into five subcases, which correspond to regions III.(1)-III.(5) in Figure 2.A.1. We now analyze each subcase as follows. First, cases (1) and (2) correspond to $t < \frac{2(1-\gamma)N}{(2\gamma-1)^+}$.

Case (1): $t < \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t \leq \tilde{t}$. This corresponds to region III.(1) in Figure 2.A.1. In this case, the platforms offer bonuses to achieve full coverage of the supply pool in the competitive equilibrium, and do not offer bonuses in the coordinated equilibrium. Thus, $b^\dagger < b^*$.

Case (2): if $t < \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t > \tilde{t}$. This corresponds to region III.(2) in Figure 2.A.1. In this case, the platforms offer bonuses to achieve partial coverage of the supply pool in the competitive equilibrium, and do not offer bonuses in the coordinated equilibrium. Thus, $b^\dagger < b^*$.

Next, consider $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$. The comparison of $\tilde{t}$ and $a-2N$ shows that $\tilde{t} = a-2N$

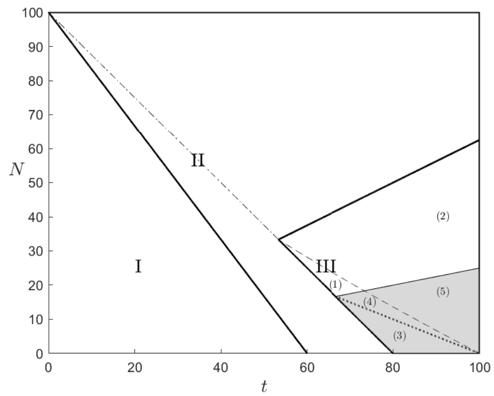


Figure 2.A.1: Comparison of Coordinated and Competitive Bonus Equilibria ($a = 100, \gamma = 0.6$)

when $N = 0$. Moreover, $\frac{\partial \tilde{t}}{\partial N}|_{N=0} = -\frac{3}{2}$, $\frac{\partial^2 \tilde{t}}{\partial N^2} = \frac{8(1-\gamma)a^2}{\left(4N^2\gamma^2-4N(2a+N)\gamma+4a^2-4aN+9N^2\right)^{3/2}} > 0$, and hence $\frac{\partial \tilde{t}}{\partial N} > \frac{\partial(a-2N)}{\partial N} = -2$ for any N. Thus, we always have $\tilde{t} > a - 2N$. Consequently, the case of $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ contains three subcases as follows.

Case (3): $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t \leq a-2N$. This corresponds to region III.(3) in Figure 2.A.1. In this case, the platforms offer bonuses to achieve full market coverage of the supply pool in both competitive and coordinated equilibria. Moreover, it follows from Proposition 3 and Proposition 5 that $b^\dagger = b^*$.

Case (4): $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $a - 2N < t \leq \tilde{t}$. This corresponds to region III.(4) in Figure 2.A.1. In this case, the platforms offer bonuses to achieve full coverage of the supply pool in the competitive equilibrium, and offer bonuses to achieve partial coverage of the supply pool in the coordinated equilibrium. Moreover, $b^* = t - 2(1-\gamma)(a-N)$ and $b^\dagger = \frac{(2\gamma-1)at-2(1-\gamma)aN}{2N+t}$. We have $b^\dagger < b^*$ since $t > a - 2N$.

Case (5): $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t > \tilde{t}$. This corresponds to region III.(5) in Figure 2.A.1. In this case, the platforms offer bonuses to achieve partial coverage of the supply pool in both competitive and coordinated equilibria. Moreover, $b^* = \frac{2\left((2\gamma-1)t-2(1-\gamma)^2N\right)at}{2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2}$ and $b^\dagger = \frac{(2\gamma-1)at-2(1-\gamma)aN}{2N+t}$, and we obtain that $b^\dagger - b^* = -\frac{aN(t+2(1-\gamma)N)^2}{(2N+t)\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)} < 0$. Combining all the cases analyzed above, we obtain that $b_i^\dagger \leq b_i^*$ always holds, and $b_i^\dagger < b_i^*$ if and only if $t < \gamma(a-N)$ or $t > \max(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}, a-2N)$.

We next prove part (ii) of the proposition. In the coordinated bonus game, the platforms' total profit function is shown in Eq. (2.A.7). Notice that we obtain the competitive equilibrium total profit $2\Pi_i^*$ by plugging $b = b_A^*$ (or b_B^*) into

Eq. (2.A.7). Since $b = b_A^\dagger$ (or $b_B^\dagger$) is the maximizer of the profit function in Eq. (2.A.7), we obtain that $\Pi_i^* \geq \Pi_i^\dagger$, and $\Pi_i^\dagger > \Pi_i^*$ if and only if $b_i^\dagger < b_i^*$. Finally, from Proposition 4, it is easy to see that as $t \to 0$, $\Pi_i^\dagger = \gamma(a - N) > 0$. ■

2.8.1.5 Proof of Proposition 6

From Lemma 1, we know in equilibrium $x_1^i = x_2^i$ and $p_1 = p_2$ where $i = \{A, B\}$. To begin with, we plug $x_1^i = x_2^i = x^i$ and $p_1 = p_2 = p$ where $i = \{A, B\}$ into Eq. (2.5) and Eq. (2.6), and we obtain

$$\begin{aligned} U_P &= (2(1-\gamma)p - tx_A + b_A)x_A N + (2(1-\gamma)p - t(1-x_B) + b_B)(1-x_B)N, \\ U_C &= (x_A + 1 - x_B)^2 N^2. \end{aligned} \tag{2.A.8}$$

From Lemma 1, when $b_A = b_B = 0$, if $t \leq 2(1-\gamma)(a-N)$, the supply pool is fully utilized with $x_A^0 = x_B^0 = \frac{1}{2}$ and $p^0 = a - N$; if $t > 2(1-\gamma)(a-N)$, the supply pool is partially utilized with $x_A^0 = \frac{(1-\gamma)a}{2(1-\gamma)N+t}$, $x_B^0 = 1 - \frac{(1-\gamma)a}{2(1-\gamma)N+t}$, and $p^0 = \frac{at}{2(1-\gamma)N+t}$. Plugging in these expressions into Eq. (2.A.8), we obtain the provider surplus U_P^0, the consumer surplus U_C^0, and the social welfare W^0 when $b_A = b_B = 0$ in Eq. (2.A.9).

$$\begin{aligned} &U_P^0 = \left(2(1-\gamma)(a-N) - \tfrac{t}{2}\right)N,\ U_C^0 = N^2,\ W^0 = 2aN - N^2 - \tfrac{Nt}{2}, && \text{if } t \leq 2(1-\gamma)(a-N), \\ &U_P^0 = \tfrac{2(1-\gamma)^2 a^2 Nt}{(2(1-\gamma)N+t)^2},\ U_C^0 = \tfrac{4(1-\gamma)^2 a^2 N^2}{(2(1-\gamma)N+t)^2},\ W^0 = \tfrac{2(1-\gamma)a^2 N\left(2N+t+\gamma(t-2N)\right)}{(2(1-\gamma)N+t)^2}, && \text{if } t > 2(1-\gamma)(a-N). \end{aligned} \tag{2.A.9}$$

First, consider the case of $t < \gamma(a-N)$. In the competitive equilibrium, we know from Proposition 2 that $b_A^* = b_B^* = 2(\gamma(a-N) - t)$, $x_A^* = x_B^* = \frac{1}{2}$, and $p^* = a - N$. Plugging these expressions into Eq. (2.A.8), we obtain that $U_P^* = \left(2(a-N) - \frac{5t}{2}\right)N$, $U_C^* = N^2$, and $W^* = 2aN - N^2 - \frac{Nt}{2}$. In the coordinated equilibrium, since $b_A^\dagger = b_B^\dagger = 0$, the surplus and welfare expressions are shown in the first case of Eq. (2.A.9). Thus, we obtain that $U_P^\dagger < U_P^*$, $U_C^\dagger = U_C^*$, and $W^\dagger = W^*$.

Second, consider the case of $\gamma(a-N) \leq t \leq \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}\right)$. In this case, since $b_A^\dagger = b_A^* = b_B^\dagger = b_B^* = 0$, we obtain that $U_P^\dagger = U_P^*$, $U_C^\dagger = U_C^*$, and $W^\dagger = W^*$.

Third, consider the case of $t > \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2 N}{(2\gamma-1)^+}\right)$. As in the proof of Proposition 5, we consider the following five subcases, which corresponds to regions III.(1)-III.(5) in Figure 2.A.1.

Case (1): $t < \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t \leq \tilde{t}$. In this case, the platforms offer bonuses $b_A^* = b_B^* = t - 2(1-\gamma)(a-N)$ to achieve full coverage of the supply pool in the competitive equilibrium. Thus, from Eq. (2.A.8), $U_P^* = \frac{Nt}{2}$, $U_C^* = N^2$, and $W^* = 2aN - N^2 - \frac{Nt}{2}$. The platforms do not offer bonuses in the coordinated equilibrium, and the surplus and welfare expressions are shown in the second case of Eq. (2.A.9). Thus, we obtain that $U_P^* - U_P^\dagger = U_C^* - U_C^\dagger = \frac{Nt\left(t-2(1-\gamma)(a-N)\right)\left(2(1-\gamma)(a+N)+t\right)}{2\left(2(1-\gamma)N+t\right)^2} > 0$. Moreover, $W^* - W^\dagger = -\frac{N\left(t-2(1-\gamma)(a-N)\right)\left(t^2+\left(4N-2a-2\gamma(a+N)\right)t-4N(1-\gamma)(a-N)\right)}{2\left(t+2(1-\gamma)N\right)^2}$, and hence $W^* > W^\dagger$ if and only if $g(t) = t^2 + (4N - 2a - 2\gamma(a+N))t - 4(1-\gamma)N(a-N) < 0$. Since $g(t)$ is a upward opening quadratic function of t, and $g(0) = -4(1-\gamma)N(a-N) < 0$, we obtain that the two roots of $g(t)$ satisfy $t_g^- < 0 < t_g^+$, and $g(t) > 0$ if and only if $t > t_g^+$ where

$$t_g^+ = a - 2N + \gamma(a+N) + \sqrt{\gamma^2(a+N)^2 + 2\gamma a(a-3N) + a^2}. \qquad (2.A.10)$$

From Eq. (2.A.5) and (2.A.10), $t_g^+ - \tilde{t} = \frac{1}{2}(a - \frac{3}{2}N) + \gamma(a + \frac{N}{2}) + \sqrt{\gamma^2(a+N)^2 + 2\gamma a(a-3N) + a^2} - \frac{1}{4}\sqrt{4\gamma^2N^2 - 4\gamma N(2a+N) + 4a^2 - 4aN + 9N^2}$. Moreover, $\left(\sqrt{\gamma^2(a+N)^2 + 2\gamma a(a-3N) + a^2}\right)^2 - \left(\frac{1}{4}\sqrt{4\gamma^2N^2 - 4\gamma N(2a+N) + 4a^2 - 4aN + 9N^2}\right)^2 = N^2 h(\frac{a}{N})$ where $h(\frac{a}{N}) = (\gamma^2 + 2\gamma + \frac{3}{4})(\frac{a}{N})^2 + (2\gamma^2 - \frac{11}{2}\gamma + \frac{1}{2})\frac{a}{N} + \frac{3\gamma^2}{4} + \frac{\gamma}{4} - \frac{9}{16}$ is an upward opening quadratic function of $\frac{a}{N}$. First, the discriminant $\Delta_h = \gamma^4 - 29\gamma^3 + \frac{117}{4}\gamma^2 + \gamma + \frac{7}{4} > 0$. Second, the larger root is $\frac{-8\gamma^4+22\gamma-1+2\sqrt{4\gamma^4-116\gamma^3+117\gamma^2+4\gamma+7}}{2(4\gamma^2+8\gamma+3)} < 2$. Since $\frac{a}{N} > 2$ in this equilibrium, we obtain that $h(\frac{a}{N}) > 0$, i.e., $t_g^+ > \tilde{t}$, and hence $t < t_g^+$ holds and $W^\dagger < W^*$. Thus, in Case (1), $U_P^\dagger < U_P^*$, $U_C^\dagger < U_C^*$, and $W^\dagger < W^*$.

Case (2): $t < \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t > \tilde{t}$. In this case, the platforms offer bonuses to achieve partial coverage of the supply pool in the competitive equilibrium. Thus, from Eq. (2.A.8), $U_P^* = \frac{2a^2Nt\left(t+(1-\gamma)N\right)^2}{\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)^2}$, $U_C^* = \frac{4a^2N^2\left(t+(1-\gamma)N\right)^2}{\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)^2}$, and $W^* = \frac{2a^2N\left(t+(1-\gamma)N\right)\left(3t^2+(7-3\gamma)Nt+2(1-\gamma)N^2\right)}{\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)^2}$. The platforms do not offer bonuses in the coordinated equilibrium, and the surplus and welfare expressions are shown in the second case of Eq. (2.A.9). Thus, we obtain that $U_P^* - U_P^\dagger = \frac{1}{2}(U_C^* - U_C^\dagger) = -\frac{2a^2Nt^2\left((1-2\gamma)t+2(1-\gamma)^2N\right)\left((3-2\gamma)t^2+2(1-\gamma)(4-\gamma)Nt+4(1-\gamma)^2N^2\right)}{\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)^2\left(t+2(1-\gamma)N\right)^2} > 0$ and $W^* - W^\dagger = -\frac{2a^2Nt^2\left((1-2\gamma)t+2(1-\gamma)^2N\right)\left((1+2\gamma)t^2+2(1+\gamma)(2-\gamma)Nt+4(1-\gamma)N^2\right)}{\left(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2\right)^2\left(t+2(1-\gamma)N\right)^2} > 0$. Thus, in Case (2), $U_P^\dagger < U_P^*$, $U_C^\dagger < U_C^*$, and $W^\dagger < W^*$.

Case (3): $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t \leq a - 2N$. In this case, the platforms offer bonuses to achieve full coverage of the supply pool in both competitive and coordinated equilibria, and $b_i^* = b_i^\dagger$ where $i = \{A, B\}$. Thus, $U_P^\dagger = U_P^*$, $U_C^\dagger = U_C^*$, and $W^\dagger = W^*$.

Case (4): $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $a-2N < t \le \tilde{t}$. In this case, the platforms offer bonuses $b_A^* = b_B^* = t - 2(1-\gamma)(a-N)$ to achieve full coverage of the supply pool in the competitive equilibrium, and hence U_P^*, U_C^*, and W^* are of the same expressions as in Case (1). The platforms offer smaller bonuses $b_A^\dagger = b_B^\dagger = \frac{(2\gamma-1)at-2(1-\gamma)aN}{2N+t}$ to achieve partial coverage of the supply pool in the coordinated equilibrium. From Eq. (2.A.8), we obtain that $U_P^\dagger = \frac{a^2Nt}{2(2N+t)^2}$, $U_C^\dagger = \frac{a^2N^2}{(2N+t)^2}$, and $W^\dagger = \frac{3a^2N}{2(2N+t)}$. Thus, we obtain that $U_P^\dagger - U_P^* = \frac{a^2Nt}{2(2N+t)^2} - \frac{Nt}{2} = -\frac{Nt(2N+a+t)(t-(a-2N))}{2(2N+t)^2} < 0$, $U_C^\dagger - U_C^* = \frac{a^2N^2}{(2N+t)^2} - N^2 = -\frac{N^2(2N+a+t)(t-(a-2N))}{(2N+t)^2} < 0$, and $W^\dagger - W^* = \frac{N(t-3a+2N)(t-a+2N)}{2(2N+t)}$. Thus, $W^\dagger > W^*$ if and only if $t > 3a-2N$. Moreover, $\tilde{t} - (3a-2N) = \frac{\sqrt{4a^2+(4\gamma^2-4\gamma+9)N^2-4(1+2\gamma)aN}-(10a-(3+2\gamma)N)}{4}$. Since $\gamma < 1$ and $\frac{a}{N} > 2$, $10a-(3+2\gamma)N > 0$. Hence, $\tilde{t} < 3a-2N$ if and only if $\left(4a^2+(4\gamma^2-4\gamma+9)N^2-4(1+2\gamma)aN\right)^2-(10a-(3+2\gamma)N)^2 = -8(12(\frac{a}{N})^2-(7+4\gamma)\frac{a}{N}+2\gamma) < 0$. For the quadratic function $12(\frac{a}{N})^2-(7+4\gamma)\frac{a}{N}+2\gamma$, $\Delta = 16\gamma^2-40\gamma+49 > 0$, and the two roots are positive but less than 2. Hence, when $\frac{a}{N} > 2$, $\tilde{t} < 3a-2N$, and $t \le \tilde{t}$ implies $t < 3a-2N$, and $W^\dagger < W^*$. Thus, in Case (4), $U_P^\dagger < U_P^*$, $U_C^\dagger < U_C^*$, and $W^\dagger < W^*$.

Case (5): $t > \frac{2(1-\gamma)N}{(2\gamma-1)^+}$ and $t > \tilde{t}$. In this case, the platforms offer bonuses to achieve partial coverage of the supply pool in the competitive equilibrium, and offer smaller bonuses to achieve less partial coverage of the supply pool in the coordinated equilibrium. Thus, U_P^*, U_C^*, and W^* are of the same expressions as in Case (2), and $U_P^\dagger$, $U_C^\dagger$, and $W^\dagger$ are of the same expressions as in Case (4). We then obtain that $U_P^\dagger - U_P^* = \frac{a^2Nt}{2(2N+t)^2} - \frac{2a^2Nt(t+(1-\gamma)N)^2}{(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2)^2} = \frac{a^2N^2t(t+2(1-\gamma)N)(-4t^2+(4\gamma-11)Nt-6(1-\gamma)N^2)}{2(2N+t)^2(2\gamma N^2+2\gamma Nt-2N^2-5Nt-2t^2)^2}$, and $U_C^\dagger - U_C^* = \frac{a^2N^2}{(2N+t)^2} - \frac{4a^2N^2(t+(1-\gamma)N)^2}{(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2)^2} = \frac{a^2N^3(t+2(1-\gamma)N)(-4t^2+(4\gamma-11)Nt-6(1-\gamma)N^2)}{2(2N+t)^2(2\gamma N^2+2\gamma Nt-2N^2-5Nt-2t^2)^2}$. Since $-4t^2+(4\gamma-11)Nt-6(1-\gamma)N^2 < 0$, we obtain that $U_P^\dagger < U_P^*$ and $U_C^\dagger < U_C^*$. Moreover, $W^\dagger - W^* = \frac{3a^2N}{2(2N+t)} - \frac{2a^2N(t+(1-\gamma)N)(3t^2+(7-3\gamma)Nt+2(1-\gamma)N^2)}{(2t^2+(5-2\gamma)Nt+2(1-\gamma)N^2)^2} = \frac{a^2N^2(t+2(1-\gamma)N)(-4t^2+(4\gamma-9)Nt-2(1-\gamma)N^2)}{2(2N+t)(2N^2\gamma+2\gamma Nt-2N^2-5Nt-2t^2)^2}$. Since $-4t^2+(4\gamma-9)Nt-2(1-\gamma)N^2 < 0$, we obtain that $W^\dagger < W^*$. Thus, in Case (5), $U_P^\dagger < U_P^*$, $U_C^\dagger < U_C^*$, and $W^\dagger < W^*$. Combining all the cases analyzed above, we obtain the following: 1) $U^\dagger < U_P^*$ always holds as long as $b_i^\dagger < b_i^*$; 2) $U_C^\dagger = U_C^*$ and $W^\dagger = W^*$ if $t < \gamma(a-N)$, and $U_C^\dagger < U_C^*$ and $W^\dagger < W^*$ if $t > \max\left(2(1-\gamma)(a-N), \frac{2(1-\gamma)^2N}{(2\gamma-1)^+}, a-2N\right)$. ■

2.8.1.6 Proof of Proposition 7

We first derive segmentation points (α_A, α_B), (β_A, β_B), (η_A, η_B), and market price p_2 in period 2. We then derive the expression of (z_A, z_B) in period 1. Finally, we analyze the platforms' equilibrium bonuses and welfare comparison.

- **Providers' Decisions in Period** 2

For a provider on line α in period 2, he prefers participating on platform A than platform B if $p_2(1-\gamma) - t\alpha + b_A \geq p_2(1-\gamma) - t(1-\alpha)$, i.e., $\alpha \leq \frac{1}{2} + \frac{b_A}{2t}$. When a provider at $\alpha = \frac{1}{2} + \frac{b_A}{2t}$ participates, his utility in period 2 is $u_2 = (1-\gamma)p_2 + \frac{b_A}{2} - \frac{t}{2}$. Thus, we obtain that (i) if $b_A \geq t - 2(1-\gamma)p_2$, line α is fully utilized, and $\alpha_A = \alpha_B = \min(1, \frac{1}{2} + \frac{b_A}{2t})$; (ii) if $b_A < t - 2(1-\gamma)p_2$, line α is partially utilized, and $\alpha_A = \frac{(1-\gamma)p_2 + b_A}{t}$ and $\alpha_B = 1 - \frac{(1-\gamma)p_2}{t}$, where $0 < \alpha_A < \alpha_B < 1$. Similarly, we obtain that (iii) if $b_B \geq t - 2(1-\gamma)p_2$, line β is fully utilized, and $\beta_A = \beta_B = \max(0, \frac{1}{2} - \frac{b_B}{2t})$; (iv) if $b_B < t - 2(1-\gamma)p_2$, line β is partially utilized, and $\beta_A = \frac{(1-\gamma)p_2}{t}$ and $\beta_B = 1 - \frac{(1-\gamma)p_2 + b_B}{t}$, where $0 < \beta_A < \beta_B < 1$. (v) if $t \leq 2(1-\gamma)p_2$, line η is fully utilized, and $\eta_A = \eta_B = \frac{1}{2}$; (vi) if $t > 2(1-\gamma)p_2$, line η is partially utilized, and $\eta_A = \frac{(1-\gamma)p_2}{t}$ and $\eta_B = 1 - \frac{(1-\gamma)p_2}{t}$, where $0 < \eta_A < \eta_B < 1$. In period 2, platform A covers $z_A\alpha_A + (z_B - z_A)\eta_A + (1 - z_B)\beta_A$ proportion of the supply pool, and platform B covers $z^A(1-\alpha^B) + (z^B - z^A)(1-\eta^B) + (1-z^B)(1-\beta^B)$ proportion of the supply pool. The market price in period 2 is

$$p_2 = a - N(1 - z_A(\alpha_B - \alpha_A) - (1 - z_B)(\beta_B - \beta_A) - (z_B - z_A)(\eta_B - \eta_A)). \qquad (2.A.11)$$

- **Providers' Decisions in Period** 1

In period 1, from Eq. (2.7) - (2.9), the providers' expected total utility from both periods can be simplified to

$$u(z) = \begin{cases} (1-\gamma)p_1 - tz + (1-\gamma)p_2(1 + \alpha_A - \alpha_B) + b_A\alpha_A - \frac{t(\alpha_A^2 + (1-\alpha_B)^2)}{2} & \text{if } z \leq z_A, \\ (1-\gamma)p_1 - t(1-z) + (1-\gamma)p_2(1 + \beta_A - \beta_B) + b_B(1-\beta_B) - \frac{t(\beta_A^2 + (1-\beta_B)^2)}{2} & \text{if } z \geq z_B. \end{cases} \qquad (2.A.12)$$

If the supply pool is fully utilized in period 1, then we obtain (z_A, z_B) from setting $u(z_A) = u(z_B)$ and $z_A = z_B$. Next, we show that the supply pool is fully utilized when $t \leq 2(1-\gamma)(a-N)$, which is implied when $t \leq \gamma(a-N)$.

If $t \leq 2(1-\gamma)(a-N)$, then $t \leq 2(1-\gamma)p_2$ holds since $p_2 \geq a - N$. Thus, line η is fully utilized with $\eta_A = \eta_B = \frac{1}{2}$. Moreover, since $b_A \geq 0$ and $b_B \geq 0$, $b_A \geq t - 2(1-\gamma)p_2$ and $b_B \geq t - 2(1-\gamma)p_2$ hold. Thus, line α is fully utilized with $\alpha_A = \alpha_B = \min(1, \frac{1}{2} + \frac{b_A}{2t})$, and line β is fully utilized with $\beta_A = \beta_B = \max(0, \frac{1}{2} - \frac{b_B}{2t})$. Thus, in period 2, all providers participate, and $p_2 = a - N$. Table 2.A.1 shows all possible cases for the segmentation points in period 2.

Due to the symmetry of the two platforms, in equilibrium $b_A = b_B$, and hence only Case (i) and Case (iv) can occur. If Case (i) occurs, then $u(z)$ simplifies to

$$u(z) = \begin{cases} (1-\gamma)p_1 - tz + (1-\gamma)(a-N) - \frac{t}{4} + \frac{b_A}{2} + \frac{b_A^2}{4t} & \text{if } z \leq z_A, \\ (1-\gamma)p_1 - t(1-z) + (1-\gamma)(a-N) - \frac{t}{4} + \frac{b_B}{2} + \frac{b_B^2}{4t} & \text{if } z \geq z_B. \end{cases}$$

	$b_B \le t$	$b_B > t$
$b_A \le t$	(i) $\alpha_A = \alpha_B = \frac{1}{2} + \frac{b_A}{2t}, \beta_A = \beta_B = \frac{1}{2} - \frac{b_B}{2t}$	(ii) $\alpha_A = \alpha_B = \frac{1}{2} + \frac{b_A}{2t}, \beta_A = \beta_B = 0$
$b_A > t$	(iii) $\alpha_A = \alpha_B = 1, \beta_A = \beta_B = \frac{1}{2} - \frac{b_B}{2t}$	(iv) $\alpha_A = \alpha_B = 1, \beta_A = \beta_B = 0$

Table 2.A.1: Second-period Segmentation Points If $t \le 2(1-\gamma)(a-N)$

Suppose the supply pool is fully utilized in period 2. We obtain that $z_A = z_B = \frac{1}{2} + \frac{(b_A+b_B+2t)(b_A-b_B)}{8t^2}$, and $Eu(z_A) = Eu(z_B) = \frac{b_A^2+b_B^2+2t(b_A+b_B)-6t^2+16(a-N)(1-\gamma)t}{8t}$. Since $t \le 2(1-\gamma)(a-N)$, $b_A^2 + b_B^2 + 2t(b_A+b_B) - 6t^2 + 16(a-N)(1-\gamma)t = b_A^2 + b_B^2 + 2t(b_A + b_B) - 6t(t - \frac{8}{3}(a-N)(1-\gamma)) \ge 0$. Thus, the supply pool is indeed fully utilized in period 1 and $p_1 = a - N$.

If Case (iv) occurs, then $u(z)$ simplifies to

$$u(z) = \begin{cases} (1-\gamma)p_1 - tz + (1-\gamma)(a-N) + b_A - \frac{t}{2} & \text{if } z \le z_A, \\ (1-\gamma)p_1 - t(1-z) + (1-\gamma)(a-N) + b_B - \frac{t}{2} & \text{if } z \ge z_B. \end{cases}$$

Suppose the supply pool is fully utilized in period 2, We obtain that $z_A = z_B = \frac{1}{2} + \frac{b_A-b_B}{2t}$, and $u(z_A) = u(z_B) = 2(1-\gamma)(a-N) - t + \frac{b_A+b_B}{2} \ge 0$ since $t \le 2(1-\gamma)(a-N)$. Thus, the supply pool is indeed fully utilized in period 1 and $p_1 = a - N$.

- **Platforms' Decisions**

Based on previous analysis, two scenarios can occur: (i) if $b_A \le t$ and $b_B \le t$, then $\alpha_A = \alpha_B = \frac{1}{2} + \frac{b_A}{2t}, \beta_A = \beta_B = \frac{1}{2} - \frac{b_B}{2t}, \eta_A = \eta_B = \frac{1}{2}, z_A = z_B = \bar{z} = \frac{1}{2} + \frac{(b_A+b_B+2t)(b_A-b_B)}{8t^2}$, $p_1 = p_2 = a - N$; (ii) if $b_A > t$ and $b_B > t$, then $\alpha_A = \alpha_B = 1, \beta_A = \beta_B = 0, \eta_A = \eta_B = \frac{1}{2}$, $z_A = z_B = \bar{z} = \frac{1}{2} + \frac{b_A-b_B}{2t}$, $p_1 = p_2 = a - N$.

Suppose scenario (i) occurs, then the profit function Π_A is

$$\Pi_A = N\left(\gamma(a-N) + \frac{b_B}{2t}\gamma(a-N) - \frac{1}{2t}b_A^2 + \frac{\gamma(a-N)-t}{2t}b_A\right)\bar{z} + \gamma(a-N)N\left(\frac{1}{2} - \frac{b_B}{2t}\right). \tag{2.A.13}$$

Denote $g(b_A) = (a-N)\gamma + \frac{b_B}{2t}(a-N)\gamma - \frac{1}{2t}(b_A)^2 + \frac{(a-N)\gamma-t}{2t}b_A$ and we obtain $\frac{\partial g(b_A)}{\partial b_A} = -\frac{1}{t}b_A + \frac{\gamma(a-N)-t}{2t}$. Thus, $\frac{\partial g(b_A)}{\partial b_A} \ge 0$ if $b_A \le \frac{\gamma(a-N)-t}{2}$ and $\frac{\partial g(b_A)}{\partial b_A} \le 0$ if $b_A \ge \frac{\gamma(a-N)-t}{2}$. If $t \le \gamma(a-N)$, we obtain that the axis of symmetry of $g(b_A)$ is positive. Since $g(0) = (a-N)\gamma + \frac{b_B}{2t}(a-N)\gamma > 0$, we obtain that $g(b_A)$ is increasing and positive for $b_A \in [0, \frac{\gamma(a-N)-t}{2}]$. Setting $g(b_A) = 0$, we obtain that the larger root is $\frac{\gamma(a-N)-t}{2} + \frac{\sqrt{\gamma^2(a-N)^2+6\gamma(a-N)t+4\gamma(a-N)b_B+t^2}}{2}$. Thus, $g(b_A)$ is decreasing in b_A and positive for $b_A \in [\frac{\gamma(a-N)-t}{2}, \frac{\gamma(a-N)-t}{2} + \frac{\sqrt{\gamma^2(a-N)^2+6\gamma(a-N)t+4\gamma(a-N)b_B+t^2}}{2}]$, and decreasing in b_A and negative for $b_A \in [\frac{\gamma(a-N)-t}{2} + \frac{\sqrt{\gamma^2(a-N)^2+6\gamma(a-N)t+4\gamma(a-N)b_B+t^2}}{2}, \infty)$. Moreover, $\bar{z}$ is increasing in b_A and positive for $b_A \geq 0$ as $\frac{\partial \bar{z}}{\partial b_A} = \frac{t+b_A}{4t^2} > 0$. From the rule of multiplication, we obtain that Π^A is increasing in b_A for $b_A \in [0, \frac{\gamma(a-N)-t}{2}]$ and decreasing in b_A for $b_A \in [\frac{\gamma(a-N)-t}{2} + \frac{\sqrt{\gamma^2(a-N)^2+6\gamma(a-N)t+4\gamma(a-N)b_B+t^2}}{2}, \infty)$. Therefore, the best response of platform A will satisfy $b_A(b_B) \in [\frac{\gamma(a-N)-t}{2}, \frac{\gamma(a-N)-t}{2} + \frac{\sqrt{\gamma^2(a-N)^2+6\gamma(a-N)t+4\gamma(a-N)b_B+t^2}}{2})$. First, the lower bound of $b_A(b_B)$ is strictly positive, i.e., $\frac{\gamma(a-N)-t}{2} > 0$. Second, for a function $h(x) = \sqrt[n]{\alpha + \beta x}$ with $\alpha > 0$ and $\beta > 0$, we have $h'(x) = \frac{\beta}{2\sqrt{\alpha+\beta x}} > 0$ and $h''(x) = -\frac{\beta^2}{4}(\alpha + \beta x)^{-\frac{3}{2}} < 0$. Thus, $h(x)$ is increasing in x at a decreasing rate, and $h'(x) \to 0$ if $x \to \infty$. Thus, we obtain that the upper bound of $b_A(b_B)$ is increasing in b_B at a decreasing rate, and $\frac{\partial}{\partial b_B}\left(\frac{\gamma(a-N)-t}{2} + \frac{\sqrt{\gamma^2(a-N)^2+6\gamma(a-N)t+4\gamma(a-N)b_B+t^2}}{2}\right) \to 0$ if $b_B \to \infty$. This implies that the upper bound of $b_A(b_B)$ will intersect with the 45 degree line at a sufficiently large b_B. Thus, $b_A(b_B) > 0$ and intersects with the 45 degree line. Similarly, we can obtain that $b_B(b_A) > 0$ and intersects with the 45 degree line. Due to symmetry, we obtain that $b_A^* = b_B^* > 0$ in scenario (i).

Suppose scenario (ii) occurs, then the profit function Π_A is

$$\Pi_A = \gamma(a-N)N + \left(\frac{\gamma(a-N)N}{t} - \frac{N}{2} + \frac{Nb_B}{2t}\right)b_A - \frac{N}{2t}b_A^2 - \frac{\gamma(a-N)N}{t}b_B. \quad (2.A.14)$$

The first-order derivative is $\frac{\partial \Pi_A}{\partial b_A} = \frac{\gamma(a-N)N}{t} - \frac{N}{2} - \frac{N}{t}b_A + \frac{Nb_B}{2t}$. Thus, the unconstrained optimizer is $b_A(b_B) = \gamma(a-N) - \frac{t}{2} + \frac{b_B}{2} > 0$. Similarly, $b_B(b_A) = \gamma(a-N) - \frac{t}{2} + \frac{b_A}{2}$. Thus, $b_A^* = b_B^* = 2\gamma(a-N) - t > 0$, and $b_A^* = b_B^* > t$ since $t < \gamma(a-N)$.

Thus, when $t < \gamma(a-N)$, the platforms offer bonuses in the competitive equilibrium, i.e., $b_A^* = b_B^* > 0$. When the platforms coordinate, they choose $b_A = b_B = b$ to maximize total profits $\Pi_A + \Pi_B$, where $\Pi_A + \Pi_B = 2N(\gamma(a-N) - \frac{1}{4t}b^2 - \frac{1}{4}b)$. Since $\frac{\partial(\Pi_A+\Pi_B)}{\partial b} = -\frac{1}{2}(\frac{1}{2} + \frac{b}{t}) < 0$, we obtain that $b_A^\dagger = b_B^\dagger = 0$.

- **Platform Profit and Welfare Comparison**

Since $b_A = b_B$ in both competitive equilibrium and coordinated equilibria, we obtain that $z_A = z_B = \frac{1}{2}$, $\alpha_A = \alpha_B = \min(1, \frac{1}{2} + \frac{b}{2t})$, $\beta_A = \beta_B = \max(0, \frac{1}{2} - \frac{b}{2t})$ when $b_A = b_B = b$. We derive the comparison of the competitive and coordinated equilibria in the following two cases.

Case (1): $b < t$. In this case, $\alpha_A = \alpha_B = \frac{1}{2} + \frac{b}{2t}$ and $\beta_A = \beta_B = \frac{1}{2} - \frac{b}{2t}$. The platform profits simplify to $\Pi_A = \Pi_B = N(\gamma(a-N) - \frac{1}{4t}b^2 - \frac{1}{4}b)$ in both equilibria. Since $b_i^* > b_i^\dagger = 0$, and $b = 0$ is each platform's profit optimizer, we obtain that $\Pi_i^* < \Pi_i^\dagger$. From Eq. (2.10), the provider surplus simplifies to

$U_P = N\left(\frac{3(1-\gamma)(a-N)}{2} - \frac{3t}{8} + \frac{b}{2} + \frac{b^2}{4t}\right)$. The consumer surplus simplifies to $U_C = N^2$, and the social welfare simplifies to $W = N\left(\frac{(\gamma+3)(a-N)}{2} - \frac{3t}{8} - \frac{b^2}{4t}\right)$. Since $\frac{\partial U_P}{\partial b} = \frac{1}{2} + \frac{b}{2t} > 0$, we obtain $U_P^* > U_P^\dagger$. U_C is a constant and hence $U_C^* = U_C^\dagger$. Since $\frac{\partial W}{\partial b} = -\frac{1}{4t} < 0$, we obtain $W^* < W^\dagger$.

Case (2): $b > t$. In this case, $\alpha_A = \alpha_B = 1$ and $\beta_A = \beta_B = 0$. The platform profits simplify to $\Pi_A = \Pi_B = \gamma(a-N)N - \frac{bN}{2}$ in both equilibria. Since $b_i^* > b_i^\dagger = 0$, and $b = 0$ is each platform's profit optimizer, we obtain that $\Pi_i^* < \Pi_i^\dagger$. From Eq. (2.10), the provider surplus simplifies to $U_P = 2(1-\gamma)(a-N)N - \frac{3Nt}{4} + \frac{Nb}{2}$. The consumer surplus simplifies to $U_C = N^2$, and the social welfare simplifies to $W = 2(a-N)N - -\frac{3Nt}{4} + N^2 - \frac{Nb}{2}$. Since $\frac{\partial U_P}{\partial b} = \frac{N}{2} > 0$, we obtain $U_P^* > U_P^\dagger$. U_C is a constant and hence $U_C^* = U_C^\dagger$. Since $\frac{\partial W}{\partial b} = -\frac{N}{2} < 0$, we obtain $W^* < W^\dagger$.

Combining both cases, we obtain that $\Pi_i^* < \Pi_i^\dagger$, $U_P^* > U_P^\dagger$, $U_C^* = U_C^\dagger$, $W^* < W^\dagger$. ■

2.8.2 Model Extensions: Endogenous Commission Rates

The commission rates are assumed to be exogenous in the discussion heretofore. This assumption abstracts away the various pricing levers that platforms may use in the real world such as commission rates and fares charged to consumers. The merits of this assumption is that it helps more clearly convey the main points of the paper, since the bonus variable fully captures the subsidy level. Anecdotal evidence also suggests that platforms adjust their commission rates infrequently, and their values are typically close to a market benchmark (e.g., 20%-30% for ride-hailing and 15%-20% for domestic tasks.) [6] This section discusses some results and insights when the commission rates are endogenous in the model. Considering endogenous commission rates can explore the relationship between different pricing levers.

Let γ_A and γ_B be the commission rate charged by platform A and B, respectively. Let b_A and b_B be the bonus offered by platform A and B, respectively. To capture the industry practice that the commission rates are a relatively long-term decision that are adjusted infrequently, the decision sequence in the model is that the competing platforms decide their commission rates prior to their bonuses in period 1. Then the providers make their participation decisions and the marketplace is cleared in period 2. Lemma 2 shows the conditions for the supply pool utilizatoin.

[6](see [88], `https://support.taskrabbit.com/hc/en-us/articles/204411610-What-is-the-TaskRabbit-Service-Fee-` and `https://www.fastcompany.com/3042248/the-gig-economy-wont-last-because-its-being-sued-to-death`).

Lemma 2 *The supply pool achieves full utilization if* $b_A + b_B \geq t - (a-N)(2-(\gamma_A+\gamma_B))$ *and partial utilization otherwise.*

Lemma 2 shows that the supply pool will be fully utilized when the providers are sufficiently compensated, either by a high bonus or a low commission rate as provided by the competing platforms. Proposition 8 further shows that commission rates and bonuses are substitutable pricing levers.

Proposition 8 *The platforms can achieve the same profit level with a menu of commission rates and bonus schemes. Specifically, the following two menus can emerge when the platforms' profits are* $\Pi_A = \Pi_B = \frac{Nt}{2}$*:*

(i) $\gamma_A^* = \gamma_B^* = \frac{t}{a-N}$*, and* $b_A^* = b_B^* = 0$.

(ii) $\gamma_A^* > \frac{t}{a-N}$ *and* $\gamma_B^* > \frac{t}{a-N}$*, and* $b_A^* = \gamma_A(a-N) - t$ *and* $b_B^* = \gamma_B(a-N) - t$.

Proposition 8 shows that when the platforms have the flexibility to adjust both commission rates and bonuses, there exist multiple states of the world. As shown in Proposition 8, with the same profitability, platforms can choose not to offer bonuses and entirely rely on commission rates to coordinate the marketplace. Platforms can also choose to offer bonuses, in which case they would charge a relatively higher commission rates. Moreover, the competing platforms may opt for different pricing schemes and still achieve the same profitability level. The case (ii) in Proposition 8 shows that b_A^* is higher than b_B^* when γ_A^* is lower than γ_B^*. The technical details of section 2.8.2 are presented in section 2.8.2.1.

2.8.2.1 Technical Details of the Equilibrium Analysis

2.8.2.1.1 Providers' Decisions and Lemma 2 A provider at x obtains utility u_A if he participates on platform A, and u_B if he participates on platform B, and

$$\begin{aligned} u_A &= p(1-\gamma_A) - tx + b_A, \\ u_B &= p(1-\gamma_B) - t(1-x) + b_B. \end{aligned}$$

First, suppose that the supply pool is fully utilized. The sufficient and necessary conditions for this case is that there exists an x_0 such that $u_A(x_0) = u_B(x_0) \geq 0$. From $u_A(x) = u_B(x)$, i.e., $p(1-\gamma_A) - tx + b_A = p(1-\gamma_B) - t(1-x) + b_B$, where $p = a - N$, x_0 is derived such that

$$x_0 = \frac{1}{2} + \frac{b_A - b_B}{2t} - \frac{(a-N)(\gamma_A - \gamma_B)}{2t}.$$

Hence, the provider's utility at x_0 is

$$u_0 = (a - N)(1 - \frac{\gamma_A + \gamma_B}{2}) - \frac{t}{2} + \frac{b_A + b_B}{2}.$$

Thus, $u_0 \geq 0$ is equivalent to

$$b_A + b_B \geq t - (a - N)(2 - (\gamma_A + \gamma_B)). \tag{2.A.15}$$

Eq. (2.A.15) is the condition for full supply pool utilization and is officially shown in Lemma 2.

Second, suppose that the supply pool is partially utilized. Denote the cutoff points as x_A and x_B. Then from $u_A(x_A) = 0$, $x_A = \frac{p(1-\gamma_A)+b_A}{t}$. From $u_B(x_B) = 0$, $1 - x_B = \frac{p(1-\gamma_B)+b_B}{t}$. Combined with $p = a - (x_A + 1 - x_B)N$, it can be derived that

$$p = \frac{at - (b_A + b_B)N}{(2 - (\gamma_A + \gamma_B))N + t}, \tag{2.A.16}$$

$$x_A = \frac{(1 - \gamma_A)a + N\frac{(1-\gamma_B)b_A-(1-\gamma_A)b_B}{t} + b_A}{(2 - \gamma_A - \gamma_B)N + t}, \tag{2.A.17}$$

$$1 - x_B = \frac{(1 - \gamma_B)a + N\frac{(1-\gamma_A)b_B-(1-\gamma_B)b_A}{t} + b_B}{(2 - \gamma_A - \gamma_B)N + t}. \tag{2.A.18}$$

Here, partial utilization requires that $x_A + 1 - x_B < 1$, i.e., $b_A + b_B < t - (a - N)(2 - \gamma_A - \gamma_B)$.

2.8.2.1.2 Platform's Profit Functions per Supply Pool Utilization

Platform i's profit function (i = A or B) is $\Pi_i = p\gamma_i Q_i - b_i Q_i$.

First, suppose that the supply pool is fully utilized, i.e., $b_A + b_B \geq t - (a - N)(2 - (\gamma_A + \gamma_B))$. In this case, $Q_A = x_0 N = (\frac{1}{2} + \frac{b_A - b_B}{2t} - \frac{(a-N)(\gamma_A-\gamma_B)}{2t})N$, and $Q_B = (1 - x_0)N = (\frac{1}{2} + \frac{b_B - b_A}{2t} - \frac{(a-N)(\gamma_B-\gamma_A)}{2t})N$. Thus, the platforms' profit functions are

$$\begin{aligned} \Pi_A^F/N &= ((a - N)\gamma_A - b_A)(\tfrac{1}{2} + \tfrac{b_A - b_B}{2t} - \tfrac{(a-N)(\gamma_A-\gamma_B)}{2t}), \\ \Pi_B^F/N &= ((a - N)\gamma_B - b_B)(\tfrac{1}{2} + \tfrac{b_B - b_A}{2t} - \tfrac{(a-N)(\gamma_B-\gamma_A)}{2t}). \end{aligned}$$

Second, suppose that the supply pool is partially utilized, i.e., $b_A + b_B < t - (a - N)(2 - (\gamma_A + \gamma_B))$. The platform's profit functions are $\Pi_A/N = x_A(p\gamma_A - b_A)$ and $\Pi_B/N = (1 - x_B)(p\gamma_B - b_B)$, i.e.,

$$\Pi_A^P/N = \frac{(1 - \gamma_A)a + N\frac{(1-\gamma_B)b_A-(1-\gamma_A)b_B}{t} + b_A}{(2 - \gamma_A - \gamma_B)N + t} \times \left(\frac{at - (b_A + b_B)N}{(2 - (\gamma_A + \gamma_B))N + t}\gamma_A - b_A\right),$$

$$\Pi_B^P/N = \frac{(1 - \gamma_B)a + N\frac{(1-\gamma_A)b_B-(1-\gamma_B)b_A}{t} + b_B}{(2 - \gamma_A - \gamma_B)N + t} \times \left(\frac{at - (b_A + b_B)N}{(2 - (\gamma_A + \gamma_B))N + t}\gamma_B - b_B\right).$$

In the remainder of the equilibrium analysis, we omit constant N for ease of exposition.

2.8.2.1.3 Bonus decisions b_A and b_B In the equilibrium analysis, backward induction first derives the equilibrium of each platform's bonus decision.

- **Best Responses**

The equilibrium analysis first derives the best responses of each platform's bonus. For ease of exposition, the remainder of the proof uses platform A's best response $b_A(b_B)$ to illustrate. Lemma 2 shows that the supply pool is fully utilized if $b_A \geq \tilde{b}_A$ where $\tilde{b}_A = t-(a-N)(2-\gamma_A-\gamma_B)-b_B$. Thus, Π_A is a piecewise function such that $\Pi_A = \Pi_A^P$ if $b_A < \tilde{b}_A$ and $\Pi_A = \Pi_A^F$ if $b_A \geq \tilde{b}_A$. The first-order derivatives of Π_A w.r.t. b_A under full utilization and under partial utilization are in Eq. (2.A.19) and Eq. (2.A.20), and

$$\frac{\partial \Pi_A^F}{\partial b_A} = -\frac{1}{2} - \frac{b_A}{t} + \frac{b_B}{2t} + \frac{(a-N)(2\gamma_A-\gamma_B)}{2t} \tag{2.A.19}$$

$$\frac{\partial \Pi_A^P}{\partial b_A} = -\frac{2((1-\gamma_B)N+t)((2-\gamma_B)N+t)b_A - ((2-3\gamma_A-\gamma_B+2\gamma_A\gamma_B)N+(1-2\gamma_A)t)(Nb_B-at)}{t((2-\gamma_A-\gamma_B)N+t)^2} \tag{2.A.20}$$

From setting $\frac{\partial \Pi_A^F}{\partial b_A}=0$, the unconstrained optimizer of Π_A^F is $b_A^F = \frac{(2\gamma_A-\gamma_B)(a-N)}{2} - \frac{t}{2} + \frac{b_B}{2}$. When evaluated at $b_A = \tilde{b}_A$, $\frac{\partial \Pi_A^F}{\partial b_A}|_{b_A=\tilde{b}_A} = \frac{3}{2t}b_B + \frac{(4-3\gamma_B)(a-N)}{2t} - \frac{3}{2}$. From setting $\frac{\partial \Pi_A^P}{\partial b_A} = 0$, the unconstrained optimizer of Π_A^P is $b_A^P = \frac{(Nb_B-at)\left((2-3\gamma_A-\gamma_B+2\gamma_A\gamma_B)N+(1-2\gamma_A)t\right)}{2\left((1-\gamma_B)N+t\right)\left((2-\gamma_B)N+t\right)}$. When evaluated at $b_A = \tilde{b}_A$, $\frac{\partial \Pi_A^P}{\partial b_A}|_{b_A=\tilde{b}_A} = \frac{(3-2\gamma_B)N+2t}{t((2-\gamma_A-\gamma_B)N+t)}b_B - \frac{2t^2-(3-2\gamma_B)(a-2N)t-2(2-\gamma_B)(1-\gamma_B)(a-N)N}{t((2-\gamma_A-\gamma_B)N+t)}$. Five thresholds of b_B can affect the shape of Π_A, and they are $b_0 = t-(a-N)(2-\gamma_A-\gamma_B)$, $b_1 = t - \frac{(4-3\gamma_B)(a-N)}{3}$, $b_2 = \frac{2t^2-(3-2\gamma_B)(a-2N)t-2(2-\gamma_B)(1-\gamma_B)(a-N)N}{(3-2\gamma_B)N+2t}$, $b_3 = t-(2\gamma_A-\gamma_B)(a-N)$, $b_4 = \frac{at}{N}$.

First, threshold b_0 determines the number of pieces in the profit function Π_A for $b_A \geq 0$. If $b_B \geq b_0$, there is one piece, and $\Pi_A = \Pi_A^F$. If $b_B < b_0$, there are two pieces, and $\Pi_A = \Pi_A^P$ for $b_A < b_0 - b_B$ and $\Pi_A = \Pi_A^F$ for $b_A \geq b_0 - b_B$. Note that $b_0 \leq 0$ implies that $b_B \geq b_0$. Second, thresholds b_1 and b_2 determine the signs of the first-order derivatives of Π_A^F and Π_A^P at $\tilde{b}_A$, respectively. $\frac{\partial \Pi_A^F}{\partial b_A}|_{b_A=\tilde{b}_A} \geq 0$ if and only if $b_B \geq b_1$ and $\frac{\partial \Pi_A^P}{\partial b_A}|_{b_A=\tilde{b}_A} \geq 0$ if and only if $b_B \geq b_2$. Third, thresholds b_3 and b_4 determine the signs of the unconstrained optimizers of Π_A^F and Π_A^P, respectively. $b_A^F \geq 0$ if and only if $b_B \geq b_3$. If $(2-3\gamma_A-\gamma_B+2\gamma_A\gamma_B)N+(1-2\gamma_A)t \geq 0$, $b_A^P \geq 0$ if and only if $b_B \geq b_4$; if $(2-3\gamma_A-\gamma_B+2\gamma_A\gamma_B)N+(1-2\gamma_A)t < 0$, $b_A^P \geq 0$ if and only if $b_B \leq b_4$. Denote $y = f(\gamma_A,\gamma_B) = 2\gamma_A\gamma_B - 3\gamma_A - \gamma_B + 2$. Since $\frac{\partial f(\gamma_A,\gamma_B)}{\partial \gamma_A} = 2\gamma_B - 3 < 0$,

$f(\gamma_A, \gamma_B)$ decreases in γ_A. Since $\frac{f(\gamma_A,\gamma_B)}{\partial \gamma_B} = 2\gamma_A - 1$, $f(\gamma_A, \gamma_B)$ decreases in γ_B if $\gamma_A < \frac{1}{2}$ and increases in γ_B if $\gamma_A \geq \frac{1}{2}$. Thus, $1 - \gamma_A \leq y \leq 2 - 3\gamma_A$ if $0 \leq \gamma_A < \frac{1}{2}$, and $2 - 3\gamma_A \leq y \leq 1 - \gamma_A$ if $\frac{1}{2} < \gamma_A \leq 1$. Thus, $y \geq 0$ holds if $\gamma_A \leq \frac{2}{3}$. If $\gamma_A > \frac{2}{3}$, both $y < 0$ and $y > 0$ can occur. Recall that $\gamma_i \leq \frac{2}{3}$ is assumed in the model of exogenous commission rates. When $\gamma_A \leq \frac{2}{3}$, $(2-3\gamma_A-\gamma_B+2\gamma_A\gamma_B)N+(1-2\gamma_A)t \geq 0$ is equivalent to $t \leq \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$. Without loss of insights, the remainder of the equilibrium analysis continues to assume $\gamma_A \leq \frac{2}{3}$ and $\gamma_B \leq \frac{2}{3}$.

Next, the comparison among thresholds from b_0 to b_4 are shown. Note that $b_0 - b_1 = (a - N)(\gamma_A - \frac{2}{3})$, and hence $b_0 \leq b_1$ if $\gamma_A \leq \frac{2}{3}$. Note that $b_0 - b_2 = -\frac{(a-N)\left((2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N+(1-2\gamma_A)t\right)}{(3-2\gamma_B)N+2t}$, and hence $b_0 \leq b_2$ if and only if $(2\gamma_A\gamma_B - 3\gamma_A - \gamma_B + 2)N + (1 - 2\gamma_A)t \geq 0$, i.e., $\gamma_A \leq \frac{2}{3}$ and $t \leq \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$. Note that $b_0 - b_3 = (a - N)(3\gamma_A - 2)$, and hence $b_0 \leq b_3$ if and only if $\gamma_A \leq \frac{2}{3}$. Note that $b_0 - b_4 = (1 - \frac{a}{N})t - (a - N)(2 - \gamma_A - \gamma_B) < 0$, and hence $b_0 < b_4$. Note that $b_1 - b_2 = \frac{(a-N)(t-\gamma_B N)}{(9-6\gamma_B)N+6t}$, and hence $b_1 \geq b_2$ if and only if $t \geq r_B N$. Note that $b_1 - b_3 = (a - N)(2\gamma_A - \frac{3}{4})$, and hence $b_1 \leq b_3$ if and only if $\gamma_A \leq \frac{2}{3}$. Note that $b_1 - b_4 = -\frac{(a-N)t}{N} - \frac{(4-3\gamma_B)(a-N)}{3}$, and hence $b_1 < b_4$. Note that $b_2 - b_3 = -\frac{(a-N)\left((4\gamma_A\gamma_B-6\gamma_A-3\gamma_B+4)N+(3-4\gamma_A)t\right)}{(3-2\gamma_B)N+2t}$, and hence $b_2 \geq b_3$ if and only if $(4\gamma_A\gamma_B - 6\gamma_A - 3\gamma_B + 4)N + (3 - 4\gamma_A)t \leq 0$. Note that $b_2 - b_4 = -\frac{2(a-N)\left(t^2+(3-2\gamma_B)Nt+(\gamma_B^2-3\gamma_B+2)N^2\right)}{((3-2\gamma_B)N+2t)N}$, and hence $b_2 < b_4$. Note that $b_3 - b_4 = (a - N)\frac{(\gamma_B-2\gamma_A)N-t}{N}$, and hence $b_3 \leq b_4$ if and only if $t \geq (\gamma_B - 2\gamma_A)N$.

Since $\gamma \leq \frac{2}{3}$, the comparison of the five thresholds shows that $b_4 = \max(b_0, b_1, b_2)$, and $b_3 \geq b_1 \geq b_0$. Note that $b_0 \leq b_2$ if and only if $t \leq \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$, and $b_1 \geq b_2$ if and only if $t \geq \gamma_B N$. Moreover, $\gamma_B N \leq \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$ given that $\gamma \leq \frac{2}{3}$. Thus, $b_A(b_B)$ can be derived in two cases: case (A1) $t \leq \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$; case (A2) $t > \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$.

Case (A1): $t \leq \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$. In this case, $b_0 \leq b_1 < b_2$ if $t < \gamma N$, and $b_0 \leq b_2 \leq b_1$ if $t \geq \gamma N$. (i) If $b_B < b_0$, $\Pi_A = \Pi_A^P$ and is decreasing in b_A for $0 \leq b_A < b_0 - b_B$; $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq b_0 - b_B$. Thus, $b_A(b_B) = 0$. (ii) If $b_0 \leq b_B \leq b_3$, $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq 0$. Thus, $b_A(b_B) = 0$. (iii) If $b_B > b_3$, $\Pi_A = \Pi_A^F$ and is first increasing in b_A for $0 \leq b_A \leq b_A^F$ then decreasing in b_A for $b_A \geq b_A^F$. Thus, $b_A(b_B) = b_A^F$.

Case (A2): $t > \frac{(2\gamma_A\gamma_B-3\gamma_A-\gamma_B+2)N}{(2\gamma_A-1)^+}$. In this case, $b_2 < b_0 \leq b_1$. (i) If $b_B < b_2$, $\Pi_A = \Pi_A^P$ and is first increasing in b_A for $0 \leq b_A \leq b_A^P$ then decreasing in b_A for $b_A^P \leq b_A < b_0 - b_B$; $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq b_0 - b_B$. Thus, $b_A(b_B) = b_A^P$. (ii) If $b_2 \leq b_B < b_0$, $\Pi_A = \Pi_A^P$ and is increasing in b_A for $0 \leq b_A < b_0 - b_B$; $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq b_0 - b_B$. Thus, $b_A(b_B) = b_0 - b_B$. (iii) If $b_0 \leq b_B \leq b_3$, $\Pi_A = \Pi_A^F$ and is decreasing in b_A for $b_A \geq 0$. Thus, $b_A(b_B) = 0$. (iv) If

$b_B > b_3$, $\Pi_A = \Pi_A^F$ and is first increasing in b_A for $0 \le b_A \le b_A^F$ then decreasing in b_A for $b_A \ge b_A^F$. Thus, $b_A(b_B) = b_A^F$.

Correspondingly, the best response of platform B has case (B1) and case (B2), which share the similar structure as case (A1) and case (A2). Denote $t_i = \frac{2\gamma_i\gamma_j - 3\gamma_i - \gamma_j + 2)N}{(2\gamma_i - 1)^+}$, where $i, j = A, B$ and $i \neq j$. The best responses of platform i, i.e., $b_i(b_j)$, are in Eq. (2.A.21) if $t \le t_i$ and in Eq. (2.A.22) if $t > t_i$:

$$b_i(b_j) = \begin{cases} 0 & \textit{if } b_j \le t - (2\gamma_i - \gamma_j)(a - N), \\ b_i^F = \frac{(2\gamma_i - \gamma_j)(a-N)}{2} - \frac{t}{2} + \frac{b_j}{2} & \textit{if } b_j > t - (2\gamma_i - \gamma_j)(a - N). \end{cases} \tag{2.A.21}$$

The supply pool utilization in Eq. (2.A.21) is (i) full or partial and (ii) full.

$$b_i(b_j) = \begin{cases} b_i^P = \frac{(Nb_j - at)\left((2-3\gamma_i-\gamma_j+2\gamma_i\gamma_j)N+(1-2\gamma_i)t\right)}{2\left((1-\gamma_j)N+t\right)\left((2-\gamma_j)N+t\right)} & \textit{if } b_j < \frac{2t^2-(3-2\gamma_j)(a-2N)t-2(2-\gamma_j)(1-\gamma_j)(a-N)N}{(3-2\gamma_j)N+2t}, \\ t - (2-\gamma_i-\gamma_j)(a-N) - b_j & \textit{if } \frac{2t^2-(3-2\gamma_j)(a-2N)t-2(2-\gamma_j)(1-\gamma_j)(a-N)N}{(3-2\gamma_j)N+2t} \le b_j \\ & \quad < t - (2-\gamma_i-\gamma_j)(a-N), \\ 0 & \textit{if } t - (2-\gamma_i-\gamma_j)(a-N) \le b_j \\ & \quad \le t - (2\gamma_i - \gamma_j)(a-N), \\ b_i^F = \frac{(2\gamma_i-\gamma_j)(a-N)}{2} - \frac{t}{2} + \frac{b_j}{2} & \textit{if } b_j > t - (2\gamma_i - \gamma_j)(a-N). \end{cases} \tag{2.A.22}$$

The supply pool utilization in Eq. (2.A.22) is (i) partial, (ii) full, (iii) full, and (iv) full.

- **Equilibrium**

Recall that $t_A = \frac{(2\gamma_A\gamma_B - 3\gamma_A - \gamma_B + 2)N}{(2\gamma_A - 1)^+}$ and $t_B = \frac{(2\gamma_A\gamma_B - 3\gamma_B - \gamma_A + 2)N}{(2\gamma_B - 1)^+}$. Note that $t_A = +\infty$ if $\gamma_A \le \frac{1}{2}$ and $t_B = +\infty$ if $\gamma_B \le \frac{1}{2}$. If $\gamma_A > \frac{1}{2}$ and $\gamma_B > \frac{1}{2}$, then $t_A - t_B = -\frac{2(\gamma_A - \gamma_B)(1-\gamma_A-\gamma_B+2\gamma_A\gamma_B)}{(2\gamma_A-1)(2\gamma_B-1)}$. Thus, $t_A \ge t_B$ if and only if $(\gamma_A - \gamma_B)(1-\gamma_A-\gamma_B+2\gamma_A\gamma_B) \le 0$. Denote $y = f(\gamma_A, \gamma_B) = 1 - \gamma_A - \gamma_B + 2\gamma_A\gamma_B$. Since $\frac{\partial y}{\partial \gamma_A} = 2\gamma_B - 1 > 0$ and $\frac{\partial y}{\partial \gamma_B} = 2\gamma_A - 1 > 0$, y is increasing in both γ_A and γ_B. Thus, $\min y = f(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}$, and $\max y = f(\frac{2}{3}, \frac{2}{3}) = \frac{5}{9}$. Thus, $y > 0$ holds and $t_A \ge t_B$ if and only if $\gamma_A \le \gamma_B$. Thus, the following three scenarios can emerge in equilibrium:

(1) If $\gamma_A \le \frac{1}{2}$ and $\gamma_B \le \frac{1}{2}$, then $t \le t_A = +\infty$ and $t \le t_B = +\infty$. Thus, it is case A1-B1.

(2) If $\frac{1}{2} < \gamma_B \leq \gamma_A \leq \frac{2}{3}$, then $0 < t_A \leq t_B$. Thus, it is case A1-B1 for $t \leq t_A$, case A2-B1 for $t_A < t \leq t_B$, and case A2-B2 for $t > t_B$.

(3) If $\frac{1}{2} < \gamma_A \leq \gamma_B \leq \frac{2}{3}$, then $0 < t_B \leq t_A$. Thus, it is case A1-B1 for $t \leq t_B$, case A1-B2 for $t_B < t \leq t_A$, and case A2-B2 for $t > t_A$.

Without losing the main insights, the remainder of the proof focuses on the discussion of case A1-B1, i.e., $t \leq t_A$ and $t \leq t_B$. In this case, since there are two best response candidates for both platforms, the following four equilibrium outcomes (with supply utilization level noted) can emerge.

① $b_A^* = b_B^* = 0$.

This bonus equilibrium is derived from case A1(i) - B1(i) (Partial or Full). Best responses $b_A(b_B) = 0$ and $b_B(b_A) = 0$ induce this equilibrium. This equilibrium requires that $t - (2\gamma_A - \gamma_B)(a - N) \geq 0$ and $t - (2\gamma_B - \gamma_A)(a - N) \geq 0$, i.e., $2\gamma_A - \gamma_B \leq \frac{t}{a-N}$ and $2\gamma_B - \gamma_A \leq \frac{t}{a-N}$.

② $b_A^* = \gamma_A(a - N) - t$, and $b_B^* = \gamma_B(a - N) - t$.

This bonus equilibrium is derived from case A1(ii) - B1(ii) (Full). Best responses $b_A(b_B) = \frac{(2\gamma_A-\gamma_B)(a-N)}{2} - \frac{t}{2} + \frac{b_B}{2}$ and $b_B(b_A) = \frac{(2\gamma_B-\gamma_A)(a-N)}{2} - \frac{t}{2} + \frac{b_A}{2}$ induce this equilibrium. This equilibrium requires that $b_A^* > t-(2\gamma_B-\gamma_A)(a-N)$, i.e., $\gamma_B > \frac{t}{a-N}$, and $b_B^* > t - (2\gamma_A - \gamma_B)(a - N)$, i.e., $\gamma_A > \frac{t}{a-N}$.

③ $b_A^* = 0$ and $b_B^* = \frac{(2\gamma_B-\gamma_A)(a-N)-t}{2}$.

This bonus equilibrium is derived from case A1(i) - B1(ii) (Partial or Full). Best responses $b_A(b_B) = 0$ and $b_B(b_A) = \frac{(2\gamma_B-\gamma_A)(a-N)}{2} - \frac{t}{2} + \frac{b_A}{2}$ induce this equilibrium. This equilibrium requires that $b_B^* \leq t - (2\gamma_A - \gamma_B)(a - N)$, i.e., $\gamma_A \leq \frac{t}{a-N}$, and $b_A^* > t - (2\gamma_B - \gamma_A)(a - N)$, i.e., $t - (2\gamma_B - \gamma_A)(a - N) < 0$, i.e., $2\gamma_B - \gamma_A > \frac{t}{a-N}$.

④ $b_A^* = \frac{(2\gamma_A-\gamma_B)(a-N)-t}{2}$ and $b_B^* = 0$.

This bonus equilibrium is derived from case A1(ii) - B1(i) (Partial or Full). Best responses $b_A(b_B) = \frac{(2\gamma_A-\gamma_B)(a-N)}{2} - \frac{t}{2} + \frac{b_B}{2}$ and $b_B(b_A) = 0$ induce this equilibrium. This equilibrium requires that $b_B^* > t - (2\gamma_A - \gamma_B)(a - N)$, i.e., $2\gamma_A - \gamma_B > \frac{t}{a-N}$, and $b_A^* \leq t - (2\gamma_B - \gamma_A)(a - N)$, i.e., $\gamma_B \leq \frac{t}{a-N}$. Table 2.A.2 summarizes the conditions on γ_i and the corresponding bonus equilibrium.

2.8.2.1.4 Commission decisions γ_A and γ_B After obtaining the equilibrium of the bonus decisions, the equilibrium analysis next derives the commission rates decisions. Let $\bar{\gamma}$ denote a potential upper bound for the commission rates such that $\bar{\gamma} \leq \frac{2}{3}$. Due to symmetry, the bonus equilibria that can emerge are ①

	$2\gamma_A - \gamma_B \le \frac{t}{a-N}$, $2\gamma_B - \gamma_A \le \frac{t}{a-N}$	$\gamma_A > \frac{t}{a-N}$, $\gamma_B > \frac{t}{a-N}$	$\gamma_A \le \frac{t}{a-N}$, $2\gamma_B - \gamma_A > \frac{t}{a-N}$	$2\gamma_A - \gamma_B > \frac{t}{a-N}$, $\gamma_B \le \frac{t}{a-N}$
$\gamma_A \le \frac{1}{2}$ and $\gamma_B \le \frac{1}{2}$	①	②	③	④
$\frac{1}{2} < \gamma_B \le \gamma_A \le \frac{2}{3}$ and $t \le t_A$	①	②	③	④
$\frac{1}{2} < \gamma_A \le \gamma_B < \frac{2}{3}$ and $t \le t_B$	①	②	③	④

Table 2.A.2: Conditions on γ_i for bonus equilibrium

and ②, and they are plugged into the platforms' profit functions to derive the commission rates.

① $b_A^* = b_B^* = 0$

The constraints on γ_A and γ_B for this bonus equilibrium are shown in Table 2.A.2. From Eq. (2.A.15), the supply pool is fully utilized if $\gamma_A + \gamma_B \le 2 - \frac{t}{a-N}$ and partially utilized if $\gamma_A + \gamma_B > 2 - \frac{t}{a-N}$. The platforms' profit functions under full and partial supply pool utilization are

$$\Pi_A^F/N = (a-N)\gamma_A(\tfrac{1}{2} - \tfrac{(a-N)(\gamma_A-\gamma_B)}{2t}), \quad \Pi_B^F/N = (a-N)\gamma_B(\tfrac{1}{2} - \tfrac{(a-N)(\gamma_B-\gamma_A)}{2t}),$$
$$\Pi_A^P/N = \tfrac{\gamma_A(1-\gamma_A)a^2t}{((2-\gamma_A-\gamma_B)N+t)^2}, \qquad \Pi_B^P/N = \tfrac{\gamma_B(1-\gamma_B)a^2t}{((2-\gamma_A-\gamma_B)N+t)^2}.$$

For ease of exposition, the remainder of the proof uses platform A's best response $\gamma_A(\gamma_B)$ to illustrate. Lemma 2 implies that when $b_A^* = b_B^* = 0$, the supply pool is fully utilized if $\gamma_A \le \tilde{\gamma}_A$, where $\tilde{\gamma}_A = 2 - \frac{t}{a-N} - \gamma_B$. Thus, Π_A is a piecewise function such that $\Pi_A = \Pi_A^F$ if $\gamma_A \le \tilde{\gamma}_A$ and $\Pi_A = \Pi_A^P$ if $\gamma_A > \tilde{\gamma}_A$. The first-order derivative of Π_A w.r.t. γ_A under full utilization and under partial utilization are shown as in Eq. (2.A.23) and Eq. (2.A.24):

$$\frac{\partial \Pi_A^F}{\partial \gamma_A} = -\frac{(a-N)(2(a-N)\gamma_A - (a-N)\gamma_B - t)}{2t}N \tag{2.A.23}$$

$$\frac{\partial \Pi_A^P}{\partial \gamma_A} = \frac{a^2 t(N(2-3\gamma_A-\gamma_B+2\gamma_A\gamma_B) + (1-2\gamma_A)t)}{(t+(2-\gamma_A-\gamma_B)N)^3}N \tag{2.A.24}$$

From setting $\frac{\partial \Pi_A^F}{\partial \gamma_A} = 0$, the unconstrained optimizer of Π_A^F is $\gamma_A^F = \frac{(a-N)\gamma_B+t}{2(a-N)}$. Thus, $\gamma_A^F \ge 0$ holds for any $\gamma_B \ge 0$. When evaluated at $\tilde{\gamma}_A$, $\frac{\partial \Pi_A^F}{\partial \gamma_A}|_{\gamma_A=\tilde{\gamma}_A} = \frac{(a-N)(3t-(4-3\gamma_B)(a-N))}{2t}$. Thus, $\frac{\partial \Pi_A^F}{\partial \gamma_A}|_{\gamma_A=\tilde{\gamma}_A} \ge 0$ if and only if $\gamma_B \ge \frac{4}{3} - \frac{t}{a-N}$. Since $N(2-3\gamma_A-\gamma_B+2\gamma_A\gamma_B) + (1-2\gamma_A)t \ge 0$, $\frac{\partial \Pi_A^P}{\partial \gamma_A} \ge 0$. Π_A^P is increasing in γ_A. The two thresholds for γ_B are $\gamma_0 = 2 - \frac{t}{a-N}$ and $\gamma_1 = \frac{4}{3} - \frac{t}{a-N}$.

First, threshold γ_0 determines the number of pieces in the profit function Π_A for $\gamma_A \ge 0$. If $\gamma_B \ge 2 - \frac{t}{a-N}$, $\tilde{\gamma}_A \le 0$, and hence there is only one piece, and $\Pi_A = \Pi_A^P$.

If $\gamma_B < 2 - \frac{t}{a-N}$, there are two pieces, and $\Pi = \Pi_A^F$ for $\gamma_A \le \gamma_0 - \gamma_B$ and $\Pi_A = \Pi_A^P$ for $\gamma_A > \gamma_0 - \gamma_B$. Second, threshold γ_1 determines the sign of the first-order derivative of π_A^F at $\tilde{\gamma}_A$. Note that $\gamma_1 < \gamma_0$. Note that $2\bar{\gamma} - \frac{t}{a-N} < \frac{4}{3} - \frac{t}{a-N} < 2 - \frac{t}{a-N} - \bar{\gamma}$ if $\bar{\gamma} < \frac{2}{3}$, and $2\bar{\gamma} - \frac{t}{a-N} = \frac{4}{3} - \frac{t}{a-N} = 2 - \frac{t}{a-N} - \bar{\gamma}$ if $\bar{\gamma} = \frac{2}{3}$.

(BR.1) Suppose $\frac{4}{3} - \frac{t}{a-N} \le \gamma_B \le 2 - \frac{t}{a-N}$. In this case, $\tilde{\gamma}_A \ge 0$, $\Pi_A = \Pi_A^F$ increases in γ_A for $\gamma_A \le \tilde{\gamma}_A$, and $\Pi_A = \Pi_A^P$ increases in γ_A for $\gamma_A \ge \tilde{\gamma}_A$. Thus, $\gamma_A(\gamma_B) = \bar{\gamma}$. Moreover, if $\gamma_B \ge 2 - \frac{t}{a-N} - \bar{\gamma}$ (i.e., $\tilde{\gamma}_A \le \bar{\gamma}$), the supply pool is partially utilized; if $\gamma_B \le 2 - \frac{t}{a-N} - \bar{\gamma}$, the supply pool is fully utilized. First, if $\bar{\gamma} = \frac{2}{3}$, then this best response is partial utilization over $\frac{4}{3} - \frac{t}{a-N} \le \gamma_B \le 2 - \frac{t}{a-N}$. Second, if $\bar{\gamma} < \frac{2}{3}$, then this best response is full utilization over $\frac{4}{3} - \frac{t}{a-N} \le \gamma_B \le 2 - \frac{t}{a-N} - \bar{\gamma}$, and partial utilization over $2 - \frac{t}{a-N} - \bar{\gamma} < \gamma_B \le 2 - \frac{t}{a-N}$.

(BR.2) Suppose $\gamma_B < \frac{4}{3} - \frac{t}{a-N}$. (This requires that $\frac{t}{a-N} < \frac{4}{3}$.) In this case, $\tilde{\gamma}_A \ge 0$, $\Pi_A = \Pi_A^F$ first increases in γ_A and then decreases in γ_A for $\gamma_A \le \tilde{\gamma}_A$, and $\Pi_A = \Pi_A^P$ increases in γ_A for $\gamma_A \ge \tilde{\gamma}_A$. Thus, (i) if $\bar{\gamma} < \gamma_A^F$, i.e., $\gamma_B > 2\bar{\gamma} - \frac{t}{a-N}$, then $\gamma_A(\gamma_B) = \bar{\gamma}$. This best response requires that $\bar{\gamma} < \frac{2}{3}$ and $2\bar{\gamma} - \frac{t}{a-N} < \gamma_B < \frac{4}{3} - \frac{t}{a-N}$. (ii) If $\gamma_A^F \le \bar{\gamma} < \tilde{\gamma}_A$, i.e., $\gamma_B \le 2\bar{\gamma} - \frac{t}{a-N}$ and $\gamma_B < 2 - \frac{t}{a-N} - \bar{\gamma}$, then $\gamma_A(\gamma_B) = \gamma_A^F$. First, if $\bar{\gamma} = \frac{2}{3}$, then this best response requires that $\gamma_B < \frac{4}{3} - \frac{t}{a-N}$. Second, if $\bar{\gamma} < \frac{2}{3}$, then this best response requires that $\gamma_B \le 2\bar{\gamma} - \frac{t}{a-N}$. Note that $\bar{\gamma} > \tilde{\gamma}_A$ cannot occur as it requires $\gamma_B > 2 - \frac{t}{a-N} - \bar{\gamma} \ge \frac{4}{3} - \frac{t}{a-N}$, contradiction.

(BR.3) Suppose $\gamma_B > 2 - \frac{t}{a-N}$. In this case, $\Pi = \Pi_A^P$ and increases for any $\gamma_A \ge 0$. Thus, $\gamma_A(\gamma_B) = \bar{\gamma}$.

In summary, the best responses of the commission rates if $\bar{\gamma} < \frac{2}{3}$ are

$$\gamma_A(\gamma_B) = \begin{cases} \gamma_A^F = \frac{(a-N)\gamma_B + t}{2(a-N)} & \text{if } \gamma_B \le 2\bar{\gamma} - \frac{t}{a-N}, \\ \bar{\gamma} & \text{if } \gamma_B > 2\bar{\gamma} - \frac{t}{a-N}. \end{cases} \tag{2.A.25}$$

The supply pool utilization in Eq. (2.A.25) is (i) full, and (ii) full for $2\bar{\gamma} - \frac{t}{a-N} < \gamma_B \le 2 - \frac{t}{a-N} - \bar{\gamma}$ and partial for $\gamma_B > 2 - \frac{t}{a-N} - \bar{\gamma}$. (Note that $2\bar{\gamma} - \frac{t}{a-N} < \frac{4}{3} - \frac{t}{a-N} < 2 - \frac{t}{a-N} - \bar{\gamma} < 2 - \frac{t}{a-N}$ if $\bar{\gamma} < \frac{2}{3}$.) The best responses of the commission rates if $\bar{\gamma} = \frac{2}{3}$ are

$$\gamma_A(\gamma_B) = \begin{cases} \gamma_A^F = \frac{(a-N)\gamma_B + t}{2(a-N)} & \text{if } \gamma_B \le \frac{4}{3} - \frac{t}{a-N}, \\ \bar{\gamma} & \text{if } \gamma_B > \frac{4}{3} - \frac{t}{a-N}. \end{cases} \tag{2.A.26}$$

The supply pool utilization in Eq. (2.A.26) is (i) full, and (ii) partial.

The discussion for the commission equilibrium focuses on the case when $\bar{\gamma} < \frac{2}{3}$ (Note that the results below carry through if $\bar{\gamma} = \frac{2}{3}$.).

(E.1) $\gamma_A^* = \gamma_B^* = \frac{t}{a-N}$. This equilibrium is derived from best responses $\gamma_A(\gamma_B) =$

$\frac{(a-N)\gamma_B+t}{2(a-N)}$ and $\gamma_B(\gamma_A) = \frac{(a-N)\gamma_A+t}{2(a-N)}$. This equilibrium requires that $\gamma_A^* \leq 2\bar{\gamma} - \frac{t}{a-N}$ (or $\gamma_B^* \leq 2\bar{\gamma} - \frac{t}{a-N}$), i.e., $\frac{t}{a-N} \leq \bar{\gamma}$. Under this equilibrium, platform profits are $\Pi_A^* = \Pi_B^* = \frac{Nt}{2}$. Plugging $\gamma_A^* = \gamma_B^* = \frac{t}{a-N}$ into the conditions shown in Table 2.A.2, this equilibrium emerges if $\frac{t}{a-N} \leq \min(\frac{1}{2}, \bar{\gamma})$ and $-\frac{-2(a-2N)}{(a-N)^2}t^2 + \frac{a-5N}{a-N}t + 2N \geq 0$.

(E.2) $\gamma_A^* = \gamma_B^* = \bar{\gamma}$**.** This equilibrium is derived from best responses $\gamma_A(\gamma_B) = \gamma_B(\gamma_A) = \bar{\gamma}$. This equilibrium requires that $2\bar{\gamma} - \frac{t}{a-N} < \bar{\gamma}$, i.e., $\frac{t}{a-N} > \bar{\gamma}$. Under this equilibrium, platform profits are $\Pi_A^* = \Pi_B^* = \frac{(a-N)\bar{\gamma}}{2}N$ for full utilization and $\Pi_A^* = \Pi_B^* = \frac{\bar{\gamma}(1-\bar{\gamma})a^2 t}{(2(1-\bar{\gamma})N+t)^2}N$ for partial utilization. Plugging $\gamma_A^* = \gamma_B^* = \bar{\gamma}$ into the conditions shown in Table 2.A.2, this equilibrium emerges if $(a-N)\bar{\gamma} \leq t \leq \frac{2(\bar{\gamma}-1)^2 N}{(2\bar{\gamma}-1)^+}$.

Note that the combination of asymmetric best responses cannot form an equilibrium. For example, $\gamma_A(\gamma_B) = \frac{(a-N)\gamma_B+t}{2(a-N)}$ and $\gamma_B(\gamma_A) = \bar{\gamma}$ would derive $\gamma_A^* = \frac{(a-N)\bar{\gamma}+t}{2(a-N)}$ and $\gamma_B^* = \bar{\gamma}$. From $\gamma_B^* \leq 2\bar{\gamma} - \frac{t}{a-N}$, we obtain $\bar{\gamma} \geq \frac{t}{a-N}$. From $\gamma_A^* > 2\bar{\gamma} - \frac{t}{a-N}$, we obtain $\bar{\gamma} < \frac{t}{a-N}$, contradiction. Thus, (E.1) and (E.2) are the only equilibria can be obtained from $b_A^* = b_B^* = 0$ using backward induction.

② $b_A^* = \gamma_A(a-N) - t$ and $b_B^* = \gamma_B(a-N) - t$

In this case, the supply pool is fully utilized, and $\Pi_A = \Pi_B = \frac{Nt}{2}$. Multiple equilibria exist for the commission rates such that any γ_A and γ_B that satisfy the conditions in Table 2.A.2 for ② is an equilibrium. ■

CHAPTER 3

OPERATIONAL DISTORTION: COMPOUND EFFECTS OF SHORT-TERMISM AND COMPETITION

3.1 Introduction

A firm's interest in its short-term valuation (short-termism) can induce operations managers to distort their operational choices, including capacity investment levels. Warren Buffet, Chairman and CEO of Berkshire Hathaway, and Jamie Dimon, Chairman and CEO of JPMorgan Chase, provide examples of such distortions in a recent *Wall Street Journal* editorial highlighting the deleterious impact of short-termism – "Companies frequently hold back on technology spending, hiring, and research and development to meet quarterly earnings forecast ..." [63]. Recent empirical research supports the anecdotal evidence that short-termism can lead to under-investment. For example, Asker et al. [6] compare a matched sample of public and private firms and find that public firms make lower long-term investments and are less responsive to investment opportunities. Flammer and Bansal [69] exploit exogenous changes in executives' long-term incentives and find that firms with long-term compensation schemes made larger investments in R&D.

Anecdotal evidence also suggests that competition can interact with short-termism to influence a firm's decisions [116, 9]. A firm may make operational choices that signal an optimistic market outlook to investors in order to obtain a high short-term market valuation, but an optimistic signal may also invite a stronger competitive response because it provides competitors with information about the opportunities in the market. Starvish [150] reinforces this point by observing that "communicating with public investors is often hard to do, none the least because public-firm managers are not allowed to talk to investors privately. And once information conveyed to investors becomes public, it is also available to competitors." In another example, GlaxoSmithKline's vice president of investor relations, Mel Foster-Hawes, recently commented, "When sharing long-term information [with investors], companies need to balance the benefits of disclosure with the risks of sharing competitive insights" [9]. Palepu et al. [132] identify the competitive dynamics in product markets as an important constraint on companies' decisions to disclose information with investors, and express that "Disclosure of proprietary information on strategies and their expected economic consequences may hurt the firm's competitive position. Managers then face a trade-off between providing information that is useful to investors in assessing the firm's economic performance, and withholding information to maximize the firm's product market advantage." Based on a survey of over 1,000 C-level executives and board members, McKinsey & Company identify competition as a critical factor that can exacerbate the effects of short-termism [14]. This anecdotal evidence that short-termism and competition interact in some way to influence a firm's decisions is useful, but falls short of an analytical model that can be rigorously studied and serve as the basis for developing testable hypotheses. Developing such a model and analyzing the implications of short-termism and competition on operational decisions and the firm's long-term value are the primary objectives of this research.

Understanding how short-termism and competition can interact to distort operational decisions is relevant to operations managers. First, there is significant anecdotal evidence that investors use a firm's operational information (such as capacity investment, procurement orders, and inventory levels) to in-form the valuation that they assign to the firm. For instance, Apple shares closed down 5 percent on November 12, 2018 after one of its facial recognition suppliers cut its outlook, citing a reduced shipment request from a large customer, which the market interpreted to mean Apple [126]. In another example, Tesla Inc. shares closed down 3 percent on April 11, 2019 after the Nikkei Asian Review reported that the Silicon Valley car maker and Panasonic Corporation would delay adding capacity at their joint battery manufacturing facility [8]. The stock market's responsiveness to such operational announcements highlights why a firm exhibiting short-termism may manipulate its operational decisions for a desirable short-term valuation. However, such catering can erode the firm's long-term value [15, 41].

Second, a firm's operational decisions can signal the firm's future market demand to potential competitors. Those competitors may target the market if an incumbent's investment in capacity signals that the firm anticipates strong demand [75]. Rasmusen [135] depict the underlying rational, "I might be able to [deter entry] by persuading [competitors] that the market is not big enough for two firms to operate in. I must not appear too profitable. I may purposely keep my sales and profits small to make the market appear unprofitable."

These issues are also of great importance to the broader business community, as evidenced by several initiatives that have recently been instituted to help combat short-termism. For example, in 2013 McKinsey & Company, BlackRock, and the Canada Pension Plan Investment Board created the Focusing Capital on the Long Term (FCLT) initiative to develop practical approaches to reduce corporate short-termism.[1] The Strategic Investor Initiative (SII) of Chief Executives for Corporate Purpose (CECP) has also been convening forums to facilitate con-

[1]https://www.fcltglobal.org/about/our-work

versations between CEOs and investors on long-term planning and re-orienting capital towards the long-term. In August 2019, the Business Roundtable, an association of nearly 200 CEOs of the largest corporations in the U.S., published their "Statement of Purpose of a Corporation" to encourage companies to re-orient from focusing primarily to serve investors and instead toward long-term value for all their stake holders [37].

To foster a better understanding of these issues, we study the combined effect of short-termism and competition on a firm's operational decisions and long-term profit. While the academic literature has examined the impact of competition separately from the impact of short-termism under information asymmetry, less is known about their interaction and the compound effects on the firms' operational decisions and long-term performance. Our research seeks to shed light on three related questions. How do short-termism and competition affect information dissemination? How do they interact to induce firms to distort their operational decisions? What is their impact on the long-term performance of a firm? We also discuss the implications of these findings.

We analyze these questions with a signaling game among a firm with private information, a competitive entrant, and an investor. We conduct our analysis as a dynamic game with incomplete information and use perfect Bayesian Nash equilibrium (PBE) as our equilibrium construct. In our main model, the firm and the competitor engage in a Stackelberg competition in which the firm makes a capacity decision before the competitor. The firm has an interest in its capital market valuation, and its utility is a weighted average of its short-term valuation assigned by the investor and its long-term profit. The weight on its short-term valuation captures the firm's level of short-termism. The firm has private information that the market demand is either high (an H-type firm) or low (an L-type firm). The investor and the entrant may interpret the firm's capacity investment as a signal of the market demand. To more clearly convey our main points, we assume that the investor is risk-neutral and the entrant has no short-termism. To confirm that these assumptions are mild, we show in Section 3.9.2 of the Appendix that the primary findings from our analysis are robust to modeling the investor with different risk attitudes and the entrant with short-termism. In Section 3.7, we extend our analysis by considering two established competitors in a simultaneous move game.

Our model provides a rigorous framework to examine the impact of short-termism in the face of competition. With this model, we make several contributions. First, we show how information sharing by the firm to the uninformed players is influenced by the level of short-termism. We find that the firm will reveal her private information of the market demand to the investor and the competitive entrant when the level of short-termism is moderate. When the level of short-termism is sufficiently low or sufficiently high, the incumbent firm will conceal her private information of the market demand from the investor and the competitive entrant. This is because the cost of signaling a high (resp. low) demand borne by the firm is too large at sufficiently high (resp. low) level of short-termism.

Second, we show that the effect of short-termism on operational distortion is not monotonic. In particular, operational distortions emerges not only when short-termism is high, but also when it is low. When the level of short-termism is high, there may exist a costless separating PBE (in which both firm types invest optimally), a costly separating PBE (in which the H-type over-invests and the L-type invests optimally) and a Pareto dominant pooling PBE (in which the H-type under-invests and the L-type over-invests). When the level of short-termism is low, there may exist a costless separating PBE, a costly separating PBE (in which the H-type invests optimally and the L-type under-invests) and a Pareto dominant pooling PBE (in which both types *under-invest*). When the level of short-termism is moderate, operational distortion is absent for both firm types.

Third, we show that short-termism is not necessarily harmful to a firm's long-term value. In fact, we find that firms may achieve the highest long-term profit over a wide range of short-termism. Even if higher short-termism induces lower long-term investment, it can still benefit the firm's long-term value because lower investment levels send a negative signal that can dissuade competitors from entering the market aggressively.

Collectively, our findings indicate that firms can benefit by embracing short-termism in some contexts rather than endeavoring to systematically reduce it, as other sources advocate [90, 41].

3.2 Literature Review

Our study is related to several streams of literature. First, our work is related to research on the impact of short-termism on managerial decision making [see 154, and cites therein]. For instance, Stein [151] shows that the threat of takeover can induce managers to inflate short-term earnings at the expense of long-term value. Stein [152] adopts a signal-jamming model and shows that firms with short-termism may forsake long-term profits by liquidating immature investments to inflate short-term earnings. Bebchuk and Stole [19] employ a signaling model and show that a highly productive firm may over-invest when the firm's investment decision is observable and the productivity of the long-term project is unobservable. More recent empirical research investigates the magnitude of short-termism (e.g., Brochet et al. 34, Sampson and Shi 140) and its effect on earnings management (e.g., Bergstresser and Philippon 24, Call et al. 40) and R&D investment (e.g., Bushee 36, Asker et al. 6.).

Second, our work contributes to recent research in operations management that studies the impact of short-termism on a firm's operational decisions (see Lai and Xiao 109 for a survey). Lai et al. [111] analyze a setting in which firms may deliberately inflate their sales by channel stuffing. Lai et al. [112] show that a downstream buyer may distort her stocking decision under a single buyback contract, and that a menu of buyback contracts can restore efficiency in the supply chain. Schmidt et al. [144] show the emergence of a pooling equilibrium, under which the high-quality firm under-invests and the low-quality firm over-invests, when capacity levels are discrete or the Undefeated refinement is applied. Lai and Xiao [110] show that a firm will manipulate her inventory decision when investors are less informed about the firm's demand uncertainty. They also show that when the manager can send a signal to the capital market by manipulating her short-termism, the first-best inventory decisions can be achieved in equilibrium.

Third, we contribute to research that has studied information sharing under competition. Li [115] studies vertical information sharing by analyzing a supply chain consisting of a manufacturer and multiple retailers that engage in Cournot competition and have private information. The threat of information leakage discourages the retailers from sharing their demand information with the manufacturer. Zhang [167] extends this work by examining both Cournot and Bertrand competition between two downstream retailers, and their goods can be either substitutes or complements. These papers assume that private information is always shared truthfully and credibly. Anand and Goyal [3] consider a supply chain consisting of two manufacturers engaging in a Cournot competition and sourcing from one common supplier. They show that the incumbent manufacturer's private demand information may be revealed to the supplier by its order quantity, and the supplier in turn leaks this private information to the uninformed competing manufacturer. Kong et al. [107] analyze a similar setting and show that a revenue-sharing contract can counter information leakage. Zhao et al. [169] analyze a setting in which two outsourced service suppliers compete for a contract under information asymmetry with the client about the service costs. They find that the strength of the signal affects the level of competition, but reducing information asymmetry may not benefit the client. There are also papers on information sharing in two competing supply chains [78, 79, 146].

Finally, our research relates to the broader signaling game work in the operations management literature. Cachon and Lariviere [38] investigate credibly sharing demand forecasts in a setting where the manufacturer has an incentive to inflate her demand forecast to induce the supplier to build more capacity. They study contracts that can ensure credible information sharing under both forced compliance and voluntary compliance regimes. Chod et al. [55] show that signaling a firm's fundamental quality to lenders through inventory transactions can be more efficient than signaling through loan requests. They advocate that the adoption of blockchain technology can facilitate signaling through inventory. Chod et al. [54] show that supplier diversification can be attributed to buyer risk, which induces buyers to signal creditworthiness in an effort to secure more favorable terms. Belavina et al. [21] analyze designs based on deferred payments to deter misconduct of crowdfunding platforms. Chakraborty and Swinney [44] study how an entrepreneur can signal her product quality to a pool of small and uninformed investors by the design of the crowdfunding campaign, including the price of the reward and the funding target.

Our paper contributes to these literature streams by investigating the impact of short-termism and competition on operational decisions and modeling the firm's conflicting incentives for information sharing in that environment. In doing so, we account for the interaction of short-termism and competition, provide deeper insights into its impact on operational distortion, and reveal when short-termism can be beneficial to the firm's long-term value.

3.3 Model Setup

We study a signaling game with two periods and three players, F, I, and E. Player F (she) is the incumbent firm (firm hereafter). Player I is a risk-neutral investor representing a perfectly competitive capital market. Player E (he) is a competitive entrant and offers a perfect substitute to the firm's product. The firm and the entrant engage in a Stackelberg competition. We follow [3] and model the demand for the product using the inverse demand function given by $p = A_t - q_F - q_E$, where q_F and q_E are the supply quantities of the firm and the entrant, respectively. The market demand, and therefore the firm's type, can be either high ($t = H$) or low ($t = L$). It is common knowledge to all players that

$A_t = A_H$ with probability $h \in (0, 1)$ and $A_t = A_L$ with probability $(1-h)$. We denote the weighted average demand as $\mu = hA_H + (1-h)A_L$, and the ratio of firm types as $\theta = \frac{A_H}{A_L} > 1$. θ implies the level of dispersion in market demand such that a larger θ corresponds to a larger difference between the two firm types and more variability in market demand.

The sequence of events is shown in Figure 3.1. In the beginning of period 1, the firm receives private information that identifies her true demand type. The firm decides her capacity investment level q_F, which is observable to the investor and the entrant. Upon observing the firm's capacity investment, the investor and the entrant update their belief about the firm's type. The investor then assigns a short-term valuation v to the firm, and the entrant chooses his capacity investment q_E. In period 2, market demand is realized, and the long-term profits of the firm and the entrant are determined.

There is no private information between the entrant and the investor. We denote their shared posterior belief of the firm's type using λ, such that $t = H$ with probability λ and $t = L$ with probability $1 - \lambda$, where $\lambda \in [0, 1]$. In equilibrium, the investor and the entrant's posterior belief encompasses three cases: the two players believe that the firm is either a H-type (i.e., when $\lambda = 1$) or a L-type (i.e., when $\lambda = 0$), or they maintain their prior belief (i.e., when $\lambda = h$).

We model the firm's short-termism using $\alpha \in [0, 1]$. When α is higher, the firm exhibits a higher emphasis on her short-term valuation and a lower emphasis on her long-term profit. As in [19], [112], and [144], the firm's utility function is a linear combination of the valuation assigned by the investor in period 1 and her profit gained after demand realization in period 2. The firm's

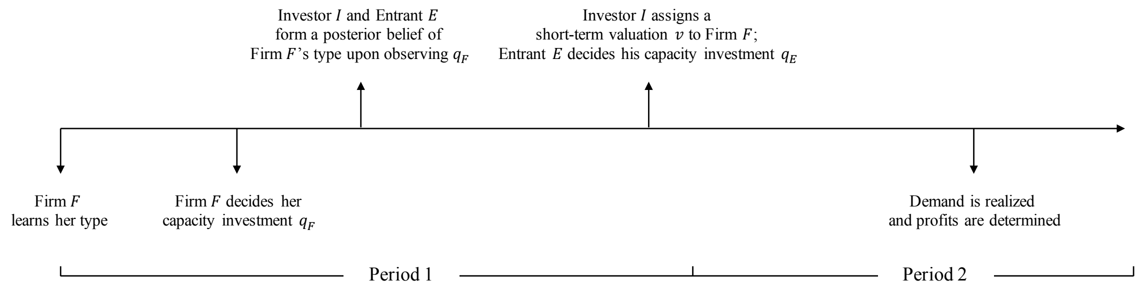


Figure 3.1: Timeline of the Model

utility is expressed as

$$U_F(t, q_F, q_E, v) = \alpha v + (1 - \alpha)\pi_F(t, q_F, q_E), \tag{3.1}$$

where v is the short-term valuation and $\pi_F(t, q_F, q_E) = (A_t - q_F - q_E)q_F$ is the firm's long-term profit.

We represent the investor's utility function by following [70]. The investor operates in a perfectly competitive capital market, and its utility is determined by the magnitude of the error in its valuation of the firm:

$$U_I(t, q_F, q_E, v) = -\left(v - \pi_F(t, q_F, q_E)\right)^2. \tag{3.2}$$

This utility function assumes that the investor is risk neutral. In Section 3.9.2 of the Appendix, we show that our main insights carry through when assuming a loss averse, risk averse, or risk seeking investor.

The entrant's utility is equal to his long-term profit obtained from the Stackelberg competition, i.e.,

$$U_E(t, q_F, q_E) = \pi_E(t, q_F, q_E) = (A_t - q_F - q_E)q_E. \tag{3.3}$$

For ease of exposition, we assume the entrant does not exhibit short-termism. Relaxing this assumption will not materially impact our equilibrium outcomes and insights. This is because the entrant does not have private information. We formally show this in Section 3.9.2 of the Appendix.

We utilize a Perfect Bayesian Nash equilibrium (PBE) to predict each player's strategy. In a PBE, each player's strategy is a best response to all the other players' strategies under reasonable posterior beliefs of the private information. The investor determines its valuation v to maximize its expected utility under the posterior belief λ of the firm's type. From Equation (3.2), we obtain the investor's best response $v^*(q_F, q_E|\lambda)$ to be

$$v^*(q_F, q_E|\lambda) = \lambda\pi_F(H, q_F, q_E) + (1-\lambda)\pi_F(L, q_F, q_E). \tag{3.4}$$

Plugging $v^*(q_F, q_E|\lambda)$ into Equation (3.1), the firm's utility function after factoring in the investor's best response is

$$U_F(t, q_F, q_E, v^*|\lambda) = \begin{cases} \alpha\lambda\pi_F(H, q_F, q_E) + (1-\alpha\lambda)\pi_F(L, q_F, q_E) & \text{for} \quad t = L, \\ \\ (1-\alpha(1-\lambda))\pi_F(H, q_F, q_E) + \alpha(1-\lambda)\pi_F(L, q_F, q_E) & \text{for} \quad t = H. \end{cases} \tag{3.5}$$

The firm's utility function is a weighted average of the long-term profits of the two firm types. When the firm only cares about her short-term valuation (i.e., $\alpha = 1$), her utility function will be the same across types. When the firm exhibits no short-termism (i.e., $\alpha = 0$) or when the firm's type is correctly inferred by the other constituents (i.e., $\lambda = 0$ if $t = L$ and $\lambda = 1$ if $t = H$), the firm's utility function is the same as her long-term profit function.

The entrant determines his capacity investment q_E to maximize his expected profit under the posterior belief of the firm's type λ. As a Stackelberg follower, the entrant's best response is shown to be

$$q_E^*(q_F|\lambda) = \frac{\lambda A_H + (1-\lambda)A_L - q_F}{2}. \tag{3.6}$$

The firm's capacity investment affects the entrant's best response in two opposing ways. First, λ is non-decreasing in q_F under a reasonable belief system, and hence a sufficiently large q_F can increase q_E because the entrant is more likely to believe there is high demand in the market. However, a larger q_F can induce a smaller q_E by suppressing the market share of the entrant as a Stackelberg follower.

Plugging the best response of the entrant into Equation (3.5), we simplify the firm's utility function to

$$U_F(t, q_F|\lambda) = \begin{cases} \dfrac{(1-2\alpha)(-\lambda)(A_H - A_L) + A_L - q_F}{2} q_F & \text{for} \quad t = L, \\ \dfrac{(1-2\alpha)(1-\lambda)(A_H - A_L) + A_H - q_F}{2} q_F & \text{for} \quad t = H. \end{cases} \tag{3.7}$$

Equation (3.7) shows that the effect of the posterior belief on the firm's utility is influenced by her short-termism. In Lemma 3, we show that α influences the impact of different equilibrium posterior beliefs on the firm's utility.

Lemma 3 *For $\lambda \in \{0, h, 1\}$ and $t \in \{H, L\}$:*

(i) If $\alpha < \frac{1}{2}$, the firm's utility decreases as λ increases (i.e., $U_F(t, q_F|\lambda = 0) > U_F(t, q_F|\lambda = h) > U_F(t, q_F|\lambda = 1)$).

(ii) If $\alpha > \frac{1}{2}$, the firm's utility increases as λ increases (i.e., $U_F(t, q_F|\lambda = 1) > U_F(t, q_F|\lambda = h) > U_F(t, q_F|\lambda = 0)$).

(iii) If $\alpha = \frac{1}{2}$, the firm's utility is independent of λ (i.e., $U_F(t, q_F|\lambda) = \frac{A_t - q_F}{2} q_F$).

Lemma 3 shows that a threshold for short-termism ($\alpha = \frac{1}{2}$) exists beyond which the impact of the posterior belief on the firm's utility will reverse. This differential impact of the posterior belief on the firm's utility is because the firm's signal of her market demand is received by both the investor and the entrant, and both form a posterior belief of the market demand. A higher λ will induce the investor to assign a higher short-term valuation to the firm, but it will also induce the entrant to increase his capacity investment. The firm's short-termism influences how her desire for a higher valuation and lower competitive entry impact her utility. If $\alpha < \frac{1}{2}$, a lower competitive entry plays a larger role. In this case, Lemma 3 shows that a posterior belief that market demand is low yields a higher utility for the firm. Thus, both firm types prefer to be perceived as a L-type. If $\alpha > \frac{1}{2}$, a higher short-term valuation plays a larger role, so both firm types prefer to be perceived as a H-type.

To simplify our exposition in light of these different preferences, we refer to the preferred firm type as the "mimicked" type and the other type as the "mimicking" type. This conveys which firm type has an incentive to mimic the preferred firm type through her capacity decisions. The L-type (H-type) is the mimicked type and the H-type (L-type) is the mimicking type when $\alpha < \frac{1}{2}$ ($\alpha > \frac{1}{2}$). If $\alpha = \frac{1}{2}$, the posterior belief no longer affects the firm's utility and each firm type prefers to be perceived as her true type.

3.4 Equilibrium Analysis

In Section 3.4.1, we show the firm's equilibrium capacity investment using no information asymmetry as a benchmark. We then show the separating PBE and the pooling PBE of the signaling game in Sections 3.4.2 and 3.4.3. In a separating equilibrium, each type of informed player chooses a distinct signal, whereas in a pooling equilibrium, all types of informed player choose the same signal [108].

3.4.1 Benchmark of Complete Information

Following the literature [3, 144], we use the firm's capacity investment under information symmetry as the benchmark outcome. In this case, our model simplifies to a Stackelberg game under complete information. Neither firm type has an incentive to mimic the other type. We show the firm's equilibrium capacity investment under complete information in Lemma 4.

Lemma 4 *If the market demand state is known by all the players, the H-type firm invests $q_H^* = \frac{A_H}{2}$ and the L-type firm invests $q_L^* = \frac{A_L}{2}$.*

Operational distortion occurs when either firm type deviates from the benchmark described in Lemma 4. We use the term "over-invest" or "under-invest" when a firm type invests more or less than her respective optimal capacity under complete information, q_H^* or q_L^*.

In the presence of information asymmetry, it is straightforward to obtain that the L-type firm invests q_L^* and the H-type firm invests q_H^* in equilibrium when $\alpha = \frac{1}{2}$. We next analyze the equilibrium outcomes when $\alpha < \frac{1}{2}$ and when $\alpha > \frac{1}{2}$. We apply the concept of strict dominance in our analysis. Strict dominance ensures a reasonable posterior belief that never assigns a positive probability to a firm type when the firm's capacity investment level is strictly dominated for that firm type. The formal definition is in Section 3.9.1 of the Appendix.

3.4.2 Separating PBE

Under a separating PBE, both firm types choose different levels of capacity investment such that the firm's type is perfectly revealed to the investor and the entrant. Lemma 3 shows that when $\alpha < \frac{1}{2}$, the H-type has an incentive to mimic the L-type, which may induce the H-type firm to invest lower than q_H^*. However, the H-type firm will not choose a quantity that is strictly dominated by q_H^*, i.e., a q_F such that

$$U_F(H, q_F, q_E^*, v^* | \lambda = 0) \leq U_F(H, q_H^*, q_E^*, v^* | \lambda = 1). \tag{3.1}$$

We identify $\bar{q}$ as the upper bound of all the quantities below q_H^* that satisfy inequality (3.1). Thus, a reasonable posterior belief should be $\lambda = 0$ for any $q_F \leq \bar{q}$.

When $\alpha > \frac{1}{2}$, the L-type has an incentive to mimic the H-type, which may induce the L-type firm to invest higher than q_L^*. However, the L-type firm will not choose a quantity that is strictly dominated by q_L^*, i.e., a q_F such that

$$U_F(L, q_F, q_E^*, v^* | \lambda = 1) \leq U_F(L, q_L^*, q_E^*, v^* | \lambda = 0). \tag{3.2}$$

We identify $\underline{q}$ as the lower bound of all the quantities above q_L^* that satisfy inequality (3.2). Thus, a reasonable posterior belief should be $\lambda = 1$ for any $q_F \geq \underline{q}$.

We focus on the least-cost separating PBE in which the mimicked type will choose an investment level strictly dominated for the mimicking type and, conditional on that, best for her own utility. We use q_L^S and q_H^S to represent the firm's investment level in the least-cost separating PBE for the L-type and H-type firm, respectively. We characterize each player's strategy in the least-cost separating

PBE in Definition 1. In Section 3.9.1 of the Appendix, we provide the technical validation for this definition, along with the mathematical formulas of $\bar{q}$ and $\underline{q}$.

Definition 1 *In the least-cost separating PBE, each player's strategy is:*

(i) If $\alpha < \frac{1}{2}$, the L-type firm invests $q_L^S \leq q_L^*$ and the H-type firm invests $q_H^S = q_H^*$. If $\alpha > \frac{1}{2}$, the L-type firm invests $q_L^S = q_L^*$ and the H-type firm invests $q_H^S \geq q_H^*$.

(ii) The investor assigns valuation $v^*(q_t^S) = \frac{A_t - q_t^S}{2} q_t^S$ for $t \in \{H, L\}$.

(iii) the entrant invests $q_E^*(q_t^S) = \frac{A_t - q_t^S}{2}$ for $t \in \{H, L\}$.

(iv) The investor and entrant's decisions are supported by their posterior belief λ and

$$\lambda = \begin{cases} 0 & \text{if } q_F \leq \bar{q}, \\ 1 & \text{otherwise} \end{cases} \text{ if } \alpha < \tfrac{1}{2}, \text{ and } \lambda = \begin{cases} 1 & \text{if } q_F \geq \underline{q}, \\ 0 & \text{otherwise} \end{cases} \text{ if } \alpha > \tfrac{1}{2}.$$

Definition 1 shows that the H-type firm invests more than the L-type firm in the least-cost separating PBE. When $\alpha < \frac{1}{2}$, the H-type firm invests optimally, and the L-type firm may under-invest to reveal her type. When $\alpha > \frac{1}{2}$, the L-type firm invests optimally, and the H-type firm may over-invest to reveal her type. The investor accurately values the firm since the firm's type is perfectly revealed in a separating PBE. The competitor also correctly perceives the market demand and chooses his quantity accordingly.

We refer to a least-cost separating PBE in which the mimicked type must deviate from q_H^* (in the case of a H-type) or q_L^* (in the case of a L-type) as a "costly" separating PBE. Otherwise, it is a "costless" separating PBE. In Proposition 1, we show the existence of the least-cost separating PBE, and identify when such a PBE will be costless. In this proposition, we use the notation θ_{LS} and θ_{HS} to define the boundaries between the costless separating PBE and the costly separating PBE. These thresholds are functions of short-termism level α, and their mathematical formulas are shown in the proof of Proposition 1. To convey the intuition, we present a visualization of θ_{LS} and θ_{HS} in Figure 3.2 and discuss them in Section 3.5.

Proposition 1 *The least-cost separating PBE always exists. It is a costless separating PBE if and only if* $\theta \geq \theta_{LS}$ *when* $\alpha < \frac{1}{2}$, *or* $\theta \geq \theta_{HS}$ *when* $\alpha > \frac{1}{2}$.

Proposition 1 shows that a costless separating PBE emerges in which the mimicked firm type can separate by choosing her optimal capacity under symmetric information, if and only if the difference between the firm types is sufficiently large. When the difference between the firm types is large, it is more costly for the mimicking type to imitate the mimicked type, leading the mimicking type to relent. If the difference between the firm types is not sufficiently large, the mimicked type can only separate by sending a costly signal.

3.4.3 Pooling PBE

Under a pooling PBE, both firm types choose the same level of capacity investment such that the firm's type is concealed from the other players. Since the firm's capacity choice is uninformative, the investor and the entrant's posterior belief will equal their prior belief, i.e. $\lambda = h$. We focus on the existence of Pareto dominant pooling PBE in which both firm types obtain a higher utility compared to their utility in the least-cost separating PBE. Such Pareto dominant pooling PBE survive the Undefeated refinement [119]. Formally, the firm's capacity investment q_F in a Pareto dominant pooling PBE satisfies:

$$U_F(L, q_F|\lambda = h) > U_F(L, q_L^S|\lambda = 0), \tag{3.3}$$

$$U_F(H, q_F|\lambda = h) > U_F(H, q_H^S|\lambda = 1). \tag{3.4}$$

Intuitively, the Undefeated refinement is appealing because both firm types should strictly prefer a Pareto-*dominant* equilibrium outcome compared to a *dominated* equilibrium outcome. This outcome makes the Undefeated refinement appropriate in our practical context in which firms will reasonably gravitate toward equilibrium that provide a superior utility. This is in contrast to the more theoretically abstract Intuitive Criterion refinement that is more widely applied in the signaling literature in economics. As noted by Mailath et al. [119], the Intuitive Criterion refinement can eliminate Pareto optimal outcomes. Mailath et al. [119] and Salanie [139] also point out that, unlike the Undefeated refinement, the Intuitive Criterion refinement is developed assuming that the buyer can convey information to the supplier and investor outside of the game. Finally, Schmidt and Buell [143] provide experimental evidence that human decision makers are more likely to pursue pooling outcomes that survive the Undefeated refinement rather than the separating outcomes predicted by the Intuitive Criterion refinement.

When multiple Pareto-dominant pooling PBE exist, we apply the concept of lexicographically maximum sequential equilibrium (LMSE) [119]. A LMSE is intuitive as it prioritizes the choice of the firm type that is being mimicked. In our setting, if $\alpha < \frac{1}{2}$ ($\alpha > \frac{1}{2}$), a Pareto-dominant pooling PBE is also a LMSE if among all Pareto-dominant pooling PBE it first maximizes the utility of a L-type (H-type) firm and conditional on that, it maximizes the utility of a H-type (L-type) firm. Using a LMSE to identify a unique PBE is intuitively and practically appealing because the mimicking type wishes to be perceived as the mimicked type rather than the other way around. To avoid revealing their true type, however, it is reasonable to assert that the mimicking type would have to behave as the mimicked type by choosing a pooling quantity that is optimal for the mimicked type. This is the exact outcome of the LMSE.

Since the utility functions of the firm are concave, there is a unique LMSE if one or more Pareto dominant pooling PBE exist. We refer to this unique pooling PBE as a pooling LMSE. We characterize the players' strategies in a pooling LMSE in Definition 2, and provide the technical validation for this definition in Section 3.9.1 of the Appendix.

Definition 2 *In a pooling LMSE each player's strategy is:*

(i) If $\alpha < \frac{1}{2}$*, both firm types invest* $q^P = q_L^P$*, where* $q_L^P = \arg\max_{q_F>0} U_F(L, q_F|\lambda = h)$ *and* $q_L^P < q_L^*$*. If* $\alpha > \frac{1}{2}$*, both firm types invest* $q^P = q_H^P$*, where* $q_H^P = \arg\max_{q_F>0} U_F(H, q_F|\lambda = h)$ *and* $q_L^* < q_H^P < q_H^*$*.*

(ii) The investor assigns valuation $v^*(q^P) = \frac{(\mu - q^P)q^P}{2}$*.*

(iii) The entrant invests $q_E^*(q^P) = \frac{\mu - q^P}{2}$*.*

(iv) The investor and entrant's decisions are supported by posterior beliefs λ *where*

$$\lambda = \begin{cases} 0 & \textit{if } q_F \le \bar{q}, \\ h & \textit{if } \bar{q} < q_F \le q^P, \\ 1 & \textit{otherwise} \end{cases} \quad \textit{if } \alpha < \tfrac{1}{2}, \textit{and } \lambda = \begin{cases} 1 & \textit{if } q_F \ge \underline{q}, \\ h & \textit{if } q^P \le q_F < \underline{q}, \\ 0 & \textit{otherwise} \end{cases} \quad \textit{if } \alpha > \tfrac{1}{2}.$$

Definition 2 shows that both firm types distort their capacity investment in the pooling LMSE compared to the capacity choice in Lemma 4. When $\alpha < \frac{1}{2}$, both types under-invest (i.e., $q_L^P < q_L^* < q_H^*$). This finding suggests that firms may systematically under-invest when the firm has relatively low short-termism. When $\alpha > \frac{1}{2}$, the L-type firm over-invests and the H-type firm under-invests in the pooling LMSE (i.e., $q_L^* < q_H^P < q_H^*$). In this case, the investor overvalues the L-type firm and undervalues the H-type firm since the firm's type is concealed. The competitor chooses his quantity based on the prior belief of the market demand, which induces the competitor to overestimate the market demand when it is low and underestimate the market demand when it is high. In Section 3.9.2 of the Appendix, we examine other Pareto dominant pooling PBE and show that are results are not exclusive to the LMSE. These other PBE are bounded by the utility optimizers of both types under a weighted posterior belief. We describe the firm's operational distortion and profit over the range of these alternative pooling PBE in Section 3.5 and Section 3.6.

In Proposition 2, we show the existence of Pareto dominant pooling PBE. We use the notation θ_{LP} and θ_{HP} to define the boundaries between the costly separating PBE and the pooling PBE. These thresholds are functions of short-termism α and prior belief h, and their mathematical formulas are shown in the proof of Proposition 2. We present a visualization of θ_{LP} and θ_{HP} in Figure 3.2 and discuss them in Section 3.5.

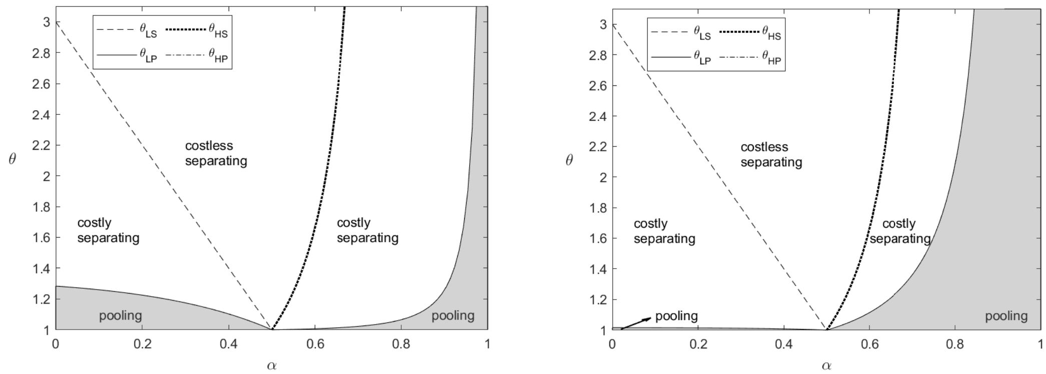
Proposition 2 *Pareto dominant pooling PBE exist if and only if $\theta < \theta_{LP}$ when $\alpha < \frac{1}{2}$, or $\theta < \theta_{HP}$ when $\alpha > \frac{1}{2}$.*

Proposition 2 shows that the Pareto dominant pooling PBE exists if and only if the difference between the firm types is sufficiently small. When the difference between the two firm types is small, it is easier for the mimicking firm type to imitate the mimicked firm type. Consequently, the mimicked firm type must distort her capacity investment more and incur more cost in order to successfully separate. At some point, the mimicked firm type will achieve a higher utility by pooling with the mimicking firm type instead of incurring a cost to separate.

3.5 Operational Distortion under Information Asymmetry

In this section, we show how the firm's operational distortion under information asymmetry changes with α. Figure 3.2 illustrates the equilibrium outcomes identified in Section 3.4 relative to α, θ, and h. The four curves in this figure correspond to the four thresholds of θ defined in Section 3.4. As described earlier, θ_{LS} and θ_{HS} are the boundaries between costless separating PBE and costly separating PBE, and θ_{LP} and θ_{HP} are the boundaries between costly separating PBE and pooling PBE. There is no operational distortion by the firm in the costless separating PBE region, but there is operational distortion in the costly separating PBE and pooling PBE regions. As Figure 3.2 clearly shows, operational distortion can be induced by either a high level of short-termism, or a low level of short-termism. Definition 3 defines four values of α that define the boundaries between the equilibrium outcomes, given θ and h. These help us evaluate the effect of α on the direction and magnitude of operational distortion.

Definition 3 *Given θ and h, there are threshold values of α that define the boundaries between equilibrium outcomes as follows:*
(i) when $\alpha \leq \underline{\alpha_P}$, a pooling PBE arises in which $q^P < q_L^* < q_H^*$;
(ii) when $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, a costly separating PBE arises in which $q_L^S < q_L^*$ and $q_H^S = q_H^*$;



(a) $h = 0.2$ (b) $h = 0.7$

Figure 3.2: Segmentation of Equilibrium Types with Competitive Entrant

(iii) when $\underline{\alpha_S} \leq \alpha \leq \overline{\alpha_S}$*, a costless separating PBE arises in which* $q_L^S = q_L^*$ *and* $q_H^S = q_H^*$*;*

(iv) when $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$*, a costly separating PBE arises in which* $q_L^S = q_L^*$ *and* $q_H^S > q_H^*$*;*

(v) when $\alpha \geq \overline{\alpha_P}$*, a pooling PBE arises in which* $q_L^* < q^P < q_H^*$*.*

Note from Figure 3.2 that $\underline{\alpha_S}$ and $\underline{\alpha_P}$ may degenerate to zero if θ is sufficiently high. We show the effect of α on operational distortion in Proposition 3.

Proposition 3 *Given* θ *and* h*, operational distortion varies with* α *as follows:*

(i) if $\alpha \leq \underline{\alpha_P}$*,* q^P *decreases as* α *decreases;*

(ii) if $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$*,* $q_H^S = q_H^*$ *and* q_L^S *decreases as* α *decreases;*

(iii) if $\underline{\alpha_S} \leq \alpha \leq \overline{\alpha_S}$*, there is no operational distortion;*

(iv) if $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$*,* $q_L^S = q_L^*$ *and* q_H^S *increases as* α *increases;*

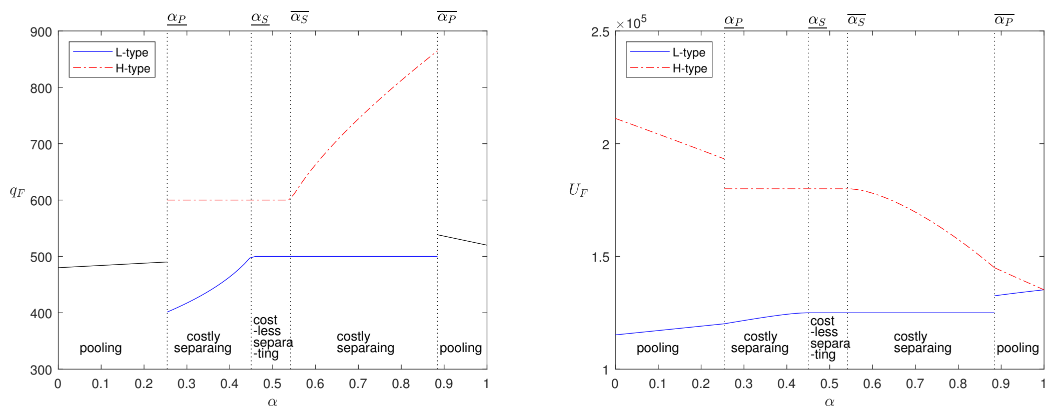
(v) if $\alpha \geq \overline{\alpha_P}$*,* q^P *decreases as* α *increases.*

Proposition 3 shows that the effect of α on operational distortion is manifold. The implication of this proposition is that the firm's short-termism influences her investment decision over the full range of α. When $\alpha \leq \underline{\alpha_P}$, both firm types under-invest in the pooling LMSE. As α decreases, q^P also decreases, representing an increase in the magnitude of under-investment by both firm types. The intuition for this is that as α decreases, both firm types are increasingly interested in their long-term value, and wish to signal to the competitor that market demand is low in order to dissuade an aggressive entry. The H-type is sufficiently motivated that is becomes too costly for the L-type to separate, that the firm types settle on pooling.

When $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, the H-type firm invests optimally and the L-type firm under-invests in the separating PBE. As α decreases, q_L^S also decreases, representing an increase in the magnitude of under-investment by the L-type firm. The intuition is that a lower α raises the firm's interest in her long-term performance and the impact of competition on it. In her desire to quell competition as α decreases, the H-type firm has a higher incentive to mimic the L-type firm. Consequently, the L-type firm must distort her capacity investment further below q_L^* in order to separate from the H-type firm.

When $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, the L-type firm invests optimally and the H-type firm over-invests in the separating PBE. As α increases, q_H^S also increases, representing an increase in the magnitude of over-investment by the H-type firm. The rationale is that a higher α raises the firm's interest in her short-term valuation. In her desire to obtain a high short-term valuation from investors, the L-type firm has a higher incentive to mimic the H-type firm. As a result, the H-type firm must distort her capacity investment further above q_H^* in order to separate from the L-type firm.

When $\alpha \geq \overline{\alpha_P}$, the L-type firm over-invests and the H-type firm under-invests in the pooling LMSE. As α increases, q^P decreases, representing a decrease in the magnitude of over-investment by the L-type firm and an increase in the magnitude of under-investment by the H-type firm. The intuition for this outcome is that an increasing α encourages both firms to focus on their short-term share price by signaling high demand to the investor. The L-type is sufficiently motivated that is becomes too costly for the H-type to separate, that the firm types settle on pooling. In Section 3.9.2 of the Appendix, we show that the insights from Proposition 3 hold when we consider the entire range of the Pareto dominant pooling PBE for $\alpha \leq \underline{\alpha_P}$ and $\alpha \geq \overline{\alpha_P}$.



(a) Capacity Investment

(b) Utility

Figure 3.3: Impact of α on the Firm's Capacity Investment q_F and Utility U_F with Competitive Entrant

To illustrate Proposition 3, we present the firm's equilibrium capacity investment and utility with respect to α in Figure 3.3 when $\theta = 1.2$, $h = 0.2$, and $A_L = 1000$. Notice that we obtain Figure 3.3 by taking a cross section of Figure 3.2(a) at $\theta = 1.2$ and adding the firm's equilibrium capacity investment and utility as the y-axes for Figures 3.3(a) and 3.3(b), respectively. The effect of α on operational distortion is not monotonic, and instead depends on the equilibrium outcome and the firm's type. Figure 3.3(a) demonstrates that as α deviates from $\frac{1}{2}$, operational distortion first occurs by the mimicked firm type in a costly separating PBE, and then by both firm types in the pooling PBE. The magnitude of operational distortion by the mimicked firm type in the costly separating PBE or pooling PBE increases as α deviates further from $\frac{1}{2}$. The operational distortion for the mimicking firm type is exacerbated in the pooling PBE as α decreases below $\frac{1}{2}$, but alleviated in the pooling PBE as α increases above $\frac{1}{2}$. In Section 3.7 of the Appendix, we consider the impact of the firm's short-termism on the capacity choice of the entrant and the overall capacity in the market.

3.6 Impact of Short-termism on Long-term Profits

The perception among many practitioners and market observers is that short-termism is harmful to a firm's long-term value and should be mitigated [13, 120, 15, 41]. Others are more cautious whether short-termism is definitively undesirable [157, 158]. In the following proposition, we contend that a certain level of short-termism can be beneficial to firms' long-term value.

Proposition 4 *The firm's long-term profits can be maximized at non-zero level(s) of short-termism, and the optimal level of short-termism can differ by firm type.*

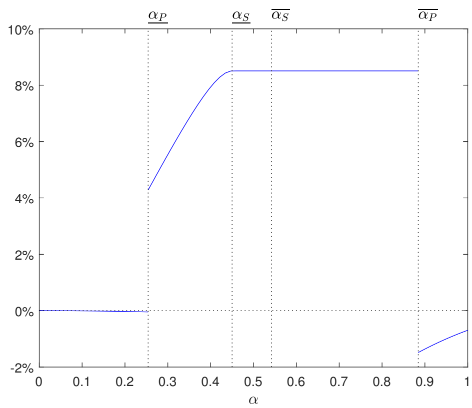
In the proof of Proposition 4 in the Appendix, we show that the optimal level of short-termism for the L-type firm can be anywhere in the range $[\underline{\alpha_S}, \overline{\alpha_P})$. In this range, the L-type firm invests q_L^* in a separating PBE. The optimal level of short-termism for the H-type firm depends on θ, the ratio of the market sizes for two firm types. We identify a threshold $\hat{\theta}$ such that if the ratio is below this threshold (i.e. $\theta < \hat{\theta}$), the optimal level of short-termism is $\overline{\alpha_P}$. At this level, the H-type firm under-invests at q_H^P in the pooling LMSE. If the ratio is above this threshold (i.e., $\theta \geq \hat{\theta}$), the optimal level of short-termism is the range $[\underline{\alpha_S}, \overline{\alpha_S}]$. In this range, the H-type firm invests q_H^* in a separating PBE.[2]

To illustrate Proposition 4, Figure 3.4 maps short-termism α against the percentage change in the firm's profit if $\theta < \hat{\theta}$, compared to a benchmark case when $\alpha = 0$.[3] In this example, $h = 0.2$ and $\theta = 1.2$, which are the same as in Figure 3.3. As shown in Figure 3.4, the long-term profit is greatest for the L-type firm when $\underline{\alpha_S} \leq \alpha < \overline{\alpha_P}$, greatest for the H-type firm when $\alpha = \overline{\alpha_P}$, and greatest for the firm if her type is unknown when $\underline{\alpha_S} \leq \alpha \leq \overline{\alpha_S}$.

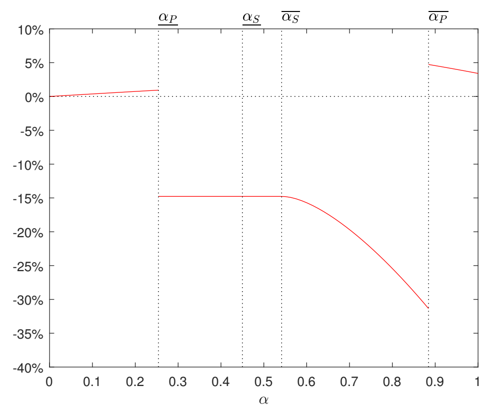
This figure makes clear that eliminating short-termism, as advocated by many practitioners, can be detrimental to a firm's long-term value. When $\theta < \max \theta_{LS}$, Definition 3 implies that $\underline{\alpha_S} > 0$. In this case, $\alpha = 0$ is *not* an optimal choice for either firm type. In the example shown in Figure 3.4, the L-type firm's long-term profit declines by approximately 8 percent when $\alpha = 0$ compared to the optimal short-termism for the L-type firm. The H-type firm's long-term profit declines roughly by 5 percent when $\alpha = 0$ compared to the optimal short-termism for the H-type firm. Figure 3.4 also shows instances in which increasing α can lead to a higher long-term profit. For example, the H-type firm has a higher profit when α is increased from 0.1 to 0.2. Practitioners have expressed concerns that short-termism leads firms to cut investments and harms their long-term value [14, 63]. While this is true in some cases, our findings also show that a reduction in long-term investment brought on by higher

[2]In the proof of Proposition 4, we also show the impact of short-termism on the firm's *expected* long-term value when she does not have private information about her type. In practice, instruments that affect the level of short-termism, such as incentive compensation plans, may be put in place by a firm's board of directors before the firm realizes her type. In this case, the optimal level of short-termism is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_S}]$. Over this wide range of α, both firm types can invest optimally in a costless separating PBE.

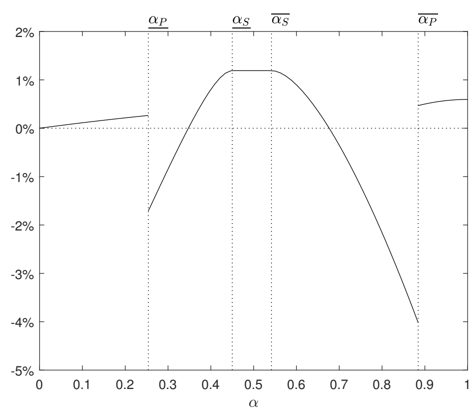
[3]The pattern in Figure 3.4 holds if we use no information asymmetry as the benchmark, as in Section 3.5. We provide an example if $\theta \geq \hat{\theta}$ in Figure 3.A.1 of the Appendix.



(a) L-type



(b) H-type



(c) Expected type

Figure 3.4: Impact of α on the Firm's Long-term Profit with Competitive Entrant

short-termism can sometimes lead to an improvement in the firm's long-term value. For example, we show in Figure 3.3(a) that the L-type firm invests less when $\alpha = 0.3$ than when $\alpha = 0.1$, and the H-type firm invests less when $\alpha = 0.9$ than when $\alpha = 0.8$. Figure 3.4 shows that the firm's long-term profit is higher in both cases when investment is lower and α is higher. In Section 3.7 of the Appendix, we consider the impact of the firm's short-termism on the long term profit of the entrant and the social welfare and consumer surplus in the market.

3.7 Impact of Short-termism in a Duopoly Setting

Our main model using a sequential move game with a less informed entrant captures a dynamic market where competitive entry is prevalent. We extend our analysis to consider a mature market with two established competitors. The firms move simultaneously and know the true market demand, so there is no information asymmetry between them and they are the same type. We continue to assume the market demand can be either high or low (i.e., two types). We also assume that both firms have the same short-termism level. This last assumption is reasonable in an established, competitive market in which firms have a similar cost of capital, capital structure, and investor pressures. The investor does not know the true market demand, and updates his posterior belief for demand upon observing the firms' capacity investment levels.

For ease of comparison to the main model, we continue to distinguish the two firms using our notation F and E, although neither firm is an entrant in this setting. At the beginning of period 1, both firm F and E learn the market demand (i.e., their type). The two firms engage in a Cournot competition, and the firms decide their capacity investment levels q_F and q_E, respectively. Next, the investor updates her belief about the firm type upon observing the firms' capacity investment. The investor then assigns short-term valuation v_F to firm F and short-term valuation v_E to firm E. In period 2, market demand is realized, and the long-term profits of the firms are determined.

The investor chooses v_F and v_E to maximize its expected utility based on its posterior belief, i.e., $(v_F^*, v_E^*) = \arg\max EU_I = \arg\max(\lambda U_I(t_H) + (1-\lambda)U_I(t_L))$, where $U_I(t) = -(v_F - \pi_F(t))^2 - (v_E - \pi_E(t))^2$. From this, we obtain that $v_i^* = \lambda \pi_i(t_H) +$

$(1-\lambda)\pi_i(t_L)$ where $i \in \{E, F\}$. This shows that the investor's short-term valuation assigned to both firms is the weighted average of the firm's long-term profits based on the investor's posterior belief. Thus, the firms' utility functions after plugging in the investor's best response (v_F^*, v_E^*) is

$$U_i(t, q_i, q_j) = \alpha(\lambda A_H + (1-\lambda)A_L - q_i - q_j)q_i + (1-\alpha)(A_t - q_i - q_j)q_i \tag{3.1}$$

where $i, j \in \{E, F\}$ and $i \neq j$. Equation (3.A.19) shows that the firms' utility is higher when $t = H$ compared to $t = L$. This indicates that the H-type is the mimicked type and the L-type is the mimicking type. This is in contrast to our main model, where the mimicked firm type will switch from the H-type to the L-type as short-termism level α decreases below $\frac{1}{2}$ due to dual signaling.

We next analyze the firms' equilibrium capacity investment. We examine the emergence of separating PBE and pooling PBE. As in our main analysis, we focus on the least-cost separating PBE and the pooling PBE that will survive the Undefeated refinement.

The firms' equilibrium capacity investment levels are obtained by finding the fixed point(s) of the best responses of the two firms. We characterize each player's strategy in the least-cost separating PBE in Definition 6. Due the concavity of the firms' utility functions, the least-cost separating PBE always exists, which we formally show in the technical details of Definition 6 in Section 3.9.1 of the Appendix. We use $q_1(\cdot)$ to denote the boundary for a L-type posterior belief, and its mathematical formula is shown in the technical details of Definition 6.

Definition 4 *In the least-cost separating PBE, each player's strategy is:*

(i) The L-type firms invest $q_L^S = \frac{A_L}{3}$*; the H-type firms invest* $q_H^S > \frac{A_H}{3}$ *if* $\theta < \frac{3}{(3-4\alpha)^+}$

and $q_H^S = \frac{A_H}{3}$ *if* $\theta \geq \frac{3}{(3-4\alpha)^+}$.

(ii) The investor assigns valuation $v_E^* = v_F^* = (A_t - 2q_t^S)q_t^S$ *for* $t \in \{H, L\}$.

(iii) The equilibrium is supported by the investor's posterior belief λ *such that* $\lambda = 1$ *for region* $\{(q_F, q_E) : q_F \geq q_1(q_E)$ *and* $q_E \geq q_1(q_F)\}$ *and* $\lambda = 0$ *otherwise.*

Quantities $\frac{A_H}{3}$ and $\frac{A_L}{3}$ are the benchmark quantities of the firms when there is complete information. It is straightforward to obtain that these quantities are the equilibrium capacity investment when $\alpha = 0$. Definition 6 shows that there are deviations from these benchmarks when $\alpha > 0$. The H-type firms will invest more than $\frac{A_H}{3}$ in the least-cost separating PBE when θ is sufficiently small. In contrast, the L-type will invest $\frac{A_L}{3}$. The condition $\theta < \frac{3}{(3-4\alpha)^+}$ is equivalent to $\alpha > \frac{3}{4} - \frac{3}{4\theta}$, so the H-type firms will over-invest when short-termism level α is sufficiently large. As shown in the technical details of Definition 6 in Section 3.9.1 of the Appendix, the H-type firms' separating quantity q_H^S increases in α when $\theta < \frac{3}{(3-4\alpha)^+}$.

The pooling PBE is obtained by finding the fixed point(s) of the pooling response(s) of each firm. Under the Undefeated refinement, a pooling PBE exists when both firm types for both firms obtain a higher utility when they pool at q^P compared to their utilities under the least-cost separating PBE. As in our main model, we refine the set of PBE using the LMSE when multiple pooling PBE exist. Due the concavity of the firms' utility functions, a unique LMSE exists in our setting. We characterize each player's strategy in the pooling LMSE in Definition 7. We use $q_2(\cdot)$ to denote the boundary for a weighted posterior belief, and its mathematical formula is shown in the technical details of Definition 7 in Section 3.9.1 of the Appendix, along with the conditions for the existence of the pooling LMSE.

Definition 5 *In the pooling LMSE, each player's strategy is:*

(i) Both firm types invest q^P *where* $\frac{A_L}{3} < q^P < \frac{A_H}{3}$

(ii) The investor assigns valuation $v_E^* = v_F^* = (\mu - 2q^P)q^P$ *for* $t \in \{H, L\}$.

(iii) The equilibrium is supported by the investor's posterior belief λ *such that* $\lambda = 1$ *for region* $\{(q_F, q_E) : q_F \geq q_1(q_E)$ *and* $q_E \geq q_1(q_F)\}$, $\lambda = h$ *for region* $\{(q_F, q_E) : q_2(q_E) \leq q_F < q_1(q_E)$ *and* $q_2(q_F) \leq q_E < q_1(q_F)\}$, *and* $\lambda = 0$ *otherwise.*

Definition 7 shows that in the pooling LMSE, the H-type firms will under-invest and the L-type firms will over-invest to conceal their types. As shown in the technical details of Definition 7, the pooling quantity q^P decreases in α.

We further examine the impact of short-termism on operational distortion and long-term profits. Definition 6 shows that the H-type firms over-invest and the L-type firms invest optimally in the least-cost separating PBE when α is sufficiently large. Definition 7 shows that the H-type firms under-invest and the L-type firms over-invest in the pooling PBE. These operational distortions have a significant effect on the firms's long-term profits. Proposition 5 shows the impact of α on firms' long-term profits in the two types of equilibrium. We denote each firm's profit in the H-type demand and L-type demand when $\alpha = 0$ as π_H^0 and π_L^0, respectively. We employ two threshold values for α, denoted as α_0 and α_1, to describe the impact of α on the profit of the H-type firm. The formulas for these thresholds are shown in the proof of Proposition 5.

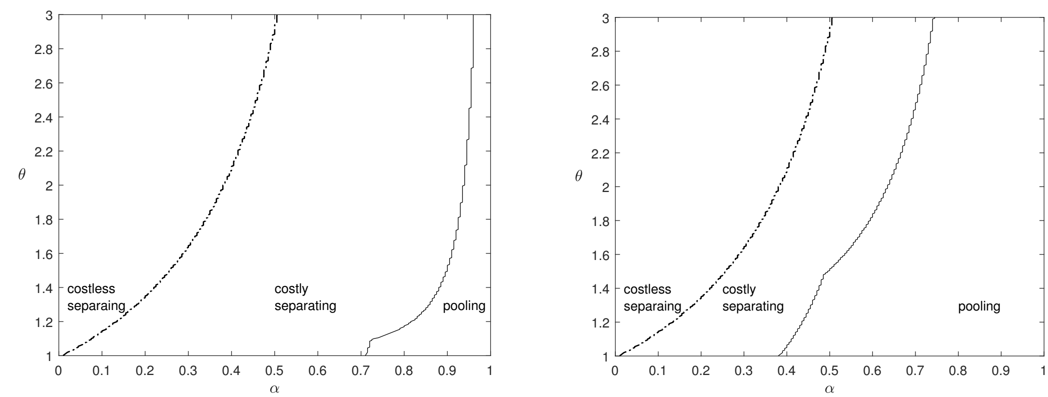
Proposition 5 *When* $\alpha > 0$, *the firms' profits exhibit the following characteristics:*

(i) In the least-cost separating PBE in which the H-type firms over-invest, the profits of the H-type firms decrease in α and are smaller than π_H^0.

(ii) In the pooling LMSE, the profits of the L-type firms increase in α but are smaller than π_L^0; the profits of the H-type firms increases in α if $\alpha \leq \alpha_0$ and decreases in α if $\alpha > \alpha_0$, and the profits of the H-type firms are larger than π_H^0 if and only if $\alpha < \alpha_1$.

Proposition 5 establishes that the impact of short-termism on the firm's long term profit depends on the firms type. In some instances the firms' long-term profits can be higher with positive short-termism compared to their profits with no short-termism. If the equilibrium is a pooling PBE when $\alpha = \alpha_0$, the H-type firm can even achieve monopoly profits when $\alpha = \alpha_0$. The L-type firm's profit does not increase with α in a duopoly setting, but it also does not decrease over wide ranges of α. These findings indicate that positive short-termism is not necessarily harmful, and in some cases it can even help the firm's long term value, depending on the firm's type.

We conduct a simulation to demonstrate the equilibrium outcomes. We let α take 201 possible values between 0 and 1 with an increment of 0.005, θ take 398 values including 1.001 and 397 possible values between 1.02 and 3 with an increment of 0.005, h take 11 values including 0.01, 0.99 and 9 possible values between 0.1 and 0.9 with an increment of 0.1. For each value of h, there are $201 \times 398 = 79{,}998$ examples due to the combinations of α and θ. We evaluate the conditions on firms' utility for all the examples to determine the type of equilibrium outcomes. Figure 3.5 shows the segmentation of equilibrium outcomes relative to α and θ at $h = 0.2$ and $h = 0.8$. The segmentation of equilibrium outcomes exhibit a similar pattern for other values of h. The segmentation boundaries are discontinuous because our simulated data are discrete. We have several ob-



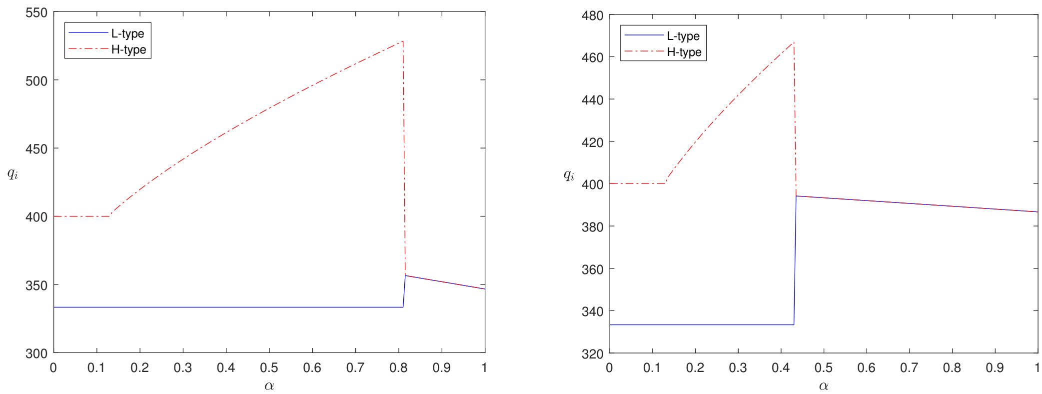
(a) $h = 0.2$ (b) $h = 0.8$

Figure 3.5: Segmentation of Equilibrium Types in Cournot Competition

servations from the simulation. First, separating PBE transits from costless to costly when α is beyond a threshold. This threshold of α can be formally obtained from Definition 6 as $\frac{3}{4} - \frac{3}{4\theta}$. Second, a pooling PBE emerges for all values of h in the simulation when α is sufficiently large and θ is sufficiently small. Moreover, the likelihood of pooling PBE increases as h increases.

We present one of the competing firms' equilibrium capacity investment with respect to α in Figure 3.6 using our simulated data when $\theta = 1.2$, $A_L = 1000$, and $h = 0.2$ or $h = 0.8$. The effect of α on operational distortion is not monotonic. For example, in Figure 3.6(a), as α increases beyond 0.13 and below 0.81, the H-type firm will over-invest in the separating PBE, and this over-investment increases in α. As α increases beyond 0.81, the H-type firm will under-invest and the L-type firm will over-invest in the pooling PBE. The under-investment of the H-type increases and the over-investment of the L-type decreases in α.

To illustrate Proposition 5, Figure 3.7 maps α against the percentage change in one competing firm's profit compared to the benchmark case when $\alpha = 0$, using the simulated data when $\theta = 1.2$ and $h = 0.2$ or $h = 0.8$. In both examples,

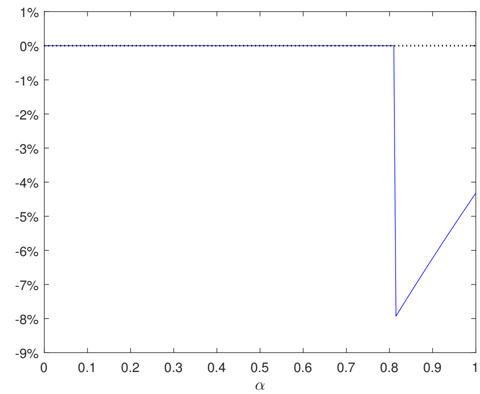


(a) $h = 0.2$ (b) $h = 0.8$

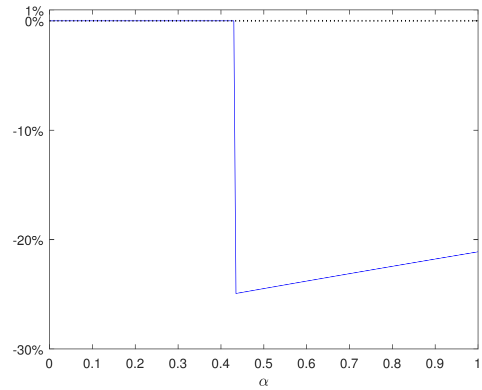
Figure 3.6: Impact of α on Firms' Capacity Investment in Cournot Competition

the L-type firms' profits are either the same or lower when $\alpha > 0$ compared to $\alpha = 0$. However, the H-type firms' profits can be higher when α is sufficiently large compared to $\alpha = 0$. For example, the H-type firms' profits are higher when α is beyond 0.81 in Figure 3.7(c) and when α is beyond 0.43 in Figure 3.7(d). The firms' expected long-term profits are either the same or lower when $\alpha > 0$. The incremental impact of α can be positive. For example, when $\alpha = 0.9$, it is beneficial for the long-term value of both firm types if the short-termism level increases.

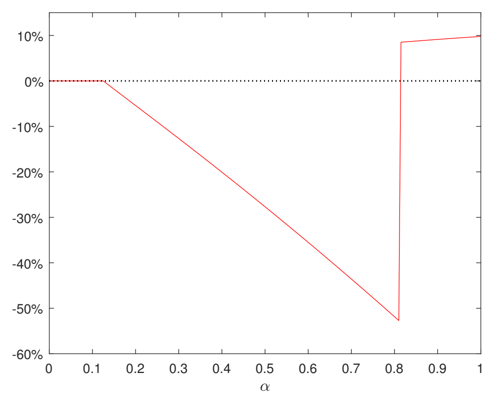
The beneficial range of short-termism for the H-type firms' long-term profits can be explained by the operational distortion. When the two firms with no short-termism engage in a Cournot competition, both firms would invest $\frac{A_t}{3}$. We also obtain that both firms would invest $\frac{A_t}{4}$ if they engage in monopolistic collusion. Had both firms invested less and at a level closer to $\frac{A_t}{4}$, they could have obtained more profits. When short-termism is sufficiently large such that pooling PBE emerges, the H-type firms under-invests, which allows the H-type firms to choose an investment level closer to $\frac{A_t}{4}$. We show in Proposition 5 that the H-type firms will choose an investment level closer to $\frac{A_t}{4}$ if $\alpha < \alpha_1$. For the



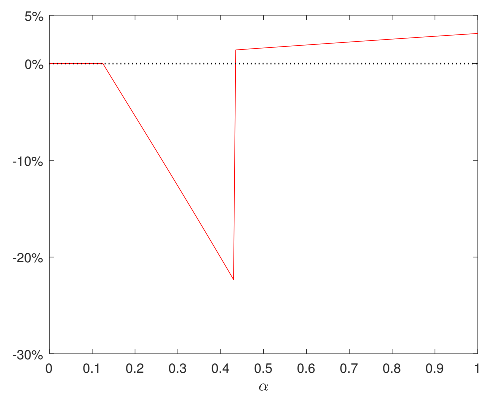
(a) h = 0.2, L-type



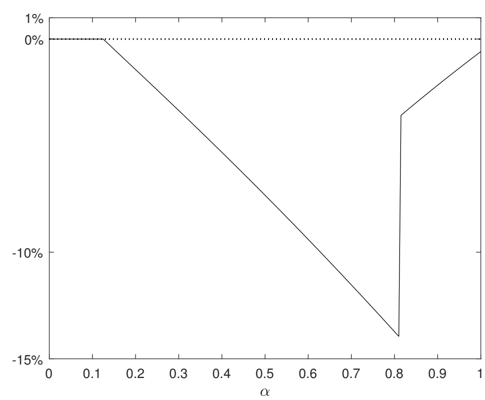
(b) h = 0.8, L-type



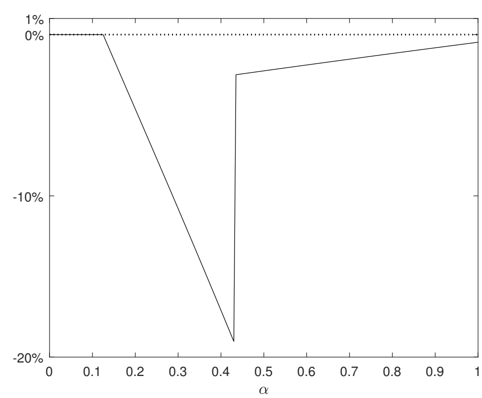
(c) h = 0.2, H-type



(d) h = 0.8, H-type



(e) h = 0.2, Expected type



(f) h = 0.8, Expected type

Figure 3.7: Impact of α on Firms' Long-term Profits in Cournot Competition

two examples in Figure 3.7, the H-type firms obtain a higher profits in the entire range of α where pooling emerges.

3.8 Implications and Conclusions

Short-termism is pervasive in the corporate world and practitioners have asserted that it is detrimental to a firm's long-term value. In this paper, we study the impact of short-termism on a firm's capacity decisions and long-term profit in the presence of competition and information asymmetry. We find that operational distortion can be induced not only at a high level of short-termism but also at a low level of short-termism. We also find that short-termism can be beneficial to a firm's long-term value under certain circumstances. This is due to the compound interaction of information asymmetry, a competitive market, and short-termism. At the risk of over-simplifying the mechanism, short-termism can induce the incumbent firm to invest optimally, to over-invest, or to restrict her investment closer to monopolistic levels, despite the presence of competition. The competitor, for his part, may be deprived of a clear signal of the market demand, and cannot take advantage of the incumbent firm's under-investment. In those cases, the net effect can be an increase in the incumbent firm's long term profit. Beyond the firm's long-term value, we document in the Section 3.7 of the Appendix that short-termism can benefit consumer and social welfare in certain circumstances.

Our paper provides new insights into the role that short-termism plays in a firm's investment behavior. Our finding that short-termism can increase the firm's long-term value challenges the generally held belief that short-termism erodes long-term value. In addition, our finding of under-investment in competitive markets when short-termism is low challenges the perception that *low* investment is definitely caused by *high* short-termism. Our model provides a clear and intuitive framework to show that the under-investment can be a rational choice, and value-enhancing for the long-term.

Although short-termism is decried as damaging a firm's long-term value and competitiveness, it is widely acknowledged to occur. The academic and practitioner literature suggests that a firm's short-termism can be influenced by stock-based managerial incentives [30, 52, 24, 168], CEO tenure [101], CEO outsider status [61], corporate culture [33], the influence of hedge funds and activist investors [7, 56], the need to raise equity capital [153], short-term financial reporting requirements and perceptions [134, 74, 63], and executives' interest in burnishing their reputations and careers [129, 82]. Some of these factors can be influenced by the company's board (compensation plans, hiring outside CEOs), regulators (reporting obligations), or investors (hedge fund and activist investor pressure). For other factors it is unclear whether any single entity can exert control. While independent actions by stand-alone entities may influence the level of short-termism at a company, several thought leaders on short-termism have highlighted the need for a coordinated response that includes company executives, regulators, and investors [14, 53, 9]. Strategic decisions aimed at manipulating these factors are beyond the scope of our study. However, if a firm's board can influence the short-term orientation of the firm, our findings provide support that it may be beneficial to the firm's long-term value. In the face of competitive pressures, boards may be justified in thoughtfully structuring managerial incentives to adjust and even increase short-termism. [4]

[4] In presence of competitive entry, restructuring managerial compensation to include more stock-based compensation is not without precedent. The academic literature provides three practical examples of this in the context of bank deregulation [58, 20, 62], airline deregulation [35], and electric utility deregulation [106]. A common empirical finding across these disparate industry settings is that companies substantially increased the stock-based compensation for executives following deregulation. As noted earlier, stock-based compensation can induce managers to cater to investors, and has been identified as a source of corporate short-termism.

3.9 Appendix

3.9.1 Proofs, Definitions, and Technical Details for Primary Model

3.9.1.1 Definition of Strict Dominance

[121] (p.469) state that a signal, q, is strictly dominated for a type $\tau_i \in T$ if there exists another signal q' such that the following inequality holds:

$$\max_{\rho \in P^*(T,q)} U(\tau_i, q, \rho) < \min_{\rho \in P^*(T,q')} U(\tau_i, q', \rho)$$

From the set $S(q)$ consisting of all types τ_i such that this inequality does not hold. Then a PBE has reasonable beliefs if for all q with $S(q) \neq \emptyset$, $\lambda(\tau_i) > 0$ only if $\tau_i \in S(q)$.

Strict Dominance states that a signal is strictly dominated for a type if the best utility that type can obtain by sending that signal is strictly lower than the worst utility that type can obtain by sending some other signal. A PBE has reasonable beliefs if those beliefs do not put a positive probability on any type sending a strictly dominated signal.

3.9.1.2 Proof of Lemma 3

For the continuous support $[0, 1]$ of λ, $\frac{\partial U_F(t,q_F|\lambda)}{\partial \lambda} = \frac{(2\alpha-1)(A_H-A_L)}{2} q_F$. Hence, $U_F(t, q_F|\lambda)$ decreases in λ if $\alpha < \frac{1}{2}$ and increases in λ if $\alpha > \frac{1}{2}$. Thus, $U_F(t, q_F|\lambda = 0) > U_F(t, q_F|\lambda = h) > U_F(t, q_F|\lambda = 1)$ if $\alpha < \frac{1}{2}$, and $U_F(t, q_F|\lambda = 1) > U_F(t, q_F|\lambda = h) > U_F(t, q_F|\lambda = 0)$ if $\alpha > \frac{1}{2}$. By Equation (3.7), $U_F(t_H, q_F|\lambda) = \frac{A_H - q_F}{2} q_F$ and $U_F(t_L, q_F|\lambda) = \frac{A_L - q_F}{2} q_F$ if $\alpha = \frac{1}{2}$. ■

3.9.1.3 Proof of Lemma 4

Under complete information, the firm's type t is common knowledge. From Equation (3.2), the investor's best response $v^*(t, q_F, q_E) = \pi_F(t, q_F, q_E)$. Thus,

$U_F(t, q_F, q_E, v^*) = \pi_F(t, q_F, q_E)$ from Equation (3.1). Since the entrant knows the firm's type, the entrant's best response is $q_E^*(t_L, q_F) = \frac{A_L - q_F}{2}$ or $q_E^*(t_H, q_F) = \frac{A_H - q_F}{2}$. Thus, the L-type firm will choose the maximizer of $U_F(t_L, q_F, q_E^*, v^*) = (A_L - q_F - \frac{A_L - q_F}{2})q_F$, i.e., $\frac{A_L}{2}$. The H-type firm will choose the maximizer of $U_F(t_H, q_F, q_E^*, v^*) = (A_H - q_F - \frac{A_H - q_F}{2})q_F$, i.e., $\frac{A_H}{2}$. ■

For ease of exposition, we add the following notations for the remaining sections: $g = 1 - h$, $\theta = \frac{A_H}{A_L}$, $\gamma = \frac{A_L}{A_H}$, $x = \theta - 1$, $y = 1 - \gamma$, $a = 1 - 2\alpha$, and $b = 2\alpha - 1$. Note that $\theta > 1$, $0 < \gamma < 1$, $x > 0$, $0 < y < 1$, $0 < a \leq 1$ if $0 \leq \alpha < \frac{1}{2}$, and $0 < b \leq 1$ if $\frac{1}{2} < \alpha \leq 1$.

3.9.1.4 Technical Details of Definition 1 and Proof of Proposition 1

We show the existence of and each player's strategy in the least-cost separating PBE in two cases: (i) $\alpha < \frac{1}{2}$, (ii) $\alpha > \frac{1}{2}$. For convenience, we suppress the best responses of the investor and the entrant, i.e., $U_F(t, q_F|\lambda) = U_F(t, q_F, q_E^*, v^*|\lambda)$.

Case 1. $\alpha < \frac{1}{2}$. In this case, the H-type mimics the L-type.

We first show the set of strict dominated quantities by the H-type defined in Equation (3.1) is non-empty. Note that $U_F(t_H, q_H^*|\lambda = 1) = \frac{A_H^2}{8}$ and $U_F(t_H, q_F = 0|\lambda = 0) = 0 < \frac{A_H^2}{8}$. Lemma 3 implies that $U_F(t_H, q_H^*|\lambda = 0) > U_F(t_H, q_H^*|\lambda = 1)$. As $U_F(t_H, q_F|\lambda = 0)$ increases in q_F for $q_F < q_H^*$, there exists a unique $0 < \bar{q} < q_H^*$ such that $U_F(t_H, \bar{q}|\lambda = 0) = U_F(t_H, q_H^*|\lambda = 1)$ and $U_F(t_H, q_F|\lambda = 0) < U_F(t_H, q_H^*|\lambda = 1)$ for any $q_F < \bar{q}$. Thus, the quantities in the non-empty set $\{q_F : q_F < \bar{q}\}$ are strictly dominated by the H-type firm. We solve that

$$\bar{q} = \frac{(2 - 2\alpha)A_H - (1 - 2\alpha)A_L - \sqrt{((3 - 2\alpha)A_H - (1 - 2\alpha)A_L)(1 - 2\alpha)(A_H - A_L)}}{2}. \tag{3.A.1}$$

Second, we compare $\bar{q}$ and q_L^*. Since the quantity in the set $\{q_F : q_F < \bar{q}\}$ that maximizes $U_F(t_L, q_F|\lambda = 0)$ is the least-cost separating quantity for the L-type, $q_L^S = \min(\bar{q}, q_L^*)$. Note that

$$\bar{q} - q_L^* = \frac{(\theta - 1)(\theta - (3 - 4\alpha))A_L}{(2 - 2\alpha)(\theta - 1) + \sqrt{((3 - 2\alpha)\theta - (1 - 2\alpha))(1 - 2\alpha)(\theta - 1)}}.$$

Thus, $q_L^S = q_L^*$ if $\theta \geq \theta_{LS}$ and $q_L^S = \bar{q}$ if $\theta < \theta_{LS}$, where

$$\theta_{LS} = 3 - 4\alpha. \tag{3.A.2}$$

Next, we show that the L-type firm has no incentive to deviate from q_L^S to be perceived as an H-type firm, i.e., $U_F(t_L, q_L^S|\lambda = 0) > \max U_F(t_L, q_F|\lambda = 1)$. The unconstrained maximizer of $U_F(t_L, q_F|\lambda = 1)$ is $q_F^*(t_L|\lambda = 1) = \frac{-(1-2\alpha)A_H+(2-2\alpha)A_L}{2}$. Note that $q_F^*(t_L|\lambda = 1) \leq 0$ if $\theta \geq \frac{2-2\alpha}{1-2\alpha}$. We discuss the following three cases. (i) If $\theta \geq \frac{2-2\alpha}{1-2\alpha}$, then $\max U_F(t_L, q_F|\lambda = 1) = 0 < U_F(t_L, q_L^S|\lambda = 0)$. (ii) If $\theta_{LS} \leq \theta < \frac{2-2\alpha}{1-2\alpha}$, then $\max U_F(t_L, q_F|\lambda = 1) = U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 1)$ and $q_L^S = q_L^*$. Note that $U_F(t_L, q_L^*|\lambda = 0) \geq U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 0)$. Lemma 3 implies that $U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 0) > U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 1)$. Thus, $U_F(t_L, q_L^S|\lambda = 0) > \max U_F(t_L, q_F|\lambda = 1)$. (iii) If $\theta < \min(\theta_{LS}, \frac{2-2\alpha}{1-2\alpha})$ (i.e., $x < \min(2a, \frac{1}{a})$), then $\max U_F(t_L, q_F|\lambda = 1) = U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 1)$ and $q_L^S = \bar{q}$. Plug $\bar{q}$ into Equation (3.7), we obtain that

$$U_F(t_L, \bar{q}|\lambda = 0) = \frac{A_L^2}{8}\left(1 - 2ax - (2a^2 + 4a + 1)x^2 + 2(1 + a)x\sqrt{((a + 2)x + 2)ax}\right). \tag{3.A.3}$$

Plug $q_F^*(t_L|\lambda = 1)$ into Equation (3.7), we obtain that $U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 1) = \frac{A_L^2}{8}(-ax + 1)^2$, and hence $U(t_L, \bar{q}|\lambda = 0) - U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 1) = -\frac{A_L^2(a+1)x^2\left((5a^2-2a+1)x-8a\right)}{8\left(x+3ax+2\sqrt{((a+2)x+2)ax}\right)}$. Note that $5a^2 - 2a + 1 > 0$ for $0 < a \leq 1$, and $(5a^2 - 2a + 1)x - 8a < (5a^2 - 2a + 1)2a - 8a = 2a(5a + 3)(a - 1) \leq 0$. Thus, $U(t_L, \bar{q}|\lambda = 0) > U_F(t_L, q_F^*(t_L|\lambda = 1)|\lambda = 1)$.

In summary, when $\alpha < \frac{1}{2}$, the least-cost separating PBE exists in which the L-type firm invests $q_L^S = \min(\bar{q}, q_L^*)$ and the H-type firm invests $q_H^S = q_H^*$. By strict dominance, the posterior belief λ in equilibrium satisfies that $\lambda = 0$ if $q_F \leq \bar{q}$ and $\lambda = 1$ if $q_F > \bar{q}$.

Case 2. $\alpha > \frac{1}{2}$. In this case, the L-type mimics the H-type.

We first show the set of strict dominated quantities by the L-type defined in Equation (3.3) is non-empty. Note that $U_F(t_L, q_L^*|\lambda = 0) = \frac{A_L^2}{8}$. Lemma 3 implies that $U_F(t_L, q_L^*|\lambda = 1) > U_F(t_L, q_L^*|\lambda = 0)$. As $U_F(t_L, q_F|\lambda = 1)$ is concave, there exists a unique $\underline{q} > q_L^*$ such that $U_F(t_L, \underline{q}|\lambda = 1) = U_F(t_L, q_L^*|\lambda = 0)$ and $U_F(t_L, q_F|\lambda = 1) > U_F(t_L, q_L^*|\lambda = 0)$ for any $q_F > \underline{q}$. Thus, the quantities in the non-empty set $\{q_F : q_F > \underline{q}\}$ are strictly dominated by the L-type firm. We solve that

$$\underline{q} = \frac{(2\alpha - 1)A_H + (-2\alpha + 2)A_L + \sqrt{((2\alpha - 1)A_H + (-2\alpha + 3)A_L)(2\alpha - 1)(A_H - A_L)}}{2}. \tag{3.A.4}$$

Second, we compare $\underline{q}$ and q_H^*. Since the quantity in the set $\{q_F : q_F > \underline{q}\}$ that maximizes $U_F(t_H, q_F|\lambda = 1)$ is the least-cost separating quantity for the H-type, $q_H^S = \max(\underline{q}, q_H^*)$. Note that

$$\underline{q} - q_H^* = A_L \frac{(\theta - 1)(1 - (3 - 4\alpha)\theta)}{(2 - 2\alpha)(\theta - 1) + \sqrt{((2\alpha - 1)\theta + (-2\alpha + 3))(2\alpha - 1)(\theta - 1)}}.$$

Thus, if $\alpha \geq \frac{3}{4}$, or if $\alpha < \frac{3}{4}$ and $\theta < \frac{1}{3-4\alpha}$, $q_H^S = \bar{q}$. If $\alpha < \frac{3}{4}$ and $\theta \geq \frac{1}{3-4\alpha}$, $q_H^S = q_H^*$. That is, $q_H^S = \bar{q}$ if $\theta < \theta_{HS}$, and $q_H^S = q_H^*$ if $\theta \geq \theta_{HS}$, where

$$\theta_{HS} = \frac{1}{(3-4\alpha)^+}. \tag{3.A.5}$$

Next, we show that the H-type firm has no incentive to deviate from q_H^S to be perceived as an L-type firm, i.e., $U_F(t_H, q_H^S|\lambda = 1) > \max U_F(t_H, q_F|\lambda = 0)$. The unconstrained maximizer of $U_F(t_H, q_F|\lambda = 0)$ is $q_F^*(t_H|\lambda = 0) = \frac{(2\alpha-1)A_L+(-2\alpha+2)A_H}{2} > 0$. We discuss the following two cases. (i) If $\theta \geq \theta_{HS}$, then $q_H^S = q_H^*$. Note that $U_F(t_H, q_H^*|\lambda = 1) > U_F(t_H, q_F^*(t_H|\lambda = 0)|\lambda = 1)$. Lemma 3 implies that $U_F(t_H, q_F^*(t_H|\lambda = 0)|\lambda = 1) > U_F(t_H, q_F^*(t_H|\lambda = 0)|\lambda = 0)$. Thus, $U_F(t_H, q_H^S|\lambda = 1) > \max U_F(t_H, q_F|\lambda = 0)$. (ii) If $\theta < \theta_{HS}$ (i.e., $b \geq \frac{1}{2}$, or $b < \frac{1}{2}$ and $y < 2b$), then $q_H^S = \underline{q}$. Plug $\underline{q}$ into Equation (3.7), we obtain that

$$U_F(t_H, \underline{q}|\lambda = 1) = \frac{A_H^2}{8}\left(1 - 2by - (2b^2 - 4b + 1)y^2 + 2y(1-b)\sqrt{((b-2)y+2)by}\right). \tag{3.A.6}$$

Plug $q_F^*(t_H|\lambda = 0)$ into Equation (3.7), we obtain that $U_F(t_H, q_F^*(t_H|\lambda = 0)|\lambda = 0) = \frac{A_H^2}{8}(-by + 1)^2$, and hence $U_F(t_H, \underline{q}|\lambda = 1) - U_F(t_H, q_F^*(t_H|\lambda = 0)|\lambda = 0) = \frac{A_H^2}{8}(1 - b)y\left(-y + 3by + 2\sqrt{((b-2)y+2)\,by}\right)$. If $b \geq \frac{1}{3}$, $U_F(t_H, \underline{q}|\lambda = 1) > U_F(t_H, q_F^*(t_H|\lambda = 0)|\lambda = 0)$. If $b < \frac{1}{3}$, $-y + 3by < 0$. Note that $(2\sqrt{((b-2)y+2)\,by})^2 - (-y + 3by)^2 = (8b - (5b^2 + 2b + 1)y)y > (8b - (5b^2 + 2b + 1)2b)y = 2b(3 - 5b)(b + 1)y > 0$. Thus, $U_F(t_H, \underline{q}|\lambda = 1) > U_F(t_H, q_F^*(t_H|\lambda = 0)|\lambda = 0)$.

In summary, when $\alpha > \frac{1}{2}$, the least-cost separating PBE exists in which the H-type firm invests $q_H^S = \max(\underline{q}, q_H^*)$ and the L-type firm invests $q_L^S = q_L^*$. By strict dominance, the posterior belief λ in equilibrium satisfies that $\lambda = 1$ if $q_F \geq \underline{q}$ and $\lambda = 0$ if $q_F < \underline{q}$. ■

3.9.1.5 Technical Details of Definition 2 and Proof of Proposition 2

We show the existence of and each player's strategy in the Pareto dominant pooling PBE in two cases: (i) $\alpha < \frac{1}{2}$, (ii) $\alpha > \frac{1}{2}$. For convenience, we suppress the best responses of the investor and the entrant, i.e., $U_F(t, q_F|\lambda) = U_F(t, q_F, q_E^*, v^*|\lambda)$.

Case 1. $\alpha < \frac{1}{2}$. In this case, the H-type mimics the L-type. We utilize Claim 1 to 3.

Claim 1 *Equation* (3.3) *holds if and only if* $\theta < \theta_{LP}$, *where* $\theta_{LP} = \frac{1+2(1-\sqrt{h}+h)\alpha+h(2-2\sqrt{h}+h)\alpha^2}{1+2\sqrt{h}\alpha+h(2-2\sqrt{h}+h)\alpha^2}$.

First, we show that Equation (3.3) holds only if $\theta < \min(1 + \frac{1}{(1-2\alpha)h}, 3 - 4\alpha)$. By Definition 1, $U_F(t_L, q_L^S|\lambda = 0) > 0$, and hence it requires that $U_F(t_L, q_F|\lambda = h) > 0$, i.e., $\theta < 1 + \frac{1}{(1-2\alpha)h}$. By Equation (3.7), the maximizer of $U_F(t_L, q_F|\lambda = h)$ is $q_F^*(t_L|\lambda = h) = \frac{((1-2\alpha)(-h)(\theta-1)+1)A_L}{2} < \frac{A_L}{2}$. Suppose $\theta \geq 3 - 4\alpha$, then $q_L^S = q_L^*$ and $\max U_F(t_L, q_F|\lambda = h) = \frac{(q_F^*(t_L|\lambda=h))^2}{2} < \frac{A_L^2}{8} = U_F(t_L, q_L^S|\lambda = 0)$, which violates Equation (3.3). Thus, Equation (3.3) holds only if $\theta < \min(1 + \frac{1}{(1-2\alpha)h}, 3 - 4\alpha)$, i.e., $x < \min(\frac{1}{ah}, 2a)$.

If $\theta < \min(1 + \frac{1}{(1-2\alpha)h}, 3 - 4\alpha)$, Equation (3.3) simplifies to $U_F(t_L, q_F|\lambda = h) > U_F(t_L, \bar{q}|\lambda = 0)$. Plug in $U_F(t_L, \bar{q}|\lambda = 0)$ from Equation (3.A.3) and $U_F(t_L, q_F|\lambda = h) = \frac{A_L^2}{8}(-ahx + 1)^2$, we obtain that

$$\begin{aligned} & U_F(t_L, q_F|\lambda = h) - U_F(t_L, \bar{q}|\lambda = 0) \\ = \;& \tfrac{A_L^2}{8} x \left(((h^2 + 2)a^2 + 4a + 1)x + 2a(1 - h) - 2(a + 1)\sqrt{((a + 2)x + 2)ax} \right). \end{aligned}$$

Define $f_1(x) = \left((h^2 + 2)a^2 + 4a + 1\right)x + 2a(1 - h)$, $f_2(x) = 2(a + 1)\sqrt{((a + 2)x + 2)\,ax}$, and $f_3(x) = (f_1(x))^2 - (f_2(x))^2$. Since $f_1(x) > 0$ and $f_2(x) > 0$, $U_F(t_L, q_F|\lambda = h) > U_F(t_L, \bar{q}|\lambda = 0)$ if and only if $f_3(x) = Ax^2 - Bx + C > 0$, where $A = 1 + 2a^2(1 + 4a + 2a^2)h^2 + a^4h^4$, $B = 4a(1 + (1 + 4a + 2a^2)h - a^2h^2 + a^2h^3)$, and $C = 4a^2(1 - h)^2$. Note that $A > 0$, $B > 0$, and $C > 0$. Hence, the two roots of $f_3(x)$ satisfy $0 < x_1 < x_2$ and

$$x_1 = \frac{2a\left(-1 + \sqrt{h}\right)^2}{1 + 2a\sqrt{h} + a^2h\left(2 - 2\sqrt{h} + h\right)}, \quad x_2 = \frac{2a\left(1 + \sqrt{h}\right)^2}{1 - 2a\sqrt{h} + a^2h\left(2 + 2\sqrt{h} + h\right)},$$

where $1 + 2a\sqrt{h} + a^2h(2 - 2\sqrt{h} + h) > 0$ and $1 - 2a\sqrt{h} + a^2h(2 + 2\sqrt{h} + h) > 0$. Thus, $f_3(x) > 0$ if and only if $x < x_1$ or $x > x_2$.

Next, we show that Equation (3.3) holds if and only if $x < x_1$ by proving that $x_2 > 2a$ and $x_1 < \min(\frac{1}{ah}, 2a)$. First, we prove $x_2 > 2a$. Note that $\frac{x_2}{2a} - 1 = \frac{n_1(a)}{1-2a\sqrt{h}+a^2h(2+2\sqrt{h}+h)}$, where $n_1(a) = -(2h + 2h^{3/2} + h^2)a^2 + 2\sqrt{h}a + 2\sqrt{h} + h$. Note that the two roots of $n_1(a)$ satisfy $a_1 < 0 < a_2$ and $a_2 = \frac{1}{\sqrt{h}} > 1$. Thus, $n_1(a) > 0$, i.e., $x_2 > 2a$ for any $a \in (0, 1]$. Second, we prove $x_1 < 2a$. Note that $\frac{x_1}{2a} - 1 = \frac{n_2(a)}{1+2a\sqrt{h}+a^2h(2-2\sqrt{h}+h)}$, where $n_2(a) = (-2h + 2h^{3/2} - h^2)a^2 - 2\sqrt{h}a - 2\sqrt{h} + h$. Note that $-2h + 2h^{3/2} - h^2 = -h\left((\sqrt{h} - 1)^2 + 1\right) < 0$ and $n_2(0) < 0$. Thus, $n_2(a) < 0$, i.e., $x_1 < 2a$ for any $a \in (0, 1]$. Third, we prove $x_1 < \frac{1}{ah}$. Note that $\frac{x_1}{\frac{1}{ah}} - 1 = \frac{n_3(a)}{1+2a\sqrt{h}+a^2h(2-2\sqrt{h}+h)}$, where $n_3(a) = (-2h^{3/2} + h^2)a^2 - 2\sqrt{h}a - 1$. Note that $-2h^{3/2} + h^2 = h^{3/2}(-2 + \sqrt{h}) < 0$ and $n_3(0) < 0$. Thus, $n_3(a) < 0$, i.e., $x_1 < \frac{1}{ah}$ for any $a \in (0, 1]$. Combing the conditions above, we obtain that Equation (3.3) holds if and only if $x < x_1$, i.e., $\theta < \theta_{LP}$, where $\theta_{LP} = 1 + x_1$ and hence

$$\theta_{LP} = \frac{1 + 2(1 - \sqrt{h} + h)a + h(2 - 2\sqrt{h} + h)a^2}{1 + 2\sqrt{h}a + h(2 - 2\sqrt{h} + h)a^2}. \tag{3.A.7}$$

Claim 2 *If Equation* (3.3) *holds, then Equation* (3.4) *holds.*

First, we show that if Equation (3.3) holds, then $q_F > q_L^S$. Suppose $q_F \le q_L^S$, then $U_F(t_L, q_F|\lambda = 0) \le U_F(t_L, q_L^S|\lambda = 0)$. By Lemma 3, $U_F(t_L, q_F|\lambda = h) < U_F(t_L, q_F|\lambda = 0)$. Thus, we obtain that $U_F(t_L, q_F|\lambda = h) < U_F(t_L, q_L^S|\lambda = 0)$, contradiction. Thus, $q_F > q_L^S$ if Equation (3.3) holds. Second, we show that $U_F(t_H, q_F|\lambda = h) > U_F(t_H, q_L^S|\lambda = 0)$ if Equation (3.3) holds. By Equation (3.7)

$$\begin{aligned} & U_F(t_H, q_F|\lambda = h) - U_F(t_H, q_L^S|\lambda = 0) \\ = \; & \tfrac{(1-2\alpha)(1-h)(A_H-A_L)+(A_H-A_L)+A_L-q_F}{2} q_F - \tfrac{(1-2\alpha)(A_H-A_L)+(A_H-A_L)+A_L-q_L^S}{2} q_L^S \\ = \; & \tfrac{(1-2\alpha)(-h)(A_H-A_L)+A_L-q_F}{2} q_F - \tfrac{A_L-q_L^S}{2} q_L^S + (1-\alpha)(A_H - A_L)(q_F - q_L^S) \\ = \; & U_F(t_L, q_F|\lambda = h) - U_F(t_L, q_L^S|\lambda = 0) + (1-\alpha)(A_H - A_L)(q_F - q_L^S). \end{aligned}$$

Since $q_F > q_L^S$, $U_F(t_H, q_F|\lambda = h) > U_F(t_H, q_L^S|\lambda = 0)$ holds. Note that $q_L^S = \bar{q}$, i.e., $U_F(t_H, q_L^S|\lambda = 0) = U_F(t_H, q_H^*|\lambda = 1)$ if Equation (3.3) holds. Thus, if Equation (3.3) holds, then $U_F(t_H, q_F|\lambda = h) > U_F(t_H, q_H^*|\lambda = 1) = U_F(t_H, q_H^S, |\lambda = 1)$, i.e., Equation (3.4) holds.

Combing Claim 1 and 2, we obtain that Pareto dominant pooling PBE exists if and only if $\theta < \theta_{LP}$ when $\alpha < \frac{1}{2}$.

Claim 3 *The pooling LMSE q^P is unique, $q^P = q_L^P$, and $q_L^P < q_L^* < q_H^*$.*

If $\theta < \theta_{LP}$, any q_F such that $U_F(t_L, q_F|\lambda = h) > U_F(t_L, q_L^S|\lambda = 0)$ is a Pareto dominant pooling PBE. By definition, $q^P = q_L^P = \arg\max U_F(t_L, q_F|\lambda = h)$. Note that by Equation (3.7), $U_F(t_L, q_F|\lambda = h)$ is concave and hence q_L^P is unique and

$$q_L^P = \frac{(2\alpha - 1)hA_H + ((-2\alpha + 1)h + 1)\,A_L}{2}. \tag{3.A.8}$$

Note that $q_L^P - q_L^* = -\frac{(1-2\alpha)h(A_H-A_L)}{2} < 0$. Thus, $q_L^P < q_L^* < q_H^*$. By Equation (3.4) and Equation (3.6), we further obtain that $q_E^*(q^P) = \frac{\mu - q^P}{2}$ and $v^*(q^P) = \frac{\mu - q^P}{2} q^P$. Given the existence of the pooling LMSE, the posterior belief λ in equilibrium satisfies $\lambda = 0$ if $q_F \le \bar{q}$, $\lambda = h$ if $q_F = q^P$, and $\lambda = 1$ if $q_F > \bar{q}$ and $q_F \ne q^P$.

Case 2. $\alpha > \frac{1}{2}$. In this case, the L-type mimics the H-type. We utilize from Claim 4 to 6.

Claim 4 *Equation* (3.4) *holds if and only if $\theta < \theta_{HP}$, where $\theta_{HP} = \frac{1-2\sqrt{g}b+g(2-2\sqrt{g}+g)b^2}{1-2(1-\sqrt{g}+g)b+g(2-2\sqrt{g}+g)b^2}$ if $\alpha < \alpha_0$ and $\theta_{HP} = \infty$ if $\alpha \ge \alpha_0$ and $\alpha_0 = \frac{3-2\sqrt{g}+g}{2(2-2\sqrt{g}+g)}$.*

First, we show that Equation (3.4) holds only if $\theta < \frac{1}{(3-4\alpha)^+}$. Suppose $\theta \geq \frac{1}{(3-4\alpha)^+}$, then $q_H^S = q_H^*$. By Lemma 3, $U_F(t_H, q_H^*|\lambda = 1) > U_F(t_H, q_F|\lambda = 1) > U_F(t_H, q_F|\lambda = h)$ for any $q_F > 0$, which violates Equation (3.4). Thus, Equation (3.4) holds only if $\theta < \frac{1}{(3-4\alpha)^+}$, i.e., $y < 2b$ if $b < \frac{1}{2}$, or $b \geq \frac{1}{2}$.

If $\theta < \frac{1}{(3-4\alpha)^+}$, Equation (3.4) simplifies to $U_F(t_H, q_F|\lambda = h) > U_F(t_H, \underline{q}|\lambda = 1)$. Plug in $U_F(t_H, \underline{q}|\lambda = 1)$ from Equation (3.A.6) and $U_F(t_H, q_F|\lambda = h) = \frac{A_H^2}{8}(-bgy+1)^2$, we obtain that

$$\begin{aligned} & U_F(t_H, q_F|\lambda = h) - U_F(t_H, \bar{q}|\lambda = 1) \\ = \; & \tfrac{A_H^2}{8} y \Big(((g^2+2)b^2 - 4b + 1)y + 2b(1-g) - 2(1-b)\sqrt{((b-2)y+2)\,by} \Big). \end{aligned} \tag{3.A.9}$$

Define $f_4(y) = ((g^2+2)b^2 - 4b + 1)y + 2b(1-g)$, $f_5(y) = 2(1-b)\sqrt{((b-2)y+2)\,by}$, and $f_6(y) = (f_4(y))^2 - (f_5(y))^2$. Since $f_5(y) \geq 0$, $U(t_H, q_F|\lambda = h) > U(t_H, \underline{q}|\lambda = 1)$ if and only if $f_4(y) > 0$ and $f_6(y) > 0$.

We now show the conditions for $f_4(y) > 0$ and $f_6(y) > 0$, respectively. First, denote $f_4(y) = n_4(b)y + 2b(1-g)$, where $n_4(b) = (g^2+2)b^2 - 4b + 1$. Note the two roots of $n_4(b)$ satisfy $0 < b_{n_4}^- < \frac{1}{2} < 1 < b_{n_4}^+$. Thus, $n_4(b) \geq 0$ if and only if $b \leq b_{n_4}^-$. Thus, if $b \leq b_{n_4}^-$, $f_4(y) > 0$; if $b > b_{n_4}^-$, $f_4(y) > 0$ if and only if $y < \bar{y}$ where $\bar{y} = \frac{2b(1-g)}{-(g^2+2)b^2+4b-1}$. Second, denote $f_6(y) = Ay^2 - By + C$, where $A = 1 + 2b^2g^2 - 8b^3g^2 + 4b^4g^2 + b^4g^4$, $B = 4b(1+g-4bg+2b^2g-b^2g^2+b^2g^3)$, $C = 4b^2(1-g)^2$. Note that $A = b^4g^4 + (4b^4 - 8b^3 + 2b^2)g^2 + 1 = n_5(g^2)$ and the discriminator of $n_5(g^2)$ is $\Delta_{n_5} = 16b^5(b-1)^2(b-2) < 0$. Thus, $A > 0$. Note that $B = 4bn_6(b)$, where $n_6(b) = g(g^2 - g + 2)b^2 - 4gb + g + 1$. Note that the discriminator of $n_6(b)$ is $\Delta_{n_6} = -4g(g+2)(g-1)^2 < 0$. Thus, $B > 0$. Note that $C > 0$. Hence, the two roots of $f_6(y)$ satisfy $0 < y_1 < y_2$ and

$$y_1 = \frac{2b(-1+\sqrt{g})^2}{1 - 2b\sqrt{g} + b^2g(2 - 2\sqrt{g} + g)}, \quad y_2 = \frac{2b(1+\sqrt{g})^2}{1 + 2b\sqrt{g} + b^2g(2 + 2\sqrt{g} + g)},$$

where $1 - 2b\sqrt{g} + b^2g(2 - 2\sqrt{g} + g) > 0$ and $1 + 2b\sqrt{g} + b^2g(2 + 2\sqrt{g} + g) > 0$. Thus, $f_6(y) > 0$ if and only if $y < y_1$ or $y > y_2$. Moreover, $y_1 - 1 = \frac{(1-bg)\left(-1+b(2-2\sqrt{g}+g)\right)}{1-2b\sqrt{g}+b^2g(2-2\sqrt{g}+g)}$ and $y_2 - 1 = \frac{(1-bg)(-1+b(2+2\sqrt{g}+g))}{1+2b\sqrt{g}+b^2g(2+2\sqrt{g}+g)}$. Thus, $y_1 < 1$ if and only if $b < \frac{1}{2-2\sqrt{g}+g} \in (\frac{1}{2}, 1)$, and $y_2 < 1$ if and only if $b < \frac{1}{2+2\sqrt{g}+g} \in (\frac{1}{5}, \frac{1}{2})$.

We now show that if $b < \frac{1}{2}$, Equation (3.4) holds if and only if $y < y_1$ by proving that $y_1 < 2b$, $y_2 > 2b$, and $y_1 < \bar{y}$ if $b < \frac{1}{2}$. First, $\frac{y_1}{2b} - 1 = \frac{n_7(b)}{1-2b\sqrt{g}+b^2g(2-2\sqrt{g}+g)}$, where $n_7(b) = -g\left((\sqrt{g}-1)^2 + 1\right)b^2 + 2\sqrt{g}b - 2\sqrt{g} + g$. Note that the two roots of $n_7(b)$ satisfy $1 < b_{n_7}^- < b_{n_7}^+$. Thus, $n_7(b) < 0$, i.e., $y_1 < 2b$, for any $b \in (0, 1]$. Thus, $y_1 < 1$ is implied if $b < \frac{1}{2}$. Second, $\frac{y_2}{2b} - 1 = \frac{n_8(b)}{1+2b\sqrt{g}+b^2g(2+2\sqrt{g}+g)}$, where $n_8(b) = -(2g + 2g^{3/2} + g^2)b^2 - 2\sqrt{g}b + 2\sqrt{g} + g$. Note that the two roots of $n_8(b)$

satisfy that $b_{n_8}^- < 0 < \frac{1}{2} < b_{n_8}^+$. Thus, $n_8(b) > 0$, i.e., $y_2 > 2b$ if $b < \frac{1}{2}$. Third, $\bar{y} - y_1 = \frac{4b(1-\sqrt{g})n_9(b)}{(1-2b\sqrt{g}+b^2g(2-2\sqrt{g}+g))(-1+4b-b^2(2+g^2))}$, where $n_9(b) = (1-\sqrt{g}+g)b^2 + (-2+\sqrt{g}-g)b + 1$. Note that the axis of symmetry of $n_9(b)$ satisfies $\frac{2-\sqrt{g}+g}{2(1-\sqrt{g}+g)} > \frac{1}{2}$ and $n_9(\frac{1}{2}) = \frac{1}{4}(1+\sqrt{g}-g) > 0$. Thus, $n_9(b) > 0$, i.e., $y_1 < \bar{y}$, if $b < \frac{1}{2}$.

We now show that if $b \geq \frac{1}{2}$, Equation (3.4) holds if and only if (i) $y < y_1$ when $b < \frac{1}{2-2\sqrt{g}+g}$, and (ii) any $0 < y < 1$ when $b \geq \frac{1}{2-2\sqrt{g}+g}$. Since $b \geq \frac{1}{2}$, $f_4(y) > 0$ if and only if $y < \bar{y}$. Note that $y_2 > 1$. Note that $y_1 < 1$ if and only if $b < \frac{1}{2-2\sqrt{g}+g}$. First, we show that if $\frac{1}{2} \leq b < \frac{1}{2-2\sqrt{g}+g}$, then $y_1 < \bar{y}$. Recall that $y_1 < \bar{y}$ if and only if $n_9(b) = (1-\sqrt{g}+g)b^2 + (-2+\sqrt{g}-g)b + 1 > 0$. Note that the axis of symmetry of $n_9(b)$ satisfies $\frac{1}{2} < \frac{1}{2-2\sqrt{g}+g} < \frac{2-\sqrt{g}+g}{2(1-\sqrt{g}+g)}$. Note that $n_9(\frac{1}{2-2\sqrt{g}+g}) = \frac{(1-\sqrt{g})^3}{(2-2\sqrt{g}+g)^2} > 0$. Thus, $n_9(b) > 0$, i.e., $y_1 < \bar{y}$, for any $\frac{1}{2} \leq b < \frac{1}{2-2\sqrt{g}+g}$. Second, we show that if $\frac{1}{2-2\sqrt{g}+g} \leq b \leq 1$, then $\bar{y} > 1$. Note that $\bar{y} - 1 = \frac{n_{10}(b)}{-(g^2+2)b^2+4b-1}$, where $n_{10}(b) = (g^2+2)b^2 - 2(1+g)b + 1$. Note that the discriminator of $n_{10}(b)$ is $\Delta_{n_{10}} = 4(2g-1)$. Thus, if $g \leq \frac{1}{2}$, then $n_{10}(b) > 0$. If $g > \frac{1}{2}$, note that the axis of symmetry of $n_10(b)$ satisfies $\frac{1+g}{g^2+2} < \frac{1}{2-2\sqrt{g}+g}$ and $n_{10}(\frac{1}{2-2\sqrt{g}+g}) = \frac{2(1-\sqrt{g})^2}{(2-2\sqrt{g}+g)^2} > 0$. Thus, $n_{10}(b) > 0$, i.e., $\bar{y} > 1$, for any $\frac{1}{2-2\sqrt{g}+g} \leq b \leq 1$.

Combing the scenarios when $b < \frac{1}{2}$ and when $b \geq \frac{1}{2}$ above, we obtain that Equation (3.4) holds if and only if $y < y_1$ when $b < \frac{1}{2-2\sqrt{g}+g}$ and any $0 < y < 1$ when $b \geq \frac{1}{2-2\sqrt{g}+g}$, i.e.,

$$\theta_{HP} = \begin{cases} \frac{1-2\sqrt{g}b+g(2-2\sqrt{g}+g)b^2}{1-2(1-\sqrt{g}+g)b+g(2-2\sqrt{g}+g)b^2} & \text{if } \alpha < \alpha_0, \\ \infty & \text{if } \alpha \geq \alpha_0, \end{cases} \quad \text{and } \alpha_0 = \frac{3-2\sqrt{g}+g}{2(2-2\sqrt{g}+g)}. \quad (3.A.10)$$

Claim 5 *If Equation* (3.4) *holds, then Equation* (3.3) *holds.*

First, we show that if Equation (3.4) holds, then $q_F < q_H^S$. Suppose $q_F \geq q_H^S$, then $U_F(t_H, q_F|\lambda = 1) \leq U_F(t_H, q_H^S|\lambda = 1)$. By Lemma 3, $U_F(t_H, q_F|\lambda = h) < U_F(t_H, q_F|\lambda = 1)$. Thus, we obtain $U_F(t_H, q_F|\lambda = h) < U_F(t_H, q_H^S|\lambda = 1)$, contradiction. Thus, $q_F < q_H^S$ if Equation (3.4) holds. Second, we show that $U_F(t_L, q_F|\lambda = h) > U_F(t_L, q_H^S|\lambda = 1)$ if Equation (3.4) holds. By Equation (3.7)

$$\begin{aligned} & U_F(t_L, q_F|\lambda = h) - U_F(t_L, q_H^S|\lambda = 1) \\ &= \tfrac{(1-2\alpha)(-h)(A_H-A_L)+A_L-q_F}{2} q_F - \tfrac{(1-2\alpha)(-1)(A_H-A_L)+A_L-q_H^S}{2} q_H^S \\ &= \tfrac{(1-2\alpha)(1-h)(A_H-A_L)+A_H-q_F}{2} q_F - \tfrac{A_H-q_H^S}{2} q_H^S - (1-\alpha)(A_H-A_L)(q_F-q_H^S) \\ &= U_F(t_H, q_F|\lambda = h) - U_F(t_H, q_H^S|\lambda = 1) - (1-\alpha)(A_H-A_L)(q_F-q_H^S) \end{aligned}$$

Since $q_F < q_H^S$, $U_F(t_L, q_F|\lambda = h) > U_F(t_L, q_H^S|\lambda = 1)$ holds. Note that $q_H^S = \underline{q}$, i.e., $U_F(t_L, q_H^S|\lambda = 1) = U_F(t_L, q_L^*|\lambda = 0)$, if Equation (3.4) holds. Thus, if Equation (3.4) holds, then $U_F(t_L, q_F|\lambda = h) > U_F(t_L, q_L^*|\lambda = 0) = U_F(t_L, q_L^S|\lambda = 0)$, i.e., Equation (3.3) holds.

Claim 6 *The pooling LMSE q^P is unique, $q^P = q_H^P$, and $q_L^* < q_H^P < q_H^*$.*

If $\theta < \theta_{HP}$, any q_F such that $U_F(t_H, q_F|\lambda = h) > U_F(t_H, q_H^S|\lambda = 1)$ is a Pareto dominant pooling PBE. By definition, $q^P = q_H^P = \arg\max U_F(t_H, q_F|\lambda = h)$. Note that by Equation (3.7), $U_F(t_H, q_F|\lambda = h)$ is concave and hence q_H^P is unique and

$$q_H^P = \frac{(1 - 2\alpha)(1 - h)(A_H - A_L) + A_H}{2}. \tag{3.A.11}$$

Note that $q_H^P - q_L^* = \frac{(1-(2\alpha-1)(1-h))(A_H-A_L)}{2} > 0$ and $q_H^P - q_H^* = -\frac{(2\alpha-1)(1-h)(A_H-A_L)}{2} < 0$. Thus, $q_L^* < q_H^P < q_H^*$. By Equation (3.4) and Equation (3.6), we further obtain that $q_E^*(q^P) = \frac{\mu - q^P}{2}$ and $v^*(q^P) = \frac{\mu - q^P}{2} q^P$. Given the existence of the pooling LMSE, the posterior belief λ in equilibrium satisfies $\lambda = 1$ if $q_F \geq \underline{q}$, $\lambda = h$ if $q_F = q_H^P$, and $\lambda = 0$ if $q_F < \underline{q}$ and $q_F \neq q_H^P$. ■

3.9.1.6 Proof of Proposition 3

We conduct the comparative statics of the firm's equilibrium capacity investment relative to α.

(i) If $\alpha \leq \underline{\alpha_P}$, both firm types invest $q^P = q_L^P = \frac{(2\alpha-1)hA_H + (1-(2\alpha-1)h)A_L}{2}$. Since $\frac{\partial q_L^P}{\partial \alpha} = h(A_H - A_L) > 0$, q^P decreases as α decreases.

(ii) If $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, the H-type firm invests $q_H^S = q_H^*$, and the L-type firm invests $q_L^S = \bar{q}$ as shown in Equation (3.A.1). We obtain that $\frac{\partial \bar{q}}{\partial a} = A_L \frac{x(\sqrt{(ax+2x+2)ax} - x - ax - 1)}{2\sqrt{(ax+2x+2)ax}}$. Since $(\sqrt{(ax+2x+2)ax})^2 - (x + ax + 1)^2 = -(x+1)^2 < 0$, $\frac{\partial(\bar{q})}{\partial a} < 0$. Thus, q_L^S decreases as α decreases.

(iii) If $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, the L-type firm invests $q_L^S = q_L^*$, and the H-type firm invests $q_H^S = \underline{q}$ as shown in Equation (3.A.4). We obtain that $\frac{\partial \underline{q}}{\partial b} = A_H \frac{y\left(\sqrt{(by-2y+2)by} + (b-1)y + 1\right)}{2\sqrt{(by-2y+2)by}}$. Since $(b-2)y + 2 > 0$ and $(b-1)y + 1 > 0$, $\frac{\partial \underline{q}}{\partial b} > 0$. Thus, q_H^S increases as α increases.

(iv) If $\alpha \geq \overline{\alpha_P}$, both firm types invest $q^P = q_H^P = \frac{(1-2\alpha)(1-h)(A_H-A_L)+A_H}{2}$. Since $\frac{\partial q_H^P}{\partial \alpha} = -(1-h)(A_H - A_L) < 0$, q^P decreases as α increases. ■

The intuition for thesee results is outlined in what follows. When $\alpha \leq \underline{\alpha_P}$, both firm types under-invest in the pooling LMSE. As α decreases, q^P also decreases, representing an increase in the magnitude of under-investment by both firm types. The intuition for this is that as α decreases, both firm types are increasingly interested in their long-term profit, and wish to signal to the competitor that market demand is low in order to dissuade an aggressive entry. The H-type is so motivated to mimic the L-type that it becomes too costly for the L-type to separate and the firm types settle on pooling.

When $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, the H-type firm invests optimally and the L-type firm under-invests in the separating PBE. As α decreases, q_L^S also decreases, representing an increase in the magnitude of under-investment by the L-type firm. The intuition is that a lower α raises the firm's interest in her long-term performance and the impact of competition on it. In her desire to quell competition as α decreases, the H-type firm has a higher incentive to mimic the L-type firm. Consequently, the L-type firm must distort her capacity investment further below q_L^* in order to separate from the H-type firm.

When $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, the L-type firm invests optimally and the H-type firm over-invests in the separating PBE. As α increases, q_H^S also increases, representing an increase in the magnitude of over-investment by the H-type firm. The rationale is that a higher α raises the firm's interest in her short-term valuation. This increases the L-type firm's desire to mimic the H-type firm. As a result, the H-type firm must distort her capacity investment further above q_H^* in order to separate from the L-type firm.

When $\alpha \geq \overline{\alpha_P}$, the L-type firm over-invests and the H-type firm under-invests in the pooling LMSE. As α increases, q^P decreases, representing a decrease in the magnitude of over-investment by the L-type firm and an increase in the magnitude of under-investment by the H-type firm. The intuition for this outcome is that an increasing α encourages both firms to focus on their short-term share price by signaling high demand to the investor. The L-type is sufficiently motivated to mimic the H-type that it becomes too costly for the H-type to separate, and the firm types settle on pooling.

3.9.1.7 Proof of Proposition 4

For convenience, we denote the equilibrium profit of the L-type firm, the H-type firm, and the expected equilibrium profit as π_L, π_H, and π_W, respectively in the remainder of this proof. Note that in a pooling PBE, $\pi_L = \frac{2A_L-\mu-q^P}{2}q^P$, $\pi_H = \frac{2A_H-\mu-q^P}{2}q^P$, and $\pi_W = h\frac{2A_L-\mu-q^P}{2}q^P + (1-h)\frac{2A_H-\mu-q^P}{2}q^P = \frac{\mu-q^P}{2}q^P$. Thus, the optimal q^P for π_L, π_H, and π_W is $\frac{2A_L-\mu}{2}$, $\frac{2A_H-\mu}{2}$, and $\frac{\mu}{2}$, respectively. In the following, we consider three cases: (i) $\theta < \max\theta_{LP}$, (ii) $\max\theta_{LP} < \theta < \max\theta_{LS}$, (iii) $\theta > \max\theta_{LS}$.

Case 1. $\theta < \max\theta_{LP}$. In this case, $0 < \underline{\alpha_P} < \underline{\alpha_S} < \frac{1}{2} < \overline{\alpha_S} < \overline{\alpha_P} < 1$.

First, we analyze the pooling region (1) $\alpha \le \underline{\alpha_P}$ and (2) $\alpha \ge \overline{\alpha_P}$. (1) If $\alpha \le \underline{\alpha_P}$, both firm types invest $q^P = q^P_L = \frac{(2\alpha-1)hA_H+((-2\alpha+1)h+1)A_L}{2}$. Note that $q^P_L \ge \frac{2A_L-\mu}{2}$, $q^P_L < \frac{2A_H-\mu}{2}$, and $q^P_L < \frac{\mu}{2}$ for any $\alpha \le \underline{\alpha_P}$. Since q^P_L increases as α increases, q^P_L will deviate more from $\frac{2A_L-\mu}{2}$, less from $\frac{2A_H-\mu}{2}$, and less from $\frac{\mu}{2}$ as α increases. Thus, π_L decreases, π_H increases, and π_W increases as α increases for $\alpha \le \underline{\alpha_P}$. Thus, $\pi_L < \pi_L(\alpha = 0)$, $\pi_H > \pi_H(\alpha = 0)$, $\pi_W > \pi_W(\alpha = 0)$ for any $0 < \alpha < \underline{\alpha_P}$. (2) If $\alpha \ge \overline{\alpha_P}$, both firm types invest $q^P = q^P_H = \frac{(1-2\alpha)(1-h)(A_H-A_L)+A_H}{2}$. Note that $q^P_H > \frac{2A_L-\mu}{2}$, $q^P_H < \frac{2A_H-\mu}{2}$, and $q^P_H \ge \frac{\mu}{2}$ for any $\alpha \ge \overline{\alpha_P}$. Since q^P_H decreases as α increases, q^P_H will deviate less from $\frac{2A_L-\mu}{2}$, more from $\frac{2A_H-\mu}{2}$, and less from $\frac{\mu}{2}$. Thus, π_L increases, π_H decreases, and π_W increases as α increases for $\alpha \ge \overline{\alpha_P}$. Now we compare the three profits when $\alpha \ge \overline{\alpha_P}$ with them when $\alpha = 0$. First, note that $q^P_H(\alpha = 1) > q^P_L(\alpha = 0) = \frac{2A_L-\mu}{2}$. Thus, $\pi_L < \pi_L(\alpha = 0)$ for $\alpha \ge \overline{\alpha_P}$. Second, note that $q^P_L(\alpha = 0) < q^P_H(\alpha = 1) < \frac{2A_H-\mu}{2}$. Thus, $\pi_H > \pi_H(\alpha = 0)$ for $\alpha \ge \overline{\alpha_P}$. Third, note that $q^P_L(\alpha = 0) < q^P_H(\alpha = 1) = \frac{\mu}{2}$. Thus, there may exist a unique $\tilde{\alpha} \in [\overline{\alpha_P}, 1]$ such that $\pi_W < \pi_W(\alpha = 0)$ for any $\alpha \in [\overline{\alpha_P}, \tilde{\alpha})$, and $\pi_W > \pi_W(\alpha = 0)$ for any $\alpha \in (\tilde{\alpha}, 1]$. Thus, in the pooling region ($\alpha \le \underline{\alpha_P}$ or $\alpha \ge \overline{\alpha_P}$), $\alpha = 0$ is optimal for π_L. Note that $q^P_L(\alpha = \underline{\alpha_P}) < q^P_H(\alpha = \overline{\alpha_P}) < \frac{2A_H-\mu}{2}$. Thus, $\alpha = \overline{\alpha_P}$ is optimal for π_H. Since $q^P_H(\alpha = 1) = \frac{\mu}{2}$, $\alpha = 1$ is optimal for π_W where $\pi_W = \frac{\mu^2}{8}$.

Next, we analyze the separating region $\underline{\alpha_P} < \alpha < \overline{\alpha_P}$. Note that the firm's profit is equal to her utility in the separating PBE according to Equation (3.5). First, we analyze π_L. By definition, $U_F(t_L, \bar{q}|\lambda = 0, \alpha = \underline{\alpha_P}+\epsilon) = U_F(t_L, q^P_L|\lambda = h, \alpha = \underline{\alpha_P})$. Note that $U_F(t_L, q^P_L|\lambda = h) = \frac{(q^P_L)^2}{4}$ increases in α and hence $U_F(t_L, q^P_L|\lambda = h, \alpha = \underline{\alpha_P}) > U_F(t_L, q^P_L|\lambda = h, \alpha = 0)$. Note that $\pi_L(\alpha = \underline{\alpha_P} + \epsilon) = U_F(t_L, \bar{q}|\lambda = 0, \alpha = \underline{\alpha_P} + \epsilon)$ and $\pi_L(\alpha = 0) = U_F(t_L, q^P_L|\lambda = h, \alpha = 0)$. Thus, $\pi_L(\alpha = \underline{\alpha_P} + \epsilon) > \pi_L(\alpha = 0)$. Since $\min\pi_L = \pi_L(\alpha = \underline{\alpha_P} + \epsilon)$, we obtain that $\pi_L > \pi_L(\alpha = 0)$ for any $\alpha \in (\underline{\alpha_P}, \overline{\alpha_P})$. The optimal level of α for π_L is when $q^S_L = q^*_L$, i.e., any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_P})$. Second, we analyze π_H. Note that $\pi_H(\alpha = 0) - \frac{A_H^2}{8} = \frac{1}{8A_L^2}(\theta - 1)((h^2 - 4h - 1)\theta + (-h^2 + 2h + 3))$ and hence $\pi^*_H(\alpha = 0) > \frac{A_H^2}{8}$ if $\theta < \frac{-h^2+2h+3}{-h^2+4h+1}$. Note that $\frac{-h^2+2h+3}{-h^2+4h+1} - \theta_{LP}(\alpha = 0) = \frac{8(\sqrt{h}(\sqrt{h}-1)+h\sqrt{h}(\sqrt{h}-1))}{(h^2-4h-1)(1+2\sqrt{h}+2h-2h^{3/2}+h^2)} > 0$. Since $\theta < \theta_{LP}(\alpha = 0)$, we obtain that $\theta < \frac{-h^2+2h+3}{-h^2+4h+1}$, i.e., $\pi_H(\alpha = 0) > \frac{A_H^2}{8}$. Since $\max\pi_H = \frac{A_H^2}{8}$, $\pi_H < \pi_H(\alpha = 0)$ holds for any $\alpha \in (\underline{\alpha_P}, \overline{\alpha_P})$. The optimal level of α for π_H is when $q^S_H = q^*_H$, i.e., any $\alpha \in (\underline{\alpha_P}, \overline{\alpha_S}]$. Third, we analyze π_W. Note that $\pi_W = h\frac{A_H-q^S_H}{2}q^S_H + (1-h)\frac{A_L-q^S_L}{2}q^S_L$. Thus, π_W increases in α for $\alpha \in (\underline{\alpha_P}, \underline{\alpha_S})$, decreases in α for $\alpha \in (\overline{\alpha_S}, \overline{\alpha_P})$, and the optimal level of α for π_W is when $q^S_H = q^*_H$ and $q^S_L = q^*_L$, i.e., any $\alpha \in [\overline{\alpha_S}, \overline{\alpha_S}]$.

Combining the results in pooling region and the separating region above, We obtain that α_L^* is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_P})$ and $\alpha_H^* = \overline{\alpha_P}$. Since $\frac{\mu^2}{8} < h\frac{A_H^2}{8} + (1-h)\frac{A_L^2}{8}$, we obtain that α_W^* is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_S}]$.

Case 2. $\max\theta_{LP} \leq \theta < \max\theta_{LS}$. In this case, $0 = \underline{\alpha_P} < \underline{\alpha_S} < \frac{1}{2} < \overline{\alpha_S} < \overline{\alpha_P} < 1$.

If $\alpha \in [0, \overline{\alpha_P})$, separating PBE arises. The optimal α for π_L is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_P})$ and $\max\pi_L = \frac{A_L^2}{8}$. The optimal α for π_H is any $\alpha \in [0, \overline{\alpha_S}]$ and $\max\pi_H = \frac{A_H^2}{8}$. The optimal α for π_W is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_S}]$ and $\max\pi_W = h\frac{A_H^2}{8} + (1-h)\frac{A_L^2}{8}$.

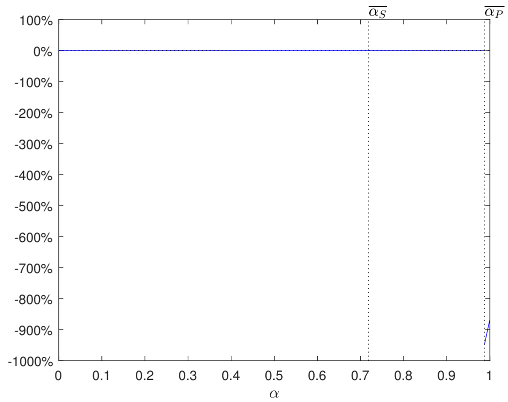
If $\alpha \in [\overline{\alpha_P}, 1]$, pooling PBE arises. By *Case* 1, π_L increases, π_H decreases, and π_W increases in α for $\alpha \in [\overline{\alpha_P}, 1]$. First, $\pi_L(\alpha = 1) - \frac{A_L^2}{8} = \frac{(4A_L - 3\mu)\mu}{8} - \frac{A_L^2}{8} = -\frac{hx(2+3hx)}{8A_L^2} < 0$. Thus, we obtain that $\alpha_L^* \in [\underline{\alpha_S}, \overline{\alpha_P})$. Second, $\pi_H(\alpha = 1) - \frac{A_H^2}{8} = \frac{(4A_H - 3\mu)\mu}{8} - \frac{A_H^2}{8} = \frac{1}{8A_L^2}(1-h)(\theta-1)((-1+3h)\theta + 3(1-h))$. Thus, $\pi_H(\alpha = 1) > \frac{A_H^2}{8}$ if and only if $(1-3h)\theta < 3(1-h)$, i.e., $h \geq \frac{1}{3}$, or $h < \frac{1}{3}$ and $\theta < \frac{3-3h}{1-3h}$. Since $\theta < \max\theta_{LS} = 3 < \frac{3-3h}{1-3h}$, $\pi_H(\alpha = 1) > \frac{A_H^2}{8}$ holds. Thus, $\pi_H > \frac{A_H^2}{8}$ for any $\alpha \in [\overline{\alpha_P}, 1]$, and $\alpha_H^* = \overline{\alpha_P}$. Third, since $\frac{\mu^2}{8} < h\frac{A_H^2}{8} + (1-h)\frac{A_L^2}{8}$, we obtain that α_W^* is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_S}]$.

Case 3. $\theta \geq \max\theta_{LS}$. In this case, $0 = \underline{\alpha_P} = \underline{\alpha_S} < \frac{1}{2} < \overline{\alpha_S} < \overline{\alpha_P} < 1$.

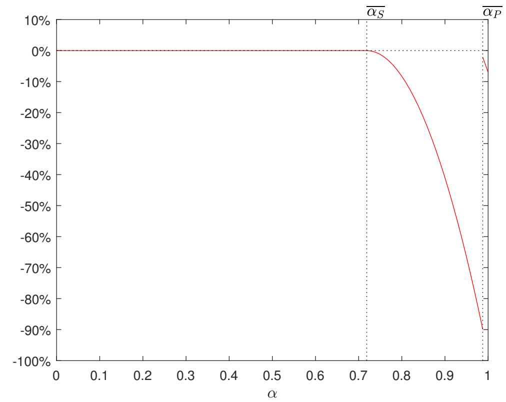
If $\alpha \in [0, \underline{\alpha_P})$, separating PBE arises. The optimal α for π_L is any $\alpha \in [0, \overline{\alpha_P})$ and $\max\pi_L = \frac{A_L^2}{8}$. The optimal α for π_H is any $\alpha \in [0, \overline{\alpha_S}]$ and $\max\pi_H = \frac{A_H^2}{8}$. The optimal α for π_W is any $\alpha \in [0, \overline{\alpha_S}]$ and $\max\pi_W = h\frac{A_H^2}{8} + (1-h)\frac{A_L^2}{8}$.

If $\alpha \in [\overline{\alpha_P}, 1]$, pooling PBE arises. First, we obtain that $\alpha_L^* \in [0, \overline{\alpha_P}]$ following the same argument in *Case* 2. Second, $\pi_H(\alpha = \overline{\alpha_P}) - \frac{A_H^2}{8} = -\frac{1}{8A_L^2}gx(-2 + (g\hat{b}^2 + 2g\hat{b} - 2)x)$, where $\hat{b} = 2\overline{\alpha_P} - 1$. Hence, $\pi_H(\alpha = \overline{\alpha_P}) < \frac{A_H^2}{8}$ if and only if $g\hat{b}^2 + 2g\hat{b} - 2 > \frac{2}{x}$. Note that $g\hat{b}^2 + 2g\hat{b} - 2$ increases in $\overline{\alpha_P}$. Note that θ_{HP} increases in α and hence $\overline{\alpha_P} = \theta_{HP}^{-1}(\theta)$ increases in θ, i.e., x. Thus, $g\hat{b}^2 + 2g\hat{b} - 2$ increases in x. Note that $\frac{x}{2}$ decreases in x. Thus, there exists a unique x such that $g\hat{b}^2 + 2g\hat{b} - 2 = \frac{2}{x}$. That is, there exists a unique $\hat{\theta} \geq \max\theta_{LS}$ such that $\pi_H(\alpha = \overline{\alpha_P}) > \frac{A_H^2}{8}$ if $\theta < \hat{\theta}$ and $\pi_H(\alpha = \overline{\alpha_P}) < \frac{A_H^2}{8}$ if $\theta > \hat{\theta}$. Third, since $\frac{\mu^2}{8} < h\frac{A_H^2}{8} + (1-h)\frac{A_L^2}{8}$, we obtain that α_W^* is any $\alpha \in [0, \overline{\alpha_S}]$.

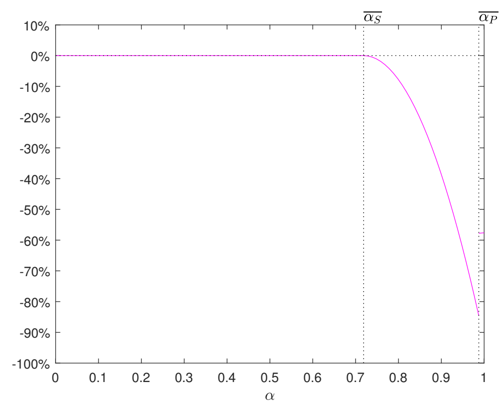
Combining *Case* 1 to 3, we obtain that α_L^* is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_P})$; $\alpha_H^* = \overline{\alpha_P}$ if $\theta < \hat{\theta}$ and α_H^* is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_S}]$ if $\theta > \hat{\theta}$, where $\hat{\theta} > \max\theta_{LS}$ and solves the equation $g(2\theta_{HP}^{-1}(\hat{\theta}) - 1)^2 + 2g(2\theta_{HP}^{-1}(\hat{\theta}) - 1) - 2 = \frac{2}{\hat{\theta}-1}$; α_W^* is any $\alpha \in [\underline{\alpha_S}, \overline{\alpha_S}]$. Figure 3.A.1 shows an example if $\theta > \hat{\theta}$ when $h = 0.2$ and $\theta = 8$.



(a) L-type



(b) H-type



(c) Expected type

Figure 3.A.1: Impact of α on Firms' Long-term Profit with Competition, $h = 0.2$, $\theta = 8$

In the example shown in Figure 3.A.1, $\underline{\alpha_S} = 0$. The long-term profit is greatest for the L-type firm when $0 \leq \alpha < \overline{\alpha_P}$, greatest for the H-type firm and the firm when her type is unknown when $0 \leq \alpha \leq \overline{\alpha_S}$. In this example, $\alpha = 0$ is one optimal choice for both firm types, which coincides with the practitioners' initiatives to eliminate short-termism. However, since the optimal range of α is wide, it is unnecessary to diminish short-termism to zero. ■

3.9.2 Robustness Checks

3.9.2.1 Analysis Using Loss Averse, Risk Averse, or Risk Seeking Investor

We show that our equilibrium outcomes and insights will remain the same if the investor is loss averse, risk averse, or risk seeking. We now assume the investor's utility function is

$$U_I(t, v) = -(\mathbb{1}_{[v \geq \pi(t)]}a + \mathbb{1}_{[v \leq \pi(t)]}b)(v - \pi(t))^2, \tag{3.A.12}$$

where $\mathbb{1}_{[\,]}$ is an indicator function. When $a > b$, the investor is loss averse or risk averse such that he incurs a higher disutility when overpricing compared to underpricing the firm's value. When $a < b$, the investor is risk seeking such that he incurs a higher disutility when underpricing compared to overpricing the firm's value. When $a = b$, the investor is risk neutral and Equation (3.A.12) is equivalent to Equation (3.2).

The investor chooses his best response v^*, given the firm and entrant's capacity investment q_F and q_E, to maximize his expected utility conditioning on his posterior belief of the market demand. First, we analyze the case when the equilibrium outcome is a separating PBE. The investor's expected utility is given by

$$EU_I(t, v) = \begin{cases} -a(v - \pi(t))^2 & \text{if } v \geq \pi(t), \\ -b(\pi(t) - v)^2 & \text{if } v \leq \pi(t). \end{cases}$$

Thus, the maximizer is $v^* = \pi(t)$. Second, we analyze the case when the equilibrium outcome is a pooling PBE. For ease of exposition, let $\pi_H = \pi(t_H)$ and $\pi_L = \pi(t_L)$. The investor's expected utility is given by $EU_I(t, v) = -h(\mathbb{1}_{[v \geq \pi_H]}a + \mathbb{1}_{[v \leq \pi_H]}b)(v - \pi_H)^2 - (1 - h)(\mathbb{1}_{[v \geq \pi_L]}a + \mathbb{1}_{[v \leq \pi_L]}b)(v - \pi_L)^2$, i.e.,

$$EU_I(t, v) = \begin{cases} -hb(v - \pi_H)^2 - (1 - h)b(v - \pi_L)^2, & \text{if } v \leq \pi_L, \\ -hb(v - \pi_H)^2 - (1 - h)a(v - \pi_L)^2, & \text{if } \pi_L \leq v \leq \pi_H, \\ -ha(v - \pi_H)^2 - (1 - h)a(v - \pi_L)^2, & \text{if } v \geq \pi_H. \end{cases}$$

Denote $\tilde{h} = \frac{hb}{hb+(1-h)a}$ and $\tilde{h} \in (0, 1)$. We discuss three cases below. (1) If $v \leq \pi_L$, $\frac{\partial EU_I(t,v)}{\partial v} = -2bv + 2b(h\pi_H + (1 - h)\pi_L)$. Since $h\pi_H + (1 - h)\pi_L > \pi_L$ and $v \leq \pi_L$, the maximizer is $v^* = \pi_L$, and $EU_I(t, v^*) = u_1 = -\tilde{h}(\pi_H - \pi_L)^2(hb + (1 - h)a)$. (2) If $\pi_L \leq v \leq \pi_H$, $\frac{\partial EU_I(t,v)}{\partial v} = -2(hb+(1-h)a)v+2(hb\pi_H+(1-h)a\pi_L)$. Thus, the maximizer is $v^* = \frac{hb}{hb+(1-h)a}\pi_H + \frac{(1-h)a}{hb+(1-h)a}\pi_L = \tilde{h}\pi_H + (1 - \tilde{h})\pi_L$, and $EU_I(t, v^*) = u_2 = -(\tilde{h}(1 - \tilde{h})^2 +$

$(1-\tilde{h})\tilde{h}^2)(\pi_H - \pi_L)^2(hb + (1-h)a)$. (3) If $v \geq \pi_H$, $\frac{\partial EU_I(t,v)}{\partial v} = -2av + 2a(h\pi_H + (1-h)\pi_L)$. Since $h\pi_H + (1-h)\pi_L < \pi_H$ and $v \geq \pi_H$, the maximizer is $v^* = \pi_H$, and $EU_I(t, v^*) = u_3 = -(1-\tilde{h})(\pi_H - \pi_L)^2(hb+(1-h)a)$. Since $u_2 - u_1 = \tilde{h}^2(\pi_H - \pi_L)^2(hb+(1-h)a) > 0$ and $u_2 - u_3 = (1-\tilde{h})^2(\pi_H - \pi_L)^2(hb+(1-h)a) > 0$, we obtain that the ultimate maximizer under pooling PBE is $v^* = \frac{hb}{hb+(1-h)a}\pi_H + \frac{(1-h)a}{hb+(1-h)a}\pi_L = \tilde{h}\pi_H + (1-\tilde{h})\pi_L$, where $\tilde{h} \in (0, 1)$ is the effective prior belief under loss aversion or risk aversion. Note that $\tilde{h}$ decreases in a and increases in b. This indicates that when the investor is more loss averse or risk averse (i.e., a increases or b decreases), the investor will lower its valuation of the firm in pooling (i.e., $\tilde{h}$ decreases). In contrast, when the investor is more risk seeking (i.e., a decreases and b increases), the investor will increase its valuation of the firm in pooling (i.e., $\tilde{h}$ increases). We also obtain that $\tilde{h} = h$ if $a = b$.

In summary, a loss-averse, risk averse or risk-seeking investor's best response v^* shares the same functional form compared to a risk-neutral investor despite the shift in the effective prior belief. Thus, the equilibrium outcomes and insights derived under our model with a risk neutral investor will carry through in a model with a loss-averse, risk-averse, or risk-seeking investor. ■

3.9.2.2 Analysis Using an Entrant with Short-termism

We show that the entrant's best response is the same as in Equation (3.6) when considering short-termism by the entrant. Suppose that the short-termism level of the entrant is $\beta \in [0, 1]$. The investor's valuation for the entrant is denoted as v_E. Thus, the entrant's utility is $U_E(t, q_F, q_E) = \beta v_E + (1-\beta)\pi_E(t, q_F, q_E)$.

The investor determines v_E to minimize the magnitude of error in his valuation of the entrant, i.e., $v_E^*(t, q_F, q_E) = \arg\min(v_E - \pi_E(t, q_F, q_E))^2$. Since the entrant has no private information than the investor about the firm's type, we consider two cases: (i) the firm's type is revealed to both the entrant and the investor, (ii) the firm's type is concealed to both the entrant and the investor.

Case (i). The investor will be error free in his valuation, i.e., $v_E^*(t, q_F, q_E) = \pi_E(t, q_F, q_E)$. Thus, $U_E(t, q_F, q_E) = \pi_E(t, q_F, q_E)$. In this case, the entrant chooses q_E to maximize his utility $U_E(t, q_F, q_E)$ and hence $q_E^*(q_F|\lambda) = \frac{\lambda A_H + (1-\lambda)A_L - q_F}{2}$, where $\lambda = 1$ for $t = H$ and $\lambda = 0$ for $t = L$.

Case (ii). The investor's valuation is $v_E^*(t, q_F, q_E) = h\pi_E(H, q_F, q_E) + (1-h)\pi_E(L, q_F, q_E)$ since the firm's type is concealed. In this case, the entrant chooses q_E to maximize his expected utility based on the prior belief, i.e., $hU_E(H, q_F, q_E) + (1-h)U_E(L, q_F, q_E) = h\beta(h\pi_E(H, q_F, q_E) + (1-h)\pi_E(L, q_F, q_E)) + h(1-\beta)\pi_E(H, q_F, q_E) + (1-h)\beta(h\pi_E(H, q_F, q_E) + (1-h)\pi_E(L, q_F, q_E)) + (1-h)(1-\beta)\pi_E(L, q_F, q_E)$. Rearranging terms, we obtain that $hU_E(H, q_F, q_E) + (1-h)U_E(L, q_F, q_E) = h\pi_E(H, q_F, q_E) + (1-$

$h)\pi_E(L, q_F, q_E)$. Thus, $q_E^*(q_F|\lambda) = \frac{\lambda A_H + (1-\lambda)A_L - q_F}{2}$, where $\lambda = h$ for both $t = H$ and $t = L$.

In summary, since the entrant's best response is the same as in Equation (3.6) when considering short-termism by the entrant, we obtain that the firm's utility function after plugging in the entrant's best response will remain the same as in Equation (3.7). Thus, the rest of the equilibrium analysis will carry through and we will obtain the same equilibrium outcomes and insights as in the current model without considering the short-termism by the entrant.■

3.9.2.3 Analysis Using the Range of the Pareto Dominant Pooling PBE

We confirm that the insights from Propositions 3 and 4 hold when we consider the full range of the Pareto dominant pooling PBE. Multiple Pareto dominant pooling PBE can occur when $\alpha \leq \underline{\alpha_P}$ and $\alpha \geq \overline{\alpha_P}$. We identify the maximum and minimum long-term profit of the firm in the range of the Pareto dominant pooling PBE. When $\alpha \leq \underline{\alpha_P}$, the profit at the LMSE is the upper bound for the L-type firm, lower bound for the H-type firm, and lower bound for the firm when her type is unknown in the profit range arising from the Pareto dominant pooling PBE. When $\alpha \geq \overline{\alpha_P}$, the profit at the LMSE is the lower bound for the L-type firm, upper bound for the H-type firm, and lower bound for the firm when her type is unknown in the profit range arising from the Pareto dominant pooling PBE.

Recall that the optimizers of the L-type firm's utility and the H-type firm's utility under $\lambda = h$ are q_L^P and q_H^P, respectively, as shown in Equation (3.A.8) and Equation (3.A.11). Recall that the optimizers of the L-type firm's profit, the H-type firm's profit, and the expected type's profit under $\lambda = h$ are $\frac{2A_L-\mu}{2}$, $\frac{2A_H-\mu}{2}$, and $\frac{\mu}{2}$, respectively, as shown in the proof of Proposition 4.

- If $\alpha \leq \underline{\alpha_P}$

We obtain the range where the L-type obtains a higher utility under pooling compared to the least-cost separating, i.e., $[\underline{q_L^P}, \overline{q_L^P}]$, by solving $U_F(t_L, q|\lambda = h) \geq U_F(t_L, \overline{q}|\lambda = 1)$, and

$$\underline{q_L^P} = A_L(\frac{1}{2} + \frac{h}{2} - ah - \frac{hs}{2} + ahs - \frac{1}{2}m_1) = q_L^P - \frac{A_L}{2}m_1, \tag{3.A.13}$$

$$\overline{q_L^P} = A_L(\frac{1}{2} + \frac{h}{2} - ah - \frac{hs}{2} + ahs + \frac{1}{2}m_1) = q_L^P + \frac{A_L}{2}m_1, \tag{3.A.14}$$

where $m_1 = ((\theta-1)(-5+12\alpha-8\alpha^2-2h+4\alpha h-h^2+4\alpha h^2-4\alpha^2h^2+7\theta-16\alpha\theta+8\alpha^2\theta+h^2\theta-4\alpha h^2\theta+4\alpha^2h^2\theta-4(1-\alpha)\sqrt{(2\alpha-1)(1+2\alpha(\theta-1)-3\theta)(\theta-1)})^{\frac{1}{2}}$. From

Claim 2 and Claim 5 in the proof for Proposition 2, we obtain that the H-type also obtains a higher utility under pooling in range $[\underline{q_L^P}, \overline{q_L^P}]$, compared to the least-cost separating. Since $\underline{q_L^P} < q_L^P < q_H^P$, the utility of both types under pooling increases in q in the range $[\underline{q_L^P}, q_L^P]$ and decreases in q in the range $[q_H^P, \overline{q_L^P}]$ (if $q_H^P < \overline{q_L^P}$). Hence, we obtain the Pareto dominant pooling range ω that survives the Undefeated refinement as

$$\omega = [q_L^P, \min(\overline{q_L^P}, q_H^P)], \tag{3.A.15}$$

after we trim ranges $[\underline{q_L^P}, q_L^P]$ and $[q_H^P, \overline{q_L^P}]$ that are Pareto dominated by the left limit and right limit of ω, respectively. Note that U_L decreases in q and U_H increases in q under pooling in ω. Hence, the LMSE is the lower bound of ω. Since $q_L^P < q_L^*$ and $q_H^P > q_H^*$, any q in range ω may exceed q_L^* or even q_H^*. We conduct a numerical analysis to investigate the upper bound of ω in the parameter space where the left-side pooling range exists. Specifically, let h take the 25 values between 0.001 and 0.999 with equal increment. For each h value, let θ take 23 values between 1 and $\theta_0(h)$ (1 and θ_0 exclusive) with equal increment. Then for each h and θ value, let α take 25 values between 0 and $\underline{\alpha_P}(h, \theta)$. This procedure generates 14,375 sample scenarios. We find that among the 14,375 scenarios, the upper bound of ω is $\overline{q_L^P}$ (i.e., $\overline{q_L^P} > q_H^P$) in 7582 scenarios, greater than q_L^* in 13384 scenarios, and greater than q_H^* in 8525 scenarios. Thus, the operational distortion can be that both types under-invest, the L-type over-invests and the H-type under-invests, or both types over-invest for the Pareto dominant pooling range when $\alpha \leq \underline{\alpha_P}$.

Next, we analyze the firm profit. According to Proposition 4, $q_L^P > \frac{2A_L - \mu}{2}$, $q_L^P < \frac{2A_H - \mu}{2}$, and $q_L^P < \frac{\mu}{2}$. Thus, considering the range ω instead of q_L^P will add quantities farther from the L-type optimum $\frac{2A_L - \mu}{2}$, closer to the H-type optimum $\frac{2A_H - \mu}{2}$, and closer to the expected-type optimum $\frac{\mu}{2}$. Thus, the L-type can achieve lower profits, the H-type can achieve higher profits, and the expected-type can achieve higher profits with range ω compared to the LMSE quantity q_L^P.

- If $\alpha \geq \overline{\alpha_P}$.

We obtain the range where the H-type obtains a higher utility under pooling compared to the least-cost separating, i.e., $[\underline{q_H^P}, \overline{q_H^P}]$, by solving $U_F(t_H, q|\lambda = h) \geq U_F(t_H, \underline{q}|\lambda = 1)$, and

$$\underline{q_H}^P = A_L(-\frac{1}{2} + \alpha + \frac{h}{2} - \alpha h + \theta - \alpha\theta - \frac{h\theta}{2} + \alpha h\theta - \frac{1}{2}m_2) = q_H^P - \frac{A_L}{2}m_2, \tag{3.A.16}$$

$$\overline{q_H}^P = A_L(-\frac{1}{2} + \alpha + \frac{h}{2} - \alpha h + \theta - \alpha\theta - \frac{h\theta}{2} + \alpha h\theta + \frac{1}{2}m_2) = q_H^P + \frac{A_L}{2}m_2, \tag{3.A.17}$$

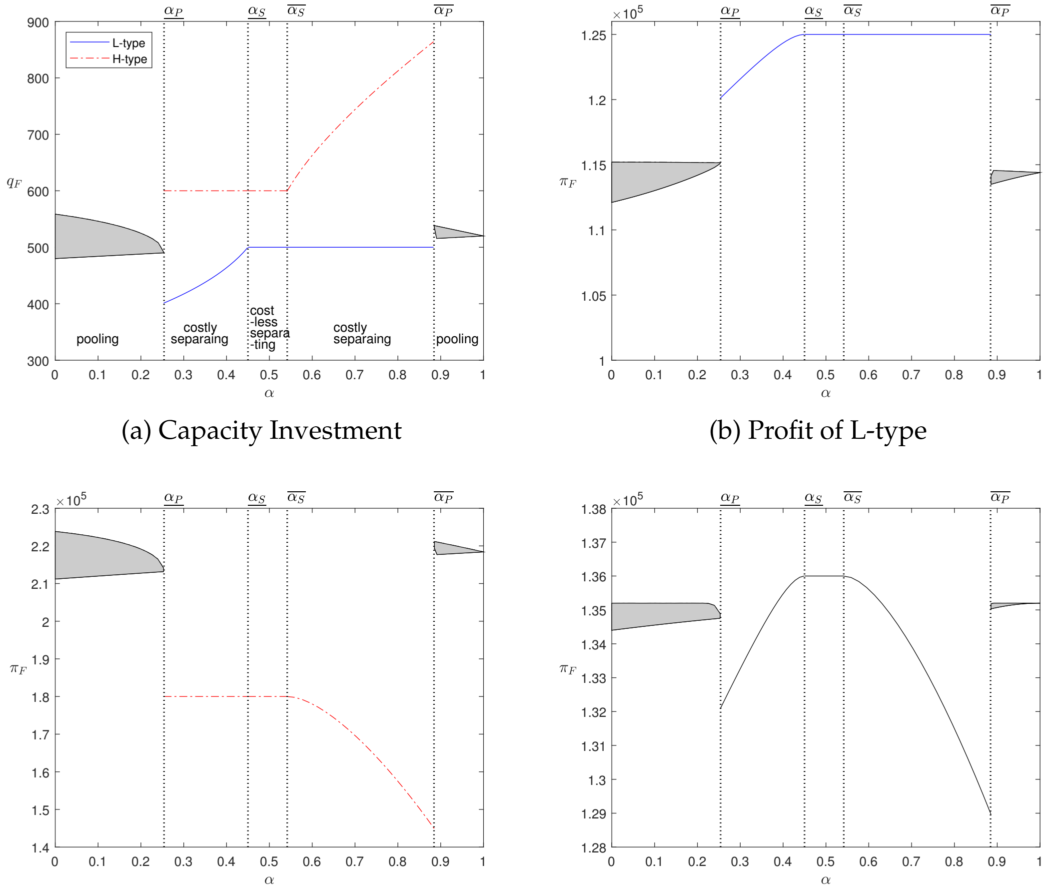
where $m_2 = ((\theta - 1)(-8 + 20\alpha - 12\alpha^2 + 2h - 8\alpha h + 8\alpha^2 h - h^2 + 4\alpha h^2 - 4a^2h^2 + 8\theta - 20\alpha\theta + 12\alpha^2\theta - 4h\theta + 12\alpha h\theta - 8\alpha^2 h\theta + h^2\theta - 4\alpha h^2\theta + 4\alpha^2 h^2\theta - 4(1 - \alpha)\sqrt{(2\alpha - 1)(3 + 2\alpha(\theta - 1) - \theta)(\theta - 1)}))^{\frac{1}{2}}$. From Claim 2 and Claim 5 in the proof for Proposition 2, we obtain that the L-type also obtains a higher utility under pooling in range $[\underline{q_H^P}, \overline{q_H^P}]$, compared to the least-cost separating. Since $\overline{q_H^P} > q_H^P > q_L^P$, the utility of both types under pooling decreases in q in the range $[q_H^P, \overline{q_H^P}]$ and increases in q in the range $[\underline{q_H^P}, q_L^P]$ (if $\underline{q_H^P} < q_L^P$) . Hence, we obtain the Pareto dominant pooling range ω that survives the Undefeated refinement as

$$\omega = [\max(\underline{q_H^P}, q_L^P), q_H^P], \tag{3.A.18}$$

after we trim ranges $[\underline{q_H^P}, q_L^P]$ and $[q_H^P, \overline{q_H^P}]$ that are Pareto dominated by the left limit and right limit of ω, respectively. Note that U_L decreases in q and U_H increases in q under pooling in ω. Hence, the LMSE is the upper bound of ω. Since $q_L^P > q_L^*$ and $q_H^P < q_H^*$ if $\alpha \geq \overline{\alpha_P}$, any q in range ω satisfies $q_L^* < q < q_H^*$. Thus, the L-type over-invests and the H-type under-invests for the Pareto dominant pooling range when $\alpha \geq \overline{\alpha_P}$.

Next, we analyze the firm profit. According to the proof of Proposition 4, $q_H^P > \frac{2A_L - \mu}{2}$, $q_H^P < \frac{2A_H - \mu}{2}$, and $q_H^P \geq \frac{\mu}{2}$. Thus, considering the range ω instead of q_H^P will add quantities closer to the L-type optimum $\frac{2A_L - \mu}{2}$, farther from the H-type optimum $\frac{2A_H - \mu}{2}$, and closer to the expected-type optimum $\frac{\mu}{2}$. Thus, the L-type can achieve higher profits, the H-type can achieve lower profits, and the expected-type can achieve higher profits with range ω compared to the LMSE quantity q_H^P.

To illustrate, we present the firm's capacity investment and profit when $A_L = 1000$, $\theta = 1.2$, and $h = 0.2$ in Figure 3.A.2. Compared to Figure 3.3(a) and Figure 3.4, Figure 3.A.2 shows that if $\alpha \leq \underline{\alpha_P}$, considering the pooling range beyond the pooling LMSE would result in higher investment capacity, lower profit for the L-type, higher profit for the H-type, and higher profit for the expected type. For example, when $\alpha = 0.2$, $\underline{q_L^P} = 451.142$, $\overline{q_L^P} = 524.858$, $q_L^P = 488$ and $q_H^P = 648$, and hence $\omega = [q_L^P, \overline{q_L^P}] = [488, 524.858]$ from Equation (3.A.15). Thus, $q < q_H^* = 600$ and either $q < q_L^* = 500$ or $q > q_L^* = 500$ holds for any q in ω in this example. If $\alpha \geq \overline{\alpha_P}$, an alternative pooling PBE can result in lower investment capacity, higher profit for the L-type, lower profit for the H-type, and higher profit for the expected type. Our main insight that short-termism can benefit the firm's long-term profit still holds when considering alternative pooling PBEs.



(a) Capacity Investment

(b) Profit of L-type

(c) Profit of H-type

(d) Profit of Expected-type

Figure 3.A.2: Pooling Range

3.9.3 Model Extensions

3.9.3.1 Impact of the Firm's Short-termism on the Entrant

The operational distortion of the firm influences the competitive entrant's capacity investment and long-term profit. Proposition 6 shows the entrant's distortion. Once again, our benchmark is complete information. For the entrant, this yields a capacity of $q^*_{EL} = \frac{A_L}{4}$ in the L-type market and $q^*_{EH} = \frac{A_H}{4}$ in the H-type market. Let q^P_E denote the entrant's investment in the pooling PBE, and let q^S_{EL} and q^S_{EH} denote the entrant's investment in the separating PBE in the L-type and H-type market, respectively.

Proposition 6 *The operational distortion of the competitive entrant varies with α as*

follows:
(i) if $\alpha \le \underline{\alpha_P}$, $q_E^P > q_{EL}^*$ and there exists a threshold α_{e1} such that $q_E^P > q_{EH}^*$ if $\alpha > \alpha_{e1}$;
(ii) if $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, $q_{EL}^S > q_{EL}^*$ and $q_{EH}^S = q_{EH}^*$;
(iii) if $\underline{\alpha_S} \le \alpha \le \overline{\alpha_S}$, $q_{EL}^S = q_{EL}^*$ and $q_{EH}^S = q_{EH}^*$;
(iv) if $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, $q_{EL}^S = q_{EL}^*$ and $q_{EH}^S < q_{EH}^*$;
(v) if $\alpha \ge \overline{\alpha_P}$, $q_E^P < q_{EH}^*$ and there exists a threshold α_{e2} such that $q_E^P > q_{EL}^*$ if $\alpha > \alpha_{e2}$.

Proposition 6 first shows that the entrant over-invests in the L-type market and may over-invest in the H-type market in the pooling PBE when $\alpha \le \underline{\alpha_P}$. In the L-type market, the higher posterior belief of the market type in the pooling PBE and the under-investment by the firm induce the entrant to over-invest. In the H-type market, the lower posterior belief of the market type in the pooling PBE and the under-investment by the firm present opposing forces. In Section 3.9.3.1.1 of Appendix 3.9.3, we show that the entrant may fully over-invest, first over-invest and then under-invest as α increases, or fully under-invest in region $\alpha \le \underline{\alpha_P}$. Second, the entrant over-invests in the L-type market as a result of the firm under-investing when $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, and the entrant under-invests in the H-type market as a result of the incumbent firm over-investing when $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$ in the costly separating PBE. In the costless separating PBE, there is no distortion in the entrant's capacity investment. Third, the entrant may over-invest in the L-type market and under-invests in the H-type market when $\alpha \ge \overline{\alpha_P}$. When $\alpha \ge \overline{\alpha_P}$ for the L-type market, the entrant may fully over-invest, or first under-invest and then over-invest as α increases in region $\alpha \ge \overline{\alpha_P}$. For the H-type market, the lower posterior belief of the market type overshadows the incumbent firm's under-investment, inducing the entrant to also under-invest.

We examine the impact of the firm's short-termism on the entrant's long-term profit and find that low levels of short-termism are not necessarily optimal. This finding is formalized in Proposition 7.

Proposition 7 *The entrant's long-term profit can be maximized at non-zero level(s) of short-termism.*

The impact of short-termism on the entrant's long-term profit is manifold. First, the entrant's long-term profit can be maximized at non-zero levels as well. The optimal level of short-termism for the entrant's profit is $\underline{\alpha_P} + 0$ in the L-type market and $\alpha = 0$ in the H-type market. At these levels of short-termism, the incumbent firm either under-invests substantially to signal her L-type or under-invests substantially to conceal her H-type, resulting in a relatively high product margin and ample room for the entrant to occupy market share as shown in Figure 3.A.4. The optimal level of short-termism for the entrant before knowing the market type is $\underline{\alpha_P} + 0$. Second, the incremental effect of increasing short-termism can also be beneficial to the entrant's long-term profit. For example, an incremental increase in short-termism induces higher profits for both types in the pooling segment $\alpha \geq \overline{\alpha_P}$. This is because this incremental increase in α leads to a higher product margin and a higher entry occupation as shown in Figure 3.A.4. We find that the optimal levels of short-termism for the entrant and the incumbent firm are not aligned. This suggests that the competitive entrant can be harmed when the short-termism level moves in a direction beneficial to the incumbent firm.

3.9.3.1.1 Technical Details on the Impact of the Firm's Short-termism on the Entrant's Capacity The entrant's investment in the separating PBE and pooling PBE is $q_E = \frac{A_t - q_S^t}{2}$ and $q_E = \frac{\mu - q^P}{2}$, respectively, according to Definition 1 and 2. Note that the numerical analysis evaluates 575 scenarios, generated by letting h take the 25 equally incremented values between 0.001 and 0.999. For each h value, let θ take 23 equally incremented values between 1 and $\theta_0(h)$ (1 and θ_0 exclusive).

- If $\alpha \leq \underline{\alpha_P}$.

In this left-side pooling region, $q_E^P = \frac{\mu - q^P}{2} = \frac{(3-2\alpha)hA_H + (1-(3-2\alpha)h)A_L}{4}$ and q_E^P decreases in α. Note that $q_E^P - \frac{A_L}{4} = \frac{(3-2\alpha)h(A_H - A_L)}{4} > 0$, i.e., $q_E^P > q_{EL}^*$. Since $q_E^P - \frac{A_H}{4} = \frac{((3-2\alpha)h-1)(A_H-A_L)}{4}$, there exists a threshold $\alpha_{e1} = \frac{3}{2} - \frac{1}{2h}$ such that $q_E^P > \frac{A_H}{4}$ if $\alpha < \alpha_{e1}$ and $q_E^P < \frac{A_H}{4}$ if $\alpha > \alpha_{e1}$. Our numerical analysis shows that among the 575 scenarios, $\alpha_{e1} > \underline{\alpha_P}$ in 322 scenarios, $0 < \alpha_{e1} < \underline{\alpha_P}$ in 69 scenarios, $\alpha_{e1} < 0 < \underline{\alpha_P}$ in 184 scenarios. Thus, the following three scenarios can emerge such that $q_E^P > q_{EH}^*$ for $\alpha \leq \underline{\alpha_P}$, $q_E^P > q_{EH}^*$ for $\alpha < \alpha_{e1}$ and then $q_E^P < q_{EH}^*$ for $\alpha_{e1} < \alpha \leq \underline{\alpha_P}$, or $q_E^P < q_{EH}^*$ for $\alpha \leq \underline{\alpha_P}$.

- If $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$.

In this costly separating region, $q_{EL}^S = \frac{A_L - \bar{q}}{2}$ where $\bar{q}$ is shown in Equation (3.A.1) and $q_{EH}^S = \frac{A_H}{4}$. Since $\bar{q} < \frac{A_L}{2}$, $q_{EL}^S > q_{EL}^*$.

- If $\underline{\alpha_S} \leq \alpha \leq \overline{\alpha_S}$.

In this costless separating region, $q_{EL}^S = \frac{A_L}{4}$ and $q_{EH}^S = \frac{A_H}{4}$.

- If $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$.

In this costly separating region, $q^S_{EH} = \frac{A_H - \underline{q}}{2}$ where $\underline{q}$ is shown in Equation (3.A.4) and $q^S_{EL} = \frac{A_L}{4}$. Since $\underline{q} > \frac{A_H}{2}$, $q^S_{EH} < q^*_{EH}$.

- If $\alpha \geq \overline{\alpha_P}$.

In this right-side pooling region, $q^P_E = \frac{\mu - q^P}{2} = \frac{(2\alpha-2+(3-2\alpha)h)A_H+(3-2\alpha-(3-2\alpha)h)A_L}{4}$ and q^P_E increases in α. Note that $q^P_E - \frac{A_H}{4} = \frac{1}{4}(2\alpha-3)(1-h)(\theta-1)A_L < 0$, i.e., $q^P_E < q^*_{EH}$ for any α. Since $q^P_E - \frac{A_L}{4} = -\frac{1}{4}(2+2\alpha(h-1)-3h)(\theta-1)A_L$, there exists a threshold $\alpha_{e2} = \frac{2-3h}{2(1-h)}$ such that $q^P_E > \frac{A_L}{4}$ if $\alpha > \alpha_{e2}$ and $q^P_E < \frac{A_L}{4}$ if $\alpha < \alpha_{e2}$. Our numerical analysis shows that among the 575 scenarios, $\alpha_{e2} < \overline{\alpha_P}$ in 429 scenarios, and $\alpha_{e2} > \overline{\alpha_P}$ in 146 scenarios (note that $\alpha_{e2} < 1$). Thus, the following two scenarios can emerge such that $q^P_E > q^*_{EL}$ for $\alpha \geq \overline{\alpha_P}$, or $q^P_E < q^*_{EL}$ for $\overline{\alpha_P} \leq \alpha < \alpha_{e2}$ and $q^P_E > q^*_{EL}$ for $\alpha > \alpha_{e2}$.

3.9.3.1.2 Technical Details on the Impact of the Firm's Short-termism on the Entrant's Long-Term Profit To provide more mathematical rigor on the impact of the firm's short-termism on the competitor's profit, we examine the pooling region when α is low, i.e., $\theta < \theta_0$. Since $q_E = \frac{A_t - q_F}{2}$ in a separating PBE, $\pi_E = \frac{(A_t - q_F)^2}{4}$. Since $q_E = \frac{\mu - q_F}{2}$ in a pooling PBE, $\pi_E = \frac{(2A_t - \mu - q_F)(\mu - q_F)}{4}$. Note that any numerical analysis in this section evaluates the 575 scenarios generated by letting h take the 25 values between 0.001 and 0.999 with equal increment. For each h value, let θ take 23 values between 1 and $\theta_0(h)$ (1 and θ_0 exclusive) with equal increment.

- L-type market

If $\alpha \leq \underline{\alpha_P}$, π_{EL} decreases in q_F since $\frac{\partial \pi_{EL}}{\partial q_F} = 2(q_F - A_L) < 0$. Thus, π_{EL} decreases in α. Moreover, $\pi_{EL} = \frac{A_L^2}{16}(-1+(-3+2\alpha)h(\theta-1))(-1+(1+2\alpha)h(\theta-1))$. If $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, $\bar{q}$ increases in α. Note that $\bar{q} < \frac{A_L}{2}$. Thus, π_{EL} decreases in α and $\pi_{EL} > \frac{A_L^2}{16}$. If $\underline{\alpha_S} < \alpha < \overline{\alpha_P}$, $\pi_{EL} = \frac{A_L^2}{16}$. If $\alpha \geq \overline{\alpha_P}$, $\pi_{EL} = \frac{1}{16}(-3-2\alpha(1-h)(\theta-1)-3h(\theta-1)+2\theta)(-3-2\alpha(1-h)(\theta-1)+h(\theta-1)+2\theta))$. Note that $\frac{\partial \pi_{EL}}{\partial \alpha} = \frac{1}{8}(1-h)(\theta-1)(6-4\alpha(-1+h)(\theta-1)+2h(\theta-1)-4\theta)$. Hence, $\frac{\partial \pi_{EL}}{\partial \alpha} > 0$ if and only if $f(\alpha) = 6-4\alpha(-1+h)(\theta-1)+2h(\theta-1)-4\theta > 0$, i.e., $\alpha > \frac{2\theta - h(\theta-1)-3}{2(\theta-1)(1-h)}$. Our numerical analysis shows that $\alpha > \frac{2\theta-h(\theta-1)-3}{2(\theta-1)(1-h)}$ holds if $\alpha > \overline{\alpha_P}$. Thus, π_{EL} increases in α. First, $\pi_{EL}(\alpha = 0) > \pi_{EL}(\alpha = 1)$ since $\pi_{EL}(\alpha = 0) - \pi_{EL}(\alpha = 1) = 4h(\theta - 1) > 0$. Second, $\pi_{EL}(\alpha = 0) - \frac{A_L^2}{16} = h(2-3h(\theta-1))(\theta-1)$. Thus, $\pi_{EL}(\alpha = 0) > \frac{A_L^2}{16}$ if and only if $\theta < \frac{2}{3h} + 1$. Our numerical analysis shows that $\theta < \frac{2}{3h} + 1$ holds if $\theta < \theta_0$. Thus, $\pi_{EL}(\alpha = 0) > \frac{A_L^2}{16}$. Third, our numerical analysis shows that $\pi_{EL}(\underline{\alpha_P} + 0) > \pi_{EL}(\alpha = 0)$. In summary, $\alpha = \underline{\alpha_P} + 0$ is optimal for π_{EL}.

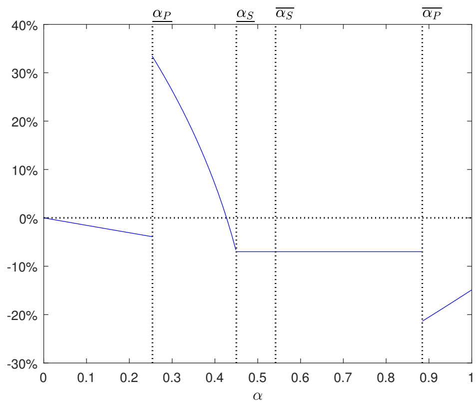
- H-type market

If $\alpha \leq \underline{\alpha_P}$, π_{EH} increases in q_E since $\frac{\partial \pi_{EH}}{\partial q_E} = A_H - \mu + 2q_E > 0$. Thus, π_{EH} decreases in α. Moreover, $\pi_{EH} = \frac{A_L^2}{16}(-1 + (-3 + 2\alpha)h(\theta - 1))(3 + (1 + 2\alpha)h(\theta - 1) - 4\theta)$. Note that $\pi_{EH} - \frac{A_H^2}{16} = \frac{A_L^2}{16}(\theta - 1)f(\alpha)$ where $f(\alpha) = 3 + (-3 - 4\alpha + 4\alpha^2)h^2(\theta - 1) - \theta - 2h(5 - 6\theta + \alpha(-2 + 4\theta))$. Since π_{EH} decreases in α, $\pi_{EH} - \frac{A_H^2}{16} > \frac{A_L^2}{16}(\theta - 1)f(0.5)$ and $f(0.5) = 3 + 8h(\theta - 1) - 4h^2(\theta - 1) - \theta > 0$. Thus, we obtain that $\pi_{EH} > \frac{A_H^2}{16}$. If $\underline{\alpha_P} < \alpha < \overline{\alpha_S}$, $\pi_{EH} = \frac{A_H^2}{16}$. If $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, $\underline{q}$ increases in α. Note that $\underline{q} > \frac{A_H}{2}$ and our numerical analysis shows that $\underline{q} < A_H$. Thus, π_{EH} decreases in α and $\pi_{EH} < \frac{A_H^2}{16}$. If $\alpha \geq \overline{\alpha_P}$, π_{EH} increases in q_E since $\frac{\partial \pi_{EH}}{\partial q_E} = A_H - \mu + 2q_E > 0$. Thus, π_{EH} increases in α. Moreover, $\pi_{EH} = \frac{A_L^2}{16}(1 + 2\alpha(-1 + h)(\theta - 1) + h(\theta - 1) - 2\theta)(-3 + 2\alpha(h - 1)(\theta - 1) - 3h(\theta - 1) + 2\theta)$. Note that $\pi_{EH} - \frac{A_H^2}{16} = \frac{A_L^2}{16}(1 - h)(\theta - 1)f(\alpha)$ where $f(\alpha) = 3 + 4\alpha(1 + h(\theta - 1)) - 4\alpha^2(-1 + h)(\theta - 1) + 3h(\theta - 1) - 5\theta$. $\pi_{EH}(\alpha = 1) > \frac{A_H^2}{16}$ since $f(1) = 3 + 3h(\theta - 1) - \theta > 0$. Note that $\pi_{EH}(\alpha = 0) - \pi_{EH}(\alpha = 1) = \frac{A_L^2}{4}h(\theta - 1)(2\theta - 1) > 0$. Thus, $\pi_{EH}(\alpha = 0) > \pi_{EH}(\alpha = 1)$. In summary, $\alpha = 0$ is optimal for π_{EH}.

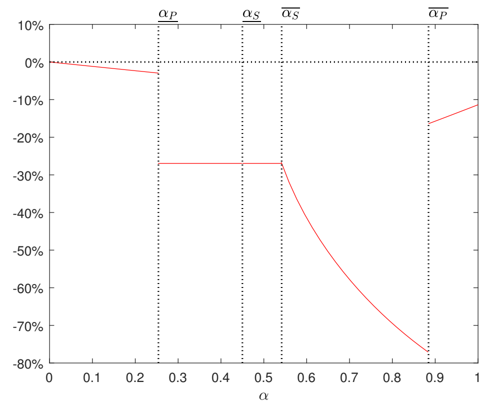
- Expected-type market

The expected profit of the entrant before the market type is realized is $\pi_{EW} = h\pi_{EH} + (1 - h)\pi_{EL}$. If $\alpha \leq \underline{\alpha_P}$, $\pi_{EW} = \frac{(\mu - q_F)^2}{4}$. Since $q_L^P < \frac{\mu}{2}$ and q_L^P increases in α, π_{EW} decreases in α. If $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$ and $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, since π_{EL} and π_{EH} decreases in α, respectively, π_{EW} decreases in α. If $\alpha \geq \overline{\alpha_P}$, $\pi_{EW} = \frac{(\mu - q_F)^2}{4}$. $q_H^P > \frac{\mu}{2}$ and decreases in α. Our numerical analysis shows that $q_H^P < \mu$. Thus, π_{EW} increases in α. First, $\pi_{EW}(\alpha = 0) > h\frac{A_H^2}{16} + (1 - h)\frac{A_L^2}{16}$ since $\pi_{EW}(\alpha = 0) - (h\frac{A_H^2}{16} + (1 - h)\frac{A_L^2}{16}) = \frac{A_L^2}{4}((h\theta + 1 - h) - \frac{1}{2}(-h\theta + h + 1))^2 - \frac{A_L^2}{16}(h\theta^2 + (1 - h)) = \frac{A_L^2 h}{16}(5 + 9h(\theta - 1) - \theta)(\theta - 1) > 0$. Second, $\pi_{EW}(\alpha = 0) > \pi_{EW}(\alpha = 1)$ since $\pi_{EW}(\alpha = 0) - \pi_{EW}(\alpha = 1) = \frac{A_L^2}{4}((h\theta + 1 - h) - \frac{1}{2}(-h\theta + h + 1))^2 - \frac{A_L^2}{4}(h\theta + 1 - h - \frac{1}{2}(-(1 - h)(\theta - 1) + \theta))^2 = h(1 + 2h(\theta - 1))(\theta - 1) > 0$. Third, our numerical analysis shows that $\pi_{EW}(\alpha = 0) < \pi_{EW}(\underline{\alpha_P} + 0)$. In summary, $\alpha = \underline{\alpha_P} + 0$ is optimal for π_{EW}.

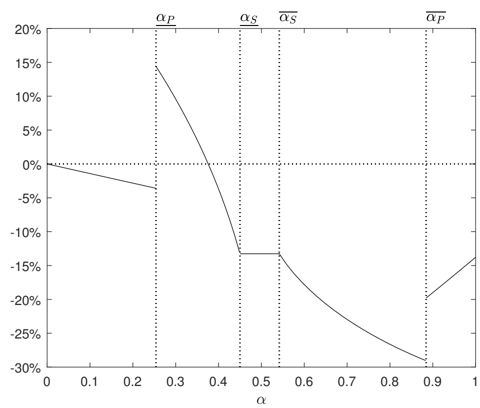
Figure 3.A.3 shows the entrant's equilibrium profit when $\theta = 1.2$ and $h = 0.2$. We clearly observe that the entrant's long-term profit is maximized at non-zero short-termism levels for the L-type and when the type is unknown.



(a) L-type



(b) H-type



(c) Expected type

Figure 3.A.3: Impact of α on the Entrant's Long-term Profit

3.9.3.2 Impact of the Firm's Short-termism on the Market

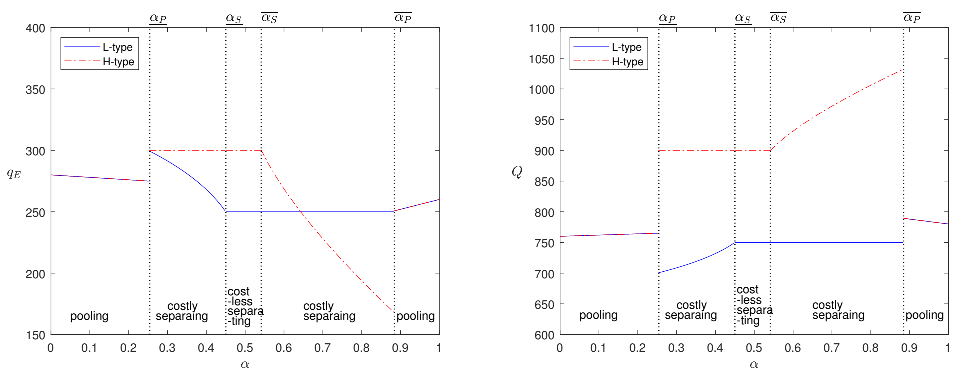
We present the impact of α on the market-level investment $Q = q_F + q_E$ in Proposition 8. Let Q^P denote the market-level investment in the pooling equilibrium, and Q_L^S and Q_H^S denote the market-level investment in the separating equilibrium in the L-type and H-type, respectively. Consistent with the discussion for the firm, we use the benchmark of complete information in which the market-level investment is $Q_L^* = \frac{3A_L}{4}$ in the L-type market and $Q_H^* = \frac{3A_L}{4}$ in the H-type market.

Proposition 8 *The operational distortion at the market level varies with α as follows:*
(i) if $\alpha \leq \underline{\alpha_P}$, $Q_L^* < Q^P < Q_H^*$;
(ii) if $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, $Q_L^S < Q_L^*$ and $Q_H^S = Q_H^*$;
(iii) if $\underline{\alpha_S} \leq \alpha \leq \overline{\alpha_S}$, $Q_L^S = Q_L^*$ and $Q_H^S = Q_H^*$;

(iv) if $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$*,* $Q_L^S = Q_L^*$ *and* $Q_H^S > Q_H^*$*;*
(v) if $\alpha \geq \overline{\alpha_P}$*,* $Q_L^* < Q^P < Q_H^*$*.*

Proposition 8 shows that the market-level investment is higher for the L-type market and lower for the H-type market in the pooling equilibrium. Particularly in the pooling PBE where the incumbent firm under-invests in the L-type market (when $\alpha \leq \underline{\alpha_P}$), the entrant is incentivized to over-invest sufficiently which induces an overall market-level over-investment. In the costly separating PBE, the market-level operational distortion is dominantly driven by the incumbent firm. We show that a market-level under-investment emerges in the L-type market when $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$ and a market-level over-investment emerges in the H-type market when $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$.

Figure 3.A.4 shows the capacity investment of the entrant and the market when $\theta = 1.2$, $h = 0.2$, and $A_L = 1000$. Compared to Figure 3.3(a), the entrant's capacity investment increases in α when the incumbent firm's capacity investment decreases in α in each equilibrium segment. Figure 3.A.4(a) shows that the entrant can invest more in the L-type market than in the H-type market in the costly separating PBE, as a result of the incumbent firm's substantial operational distortion to costly signal her type.



(a) Entrant (b) Market-level

Figure 3.A.4: Impact of α on Entrant's Capacity Investment q_E and Market-level Capacity Investment Q

We examine the impact of the firm's short-termism on social welfare and find that low levels of short-termism are not necessarily optimal. This finding is summarized in Proposition 9.

Proposition 9 *The social welfare can be maximized at non-zero level(s) of short-termism.*

Social welfare encompasses the profits of the incumbent firm and the entrant and the consumer surplus. The consumer surplus is the difference between the total amount that the consumers are willing to pay and the total amount they actually pay. We show that the consumer surplus increases in the market-level investment. Thus, non-zero short-termism levels and incremental short-termism can benefit the consumer surplus, as implied by Figure 3.A.4(b). The optimal level of short-termism for the social welfare is $\overline{\alpha_P}$ in the L-type market, either $\overline{\alpha_P} - 0$ or $\overline{\alpha_P}$ in the H-type market and in the market before the type is known. This implies that the optimal level of short-termism for the incumbent firm's profit can be socially optimal in the H-type market. We also find that the incremental effect of increasing short-termism can benefit social welfare. Thus, eliminating short-termism can be unnecessary and even harmful for social welfare.

3.9.3.2.1 Technical Details on the Impact of the Firm's Short-termism on the Market Capacity The market-level investment is $Q = q_F + q_E$. Thus, $Q = \frac{\mu+q_F}{2}$ in pooling PBE and $Q = \frac{A_t+q_F}{2}$ in separating PBE. Thus, Q increases as q_F increases in each equilibrium region.

- If $\alpha \leq \underline{\alpha_P}$.

In this left-side pooling region, $Q = \frac{A_L}{4}(3 + (1 + 2\alpha)h(\theta - 1))$ and Q increases in α. $Q - \frac{3A_H}{4} = \frac{A_L}{4}(-3 + h + 2\alpha h)(\theta - 1)$. Thus, $Q < \frac{3A_H}{4}$ since $\alpha < \frac{3-h}{2h} \in (1, \infty)$. $Q - \frac{3A_L}{4} = \frac{1}{4}(2\alpha + 1)h(\theta - 1) > 0$. Thus, $Q > \frac{3A_L}{4}$. Thus, $Q_L^* < Q < Q_H^*$ for $\alpha \leq \underline{\alpha_P}$.

- If $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$.

In this costly separating region, $Q_L^S = \frac{A_L+\bar{q}}{2}$ where $\bar{q}$ is shown in Equation (3.A.1) and $Q_H^S = \frac{3A_H}{4}$. Since $\bar{q} < \frac{A_L}{2}$, $Q_L^S < Q_L^*$.

- If $\underline{\alpha_S} \leq \alpha \leq \overline{\alpha_S}$.

In this costless separating region, $Q_L^S = \frac{3A_L}{4}$ and $Q_H^S = \frac{3A_H}{4}$.

- If $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$.

In this costly separating region, $Q_H^S = \frac{A_H+\underline{q}}{2}$ where $\underline{q}$ is shown in Equation (3.A.4) and $Q_L^S = \frac{3A_L}{4}$. Since $\underline{q} > \frac{A_H}{2}$, $Q_H^S > Q_H^*$.

- If $\alpha \geq \overline{\alpha_P}$.

In this right-side pooling region, $Q = \frac{A_L}{4}(1 - 2\alpha(1-h)(\theta-1) + h(\theta-1) + 2\theta)$ and Q decreases in α. $Q - \frac{3A_H}{4} = -\frac{A_L}{4}(1+2\alpha)(\theta-1)(1-h)$. Thus, $Q < \frac{3A_H}{4}$. $Q - \frac{3A_L}{4} = \frac{A_L}{4}(2 - 2\alpha(1-h) + h)(\theta-1)$. Thus, $Q > \frac{3A_L}{4}$ since $\alpha < \frac{h+2}{2(1-h)} \in (1, \infty)$. Thus, $Q_L^* < Q < Q_H^*$ for $\alpha \geq \overline{\alpha_P}$.

3.9.3.2.2 Technical Details on the Impact of the Firm's Short-termism on Consumer Surplus and Social Welfare The social welfare SW we consider encompasses the firm's profit π_F, the entrant's profit π_E, and the consumer surplus CS, and $SW = \pi_F + \pi_E + CS$. Note that any numerical analysis in this section evaluates the 575 scenarios generated by letting h take the 25 values between 0.001 and 0.999 with equal increment. For each h value, let θ take 23 values between 1 and $\theta_0(h)$ (1 and θ_0 exclusive) with equal increment.

- Consumer Surplus

The consumer surplus is measured by the area between the demand curve and the price level, i.e., $CS = \frac{1}{2}(A_t - p)Q = \frac{Q^2}{2}$, where Q is the market-level investment. Thus, the consumer surplus increases as the total output increases.

- Social Welfare

Social welfare $SW = \pi_F + \pi_E + CS = (A_t - Q)q_F + (A_t - Q)q_E + CS = (A_t - \frac{Q}{2})Q$. In the costless separating PBE, $SW_H = \frac{15A_H^2}{32}$ and $SW_L = \frac{15A_L^2}{32}$. Since $\frac{\partial SW}{\partial Q} = A_t - Q$, the maximizer of SW is $Q^{SW} = A_t$. First, we discuss the L-type market. If $\alpha \leq \underline{\alpha_P}$, our numerical analysis shows that $Q(\underline{\alpha_P} - 0) < A_L$. Since Q increases in α, SW_L increases in α. If $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$, SW_L increases in α since $Q < \frac{3A_L}{4}$ and Q increases in α. If $\underline{\alpha_S} < \alpha < \overline{\alpha_P}$, $SW_L = \frac{15A_L^2}{32}$. If $\alpha \geq \overline{\alpha_P}$, our numerical analysis shows that $Q(\overline{\alpha_P} + 0) < A_L$. Since Q decreases in α, SW_L decreases in α. $SW_L(\alpha = 0) - \frac{15A_L^2}{32} = -\frac{A_L^2}{32}h(-2 + h(\theta-1))(\theta-1) > 0$ since our numerical analysis shows that $\theta < \frac{2}{h} + 1$ holds if $\theta < \theta_0$. Thus, $SW_L(\alpha = 0) > \frac{15A_L^2}{32}$. Our numerical analysis shows that $SW_L(\underline{\alpha_P} - 0) < SW_L(\overline{\alpha_P} + 0)$. Thus, $\overline{\alpha_P} + 0$ is optimal for SW_L.

Second, we discuss the H-type market. If $\alpha \leq \underline{\alpha_P}$, SW_H increases in α since $Q < \frac{3A_H}{4}$ and Q increases in α. If $\underline{\alpha_P} < \alpha < \overline{\alpha_S}$, $SW_H = \frac{15A_H^2}{32}$. If $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, note that $Q > \frac{3A_H}{4}$ and increases in α. Our numerical analysis shows that

$Q(\overline{\alpha_P} - 0) < A_H$ and hence SW_H increases in α. If $\alpha \geq \overline{\alpha_P}$, SW_H decreases in α since Q¡$\frac{3A_H}{4}$ and Q decreases in α. Note that $SW_H(\alpha = 0) - SW_H(\alpha = 0.5) = A_L^2((\theta - \frac{1}{8}(3 + h(\theta - 1)))\frac{1}{4}(3 + h(1 - \theta)) - \frac{15\theta^2}{32}) = \frac{A_L^2}{32}(\theta - 1)(9 + h^2(\theta - 1) - 15\theta - 8h\theta)$. Since $\theta > \frac{9-h^2}{15+8h-h^2} \in (\frac{4}{11}, \frac{3}{5})$, $SW_H(\alpha = 0) < SW_H(\alpha = 0.5)$. Note that $SW_H(\alpha = 0) - SW_H(\alpha = 1) = A_L^2((\theta - \frac{1}{8}(3 + h(\theta - 1)))\frac{1}{4}(3 + h(1 - \theta)) - (\theta - \frac{1}{8}(1 - 2(1 - h)(\theta - 1) + h(\theta - 1) + 2\theta))\frac{1}{4}(1 - 2(1 - h)(\theta - 1) + h(\theta - 1) + 2\theta)) = \frac{A_L^2}{16}(\theta - 1)h(9 + 5h(\theta - 1) - 16\theta)$. Since $\theta > \frac{9-5h}{16-5h} \in (\frac{9}{16}, \frac{4}{11})$, $SW_H(\alpha = 0) < SW_H(\alpha = 1)$. Our numerical analysis shows that $SW_H(\overline{\alpha_P} - 0) > SW_H(\overline{\alpha_p} + 0)$ in 552 scenarios among the 575 scenarios. Our numerical analysis shows that $SW_H(\underline{\alpha_P} - 0) < \frac{15A_H^2}{32}$. Thus, $\alpha = \overline{\alpha_P} - 0$ or $\alpha = \overline{\alpha_P} + 0$ is optimal for SW_H.

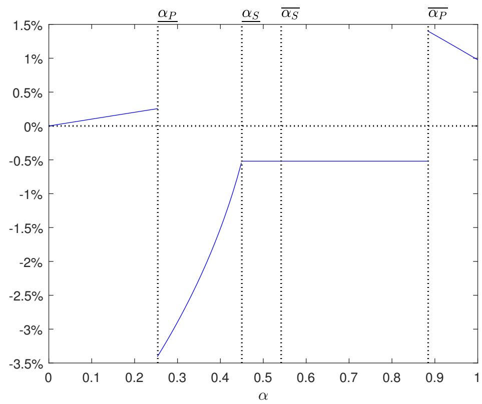
Third, we discuss the market before knowing type. In the pooling region, $SW_W = (\mu - \frac{Q}{2})Q$. If $\alpha \leq \underline{\alpha_P}$, since Q increases in α and our numerical analysis shows that $Q(\underline{\alpha_P} - 0) < A_L < \mu$, SW_W increases in α. If $\underline{\alpha_P} < \alpha < \underline{\alpha_S}$ or $\overline{\alpha_S} < \alpha < \overline{\alpha_P}$, SW_W increases in α since SW_L or SW_H increases in α, respectively. If $\alpha \geq \overline{\alpha_P}$, since Q decreases in α and our numerical analysis shows that $Q(\overline{\alpha_P} + 0) < \mu$, SW_W decreases in α. Note that $SW_W(\alpha = 0) - (h\frac{15A_H^2}{32} + (1 - h)\frac{15A_L^2}{32}) = \frac{A_L^2}{32}h(\theta - 1)(11 + 7h(\theta - 1) - 15\theta)$. Since $\theta > \frac{11-7h}{15-7h}$, $SW_W(\alpha = 0) < SW_W(\alpha = 0.5)$. Our numerical analysis shows that $SW_W(\overline{\alpha_P} - 0) > SW_W(\underline{\alpha_P} - 0)$, and $SW_W(\overline{\alpha_P} - 0) > SW_W(\underline{\alpha_P} + 0)$ in 552 scenarios among the 575 scenarios, we obtain that $\alpha = \overline{\alpha_P} - 0$ or $\alpha = \overline{\alpha_P} + 0$ is optimal for SW_W.

Figure 3.A.5 shows the social welfare when $\theta = 1.2$ and $h = 0.2$. We clearly observe that the social welfare is maximized at non-zero short-termism levels.

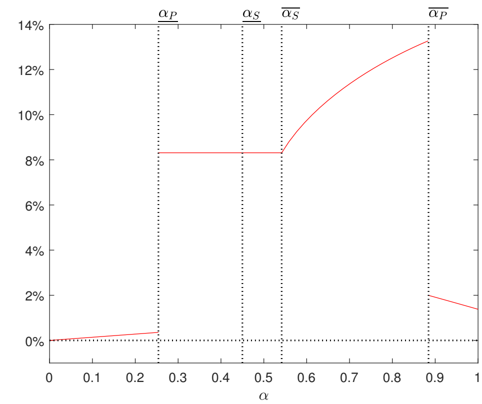
3.9.3.3 Impact of Short-termism on Operational Distortion and Profit in the Duopoly Setting

This section supports the extension of our analysis to a duopoly setting, which represents a form of established competition. The utility functions in a duopoly follow a similar logic as our main model. The investor chooses v_F and v_E to maximize its expected utility based on its posterior belief, i.e., $(v_F^*, v_E^*) = \arg\max EU_I = \arg\max(\lambda U_I(t_H) + (1 - \lambda)U_I(t_L))$, where $U_I(t) = -(v_F - \pi_F(t))^2 - (v_E - \pi_E(t))^2$. From this, we obtain that $v_i^* = \lambda\pi_i(t_H) + (1 - \lambda)\pi_i(t_L)$ where $i \in \{E, F\}$. This shows that the investor's short-term valuation assigned to both firms is the weighted average of the firm's long-term profit based on the investor's posterior belief. Thus, the firms' utility functions after plugging in the investor's best response (v_F^*, v_E^*) is

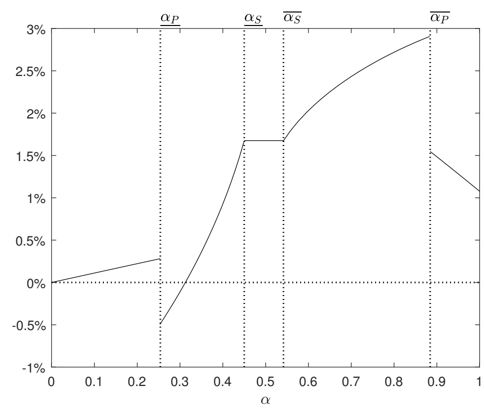
$$U_i(t, q_i, q_j) = \alpha(\lambda A_H + (1 - \lambda)A_L - q_i - q_j)q_i + (1 - \alpha)(A_t - q_i - q_j)q_i \qquad (3.A.19)$$



(a) L-type



(b) H-type



(c) Expected type

Figure 3.A.5: Impact of α on the Social Welfare

where $i, j \in \{E, F\}$ and $i \neq j$. Equation (3.A.19) shows that the firms' utility is higher when $t = H$ compared to $t = L$. T his indicates that the H-type is the mimicked type and the L-type is the mimicking type. This is in contrast to our main model, where the mimicked firm type will switch from the H-type to the L-type as short-termism level α decreases below $\frac{1}{2}$ due to dual signaling.

The firms' equilibrium capacity investment levels are obtained by finding the fixed point(s) of the best responses of the two firms. We characterize each player's strategy in the least-cost separating PBE in Definition 6. Due the concavity of the firms' utility functions, the least-cost separating PBE always exists, which we formally show in the technical details of Definition 6 in Section 3.9.3.3.1 of Appendix 3.9.3. We use $q_1(\cdot)$ to denote the boundary for an L-type posterior belief, and its mathematical formula is shown in Section 3.9.3.3.1 of Appendix 3.9.3.

Definition 6 *In the least-cost separating PBE, each player's strategy is:*

(i) The L-type firms invest $q_L^S = \frac{A_L}{3}$; the H-type firms invest $q_H^S > \frac{A_H}{3}$ if $\theta < \frac{3}{(3-4\alpha)^+}$ and $q_H^S = \frac{A_H}{3}$ if $\theta \geq \frac{3}{(3-4\alpha)^+}$.

(ii) The investor assigns valuation $v_E^* = v_F^* = (A_t - 2q_t^S)q_t^S$ for $t \in \{H, L\}$.

(iii) The equilibrium is supported by the investor's posterior belief λ such that $\lambda = 1$ for region $\{(q_F, q_E) : q_F \geq q_1(q_E)$ and $q_E \geq q_1(q_F)\}$ and $\lambda = 0$ otherwise.

Quantities $\frac{A_H}{3}$ and $\frac{A_L}{3}$ are the benchmark quantities of the firms when there is complete information. It is straightforward to obtain that these quantities are the equilibrium capacity investment when $\alpha = 0$. Definition 6 shows that there are deviations from these benchmarks when $\alpha > 0$. The H-type firms will invest more than $\frac{A_H}{3}$ in the least-cost separating PBE when θ is sufficiently small. In contrast, the L-type will invest $\frac{A_L}{3}$. The condition $\theta < \frac{3}{(3-4\alpha)^+}$ is equivalent to $\alpha > \frac{3}{4} - \frac{3}{4\theta}$, so the H-type firms will over-invest when short-termism level α is sufficiently large. As shown in the technical details of Definition 6 in Section 3.9.3.3.1 of Appendix 3.9.3, the H-type firms' separating quantity q_H^S increases in α when $\theta < \frac{3}{(3-4\alpha)^+}$.

The pooling PBE is obtained by finding the fixed point(s) of the pooling response(s) of each firm. Under the Undefeated refinement, a pooling PBE exists when both firm types for both firms obtain a higher utility when they pool at q^P compared to their utilities under the least-cost separating PBE. As in our main model, we refine the set of PBE using the LMSE when multiple pooling PBE exist. Due the concavity of the firms' utility functions, a unique LMSE exists in our setting. We characterize each player's strategy in the pooling LMSE in Definition 7. We use $q_2(\cdot)$ to denote the boundary for a weighted posterior belief, and its mathematical formula is shown in the technical details of Definition 7 in Section 3.9.3.3.2 of Appendix 3.9.3, along with the conditions for the existence of the pooling LMSE.

Definition 7 *In the pooling LMSE, each player's strategy is:*

(i) Both firm types invest q^P where $\frac{A_L}{3} < q^P < \frac{A_H}{3}$

(ii) The investor assigns valuation $v_E^* = v_F^* = (\mu - 2q^P)q^P$ for $t \in \{H, L\}$.

(iii) The equilibrium is supported by the investor's posterior belief λ such that $\lambda = 1$ for region $\{(q_F, q_E) : q_F \geq q_1(q_E)$ and $q_E \geq q_1(q_F)\}$, $\lambda = h$ for region $\{(q_F, q_E) : q_2(q_E) \leq q_F < q_1(q_E)$ and $q_2(q_F) \leq q_E < q_1(q_F)\}$, and $\lambda = 0$ otherwise.

Definition 7 shows that in the pooling LMSE, the H-type firms will under-invest and the L-type firms will over-invest to conceal their types. As shown in Section 3.9.3.3.2 of Appendix 3.9.3, the pooling quantity q^P decreases in α.

We present the manifold effect of short-termism on operational distortion in Proposition 10. Proposition 10 is the counterpart of Proposition 3 in the incumbent-entrant setting.

Proposition 10 *Given θ and h, operational distortion in the duopoly setting varies with α as follows:*

(i) if $\alpha \le \overline{\alpha_{SD}}$, there is no operational distortion;

(ii) if $\overline{\alpha_{SD}} < \alpha < \overline{\alpha_{PD}}$, $q_L^S = \frac{A_L}{3}$ and q_H^S increases as α increases;

(iii) if $\alpha \ge \overline{\alpha_{PD}}$, q^P decreases as α increases.

We present the manifold effect of short-termism on long-term profit in Proposition 11. This proposition is the counterpart of Proposition 4 in the incumbent-entrant setting.

Proposition 11 *The firm's long-term profit in the duopoly setting can be maximized at non-zero level(s) of short-termism, and the optimal level of short-termism can differ by firm type.*

The technical proofs and details of Propositions 10 and 11 are in Sections 3.9.3.3.3 and 3.9.3.3.4 of Appendix 3.9.3.

3.9.3.3.1 Technical Details of Definition 6 We derive the least-cost separating PBE. First, we claim that both firms will either be in a separating PBE or in a pooling PBE. The case where one player separates and the other player pools cannot sustain as an equilibrium because the investor would infer the true market demand from the player who separates and hence the other player could not sustain in a pooling PBE in such a case. To obtain the separating PBE, we derive the firms' best responses when their type is revealed and find the fixed point(s). Let $i, j \in \{E, F\}$ and $i \neq j$ be the subscript for the firms. We focus our discussion on the best response of firm i, $q_i(q_j)$, given firm j's strategy q_j below for ease of illustration.

For the L-type demand, firm i's best response under posterior belief $\lambda = 0$ is $\frac{A_L - q_j}{2}$, and firm i obtains utility $u_1(q_j) = \frac{(A_L - q_j)^2}{4}$. Firm i's utility under posterior belief $\lambda = 1$ is $U_i(t_L, q_i, q_j | \lambda = 1) = (\alpha A_H + (1-\alpha)A_L - q_i - q_j)q_i$. Thus, there exists a $\underline{q_i}$ such that $U_i(t_L, \underline{q_i}, q_j | \lambda = 1) = u_1(q_j)$ and $U_i(t_L, q_i, q_j | \lambda = 1) < u_1(q_j)$ for any $q_i > \underline{q_i}$. We solve for $\underline{q_i}$ and obtain that $\underline{q_i} = q_1(q_j)$ where

$$q_1(q_j) = \frac{A_L}{2}(1 - \alpha + \alpha\theta + \sqrt{\alpha}\sqrt{\theta - 1}\sqrt{2 - \alpha + \alpha\theta - 2\frac{q_j}{A_L} - \frac{q_j}{A_L}}). \qquad (3.A.20)$$

Since any $q_i > q_1(q_j)$ is strictly dominated by $\frac{A_L - q_j}{2}$, the reasonable posterior belief should be $\lambda = 1$ for any $q_i \geq q_1(q_j)$. For the H-type demand, firm i's best response under posterior belief $\lambda = 1$ is $\max(q_1(q_j), \frac{A_H - q_j}{2})$, where $q_1(q_j)$ is shown in Equation (3.A.20). Thus, the best response of firm i is $q_i(t_L, q_j | \lambda = 0) = \frac{A_L - q_j}{2}$ with L-type demand and $q_i(t_H, q_j | \lambda = 1) = \max(q_1(q_j), \frac{A_H - q_j}{2})$ with H-type demand in a separating PBE.

Next, we derive the separating PBE from the best responses. For the L-type demand, we obtain that $q_F^* = q_E^* = \frac{A_L}{3}$ from solving $q_F = \frac{A_L - q_E}{2}$ and $q_E = \frac{A_L - q_F}{2}$, For the H-type demand, there are two cases. First, from solving best responses $q_F = q_1(q_E)$ and $q_E = q_1(q_F)$, we obtain that $q_F^* = q_E^* = \frac{A_L}{18}(6 - 4\alpha + 4\alpha\theta - \sqrt{-36 + (4\alpha\theta - 4\alpha + 6)^2})$ or $q_F^* = q_E^* = \frac{A_L}{18}(6 - 4\alpha + 4\alpha\theta + \sqrt{-36 + (4\alpha\theta - 4\alpha + 6)^2})$. Note that the fixed points can sustain as PBE only if $q_1(q_j) > \frac{A_H - q_j}{2}$ holds. This condition is equivalent to $q_F^* > \frac{A_H}{3}$ and $q_E^* > \frac{A_H}{3}$ when $q_F^* = q_E^*$. If $q_i^* = \frac{A_L}{18}(6 - 4\alpha + 4\alpha\theta - \sqrt{-36 + (4\alpha\theta - 4\alpha + 6)^2})$, then $q_i^* > \frac{A_H}{3}$ if and only if $\sqrt{-36 + (4\alpha\theta - 4\alpha + 6)^2} < (\theta - 1)(4\alpha - 6) < 0$, contradictory. If $q_i^* = \frac{A_L}{18}(6 - 4\alpha + 4\alpha\theta + \sqrt{-36 + (4\alpha\theta - 4\alpha + 6)^2})$, then $q_i^* > \frac{A_H}{3}$ if and only if $\sqrt{-36 + (4\alpha\theta - 4\alpha + 6)^2} > (\theta - 1)(4\alpha - 6) > 0$, i.e., $\theta < \frac{3}{(3-4\alpha)^+}$. Second, from best responses $q_F = \frac{A_H - q_E}{2}$ and $q_E = \frac{A_H - q_F}{2}$, we obtain that $q_F^* = q_E^* = \frac{A_H}{3}$. Note that the fixed point can sustain as PBE only if $q_1(q_j) \leq \frac{A_H - q_j}{2}$ holds. Since $q_F^* = q_E^* = \frac{A_H}{3}$, this condition is equivalent to $\frac{1}{3}(\theta - 1)((4\alpha - 3)\theta + 3) \leq 0$, i.e., $\theta \geq \frac{3}{(3-4\alpha)^+}$.

In summary, both firms invest $q_L^S = \frac{A_L}{3}$ for the L-type demand. For the H-type demand, both firms invest $q_H^S = \frac{A_H}{3}$ if $\theta \geq \frac{3}{(3-4\alpha)^+}$ and $q_H^S = \underline{q_H} > \frac{A_H}{3}$ if $\theta < \frac{3}{(3-4\alpha)^+}$ where

$$\underline{q_H} = \frac{A_L}{18}(6 - 4\alpha + 4\alpha\theta + \sqrt{-36 + (4\alpha\theta - 4\alpha + 6)^2}). \tag{3.A.21}$$

Note that the threshold $\frac{3}{(3-4\alpha)^+}$ increases in α. Moreover, condition $\theta < \frac{3}{(3-4\alpha)^+}$ is equivalent to $\alpha > \frac{3}{4} - \frac{3}{4\theta}$, which always holds when $\alpha \geq \frac{3}{4}$.

Correspondingly, the investor's valuation in the separating PBE is $v_i^* = \pi_i(t) = (A_t - q_t^S - q_t^S)q_t^S = (A_t - 2q_t^S)q_t^S$. We have obtained that the reasonable posterior belief should be $\lambda = 1$ for any $q_i \geq q_1(q_j)$ from strict dominance. Thus, the posterior belief to sustain the separating PBE is $\{(q_F, q_E) : q_F \geq q_1(q_E) \textit{ and } q_E \geq q_1(q_F)\}$ and $\lambda = 0$ otherwise, where function $q_1(\cdot)$ is shown in Equation (3.A.20). ■

3.9.3.3.2 Technical Details of Definition 7 We derive the pooling PBE. To derive the pooling PBE, we derive the firms' best responses when their type is concealed and find the fixed point(s). Let $i, j \in \{E, F\}$ and $i \neq j$ be the subscript for the firms. We focus our discussion on the best response of firm i, $q_i(q_j)$, given firm j's strategy q_j below for ease of illustration.

For a pooling response to exist, firm i's utility of both types at a pooling response candidate $q_i^P(q_j)$ should be higher than at the least-cost separating response, $q_i^{LS}(q_j)$ for the L-type and $q_i^{HS}(q_j)$ for the H-type, given firm j's decision q_j, i.e.,

$$U_i(t_H, q_i^P(q_j), q_j|\lambda = h) \geq U_i(t_H, q_i^{HS}(q_j), q_j|\lambda = 1), \tag{3.A.22}$$

$$U_i(t_L, q_i^P(q_j), q_j|\lambda = h) \geq U_i(t_L, q_i^{LS}(q_j), q_j|\lambda = 0). \tag{3.A.23}$$

Note that inequality (3.A.22) holds only if $q_i^{HS}(q_j) = q_1(q_j)$ and $q_i^P(q_j) < q_1(q_j)$ due to concavity of the utility function. We further find that inequality (3.A.22) implies (3.A.23). Note that

$$\begin{aligned}
&U_i(t_L, q_i^P(q_j), q_j|\lambda = h) - U_i(t_L, q_1(q_j), q_j|\lambda = 1)\\
=&(\alpha\mu + (1-\alpha)A_L - q_i^P(q_j) - q_j)q_i^P(q_j) - (\alpha A_H + (1-\alpha)A_L - q_1(q_j) - q_j)q_1(q_j)\\
=&(\alpha\mu + (1-\alpha)A_H - q_i^P(q_j) - q_j)q_i^P(q_j) - (\alpha A_H + (1-\alpha)A_H - q_1(q_j) - q_j)q_1(q_j)\\
&+ (1-\alpha)(A_H - A_L)(q_1(q_j) - q_i^P(q_j))\\
=&U_i(t_H, q_i^P(q_j), q_j|\lambda = h) - U_i(t_H, q_1(q_j), q_j|\lambda = 1) + (1-\alpha)(A_H - A_L)(q_1(q_j) - q_i^P(q_j)).
\end{aligned}$$

Thus, inequality (3.A.22) implies that $U_i(t_H, q_i^P(q_j), q_j|\lambda = h) \geq U_i(t_H, q_1(q_j), q_j|\lambda = 1)$ and $q_1(q_j) > q_i^P(q_j)$, i.e., inequality (3.A.23).

When multiple pooling responses that satisfy inequalities (3.A.22) and (3.A.23) exist, we choose the lexicographically maximum sequential pooling response(s). In our setting, the lexicographically maximum sequential pooling response would be the maximizer of the utility function under H-type demand, i.e., $q_i^P(q_j) = q_2(q_j)$ where

$$q_2(q_j) = \frac{\alpha\mu + (1-\alpha)A_H - q_j}{2}. \tag{3.A.24}$$

Note that $q_2(q_j) < \frac{A_H - q_j}{2} < q_1(q_j)$. In the technical details of Definition 6, we have derived that the reasonable posterior belief is $\lambda = 1$ for any $q_i(q_j) \geq q_1(q_j)$. To construct monotonically consistent posterior belief, we impose that the reasonable posterior belief for the remaining region should be $\lambda = h$ for any $q_2(q_j) \leq q_i(q_j) < q_1(q_j)$, and $\lambda = 0$ for any $q_i(q_j) < q_2(q_j)$. Since $q_2(q_j)$ is the H-type firms' first-best pogoing response, and the L-type's utility under pooling decreases in $q_i(q_j)$ for any $q_2(q_j) \leq q_i(q_j) < q_1(q_j)$, the lexicographically maximum sequential pooling response $q_2(q_j)$ will emerge as the pooling best response for both types under this posterior belief.

Next, we derive the pooling LMSE by solving from the best responses $q_F = q_2(q_E)$ and $q_E = q_2(q_F)$, and we obtain that $q_F^* = q_E^* = q^P$, where

$$q^P = \frac{\alpha\mu + (1-\alpha)A_H}{3}. \tag{3.A.25}$$

Since $A_L < \mu = hA_H + (1-h)A_L < A_H$, we obtain $\frac{A_L}{3} < q^P < \frac{A_H}{3}$ and q^P decreases in α. Correspondingly, the investor's valuation in the pooling PBE is $v_i^* = h\pi_i(t_H) + (1-h)\pi_i(t_L) = h(A_H - q^P - q^P)q^P + (1-h)(A_L - q^P - q^P)q^P = (\mu - 2q^P)q^P$. The posterior belief to sustain this pooling PBE is $\lambda = 1$ for $\{(q_F, q_E) : q_F \geq q_1(q_E)\ and\ q_E \geq q_1(q_F)\}$, $\lambda = h$ for $\{(q_F, q_E) : q_2(q_E) \leq q_F < q_1(q_E)\ and\ q_2(q_F) \leq q_E < q_1(q_F)\}$, and $\lambda = 0$ otherwise, where function $q_1(\cdot)$ is shown in Equation (3.A.20) and function $q_2(\cdot)$ is shown in Equation (3.A.24).

For q^P to emerge as the pooling PBE, the following conditions must hold.

$$q_1(q^P) > \frac{A_H - q^P}{2} \tag{3.A.26}$$

$$U_i(t_H, q^P, q^P|\lambda = h) \geq U_i(t_H, q_1(q^P), q^P|\lambda = 1), \tag{3.A.27}$$

$$U_i(t_H, q^P, q^P|\lambda = h) \geq U_i(t_H, q_H^S, q_H^S|\lambda = 1), \tag{3.A.28}$$

$$U_i(t_L, q^P, q^P|\lambda = h) \geq U_i(t_L, q_L^S, q_L^S|\lambda = 0). \tag{3.A.29}$$

Conditions (3.A.26) and (3.A.27) describe that either firm has no incentive to deviate from q^P given the other firm's decision is q^P, which are derived from conditions (3.A.22) and (3.A.23). Conditions (3.A.28) and (3.A.29) describe that the pooling PBE Pareto dominate the separating PBE shown in Definition 6. ■

3.9.3.3.3 Technical Details of Proposition 10 If $\alpha \leq \overline{\alpha_{SD}}$, the L-type firm invests $\frac{A_L}{3}$ and the H-type firm invests $\frac{A_H}{3}$, and hence no operational distortion emerges. If $\overline{\alpha_{SD}} < \alpha < \overline{\alpha_{PD}}$, $q_L^S = \frac{A_L}{3}$ and $q_H^S = \underline{q_H}$ as shown in Eq. (3.A.21). Since $\frac{\partial \underline{q_H}}{\partial \alpha} = 4(\theta - 1) + \frac{2(\theta-1)(2\alpha(\theta-1)+3)}{\sqrt{(\theta-1)\alpha(\alpha(\theta-1)+3)}} > 0$, q_H^S increases in α. If $\alpha \geq \overline{\alpha_{PD}}$, q^P is shown in Eq. (3.A.25). Since $\frac{\partial q^P}{\partial \alpha} = \mu - A_H < 0$, q^P decreases in α.

3.9.3.3.4 Technical Details of Proposition 11 We denote each firm's profit in the H-type demand and L-type demand when $\alpha = 0$ as π_H^0 and π_L^0

First, we evaluate each firm's profit when they choose the same capacity investment level, i.e., when $q_i = q$. In this case, $\pi_i(t) = (A_t - 2q)q$. Thus, the profit maximizer is $q^* = \frac{A_H}{4}$ for the H-type and $q^* = \frac{A_L}{4}$ for the L-type. The expected profit $\pi_W = h\pi_H + (1-h)\pi_L = (\mu - 2q)q$, and hence the maximizer is $q^* = \frac{\mu}{4}$ for the expected profit.

Second, we evaluate each firm's profit if $\alpha \leq \overline{\alpha_{SD}}$. In this costless separating region, the L-type firm's profit equals $\pi_L^0 = \frac{A_L^2}{9}$ and the H-type firm's profit equals $\pi_H^0 = \frac{A_H^2}{9}$.

Third, we evaluate each firm's profit if $\overline{\alpha_{SD}} < \alpha < \overline{\alpha_{PD}}$. In this costly separating region, since $\underline{q_H} > \frac{A_H}{3} > \frac{A_H}{4}$, we obtain that $\pi_i(t_H, \underline{q_H}, \underline{q_H}) < \pi_i(t_H, \frac{A_H}{3}, \frac{A_H}{3}) = \pi_H^0$. This indicates that if $\theta < \frac{3}{(3-4\alpha)^+}$, the H-type firms' profits are smaller than π_H^0 in this separating region. Moreover, since $\underline{q_H}$ increases in α and π_i decreases in $\underline{q_H}$, a larger α will result in lower profits for the H-type firms. The L-type firm's profit equals $\pi_L^0 = \frac{A_L^2}{9}$.

Fourth, we evaluate each firm's profit if $\alpha \geq \overline{\alpha_{PD}}$. In this pooling region, for the L-type firm, since $q^P > \frac{A_L}{3} > \frac{A_L}{4}$, we obtain that $\pi_i(t_L, q^P, q^P) < \pi_i(t_L, \frac{A_L}{3}, \frac{A_L}{3}) = \pi_L^0$. Moreover, since q^P decreases in α and π_i decreases in q^P, a larger α will result in higher profits for the L-type. For the H-type firms, $q^P - \frac{A_H}{4} = (\frac{\alpha(h\theta+(1-h))+(1-\alpha)\theta}{3} - \frac{\theta}{4})A_L$. Thus, there exists a threshold $\alpha_0 = \frac{\theta}{4(1-h)(\theta-1)}$ such that $q^P \geq \frac{A_H}{4}$ if and only if $\alpha \leq \alpha_0$ and $q^P < \frac{A_H}{4}$ if and only if $\alpha > \alpha_0$. Note that α_0 increases in h and decreases in θ. Thus, (i) if $\alpha \leq \alpha_0$, $\frac{A_H}{4} \leq q^P < \frac{A_H}{3}$, and hence $\pi_i(t_H, q^P, q^P) \geq \pi_i(t_H, \frac{A_H}{3}, \frac{A_H}{3})$. (ii) If $\alpha > \alpha_0$, $q^P < \frac{A_H}{4}$, and we obtain that $(\frac{A_H}{4} - q^P) - (\frac{A_H}{3} - \frac{A_H}{4}) = \frac{2(1-h)(\theta-1)\alpha-\theta}{6}$. Thus, we obtain that $(\frac{A_H}{4} - q^P) \leq (\frac{A_H}{3} - \frac{A_H}{4})$, i.e., $\pi_i(t_H, q^P, q^P) \geq \pi_i(t_H, \frac{A_H}{3}, \frac{A_H}{3})$, if and only if $\alpha \leq 2\alpha_0$. In summary, in a pooling PBE, $\pi_i(t_H, q^P, q^P) \geq \pi_i(t_H, \frac{A_H}{3}, \frac{A_H}{3})$ if and only if $\alpha \leq \alpha_1$ where $\alpha_1 = \frac{\theta}{2(1-h)(\theta-1)}$. Since α_1 increases in h and decreases in θ, the case when firms obtain higher profits under pooling compared to no short-termism case is more likely with the increase of h and the decrease of θ. Moreover, since $q^P \geq \frac{A_H}{4}$ if and only if $\alpha \leq \alpha_0$ and q^P decreases in α, we obtain that $\pi_i(t_H, q^P, q^P)$ increases in α for $\alpha \leq \alpha_0$ and decreases in α for $\alpha > \alpha_0$. The H-type firms would obtain the monopoly profit when $\alpha = \alpha_0$ conditioning on that $\alpha_0 \leq 1$, i.e., $\theta \geq \frac{4(1-h)}{4(1-h)-1}$. We note that for the H-type firm, α_0 is the threshold for monotonicity and α_1 is the threshold for comparison to the zero short-termism case in the pooling region. Since $\alpha_1 = 2\alpha_0$, it is more likely to achieve higher profits in the entire pooling region than the zero short-termism case (i.e., $\alpha < \alpha_1$) compared to achieving a market-level monopoly profit (i.e., $\alpha < \alpha_0$) in the pooling region.

We further show the impact of α on the expected equilibrium profit π_W. In the costly separating region, since $\pi_H < \pi_H^0$ and $\pi_L = \pi_L^0$, we obtain that $\pi_W < \pi_W^0$. π_W decrease in α since π_H decreases in α. In the pooling region, plug in q^P from

Equation (3.A.25), we obtain that $A_L^2(\pi_W - \pi_W^0) = f = \frac{A_L^2}{9}(1-h)(\theta-1)(-2(\theta-1)(1-h)\alpha^2 + ((4-3h)\theta-(3-3h))\alpha + 1 - 2\theta)$. Note that $-2(\theta-1)(1-h) < 0$, $(4-3h)\theta-(3-3h) > 0$, and $1-2\theta < 0$. Thus, $f > 0$ if and only if $\alpha > \alpha^- > 0$ when $\Delta_f = 1 + 9h^2(\theta-1)^2 - 2h(5-9\theta+4\theta^2) > 0$, where $\alpha^- = \frac{-3-3h(\theta-1)+4\theta+\sqrt{\Delta_f}}{4(\theta-1)(1-h)}$. In a numerical analysis, we find that $\alpha^- > 1$ holds (if $\Delta_f > 0$) in all the scenarios when h takes value from 0 to 1 with increment of 0.02 and θ takes value from 1 to 3 with increment of 0.02. Thus, we obtain that $\pi_W < \pi_W^0$ for our parameter space. We also observe that π_W increases in α in the pooling region or our parameter space.

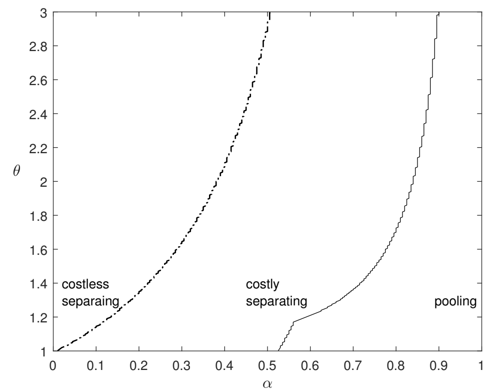
Figure 3.A.6 shows an example when the H-type firm can achieve the monopolistic collusion profit in the pooling region. We set $h = 0.45$ and $A_L = 1000$ in Figure 3.A.6 and $\theta = 2$ from Figure 3.A.6(b) to 3.A.6(f). In this example, $\overline{\alpha_S} = 0.375$ and $\overline{\alpha_P} = 0.84$. We show in Figure 3.A.6(d) that threshold $\alpha_0 = 0.91$, where the H-type firm invests at the monopolist collusion level, i.e., $\frac{A_H}{4} = 500$. The H-type firm's profit increases in α when $\alpha < 0.91$ and decreases in α when $\alpha \geq 0.91$. Beyond this curvature in the H-type firm's profit in the pooling region, we continue to observe the similar patterns in Figure 3.A.6 as in the Figures shown in Section 3.7.

■

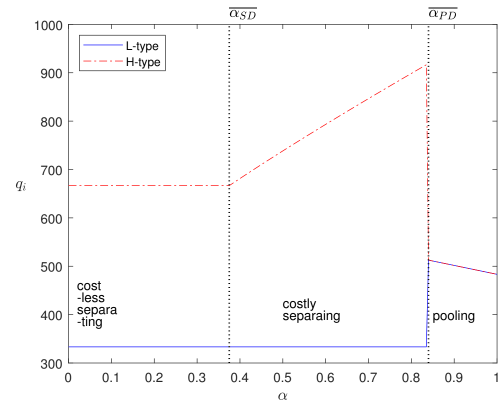
3.9.3.4 Future Research on Empirical Validation

It can be empirically challenging to validate the results generated by signaling models. We describe two potential strategies that future researchers may employ to close this gap. The first is the most straightforward – construct a set of lab experiments and randomly assign participants to the players and types described in our model. These experiments can be constructed such that the participant assignments, experimental setup, and payoffs reflect the players, model parameters, and utilities that are captured under each of the possible and predicted outcomes in our analysis. The experimental results can be compared to the predicted outcomes from the signaling model and statistical inferences can be drawn based on the likelihood of realizing those results relative to the other possible outcomes allowed for in the experimental design. This strategy has already been employed successfully in the OM literature. For instance, Schmidt and Buell [143] use lab experiments to validate the predictions generated by the signaling game model described in Schmidt et al. [144].

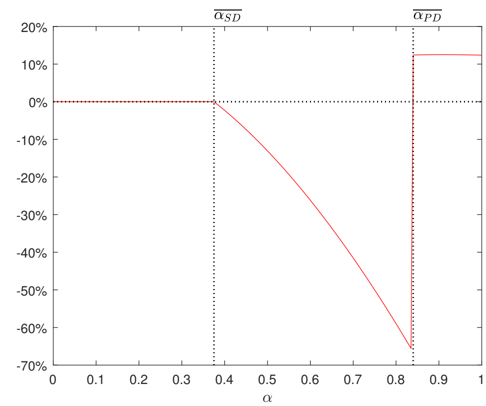
A second strategy is to develop a predictive empirical model using archival data. The objective is to test whether firms that have operated in situations reflected by our model, also behave in ways that are predicted by our model. An empirical analysis requires measures for each variable in our model –



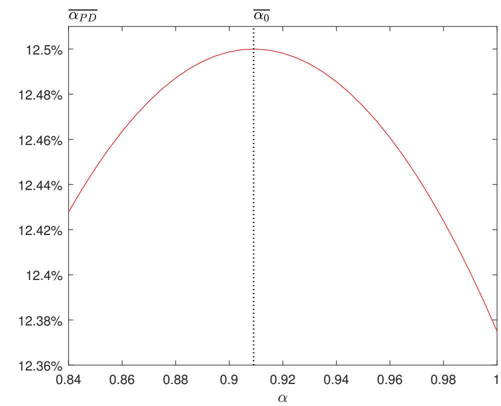
(a) Equilibrium segmentation



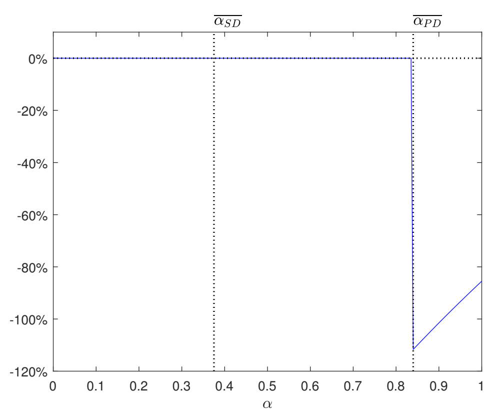
(b) Capacity investment



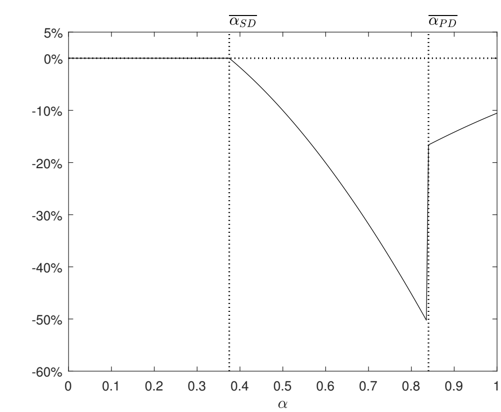
(c) H-type profit



(d) H-type profit pooling region zoomed in



(e) L-type profit



(f) Expected type profit

Figure 3.A.6: Impact of α on Equilibrium Segmentation, Capacity Investment, Profits in Cournot Competition

short-termism, abnormal investment level (i.e. over-investment and under-investment), the presence of competition, the presence of information asymmetry, the firm's long-term profit, the relative difference between high and low type market prospects, prior beliefs, and the firm's type. Although we are unaware of any empirical work that tests the combination of variables presented in our model, we have undertaken a broad literature review and identified a range of empirical works that develop and employ measures for many of the individual variables that we use.

There are a variety of variables in the academic literature to measure or detect the presence of short-termism. These include the sensitivity of the firm's share price to its current earnings per share [6], the presence (or lack) of long-term executive compensation plans [69], and the prevalence of accruals management [34]. Similarly, a variety of variables have been employed to measure abnormal investment levels, including abnormal inventory [103], abnormal inventory growth [102], abnormal research and development [138], and abnormal production [138]. Canonical measures of competition include the Herfindahl–Hirschman Index [80, 81] and the price-cost margin [131], although others have also been proposed [31]. Several empirical measures of information asymmetry have been employed in the literature, including market-to-book ratio, dispersion of analysts' earnings per share forecasts, and the bid-ask spread on the firms' securities [122, and cites therein]. Gomes and Phillips [73] compute the information asymmetry index as the average of the quintile ranking of a firm based on these six information asymmetry measures. Finally, common measures for the firm's long-term profit include economic value added [48], return on investment [32], and long-term abnormal stock returns [65].

We could not find established measures for the relative difference between high and low type market prospects (θ), prior beliefs, and the firm's type, but we do have some reasonable suggestions. One option to measure θ is to exploit the market size ranges provided by industry and securities analysts. The ratio of the top and bottom end of an analyst's range (or the averages across several analysts) can provide an estimate of θ. A measure of prior beliefs can be derived by taking the proportion of the number of buy-side analysts that rate the firm's stock a "buy" relative to the number of buy-side analysts that rate the firm's stock either a "buy" or "sell." Finally, the firm's type may be determined retrospectively based on the firm's long-term performance relative to analysts' consensus estimates.

The aforementioned measures can be derived from data available in the Institutional Brokers' Estimate System (IBES) database, the Center for Research in Security Prices (CRSP) database, and the Standard and Poor's COMPUSTAT database. Such an analysis is not, however, without its challenges. One notable challenge is that these databases aggregate data over an entire quarter (COMPUSTAT) and over many operating units of the firm (IBES, CRSP, and COMPUSTAT). This effectively represents the consolidation of many investment decisions over time and across the company. Our analytical model, however, reflects the anticipated actions of an operator on a single decision and in a single product market. It has been shown in other settings that empirical analyses using aggregated data can often mask results that would otherwise be available in more detailed data [for example, 46]. It may be that an empirical analysis in our setting will suffer a similar fate. This challenge can be overcome by collecting data on business-line or product-level decisions, or by focusing the analysis on firms that only have a single product.

BIBLIOGRAPHY

[1] Shipra Agrawal, Vashist Avadhanula, Vineet Goyal, and Assaf Zeevi. Mnl-bandit: A dynamic learning approach to assortment selection. *Operations Research*, 67(5):1453–1485, 2019.

[2] Gad Allon, Achal Bassamboo, and Eren B Çil. Large-scale service marketplaces: The role of the moderating firm. *Management Sci.*, 58(10):1854–1872, 2012.

[3] Krishnan S Anand and Manu Goyal. Strategic information management under leakage in a supply chain. *Management Science*, 55(3):438–452, 2009.

[4] Stephen Antczak. 11 things to know before becoming an Uber Or Lyft driver. *Forbes*, 2017. https://www.forbes.com/sites/nextavenue/2017/04/23/11-things-to-know-before-becoming-an-uber-or-lyft-driver/#48fc30c96579.

[5] Mark Armstrong. Competition in two-sided markets. *The RAND J. Economics*, 37(3):668–691, 2006.

[6] John Asker, Joan Farre-Mensa, and Alexander Ljungqvist. Corporate investment and stock market listing: A puzzle? *The Review of Financial Studies*, 28(2):342–390, 2014.

[7] Aspen Institute. Overcoming short-termism: A call for a more responsible approach to investment and business management, 2019. Accessed January 7, 2021, https://www.aspeninstitute.org/wp-content/uploads/files/content/docs/pubs/overcome_short_state0909_0.pdf.

[8] Claudia Assis. Tesla stock falls after report that gigafactory expansion is on hold. Accessed June 20, 2019, https://www.marketwatch.com/story/tesla-stock-falls-after-report-that-gigafactory-expansion-is-on-hold-2019-04-11, 2019.

[9] Ariel Fromer Babcock, Allen He, and Victoria Tellez. Driving the conversation: Long-term roadmaps for long-term success, 2019. *FCLT Global*; Accessed June 20, 2019, https://www.fcltglobal.org/resource/driving-the-conversation-long-term-roadmaps-for-long-term-success/.

[10] Jiaru Bai, Kut C So, Christopher S Tang, Xiqun Chen, and Hai Wang. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing Service Oper. Management*, 2018. forthcoming.

[11] Gah-Yi Ban and N Bora Keskin. Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity. *Available at SSRN 2972985*, 2020.

[12] Siddharth Banerjee, Sreenivas Gollapudi, Kostas Kollias, and Kamesh Munagala. Segmenting two-sided markets. *Proceedings of the 26th International Conference on World Wide Web*, pages 63–72, 2017.

[13] Dominic Barton. Capitalism for the long term. *Harvard Business Review*, 89(3):84–91, 2011.

[14] Dominic Barton, Jonathan Bailey, and Joshua Zoffer. Rising to the challenge of short-termism, 2016. *McKinsey Global Institute*, Accessed June 20, 2019, `https://www.fcltglobal.org/resource/rising-to-the-challenge-of-short-termism/`.

[15] Dominic Barton, James Manyika, and Timothy Koller. Measuring the economic impact of short-termism. *McKinsey Global Institute*, 2017. Accessed June 20, 2019, `https://www.fcltglobal.org/resource/measuring-the-economic-impact-of-short-termism/`.

[16] Hamsa Bastani. Predicting with proxies: Transfer learning in high dimension. *Management Science*, 2020.

[17] Hamsa Bastani and Mohsen Bayati. Online decision-making with high-dimensional covariates. *Forthcoming in Operations Research*, 2015.

[18] Hamsa Bastani and Mohsen Bayati. Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294, 2020.

[19] Lucian Arye Bebchuk and Lars A Stole. Do short-term objectives lead to under-or overinvestment in long-term projects? *The Journal of Finance*, 48 (2):719–729, 1993.

[20] David A Becher, Terry L Campbell II, and Melissa B Frye. Incentive compensation for bank directors: The impact of deregulation. *The Journal of Business*, 78(5):1753–1778, 2005.

[21] Elena Belavina, Simone Marinesi, and Gerry Tsoukalas. Rethinking crowdfunding platform design: mechanisms to deter misconduct and improve efficiency. *Management Science*, 66(11):4980–4997, 2020.

[22] Saif Benjaafar, Jian-Ya Ding, Guangwen Kong, and Terry Taylor. Labor welfare in on-demand service platforms. Working paper, University of Minnesota, 2018.

[23] Saif Benjaafar, Guangwen Kong, Xiang Li, and Costas Courcoubetis. Peer-to-peer product sharing: Implications for ownership, usage, and social welfare in the sharing economy. *Management Sci.*, 2018. forthcoming.

[24] Daniel Bergstresser and Thomas Philippon. CEO incentives and earnings management. *Journal of Financial Economics*, 80(3):511–529, 2006.

[25] Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044, 2020.

[26] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Optimization in online content recommendation services: Beyond click-through rates. *Manufacturing & Service Operations Management*, 18(1):15–33, 2016.

[27] Omar Besbes, Francisco Castro, and Ilan Lobel. Surge pricing and its spatial supply response. Working paper, Columbia University, 2018.

[28] Kostas Bimpikis, Ozan Candogan, and Daniela Saban. Spatial pricing in ride-sharing networks. Working paper, Stanford University, 2018.

[29] Kostas Bimpikis, Ozan Candogan, and Daniela Saban. Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769, 2019.

[30] John M Bizjak, James A Brickley, and Jeffrey L Coles. Stock-based incentive compensation and investment behavior. *Journal of Accounting and Economics*, 16(1-3):349–372, 1993.

[31] Jan Boone. A new way to measure competition. *The Economic Journal*, 118 (531):1245–1261, 2008.

[32] Richard A Brealey, Stewart C Myers, and Alan J Marcus. *Fundamentals of Corporate Finance*. McGraw-Hill/Irwin series in finance, insurance, and real estate. McGraw-Hill Education, 2015.

[33] Francois Brochet, George Serafeim, and Maria Loumioti. Short-termism: Don't blame investors. *Harvard Business Review* (June issue), 2012.

[34] Francois Brochet, Maria Loumioti, and George Serafeim. Speaking of the short-term: Disclosure horizon and managerial myopia. *Review of Accounting Studies*, 20(3):1122–1163, 2015.

[35] Stephen Bryan, Lee-Seok Hwang, and Steven Lilien. CEO compensation after deregulation: The case of electric utilities. *The Journal of Business*, 78 (5):1709–1752, 2005.

[36] Brian J Bushee. The influence of institutional investors on myopic R&D investment behavior. *Accounting Review*, pages 305–333, 1998.

[37] Business Roundtable. Business roundtable redefines the purpose of a corporation to promote 'an economy that serves all Americans', 2019. Accessed August 20, 2019, `https://www.businessroundtable.org`.

[38] Gérard P Cachon and Martin A Lariviere. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Science*, 47 (5):629–646, 2001.

[39] Gerard P Cachon, Kaitlin M Daniels, and Ruben Lobel. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing Service Oper. Management*, 19(3):368——384, 2017.

[40] Andrew C Call, Shuping Chen, Bin Miao, and Yen H Tong. Short-term earnings guidance and accrual-based earnings management. *Review of Accounting Studies*, 19(2):955–987, 2014.

[41] Dennis Carey, Brian Dumaine, Michael Useem, and Rodney Zemmel. Why CEOs should push back against short-termism. *Harvard Business Review* (May 31), 2018.

[42] Felipe Caro and Jérémie Gallien. Inventory management of a fast-fashion retail network. *Operations Research*, 58(2):257–273, 2010.

[43] CBSNews. Uber, valued at $62 billion, still loses money on its rides. *CBSNews*, 2018. `https://www.cbsnews.com/news/uber-valued-at-62-billion-still-loses-money-on-its-rides/`.

[44] Soudipta Chakraborty and Robert Swinney. Signaling to the crowd: Private quality information and rewards-based crowdfunding. *Manufacturing & Service Operations Management*, 23(1):155–169, 2021.

[45] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[46] Li Chen and Hau L Lee. Bullwhip effect measurement and its implications. *Operations Research*, 60(4):771–784, 2012.

[47] M Keith Chen and Michael Sheldon. Dynamic pricing in a labor market: Surge pricing and flexible work on the Uber platform. Working paper, University of California, Los Angeles, 2015.

[48] Shimin Chen and James L Dodd. Economic value added (eva): An empirical examination of a new corporate performance measure. *Journal of Managerial Issues*, pages 318–333, 1997.

[49] Xi Chen, Zachary Owen, Clark Pixton, and David Simchi-Levi. A statistical learning approach to personalization in revenue management. *Forthcoming, Management Science*, 2020.

[50] Ying-Ju Chen, Tinglong Dai, C Gizem Korpeoglu, Ersin Körpeoğlu, Ozge Sahin, Christopher S Tang, and Shihong Xiao. Innovative online platforms: Research opportunities. 2018. Working Paper, Johns Hopkins University.

[51] Yiwei Chen and Ming Hu. Pricing and matching with forward-looking buyers and sellers. Working paper, University of Toronto, 2018.

[52] Qiang Cheng and Terry D Warfield. Equity incentives and earnings management. *The Accounting Review*, 80(2):441–476, 2005.

[53] Danielle Chesebrough and Rory Sullivan. Coping, shifting, changing 2.0: Corporate and investor strategies for managing market short-termism, 2017. *United Nations Global Compact*, Accessed January 7, 2021, `https://www.unglobalcompact.org/library/5421`.

[54] Jiri Chod, Nikolaos Trichakis, and Gerry Tsoukalas. Supplier diversification under buyer risk. *Management Science*, 65(7):3150–3173, 2019.

[55] Jiri Chod, Nikolaos Trichakis, Gerry Tsoukalas, Henry Aspegren, and Mark Weber. On the financing benefits of supply chain transparency and blockchain adoption. *Management Science*, 66(10):4378–4396, 2020.

[56] John C Coffee Jr and Darius Palia. The wolf at the door: The impact of hedge fund activism on corporate governance. *Journal of Corporation Law*, 41:545, 2015.

[57] Maxime C Cohen and Renyu Philip Zhang. Competition and coopetition for two-sided platforms. Working paper, New York University, 2018.

[58] Anthony J Crawford, John R Ezzell, and James A Miles. Bank CEO pay-performance relations and the effects of deregulation. *Journal of Business*, 68(2):231–256, 1995.

[59] Yao Cui, A. Yesim Orhun, and Ming Hu. Under the same roof: Value of shared living in airbnb. Working paper, Cornell University, 2019.

[60] Zoë Cullen and Chiara Farronato. Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Forthcoming, Management Science*, 2020.

[61] Trey Cummings and Anne Marie Knott. Outside CEOs and innovation. *Strategic Management Journal*, 39(8):2095–2119, 2018.

[62] Vicente Cuñat and Maria Guadalupe. Executive compensation and competition in the banking and financial sectors. *Journal of Banking & Finance*, 33(3):495–504, 2009.

[63] Jamie Dimon and Warren Buffet. Short-termism is harming the economy. *The Wall Street Journal*, June 7, 2018; Dow Jones, Inc., 2018.

[64] Jing Dong and Rouba Ibrahim. Managing supply in the on-demand economy: Flexible workers, full-time employees, or both? *Operations Research*, 2020.

[65] Allan C Eberhart, William F Maxwell, and Akhtar R Siddique. An examination of long-term abnormal stock returns and operating performance following R&D increases. *The Journal of Finance*, 59(2):623–650, 2004.

[66] Pnina Feldman, Andrew E Frazelle, and Robert Swinney. Can delivery platforms benefit restaurants? 2019. Working paper, Boston University.

[67] Guiyun Feng, Guangwen Kong, and Zizhuo Wang. We are on the way: Analysis of on-demand ride-hailing systems. Working paper, University of Minnesota, 2018.

[68] Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2016.

[69] Caroline Flammer and Pratima Bansal. Does a long-term orientation create value? Evidence from a regression discontinuity. *Strategic Management Journal*, 38(9):1827–1847, 2017.

[70] Robert Gibbons. *Game theory for applied economists*. Princeton University Press, 1992.

[71] Chloe Kim Glaeser, Marshall Fisher, and Xuanming Su. Optimal retail location: Empirical methodology and application to practice. *Manufacturing & Service Operations Management*, 21(1):86–102, 2019.

[72] Negin Golrezaei, Hamid Nazerzadeh, and Paat Rusmevichientong. Real-time optimization of personalized assortments. *Management Science*, 60 (6):1532–1551, 2014.

[73] Armando Gomes and Gordon Phillips. Why do public firms issue private and public securities? *Journal of Financial Intermediation*, 21(4):619–658, 2012.

[74] John R Graham, Campbell R Harvey, and Shiva Rajgopal. The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, 40(1):3–73, 2005.

[75] Thomas S Gruca and D Sudharshan. A framework for entry deterrence strategy: The competitive environment, choices, and consequences. *Journal of Marketing*, 59(3):44–55, 1995.

[76] Harish Guda and Upender Subramanian. Your Uber is arriving: managing on-demand workers through surge pricing, forecast communication and worker incentives. *Management Sci.*, 65(5):1995—-2014, 2019.

[77] Itai Gurvich, Martin Lariviere, and Antonio Moreno. Operations in the on-demand economy: Staffing services with self-scheduling capacity. Technical report, Northwestern University, 2016.

[78] Albert Y Ha and Shilu Tong. Contracting and information sharing under supply chain competition. *Management Science*, 54(4):701–715, 2008.

[79] Albert Y Ha, Shilu Tong, and Hongtao Zhang. Sharing demand information in competing supply chains with production diseconomies. *Management Science*, 57(3):566–581, 2011.

[80] Orris C Herfindahl. Concentration in the U.S. steel industry. *Unpublished PhD. Dissertation, Columbia University*, 1950.

[81] Albert O Hirshman. The paternity of an index. *American Economic Review*, 54(5):761–762, 1964.

[82] Bengt Holmström. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):169–182, 1999.

[83] Dorothée Honhon, Vishal Gaur, and Sridhar Seshadri. Assortment planning and inventory decisions under stockout-based substitution. *Operations Research*, 58(5):1364–1379, 2010.

[84] Harold Hotelling. Stability in competition. In *The Collected Economics Articles of Harold Hotelling*, pages 50–63. Springer, 1990.

[85] Bin Hu, Ming Hu, and Han Zhu. Surge pricing and two-sided temporal responses in ride-hailing. Working paper, University of Texas at Dallas, 2018.

[86] Ming Hu and Yun Zhou. Price, wage and fixed commission in on-demand matching. Working paper, Unversity of Toronto, 2017.

[87] Ming Hu and Yun Zhou. Dynamic type matching. Working paper, University of Toronto, 2018.

[88] Ellen Huet. Uber tests taking even more from its drivers with 30% commission. *Forbes*, 2015. `https://www.forbes.com/sites/ellenhuet/2015/05/18/uber-new-uberx-tiered-commission-30-percent/#6b99b36143f6`.

[89] Shep Hyken. The gig economy opens the door for employment opportunities. *Forbes*, 2018. `https://www.forbes.com/sites/shephyken/2018/07/29/the-gig-economy-opens-the-door-for-employment-opportunities/#d2a91b776628`.

[90] Kate Isaacs, David Langstaff, and Russell Eisenstat. 4 ways CEOs can conquer short-termism. *Harvard Business Review* (Feb 24), 2017.

[91] Srikanth Jagabathula and Gustavo Vulcano. A partial-order-based model to estimate individual preferences using panel data. *Management Science*, 64(4):1609–1628, 2018.

[92] Baojun Jiang and Lin Tian. Collaborative consumption: Strategic and economic implications of product sharing. *Management Sci.*, 64(3):1171–1188, 2016.

[93] Alberto Jiménez-Valverde and Jorge M Lobo. Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica*, 31(3):361–369, 2007.

[94] Chen Jin, Kartik Hosanagar, and Senthil K Veeraraghavan. Do ratings cut both ways? Impact of bilateral ratings on platforms. 2019. Working paper, University of Pennsylvania.

[95] Justin M Johnson and Taghi M Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27, 2019.

[96] Ian T Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[97] Ata Kaban. Improved bounds on the dot product under random projection and random sign projection. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 487–496, 2015.

[98] Ashish Kabra, Elena Belavina, and Karan Girotra. Designing promotions to scale marketplaces. Working paper, INSEAD, Fontainebleau, France, 2016.

[99] Ashish Kabra, Elena Belavina, and Karan Girotra. Bike-share systems: Accessibility and availability. *Management Science*, 2019.

[100] Yash Kanoria and Daniela Saban. Facilitating the search for partners on matching platforms: Restricting agents' actions. Working paper, Stanford University, 2018.

[101] Steven N Kaplan and Bernadette A Minton. How has CEO turnover changed? *International review of Finance*, 12(1):57–87, 2012.

[102] Saravanan Kesavan and Vidya Mani. The relationship between abnormal inventory growth and future earnings for U.S. public retailers. *Manufacturing & Service Operations Management*, 15(1):6–23, 2013.

[103] Saravanan Kesavan, Vishal Gaur, and Ananth Raman. Do inventory and gross margin data improve sales forecasts for U.S. public retailers? *Management Science*, 56(9):1519–1533, 2010.

[104] Paul Klemperer. The competitiveness of markets with switching costs. *The RAND J. Economics*, 18(1):138–150, 1987.

[105] A Gürhan Kök and Marshall L Fisher. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55(6):1001–1021, 2007.

[106] Stacey R Kole and Kenneth M Lehn. Deregulation and the adaptation of governance structure: The case of the US airline industry. *Journal of Financial Economics*, 52(1):79–117, 1999.

[107] Guangwen Kong, Sampath Rajagopalan, and Hao Zhang. Revenue sharing and information leakage in a supply chain. *Management Science*, 59(3): 556–572, 2013.

[108] David M Kreps and Joel Sobel. Signalling. In Robert J. Aumann and Sergiu Hart, editors, *Handbook of Game Theory with Economic Applications*, volume 2, pages 849–867. Elsevier North-Holland, Amsterdam; Boston, 1992.

[109] Guoming Lai and Wenqiang Xiao. Supply chain information signaling and capital market. In *Handbook of Information Exchange in Supply Chain Management*, pages 215–233. Springer, 2017.

[110] Guoming Lai and Wenqiang Xiao. Inventory decisions and signals of demand uncertainty to investors. *Manufacturing & Service Operations Management*, 20(1):113–129, 2018.

[111] Guoming Lai, Laurens Debo, and Lin Nan. Channel stuffing with short-term interest in market value. *Management Science*, 57(2):332–346, 2011.

[112] Guoming Lai, Wenqiang Xiao, and Jun Yang. Supply chain performance under market valuation: An operational approach to restore efficiency. *Management Science*, 58(10):1933–1951, 2012.

[113] Joonkyum Lee, Vishal Gaur, Suresh Muthulingam, and Gary F Swisher. Stockout-based substitution and inventory planning in textbook retailing. *Manufacturing & Service Operations Management*, 18(1):104–121, 2016.

[114] Jun Li and Serguei Netessine. Higher market thickness reduces matching rate in online platforms: Evidence from a quasiexperiment. *Management Science*, 66(1):271–289, 2020.

[115] Lode Li. Information sharing in a supply chain with horizontal competition. *Management Science*, 48(9):1196–1212, 2002.

[116] Lincoln Liburd, Tim Koller, Werner Rehm, and Robert Palter. Straight talk for the long term, 2015. Accessed June 20, 2019, `https://www.fcltglobal.org/resource/straight-talk-for-the-long-term/`.

[117] R Duncan Luce. *Individual Choice Behavior*. John Wiley & Sons, Inc., 1959.

[118] Siddharth Mahajan and Garrett Van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351, 2001.

[119] George J. Mailath, Masahiro Okuno-Fujiwara, and Andrew Postlewaite. Belief-based refinements in signalling games. *Journal of Economic Theory*, 60(2):241–276, 1993.

[120] L. Roger Martin. Yes, short-termism really is a problem. *Harvard Business Review* (Oct 9), 2015.

[121] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[122] Pankaj K Maskara and Donald J Mullineaux. Information asymmetry and self-selection bias in bank loan announcement studies. *Journal of Financial Economics*, 101(3):684–694, 2011.

[123] Daniel McFadden. The measurement of urban travel demand. *Journal of Public Economics*, 3(4):303–328, 1974.

[124] Dev Mehta and Ann-Kristin Hamke. In-depth: B2B Ecommerce 2019. 2019.

[125] Sentao Miao and Xiuli Chao. Fast algorithms for online personalized assortment optimization in a big data regime. *Available at SSRN 3432574*, 2019.

[126] Tripp Mickle. Apple shares sink after iPhone suppliers lower outlooks, 2018. *Wall Street Journal* (November 12).

[127] Andrés Musalem, Marcelo Olivares, Eric T Bradlow, Christian Terwiesch, and Daniel Corsten. Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197, 2010.

[128] Chakravarthi Narasimhan, Purushottam Papatla, Baojun Jiang, Praveen K Kopalle, Paul R Messinger, Sridhar Moorthy, Davide Proserpio, Upender Subramanian, Chunhua Wu, and Ting Zhu. Sharing economy: Review of current research and future directions. *Customer Needs and Solutions*, 5(1-2):93–106, 2018.

[129] MP1985 Narayanan. Managerial incentives for short-term results. *The Journal of Finance*, 40(5):1469–1484, 1985.

[130] Afshin Nikzad. Thickness and competition in ride-sharing markets. Working paper, University of Southern California, 2018.

[131] Stanley I Ornstein. Empirical uses of the price-cost margin. *The Journal of Industrial Economics*, pages 105–117, 1975.

[132] Krishna Palepu, Victor Bernard, Paul Healy, and Erik Peek. *Business analysis and valuation: Text and cases*. Thomson, IFRS edition, 2007.

[133] Deuk Hee Park, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11):10059–10072, 2012.

[134] Alfred Rappaport. Beyond quarterly earnings, 2004. *The Wall Street Journal*; March 8, 2004; Dow Jones, Inc.

[135] Eric Rasmusen. Signal jamming and limit pricing: A unified approach. In Moriki Hosoe and Eric Rasmusen, editors, *Public Policy and Economic Analysis*. Kyushu University Press, 1997.

[136] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender Systems Handbook*, pages 1–35. Springer, 2011.

[137] Jean-Charles Rochet and Jean Tirole. Platform competition in two-sided markets. *J. European Economic Association*, 1(4):990–1029, 2003.

[138] Sugata Roychowdhury. Earnings management through real activities manipulation. *Journal of Accounting and Economics*, 42(3):335–370, 2006.

[139] Bernard Salanie. *The economics of contracts: A primer*. MIT Press, Cambridge, Mass., 2nd edition, 2005.

[140] Rachelle C Sampson and Yuan Shi. Are US firms becoming more short-term oriented? Evidence of shifting firm time horizons from implied discount rates, 1980–2013. *Strategic Management Journal*, 2020.

[141] Denis Sauré and Assaf Zeevi. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, 15(3): 387–404, 2013.

[142] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative filtering recommender systems. In *The Adaptive Web*, pages 291–324. Springer, 2007.

[143] William Schmidt and Ryan Buell. Experimental evidence of pooling outcomes under information asymmetry. *Management Science*, 63(5):1586–1605, 2017.

[144] William Schmidt, Vishal Gaur, Richard Lai, and Ananth Raman. Signaling to partially informed investors in the newsvendor model. *Production and Operations Management*, 24(3):383–401, 2015.

[145] Aarti Shahani. In the battle between Lyft and Uber, the focus is on drivers. *NPR*, 2016. https://www.npr.org/sections/alltechconsidered/2016/01/18/463473462/is-uber-good-to-drivers-it-s-relative.

[146] Noam Shamir and Hyoduk Shin. Public forecast information sharing in a market with competing supply chains. *Management Science*, 62(10):2994–3022, 2015.

[147] Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. Springer, 2011.

[148] Michael Sheldon. Income targeting and the ridesharing market. Working paper, University of Chicago, 2016.

[149] Gerry Shih. China taxi apps Didi Dache and Kuaidi Dache announce $6 billion tie-up. *Reuters*, 2015. https://www.reuters.com/article/us-china-taxi-merger/china-taxi-apps-didi-dache-and-kuaidi-dache-announce-6-billion-tie-up-idUSKBN0LI04420150214.

[150] Maggie Starvish. Managerial myopia: Why public companies underinvest in the future. *Forbes* (Feb 7), 2013.

[151] Jeremy C Stein. Takeover threats and managerial myopia. *Journal of Political Economy*, 96(1):61–80, 1988.

[152] Jeremy C Stein. Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The Quarterly Journal of Economics*, 104(4):655–669, 1989.

[153] Jeremy C. Stein. Agency, information and corporate investment. In George M. Constantinides, Milton Harris, and Rene M. Stulz, editors, *Handbook of the economics of finance*, volume 21, chapter 2, pages 110–163. Elsevier North-Holland, Amsterdam; Boston, 1st edition, 2003.

[154] Jeremy C Stein. Agency, information and corporate investment. In *Handbook of the Economics of Finance*, volume 1, pages 111–165. Elsevier, 2003.

[155] Daryll Strauss. The vicious cycle of Uber. *Medium*, 2016. https://medium.com/@daryll/the-vicious-cycle-of-uber-bb7f9db241bc.

[156] Xiaoyuan Su and Taghi M Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009.

[157] Lawrence Summers. Corporate long-termism is no panacea - but it is a start. *Financial Times* (Aug 9), 2015.

[158] Lawrence H. Summers. Is corporate short-termism really a problem? The jury's still out. *Harvard Business Review* (Feb 16), 2017.

[159] Terry A Taylor. On-demand service platforms. *Manufacturing Service Oper. Management*, 20(4):704–720, 2018.

[160] Terry A. Taylor. On-demand service platforms. *Manufacturing & Service Operations Management*, 20(4):704–720, 2018.

[161] Robert Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[162] Robin Van Meteren and Maarten Van Someren. Using content-based filtering for recommendation. In *Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop*, volume 30, pages 47–56, 2000.

[163] Gustavo Vulcano, Garrett Van Ryzin, and Wassim Chaar. Om practice—choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management*, 12(3):371–392, 2010.

[164] Gabriel Wildau. China anti-monopoly rules unlikely to stymie taxi app merger. *Financial Times*, 2015. `https://www.ft.com/content/d90d9af2-b67b-11e4-95dc-00144feab7de`.

[165] Alex Wilhelm. Understanding lyft's recent financial performance. *Crunchbase news*, 2018. `https://news.crunchbase.com/news/understanding-lyfts-recent-financial-perfomance/`.

[166] Dennis J Zhang, Hengchen Dai, Lingxiu Dong, Fangfang Qi, Nannan Zhang, Xiaofei Liu, Zhongyi Liu, and Jiang Yang. The long-term and spillover effects of price promotions on retailing platforms: Evidence from a large randomized experiment on Alibaba. *Management Science*, 66(6): 2589–2609, 2020.

[167] Hongtao Zhang. Vertical information exchange in a supply chain with duopoly retailers. *Production and Operations Management*, 11(4):531, 2002.

[168] Xiaomeng Zhang, Kathryn M Bartol, Ken G Smith, Michael D Pfarrer, and Dmitry M Khanin. CEOs on the edge: Earnings manipulation and stock-based incentive misalignment. *Academy of Management Journal*, 51(2):241–258, 2008.

[169] Xia Zhao, Ling Xue, and Fuqiang Zhang. Outsourcing competition and information sharing with asymmetrically informed suppliers. *Production and Operations Management*, 23(10):1706–1718, 2014.