The Amazon business model, the platform economy and executive compensation:

Three essays in search theory

Acknowledgements

Contents

List of Figures

List of Tables

1

Introduction

Why search frictions?

Economics is all about gains from trade. But before gains from trade can be realized, people must meet first. Search frictions characterize the barriers to meet. The existence of buyers and sellers, who can in principle agree on a price, is not sufficient for immediate transactions. Agents need to get involved in a costly search process to find matching partners, and ultimately must decide whether or not to trade now rather than betting for better trading opportunities in the future.

Most real-world transactions are characterized by these forms of imperfections or *search frictions* — consumers search for goods online and offline, workers search for vacancies, investors search for financial products in an exchange or over the counter, etc. Search frictions are derived from various sources, including imperfect information about trading partners, heterogeneous demand and supply, coordination failures, etc.

This thesis is among the intellectual efforts to explain the real world through the lens of search theory. However, I leave the comfort zone of search theory somewhat and use it to explain issues that are rarely touched by search theorists. In three chapters, we explore issues ranging from the Amazon business mode (Chapter 2 and Chapter 3) to executive compensation (Chapter 4).

Why does Amazon combine a middleman and a marketmaker?

Among millions of products available on Amazon, some are sold by Amazon itself, some are sold by so-called third-party sellers, and the majority are sold by both. This means, Amazon is a *middleman*, who specializes in buying and reselling products in its name, as well as a *market-maker*, who offers a marketplace (platform) for fees, where the participating buyers and sellers can search and trade with each other. We thus call Amazon a *Marketmaking Middleman*. It is not just Amazon that adopts this hybrid model. A similar business model has been observed in financial markets. For example, the New York Stock Exchange (NYSE) took an expanded platform "NYSE Arca" after a severe market share drop around 2008. In housing markets, the Trump Organization established a luxury residential real estate brokerage firm, competing with thousands of housing brokers in New York City.

Why do intermediaries use a hybrid mode? Why has the middleman sector or the platform not become the exclusive avenue of trade, despite the recent technological advancements? What determines the position of an intermediary's optimal mode in the spectrum spanning from a pure marketmaker mode to a pure middleman mode? These are the questions answered in Chapter 2 and Chapter 3.

Chapter 2 develops a directed search framework to explain these puzzles. In the framework, buyers and sellers can search for counter-parties either through an intermediated market which is operated by a monopolistic intermediary or via a decentralized market. The intermediary, who corresponds to a real-world hybrid-mode intermediary, e.g., Amazon, can make use of both a middleman and a marketmaker sector. The decentralized market represents an individual's outside option that creates competitive pressure on the overall intermediated market.

There exist search frictions in both the intermediated market and the decentralized market. It is not surprising to expect a search-frictional decentralized market. In this chapter, the decentralized market is modeled by random search. But for the intermediated market, one might argue that search frictions should vanish in platform-type intermediaries like Amazon who provide all sorts of search tools and price/capacity/review information. Indeed, search frictions in the traditional sense may fade, but coordination frictions always exist. Burdett et al. (2001) described such coordination frictions precisely: "*(Consider a market where) first sellers set prices, and then each buyer chooses which seller to visit. There is no search problem in the traditional sense because buyers know the price and the capacity of each seller with certainty. Still, in equilibrium, there is a chance that more buyers will show up at a given location than the seller can accommodate, in which case some customers get rationed; simultaneously, fewer buyers may show up at another location than the seller there can accommodate, in which case the seller gets rationed.*" This is the way we model frictions in the intermediated market.

With this framework, the intermediary faces the following trade-off between the middleman mode and market-maker mode. Compared to an individual seller, a middleman can hold a larger amount of inventory on the one hand, which reduces out-of-stock risk and delivers more transactions. On the other hand, activating a platform attracts sellers to trade in the intermediated market, and ultimately leads to fewer sellers available in the decentralized market. Buyers thus expect a lower outside value and they are willing to accept less favorable trading terms at the intermediated market. Accordingly, the intermediary can charge higher price/fees. In a nutshell, the intermediary has a trade-off between a larger transaction volume by operating as a middleman and a higher price/fee by acting as market-maker. This trade-off determines the optimal intermediation mode, and eventually a *marketmaking middleman* such as Amazon, who adopts a mixture of these two intermediation modes, can be profit-maximizing.

There are two roles that search frictions play in this trade-off. First, search frictions in the intermediated market highlight the advantage of middleman mode. A middleman with a much larger inventory can significantly decrease the coordination frictions — it can simply accommodate the uncertain demands from uncoordinated buyers. This is perhaps why Amazon invested heavily in building warehouses and expanding its network of delivery in the passing decades. Second, search frictions in the decentralized market highlight the advantage of the market-maker mode. Suppose we assume away search frictions in the decentralized market where there are abundant sellers. Then whatever the intermediated market structure is, with or without an active market-maker sector, buyers' outside value does not change. Hence, the trade-off between middleman and market-maker modes would not exist.

Our theory is not merely a thought experiment, it has strong real-world support. We examine the implications of our theory empirically. We take Amazon as the intermediated market and eBay as the decentralized market and collect data from both markets. For our chosen product category, Amazon acts as a marketmaking middleman: for 32% of the sample, Amazon acts as a middleman; for the other 68%, Amazon acts as a platform. Our empirical evidence strongly supports the model's prediction that Amazon is more likely to sell the product as a middleman when the chance of buyers to meet a seller in eBay is low, the buyers' bargaining power is low, or total demand is high.

Chapter 3 provides an important extension to the baseline model of Chapter 2: Competing intermediaries. We consider a Bertrand competition game between an incumbent intermediary who can mix a middleman mode and a marketmaker mode, and an entrant intermediary who is restricted to be a marketmaker. We find that the entrant faces the choice of being a second-source of intermediation service with high price/fees versus being a sole active source by undercutting the incumbent. However, either option would indicate a positive outside value for individuals, which determines the terms of trade an individual is willing to accept at the incumbent. Therefore, the trade-off about intermediation modes in Chapter 2 continues to hold for an intermediary (the incumbent) in a duopoly. We show that for a reasonable set of parameters, there exists an pure strategy equilibrium where a hybrid incumbent emerges. In this equilibrium, the optimal (incumbent) intermediation mode features a larger middleman when the chance of a buyer to meet a seller at the intermediary is low, or the total demand is high. We also show there exists a mixed strategy equilibrium where the incumbent activates its market-maker mode with positive probability. These analyses serve as theoretical robustness checks for Chapter 2.

Furthermore, the duopoly framework gives a number of new insights. First, we show that the intermediation structure of the incumbent is an equilibrium result, where the strategy of the entrant (or outside) intermediary also plays a role. In other words, what shapes the Amazon structure is not only Amazon but also other fee-setting intermediaries such as eBay. Second, the first-mover advantage of the incumbent does not necessarily leads to a higher market share. In the mixed strategy equilibrium, the entrant intermediary might be able to undercut and become the sole source for some category of goods. As a real-world example, the new clothing and fashion online platform Zalando can defeat incumbents like Amazon and eBay and becomes the leading online shop.

Why do larger firms pay executives more for performance?

Chapter 4 turns to another highly debated issue, the incentive compensation of top executives, e.g., CEO, CFO, president, etc. Executives are highly paid, with the majority of their wealth coming from performance-related rewards, including options and stocks. In this chapter, I aim to explain a newly documented empirical fact: The firm-size incentive premium. I show that executives' job ladder which stems from the search frictions in the managerial labor market has a point in explaining the firm-size incentive premium both qualitatively and quantitatively.

The firm-size incentive premium refers to the fact that the fraction of incentives in the executive compensation contract increases with firm size. This fact is based on the executive compensation data in S&P 1500 firms. The contract incentives are measured by pay-for-performance sensitivity, i.e., for one percent increase in firm value, how much wealth the executive receives from the compensation package (mainly through options and stocks). These incentives are believed to be necessary to motivate the executive's effort and align the interests with that of shareholders.

The chapter starts with a theoretical explanation. In a framework of on-the-job search, an executive is poached by outside firms. A dynamic incentive contract is designed to deal with both the moral hazard problem inside the firm and competition from the external executive labor market. The competition for executives increases total compensation, and more importantly, it generates a new source of incentives for executives to exert effort, called *labor market incentives*, which substitutes for performance-based incentives embedded in bonus, stocks, options, etc. In the extreme case that labor market incentives are so large, executives would take the effort (or have the aligned interest with shareholders) without any incentive pay in the contract.

I show that labor market incentives decrease with firm size. Why? There are two channels in the model. Here I would like to emphasize the *job ladder* channel. There is a job ladder where executives climb from small towards larger firms through job-to-job transitions. This is so in the data that job-hopping is prevalent and most job-hopping is towards larger firms. In the model, this happens as the executive auctions his/her labor to competing firms and the larger firm can bid higher (Postel-Vinay and Robin, 2002).

Now think of an executive in Amazon, who is probably at the top of the job ladder, and hence he receives very little incentives from the labor market. In contrast, an executive of Netflix should be lower on the job ladder, because he/she can be poached by larger firms including Amazon, so he/she has larger labor market incentives. Briefly speaking, the position on the job ladder determines how much labor market incentive one can get. Executives of larger firms are higher on the job ladder; hence they receive less from the labor market by climbing the job ladder. If so, then larger firms need to provide more incentives via the contract to assure the skin in the game. Therefore, the job ladder effect explains the firm-size incentive premium.

If search frictions are assumed away, this chapter would be a traditional career concern story where the executives are motivated to take the effort by his labor market perspective, and this motivation is identical for everyone with the same age (Gibbons and Murphy, 1992). Indeed, every executive (and every one of us) has a career concern. But the effect of search frictions clarifies that career concerns are heterogeneous, and are decreasing along the job ladder. Through the lens of search frictions, we see that even the corporate multi-millionaire class is layered, and this structure speaks to the firm-size incentive premium.

In the end of this chapter, I structurally estimate the theoretical model by the simulated method of moments. Then I simulate data based on the estimated model and predict the firm-size incentive premium using the simulated data. I find the prediction of the model is very close to the incentive premium in the real data. Since there is nothing mechanically to force them to be the same, this shows the model captures some essential feature in the real world.

If you can't beat them, join them.

proverb

2

Marketmaking Middlemen

2.1 Introduction

This chapter is based on the working paper co-authored by Pieter Gautier and Makoto Watanabe.[1] It develops a framework in which market structure is determined by the intermediation service offered to customers. There are two representative modes of intermediation that are widely used in real-life markets. In one mode, an intermediary acts as a *middleman* (or a *merchant*), who is specialized in buying and selling for his own account and typically operates with inventory holdings (e.g. supermarkets, traditional brick and mortar retailers, and dealers in financial and steel markets). In the other mode, an intermediary acts as a *marketmaker*, who offers a marketplace for fees, where the participating buyers and sellers can search and trade with each other

[1]We thank seminar and conference participants at U Essex, U Bern, U Zurich, Goergetown U, Albany, the Symposium on Jean Tirole 2014, the Search and Matching workshop in Bristol, SaM network annual conference 2015/2016 in Aix-en-Provence/Amsterdam, Toulouse School of Economics, Tokyo, Rome, the 2015 IIOC meeting in Boston, the EARIE 2015 in Munich, the 16th CEPR/JIE conference on Applied Industrial Organization, Workshop of the Economics of Platform in Tinbergen Institute, and the 2016 Summer Workshop on Money, Banking, Payments and Finance in Chicago FED for useful comments.

and at least one side of the market pays a fee for using the platform (e.g. auction sites, brokers in goods or financial markets, and many real estate agencies).

The market-making mode became more appropriate since new advanced Internet technology facilitated the use of online platforms in the late 1990s and early 2000s. In financial markets, an expanded platform sector is adopted in a specialist market, i.e., the New York Stock Exchange (NYSE),[2] and even in a typical dealers' (i.e., middlemen's) market, i.e the NASDAQ. In goods and service markets, the electronic retailer Amazon.com and the online hotel/travel reservation agency Expedia.com, who have been a pure middleman, also act as a marketmaker, by allowing other suppliers to participate in their platform as independent sellers. In housing markets, some entrepreneurs run a dealer company (developing and owning luxury apartments and residential towers) and a brokerage company simultaneously in the same market.

Common to all the above examples is that intermediaries operate both as a middleman and a marketmaker at the same time. This is what we call a *marketmaking middleman*. Hence, the first puzzle is to explain the emergence of marketmaking middlemen, i.e., why the middleman or the platform sector has not become the exclusive avenue of trade, despite the recent technological advancements.

We also observe considerable differences in the microstructure of trade in these markets. The NASDAQ is still a more "middlemen-based" market relative to the NYSE. While some intermediaries in housing markets are marketmaking middlemen, many intermediaries are brokers. Other online intermediaries, such as eBay and Booking.com, are pure marketmakers, who do not buy and sell on their own accounts, like Amazon.com and Expedia.com do. They solely concentrate on their platform business. So the second puzzle is to explain what determines the position of an intermediary's optimal mode in the spectrum spanning from a pure marketmaker mode to a pure middleman mode.

We consider a model in which the intermediated-market structure is determined endogenously as a result of the strategic choice of a monopolistic intermediary. In our

[2] In the finance literature, the following terminologies are used to classify intermediaries: brokers refer to intermediaries who do not trade for their own accounts, but act merely as conduits for customer orders, akin to our marketmakers; dealers refer to intermediaries who do trade for their own accounts, akin to our middlemen/merchants. The marketmakers (or specialists) in financial markets quote prices to buy or sell assets as well as take market positions, so they may correspond broadly to our market-making middlemen.

model, there are two markets open to agents, one is an intermediated market operated by the intermediary, and the other is a decentralized market where buyers and sellers search individually. The intermediated market combines two business modes: as a middleman, the intermediary is prepared to serve many buyers at a time by holding inventories; as a marketmaker, the intermediary offers a platform and receives fees. The intermediary can choose how to allocate the attending buyers among these two business modes.

We formulate the intermediated market as a directed search market in order to feature the intermediary's technology of spreading price and capacity information efficiently – using the search function offered in the NYSE Arca or Expedia/Amazon website or in the web-based platform for house hunters. For example, one can receive instantly all relevant information such as prices, the terms of trade and stocks of individual sellers. In this setting, each individual seller is subject to an inventory capacity of discrete units (normalized to one unit in the model), whereas the middleman is subject to an inventory capacity of a mass K. Naturally, the middleman is more efficient in matching demands with supplies in a directed search equilibrium. The decentralized market represents an individual seller's outside option that determines the lower bound of his market utility.

With this set up, we consider two situations, *single-market search* versus *multiple-market search*. Under single market search, agents have to choose which market to search in advance, either the decentralized market or the intermediated market. This implies that the intermediary needs to subsidize buyers with their expected value in the decentralized market, but once they participate, the intermediated market operates without fear of competitive pressure outside. Given that the middleman mode is more efficient in realizing transactions, the intermediary uses the middleman-mode exclusively when agents search in a single market.

When agents are allowed to search in multiple markets, attracting buyers becomes less costly compared to the single-market search case — the intermediary does not need to subsidize buyers to induce participation. However, the prices/fees charged in the intermediated market must be acceptable relative to the available option in the decentralized market. Otherwise, buyers and sellers can easily switch to the outside market. Thus, under multiple-market search, the outside option creates competitive pressure to the overall intermediated market. In deciding the optimal intermediation mode, the intermediary takes into account that a higher middleman capacity induces more buyers to buy from the middleman, and fewer buyers to search on the platform. This has two opposing effects on its profits. On the one hand, a higher capacity of the middleman leads to more transactions in the intermediated market, and consequently to larger profits. On the other hand, sellers are less likely to trade on a smaller-scaled platform and buyers are more likely to trade with a larger scaled middleman, so that more sellers are available when a buyer attempts to search in the decentralized market. Accordingly, buyers expect a higher value from the less tight decentralized market. This causes cross-markets feedback that leads to competitive pressure on the price/fees that the intermediary can charge, and a downward pressure on its profits. Hence, the intermediary trade-offs a larger quantity against lower price/fees to operate as a larger-scaled middleman. This trade-off determines the middleman's selling capacity and eventually the intermediation mode.

Single-market search may correspond to the traditional search technology for local supermarkets or brick and mortar retailers. Over the course of a shopping trip, consumers usually have to search, buy and even transport the purchased products during a fixed amount of time. Given the time constraint, they visit a limited number of shops — typically one supermarket — and appreciate the proximity provided by its inventory. In contrast, multi-market search is related to the advanced search technologies that are available in the digital economy. It allows the online-customers to search and compare various options easily. Multiple market search is also relevant in the market for durable goods such as housing or expensive items where customers are exposed to the market for a sufficiently long time to ponder multiple available options.

We show that a marketmaking middleman can emerge in a directed search equilibrium. The marketmaking middleman can outperform either extreme intermediation mode. Relative to a pure market-maker, its inventory holdings can reduce the out-of-stock risk, while relative to a pure middleman its platform can better exploit the surplus of intermediated trade. It is this trade-off that answers the two puzzles above. Somewhat surprisingly, our result suggests that an improvement in search technologies induces the intermediary to generate inefficiencies to improve profits. This occurs via the use of the frictional platform that generates unmatched buyers who then search again but are also unmatched in the frictional decentralized market.

We offer various extensions to our baseline model. First, we introduce non-linear matching functions in the decentralized market, which increases the profitability of middleman even with multi-market search. Second, we introduce the aggregate resource constraint and frictions in the wholesale market, which increases the profitability of using an active platform even with single-market search. Third, we introduce a convex inventory-holding cost function, which reduces the profitability of a middleman, and sellers' purchase/production costs that accrue prior to entering the platform, which reduce the profitability of marketmaker. However, these extensions do not alter our main insight on the emergence of marketmaking middlemen. Forth, we introduce competing intermediaries. As is consistent with the monopoly analysis, we show that an active platform of an incumbent intermediary that charges positive fees can only be profitable when agents search in multi markets and the other intermediary enters with an active platform.

Finally, we examine empirically the implication of our theory. Just like in the last extension of competing intermediaries, we take Amazon as the centralized market and eBay as the decentralized market. For our chosen product category, Amazon acts as a marketmaking middleman: for 32% of the sample, Amazon acts as a middleman; for the other 68%, Amazon acts as a platform. Our empirical evidence strongly supports the model's prediction that Amazon is more likely to sell the product as a middleman when the chance of buyers to meet a seller in eBay is low, the buyers' bargaining power is low, and total demand is high.

This paper is related to the literature of middlemen developed by Rubinstein and Wolinsky (1987).[3] Using a directed search approach, Watanabe(2010, 2018a, 2018b) provides a model of an intermediated market operated by middlemen with high inventory holdings. The middleman's high selling-capacity enables them to serve many buyers

[3]Rubinstein and Wolinsky (1987) show that an intermediated market can be active under frictions, when it is operated by middlemen who have an advantage in the meeting rate over the original suppliers. Given some exogenous meeting process, two main reasons have been considered for the middlemen's advantage in the rate of successful trades: a middleman may be able to guarantee the quality of goods (Biglaiser 1993, Li 1998), or to satisfy buyers' demand for a variety of goods (Shevchenko 2004). While these are clearly sound reasons for the success of middlemen, the buyers' search is modeled as an undirected random matching process, implying that the middlemen's capacity cannot influence buyers' search decisions in these models. See also Duffie et al. (2005), Lagos and Rocheteau (2009), Lagos et al. (2011), Weill (2007), Johri and Leach (2002), Masters (2007), Watanabe (2010), Wright and Wong (2014), Geromichalos and Jung (2018), Lagos and Zhang (2016), Awaya and Watanabe (2018b), Awaya and Watanabe (2018a), Nosal et al. (2015).

at a time, thus to lower the likelihood of stock-out, which generates a retail premium of inventories. This mechanism is adopted by the middleman in our model. Hence, if intermediation fees were not available, then our model would be a simplified version of Watanabe where we added an outside market. It is worth mentioning that in Watanabe(2010, 2018a, 2018b), the middleman's inventory is modeled as an indivisible unit, i.e., a positive integer, so that the middlemen face a non-degenerate distribution of their selling units as other sellers do. In contrast, here we model the inventory as a mass, assuming more flexible inventory technologies, so that the middleman faces a degenerate distribution of sales. This simplification allows us to characterize the middleman's profit-maximizing choice of inventory holdings — in Watanabe(2010, 2018b) the inventory level of middlemen is determined by aggregate demand-supply balancing, and in Watanabe (2018a) it is treated as an exogenous parameter. More recently, Holzner and Watanabe (2016) study a labor market equilibrium using a directed search approach to model a job-brokering service offered by Public Employment Agencies, but the choice of intermediation mode is not the scope of their paper.

Our paper is also related to the two-sided market literature.[4] The critical feature of a platform is the presence of a cross-group externality, i.e., the participants' expected gains from a platform depend positively on the number of participants on the other side of it. Caillaud and Jullien (2003) show that even when agents have a pessimistic belief on the intermediated market, the intermediary can make profits by using "divide-and-conquer" strategies, namely, subsidizing one group of participants in order to attract another group and extract the ensuing benefits. To be consistent with this literature, we develop an equilibrium with an intermediary based on similar pessimistic beliefs. Broadly speaking, if there were no middleman mode, our model would be a directed search version of Caillaud and Jullien (2003) in combination with a decentralized market. Further, our result that the intermediary sometimes induces agents to search more than they like is related to the idea of search diversion in Hagiu and Jullien (2011). They pursue this idea in a model of an information platform that has

[4]See, e.g. Rochet and Tirole (2003), Rochet and Tirole (2006), Caillaud and Jullien (2001), Caillaud and Jullien (2003), Rysman (2009), Armstrong (2006), Hagiu (2006), (Weyl, 2010). Related papers from other aspects can be found in Baye and Morgan (2001), Rust and Hall (2003), Parker and Van Alstyne (2005), Nocke et al. (2007), Galeotti and Moraga-González (2009), Loertscher and Niedermayer (2008), Edelman et al. (2015), Hagiu and Wright (2014), Condorelli et al. (2018), and Rhodes et al. (2017). Earlier contributions of this strand of literature are, e.g., Stahl (1988), Gehrig (1993), Yavaş (1994), Yavaş (1996), Spulber (1996), and Fingleton (1997). For platform studies emphasizing matching heterogeneity, see e.g., Bloch and Ryder (2000), Damiano and Li (2008) and De Fraja and Sákovics (2012).

superior information about the match between consumers and stores and that could direct consumers first to their least preferred store.

Rust and Hall (2003) develop a search model which features the coexistence of different intermediation markets.[5] They consider two types of intermediaries, one is "middlemen" whose market requires costly search and the other is a monopolistic "market maker" who offers a frictionless market. They show that agents segment into different markets depending on heterogeneous production costs and consumption values, thus these two types of intermediaries can coexist in equilibrium. Their model is very different from ours in many respects. For instance, selling capability and inventory do not play any role in their formulation of a search rule, but it is the key ingredient in our model. As Rust and Hall (2003) state: "An important function of intermediaries is to hold inventory to provide a buffer stock that offers their customers liquidity at times when there is an imbalance between supply and demand. In the securities business, liquidity means being able to buy or sell a reasonable quantity of shares on short notice. In the steel market, liquidity is also associated with a demand for immediacy so that a customer can be guaranteed of receiving shipment of an order within a few days of placement. *Lacking inventories and stock-outs, this model cannot be used to analyze the important role of intermediaries in providing liquidity* (page 401; emphasis added)." This is exactly what we emphasize in our model which incorporates Rust and Hall's observation. We show that intermediaries can pursue different types of intermediation modes even when faced with homogeneous agents.

The rest of the paper is organized as follows. Section 2 presents our model of intermediation and the benchmark case of single-market search. Section 3 extends the analysis to allow for multiple-market technologies and presents the key finding of our paper. Section 4 discusses modeling issues. Section 5 discusses some real-life applications of our theory. Section 6 presents the empirical evidence. Finally, section 7 concludes. Omitted proofs are in the Appendix I. Appendix II and III contain the extension to allow for unobservable capacity and participation fees, and additional details on the empirical analysis.

[5] See Ju et al. (2010) who extend the Rust and Hall model by considering oligopolistic market makers.

2.2 A basic model with single-market search

This section studies the choice of intermediation mode for single-market technologies that serves as a benchmark of our economy. Along the way, we introduce the environment in which the monopolistic intermediary operates.

2.2.1 The framework

Agents We consider a large economy with two populations, a mass B of buyers and a mass S of sellers. Agents of each type are homogeneous. Each buyer has unit demand for a homogeneous good, and each seller is able to sell one unit of that good. The consumption value for buyers is normalized to 1. Sellers can purchase the good from a wholesale market. We assume there exists a competitive wholesale market, and the demand of all suppliers is always satisfied with a price equal to the marginal cost $c \in [0, 1)$.[6]

Retail markets Buyers and sellers can only meet each other in a retail market. There are two retail markets, a centralized market, which is operated by a monopolistic intermediary, and a decentralized market, which gives the outside option for agents. Retail services can be exclusive or non-exclusive. We therefore consider two different search technologies of buyers/sellers. This section spells out the details of single-market search where buyers/sellers can attend only one market, and Section 2.3 is about multi-market search where agents can attend both markets sequentially. We next describe the trading mechanisms in each market.

Decentralized market The decentralized market (hereafter D market) is featured by random matching and bilateral bargaining. Suppose all buyers and sellers participate in D market, then a buyers finds a seller with probability λ^b and a seller finds a buyer with probability λ^s, satisfying $B\lambda^b = S\lambda^s, \lambda^b, \lambda^s \in (0, 1)$. If a subset of buyers $B^D \leq B$ and sellers $S^D \leq S$ participate, then the meeting probabilities become $\lambda^{b\prime} \equiv \lambda^b \frac{S^D}{S}$ and $\lambda^{s\prime} \equiv \lambda^s \frac{B^D}{B}$, respectively.[7] This matching technology, which is linear in the participants

[6] One can as well model producers in the wholesale market, but they are passive in our model. Without retail technologies producers are not able to serve buyers themselves but just supply the good in a competitive wholesale market. We show such a set-up in figure 2.1.

[7] To understand $\lambda^{b\prime}$, imagine that all sellers have a slot at D market, but not all would be available. A seller is absent from D market if he has sold the unit inventory in other markets. If the seller that a buyer is supposed to meet is not available, then the meet would fail. Given an mass of $S^D < S$ sells join D market, the probability a buyer finds an available seller is $\lambda^b \frac{S^D}{S}$. The same logic applies to $\lambda^{s\prime}$. It is easy to verify that $B^D \lambda^{b\prime} = S^D \lambda^{s\prime}$.

on the other side of the market, is a simplified way to formulate the outside option of agents. In Section 2.4.1, we show that our main insight is valid with general non-linear matching functions where the meeting rate (and the expected value) depends on the relative measures of buyers and sellers. Matched partners follow an efficient bargaining process, which yields a linear sharing of the total surplus, with a share of $\beta \in (0,1)$ for the buyer and a share of $1-\beta$ for the seller.

Centralized market The centralized market (hereafter C market) is operated by a monopolistic intermediary whose profit-maximizing mode is the focus of the model. The intermediary can perform two different intermediation activities. As a middleman, he purchases a good of mass $K \geq 0$ from the wholesale market at a cost c, and resells it to buyers at a price of $p^m \in [c,1]$. As a market-maker, he does not buy and sell but instead provides a platform where buyers and sellers can interact with each other for trade after paying fees. Denote the fees in the market-maker sector by $\{f^b, f^s\}$, where $f^b,\ f^s \in [0,1]$ is a transaction fee charged to a buyer or a seller, respectively, and satisfies $f^b + f^s \leq 1$. By choosing $\{K, p^m, f^b, f^s\}$, the intermediary fine-tunes the flows of buyers visiting the middleman and the market-maker sectors and to maximize his profits.[8]

Observing the intermediary's inventory K, price p^m, and fees $\{f^b, f^s\}$, buyers and sellers choose which market to participate. Suppose that a mass of $B^C \in (0,B]$ buyers and a mass of $S^C \in [0,S]$ sellers have decided to participate in the C market, then trade in C market follows a directed search game, which consists of the following two stages.[9]

1. In the first stage, all participating sellers with the unit selling-capacity simultaneously post a price which they are willing to sell at. Thanks to the advanced matching technology owned by the intermediary, the prices and capacities of all the individual suppliers (sellers and the middleman) are publicly observable within the C market.

2. Observing the price and capacity information, in the second stage, all buyers simultaneously decide which supplier to visit. As is standard in the literature, we

[8] Allowing for participation fees/subsidies, which accrue irrespective of transactions in the C market, will not affect our main result. In Appendix II, we offer such an extended model.

[9] The intermediary's announcement and the mass of participating agents are common knowledge.

assume that each buyer can visit one supplier, one of the sellers or the middleman.

Directed search equilibrium in C market Since buyers cannot coordinate their actions over which supplier to visit, C market is subject to coordination frictions. There is a chance that more buyers may show up at a given supplier than the supplier can accommodate, in which case some buyers get rationed (a supplier selects at random); at the same time, fewer buyers may show up at another supplier than the supplier can accommodate, in which case the supplier gets rationed. Therefore, we investigate a symmetric equilibrium where all individual sellers post the same price and all buyers use an identical mixed strategy to direct for a supplier given any configuration of the announced prices/fees. We do not consider equilibria in which buyers follow asymmetric strategies since this would require unrealistic implicit coordination among them. In the equilibrium, each individual seller (if any) has an expected queue x^s of buyers with an equilibrium price p^s and the middleman has an expected queue x^m of buyers with an equilibrium price p^m.

Since the middleman has a capacity of K, the expected profit from the middleman sector is given by $\min\{K, x^m\}p^m$. The expected value of buyers visiting the middleman is

$$V^m(x) = \min\{\frac{K}{x^m}, 1\}(1 - p^m),$$

where $\min\{\frac{K}{x^m}, 1\}$ is the matching probability of a buyer at the middleman.

The matching probability at each individual seller follows an urn-ball matching function. The number of buyers visiting an individual seller is a random variable, denoted by N, which follows a Poisson distribution, $\text{Prob}[N = n] = \frac{e^{-x}x^n}{n!}$, with an expected queue of buyers $x \geq 0$.[10] With limited selling capacity, each seller is able to serve only one buyer. A seller with an expected queue $x^s \geq 0$ has a probability $1 - e^{-x^s}$ ($= \text{Prob}[N \geq 1]$) of successfully selling, while each buyer has a probability $\eta^s(x^s) = \frac{1-e^{-x^s}}{x^s}$ of successfully buying. Hence, the expected value of a seller in the

[10] This is due to coordination frictions. Suppose there are b buyers and s sellers. If each buyer visit each seller with equal probability, any seller gets a buyer with probability $1 - (1 - \frac{1}{s})^b$. Taking the limit as b and s go to infinity and $x^s = b/s$ fixed, in a large market as we propose here, a fraction $1 - e^{-x^s}$ of sellers get a buyer. This process is called an urn-ball matching function. See the seminal work by Peters (1991, 2000).

platform with a price p^s and an expected queue x^s is given by

$$W(x^s) = x^s \eta^s(x^s)(p^s - f^s - c),$$

while the expected value of a buyer who visits the seller is

$$V^s(x^s) = \eta^s(x^s)(1 - p^s - f^b).$$

In the equilibrium, the expected queue length x^s and x^m should satisfy two requirements. The first requirement is a standard accounting identity,

$$S^C x^s + x^m = B^C, \tag{2.1}$$

which states that the number of buyers visiting individual sellers $S^C x^s$ and the middleman x^m should sum up to the total population of participating buyers B^C. The second requirement is that buyers search optimally:

$$x^m = \begin{cases} B^C & \text{if } V^m(B^C) \geq V^s(0) \\ (0, B^C) & \text{if } V^m(x^m) = V^s(x^s) \\ 0 & \text{if } V^m(0) \leq V^s(\frac{B^C}{S^C}), \end{cases} \tag{2.2}$$

where $V^i(x^i)$ is the equilibrium value of buyers in the C market of visiting a seller if $i = s$ and the middleman if $i = m$. Notice, the third case of (2.2) happens only if $S^C > 0$. Combining (2.1) and (2.2) gives the counterpart for $x^s \in [0, \frac{B^C}{S^C}]$. Based on the mass of buyers visiting middleman and marketmaker, we define the intermediation mode.

Definition 2.1 (Intermediation Mode)**.** *Suppose $B^C \in (0, B]$ buyers and $S^C \in [0, S]$ sellers participate in C market. Then, given the equilibrium search conditions (2.1) and (2.2), we say that the intermediary acts as:*

- *a pure middleman if $x^m = B^C$;*
- *a market-making middleman if $x^m \in (0, B^C)$;*
- *a pure market-maker if $x^m = 0$.*

Pessimistic beliefs and participating in C market Now we fold backward one more step and describe the participation rule under single-market search. That is, how B^C and S^C are determined. Consistent with Caillaud and Jullien (2003), we assume agents hold pessimistic beliefs on the participation decision of agents on the other side of the C market. Under such beliefs, buyers and sellers coordinate on a participation rule so that C market is empty whenever possible. In the framework of Caillaud and Jullien (2003), a divide-and-conquer strategy is the only way to get a positive market share out of pessimistic beliefs. In our model, however, even without the participation of sellers, there is supply in the C market — the inventory K of the middleman. Therefore, to break the pessimistic beliefs, the intermediary must convince buyers that if they do join C market, the expected value is higher than if they join D market. That is,[11]

$$\max\{V^m(x^m), V^s(x^s)\} \geq \lambda^b \beta(1-c). \tag{2.3}$$

Whenever condition (2.3) holds, all buyers join C market $B^C = B$.

Timing We summarize the set-up of the model by the timing structure.

1. Two retail markets, a C market and a D market, open. In the C market, the intermediary decides whether or not to activate the middleman sector and/or the platform. The intermediary announces middleman capacity K and price p^m, and a set of fees $\{f^b, f^s\}$ in the platform sector.

2. Observing the announced capacity, price and fees, buyers and sellers simultaneously decide whether to participate in the C market under pessimistic beliefs.

3. In the C market, the participating buyers, sellers and middleman are engaged in a directed search game. In the D market, agents search randomly and follow the efficient sharing rule for the trade surplus.

Figure 2.1 gives an overview of our model. Sellers are represented by red dots and buyers are represented by blue dots. They may engage in a random match in the Decentralized Market. Alternatively, they may meet in the Centralized Market. The market-maker sector of the C market is consisted of sellers. The intermediary can also establish a middleman sector with a continuum of inventory. Buyers who visit the C market can choose which supplier to visit and are subject to coordination frictions.

[11]Under such a belief, Caillaud and Jullien (2003) propose that the intermediary can charge negative fees (give subsidy) to induce the participation of agents. In our setup, the middleman's capacity advantage reveals that supply K is available in the C market, irrespective of the number of sellers participating. Hence, the intermediary can induce buyers' participation under those beliefs, as long as condition (2.3) is satisfied. When the middleman' supply is not observable in the participation stage, we will offer a version of our model with participation fees/subsidies in Appendix II, and show that our main result is still valid.

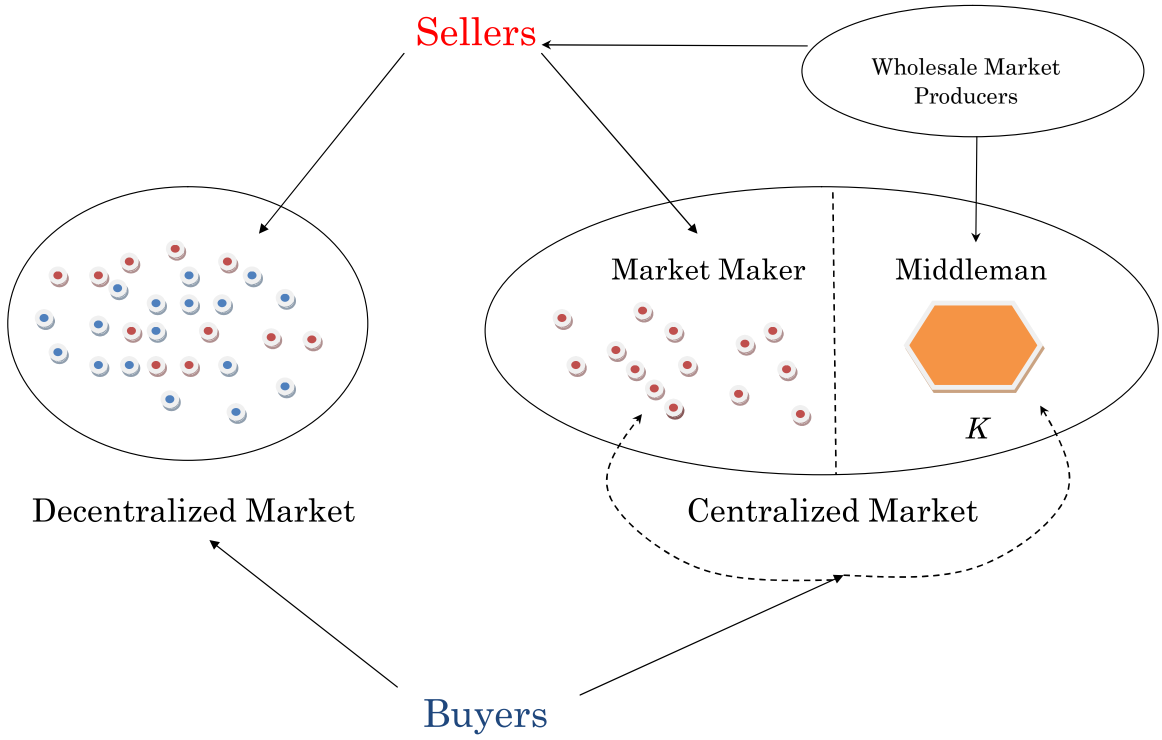

Figure 2.1: Overview

Note: The figure depicts the model abstract from the timing and the price/fee structure. Buyers are represented by blue dots. Sellers are represented by red dots. They can participate in D market where they are matched following a linear matching technology. Alternatively, they can participate in C market which has two sectors: A market-maker sector which consists of individual sellers and a middleman sector which is supplied by the intermediary. A buyer can choose one of the sellers or the middleman to visit taking into account the price and matching probability.

Given the characterization of the directed search equilibrium in C market and the random matching in D market, working backward, we proceed to characterize the profit-maximizing intermediation mode arising in this game.

2.2.2 Optimal intermediation mode under single-market search

Given the directed search equilibrium, the intermediary chooses an inventory capacity K, price/fees p^m, f^b, f^s to tune his business mode, and ultimately maximize his profits. In what follows, we show that if agents have a single-market search technology, then the intermediary will not open the platform, inducing $S^C = 0$, and will serve all buyers $B^C = B$ as a pure middleman with $K = B$ and $x^m = B$. This leads to the following pure middleman profits,

$$\Pi = B(p^m - c),$$

subject to the participation constraint of buyers in the C market,

$$V^m(x^m) = 1 - p^m \geq \lambda^b \beta(1-c). \tag{2.4}$$

The middleman sets $p^m = 1 - \lambda^b \beta(1-c)$. Note here that the outside value of buyers is given by $\lambda^b \beta(1-c)$, which is supported by their belief that the D market is non-empty. We make the usual tie-breaking assumption that agents choose to participate in a market if they are indifferent.

Now, we show that creating an active platform is not profitable. Suppose that the intermediary opens a platform with $S^C = S$ sellers and intermediation fees $f = f^b + f^s \leq 1$. Then, the platform generates a non-negative trade surplus $1 - f \geq 0$. In any active platform under pessimistic beliefs, it must be that $V^s(x^s) \geq \lambda^b \beta(1-c)$ and $p^s \geq f^s + c$. These conditions imply $f = f^s + f^b < (1 - \lambda^b \beta)(1-c)$. On the other hand, to have an active middleman, it must be that the inventory price $p \leq 1 - \lambda^b \beta(1-c)$. These are the constraints confronted by the intermediary to break pessimistic beliefs. Then, the intermediary's expected profits consist the revenue of platform fees, $S(1-e^{-x^s})f$, and the revenue of inventory sales minus inventory cost, $\min\{K, x\}p - Kc$. Without going into the optimization problem, we observe that

$$\begin{aligned}
\Pi(x, f, p, K) &= S(1-e^{-x^s})f + \min\{K, x\}p - Kc \\
&< Sx^s f + xp - Kc \\
&\leq (Sx^s + x)\max\{f, p\} - Kc \\
&\leq B(1 - \lambda^b \beta)(1-c) = \Pi,
\end{aligned}$$

for all $x^s \in (0, \frac{B}{S}]$. Hence, opening the platform is not profitable.

The intuition behind the occurrence of a pure middleman is as follows. Given the frictions on the platform, a larger middleman sector creates more transactions. To achieve the highest possible number of transactions, the intermediary shuts down the platform. In a nutshell, the middleman's capacity is the best way to distribute the good and, if agents search within a single market, the intermediary is guaranteed with the highest possible surplus of it. The allocation characterized here serves as a benchmark for our economy.

Proposition 2.1 (Pure middleman)**.** *Given single-market search technologies, the intermedi-*

ary will not open the platform and will act as a pure middleman with $x^m = K = B$, serving all buyers for sure.

2.3 Multi-market search

In this section, we extend our analysis to multiple-market search technologies where agents can search in both C market and D market.

Opening sequence To facilitate the presentation of our key idea, we assume that C market opens prior to D market.[12] Apart from the fact that this appears to be the most natural setup in our economy, we are motivated by the first mover advantage of the intermediary: its expected profit is higher if C market opens before, rather than after, D market. Hence, our setup will arise endogenously if the intermediary is allowed to select the timing of the market sequence.

As a real-life correspondence of this sequence, many intermediaries are enthusiastic about providing a wide range of information on merchandise and offering personalized service such as special offer emails tailored to a customer's interest. These efforts would enhance their chance to become a first-mover.[13]

Participating in C market Under multi-market search, participating in the C market does not rule out the possibility of trading in D market. Hence, for the intermediary, convincing agents to participate is not difficult. Indeed, the opening sequence makes it nature that all agents visit C market and then D market. What is still difficult is to attract transactions: The intermediary needs to convince agents that trade at C market is better than continuing search in D market. The competition pressure from the D market is reflected in the constraints (2.5), (2.6) and (2.7) on price/fees below.[14]

2.3.1 Directed search equilibrium under multi-market search

We work backward and first describe the directed search equilibrium for the C market, given the market-maker sector is active and such an equilibrium exists under

[12] If the two markets opened at the same time, we would have to deal with the agents' beliefs about what other agents would choose when they turn out to be matched in both markets. This would give rise to the multiplicity of equilibria which complicates the analysis significantly. Our sequential setup avoids this issue. In an infinite horizon model, one can construct a stationary equilibrium relatively easily where the order of the markets does not matter, see Watanabe (2018a).

[13] In a recent study without intermediation, Armstrong and Zhou (2015) show that a seller often makes it harder or more expensive to buy its product later than at the first opportunity.

[14] Note further that irrespective of agents' belief, empty D market cannot occur in equilibrium. This is because even when buyers are extremely pessimistic about the D market so that sellers are indifferent between entering and not entering, sellers will choose to enter the D market.

the announcement of the intermediary.

Incentive constraints to trade in C market As in single-market search, any directed search equilibrium in the C market has to satisfy (2.1) and (2.2). Given the multiple-market search technology, what is new here is that agents expect a non-negative value for the D market when deciding whether or not to accept an offer in the C market. Whenever the platform is active, that is $x^s > 0$ and $S^C = S$, it must satisfy the incentive constraints:

$$1 - p^s - f^b \geq \lambda^b e^{-x^s} \beta (1 - c), \tag{2.5}$$

$$p^s - f^s - c \geq \lambda^s \xi (x^m, K) (1 - \beta) (1 - c). \tag{2.6}$$

The constraint of buyers (2.5) states that the offered price/fee in the platform is acceptable only if the offered payoff, $1 - p^s - f^b$, is no less than the expected value in the D market: the outside payoff is $\beta(1 - c)$ if the buyer is matched with a seller who has failed to trade in the C market. This happens with probability $\lambda^b e^{-x^s}$. Hence, the larger the platform size x^s, the higher the chance that a seller trades in the C market, and the lower the chance that a buyer can trade successfully in the D market and the lower his expected outside payoff.

The constraint of sellers (2.6) states that the payoff in the C market $p^s - f^s - c$ should be no less than the expected payoff in the D market. This depends on a seller's chance of engaging in a trade in the D market $\lambda^s \xi (x^m, K)$, where $\xi (x^m, K)$ represents the probability that a buyer has failed to trade in the C market and is given by

$$\xi (x^m, K) \equiv 1 - \frac{1}{B} \left(\min \{K, x^m\} + S \left(1 - e^{-\frac{B - x^m}{S}} \right) \right).$$

The buyer visits the middleman sector with probability $\frac{x^m}{B}$ and is served with probability $\min \left\{ \frac{K}{x^m}, 1 \right\}$, or he visits the platform with probability $\frac{Sx^s}{B}$ and is served with probability $\eta^s(x^s) = \frac{1 - e^{-x^s}}{x^s}$. Hence, in the above expression, the second term represents the expected chance of the buyer to trade in the C market.

We have a similar condition of buyers for the middleman sector:

$$1 - p^m \geq \lambda^b e^{-x^s} \beta (1 - c), \tag{2.7}$$

where the middleman's price must be acceptable for buyers relative to the expected payoff in the D market.

It is important to point out that under these incentive constraints, all buyers/sellers find it at least better to trade at C market. Hence, under the tie-breaking assumption, all matched pairs trade at C. That is, $B^C = B, S^C = S$. However, the tie-breaking assumption is not crucial. This is because the intermediary can lower $f^i, i = b, s$ slightly so that (2.5) to (2.7) hold strictly, and $B^C = B, S^C = S$.

In the above formulation, all agents are supposed to stay in the D market so that some meetings are successful and others are not. This creates an interaction between the C market and the D market. With non-linear matching functions, this assumption can be relaxed. We will clarify this point in Section 2.4.1.

Equilibrium values Given the outside option of the D market, the equilibrium value of buyers in the C market is $V = \max\{V(x^s), V(x^m)\}$, where

$$V^s\left(x^s\right) = \eta^s(x^s)\left(1 - p^s - f^b\right) + \left(1 - \eta^s(x^s)\right)\lambda^b e^{-x^s}\beta\left(1 - c\right) \tag{2.8}$$

for an active platform $x^s > 0$ and

$$V^m\left(x^m\right) = \min\{\frac{K}{x^m}, 1\}\left(1 - p^m\right) + \left(1 - \min\{\frac{K}{x^m}, 1\}\right)\lambda^b e^{-x^s}\beta\left(1 - c\right) \tag{2.9}$$

for an active middleman sector $x^m > 0$. Here, if a buyer visits a seller (or a middleman), then he gets served with probability $\eta^s(x^s)$ (or $\eta^m(x^m)$) and his payoff is $1 - p^s - f^b$ (or $1 - p^m$). If not served in the C market, then he enters the D market and he can find an available seller with probability $\lambda^b e^{-x^s}$. Similarly, the equilibrium value of active sellers in the platform is given by

$$W(x^s) = x^s\eta^s(x^s)\left(p^s - f^s - c\right) + \left(1 - x^s\eta^s\left(x^s\right)\right)\lambda^s\xi\left(x^m, K\right)\left(1 - \beta\right)\left(1 - c\right). \tag{2.10}$$

A seller trades successfully in the C market platform with probability $x^s\eta^s(x^s)$ and receives $p^s - f^s - c$. If not successful, he enters the D market where he meets a buyer with probability $\lambda^s\xi\left(x^m, K\right)$.

We now need to determine the equilibrium price p^s. We follow the standard procedure in the directed search literature. Suppose a seller deviates to a price $p \neq p^s$ that attracts an expected queue $x \neq x^s$ of buyers. Note that given the limited selling-capacity, this deviation has measure zero and does not affect the expected utility in the C market, V. Since buyers must be indifferent between visiting any seller (including the deviating seller), the equilibrium market-utility should satisfy

$$V = \eta^s(x)\left(1 - p - f^b\right) + (1 - \eta^s(x))\,\lambda^b e^{-x^s}\beta(1-c), \tag{2.11}$$

where $\eta^s(x) \equiv \frac{1-e^{-x}}{x}$ is the probability that a buyer is served by this deviating seller. If not served, which occurs with probability $1 - \eta^s(x)$, his expected utility in the D market is $\lambda^b e^{-x^s}\beta(1-c)$. Given market utility V, (2.11) determines the relationship between x and p, which we denote by $x = x(p|V)$. This yields a downward sloping demand faced by the seller: when the seller raises his price p, the queue length of buyers x becomes smaller, and vice versa.

Given the search behavior of buyers described above and the market utility V, the seller's optimal price should satisfy

$$p^s(V) = \arg\max_p \left\{ \begin{array}{c} \left(1 - e^{-x(p|V)}\right)(p - f^s - c) \\ +e^{-x(p|V)}\lambda^s\xi(x^m, K)(1-\beta)(1-c) \end{array} \right\}.$$

Substituting out p using (2.11), the sellers' objective function can be written as

$$W(x) = \left(1 - e^{-x}\right)(v(x^m, K) - f) - x\left(V - \lambda^b e^{-x^s}\beta(1-c)\right) + \lambda^s\xi(x^m, K)(1-\beta)(1-c),$$

where $x = x(p|V)$ satisfies (2.11) and

$$v(x^m, K) \equiv \left[1 - \lambda^b e^{-\frac{B-x^m}{S}}\beta - \lambda^s\xi(x^m, K)(1-\beta)\right](1-c)$$

is the intermediated trade surplus, i.e., the total trading surplus in the C market net of the outside options. Since choosing a price is isomorphic to choosing a queue, the first order condition is

$$\frac{\partial W(x)}{\partial x} = e^{-x}(v(x^m, K) - f) - \left(V - \lambda^b e^{-x^s}\beta(1-c)\right) = 0.$$

The second order condition can be easily verified. Arranging the first order condition using (2.11) and evaluating it at $x^s = x(p^s|V)$, we obtain the equilibrium price $p^s = p^s(V)$ which can be written as

$$p^s - f^s - c = \left(1 - \frac{x^s e^{-x^s}}{1 - e^{-x^s}}\right)(v(x^m, K) - f) + \lambda^s\xi(x^m, K)(1-\beta)(1-c). \tag{2.12}$$

Incentive constraints revisited We can now rewrite the incentive constraints (2.5) and (2.6) by substituting in (2.12). This yields

$$f \leq v(x^m, K), \tag{2.13}$$

which states that for the platform to be active $x^s > 0$, the total transaction fee f should not be greater than the intermediated trade surplus, $v\left(x^m, K\right)$. Whenever (2.5) and (2.6) are satisfied, (2.13) must hold, and whenever (2.13) is satisfied, (2.5) and (2.6) must hold. Hence, we can say that the market maker faces the constraint (2.13) for an active platform.

Observe that for $K < x^m$, we have

$$v\left(x^m, K\right) = \left[1 - \lambda^b e^{-\frac{B-x^m}{S}}\beta - \lambda^s\left(1 - \frac{K + S(1 - e^{-\frac{B-x^m}{S}})}{B}\right)(1-\beta)\right](1-c),$$

which is decreasing in x^m. This occurs because a larger sized platform (i.e., a lower x^m) crowds out the D market transactions and lowers the outside value.

2.3.2 Intermediation mode

Our next step is to determine the profit of each intermediation mode, denoted by $\tilde{\Pi}(x^m)$.

Pure middleman: If the intermediary does not open the platform then $x^m = B$ and any encountered seller in the D market is always available for trade. Hence, as before, the middleman selects capacity $K = B$, serves all buyers at a price $p^m = 1 - \lambda^b\beta(1-c)$, satisfying (2.7), and unit cost (or wholesale price) c, and makes profits

$$\tilde{\Pi}(B) = B(1 - \lambda^b\beta)(1-c). \tag{2.14}$$

Pure market-maker: When the middleman sector is not open, $x^s = \frac{B}{S}$. Given the equilibrium price p^s in the platform in (2.12), the intermediary charges a fee $f = f^b + f^s$ in order to maximize

$$S\left(1 - e^{-\frac{B}{S}}\right)f,$$

subject to the constraint (2.13). The constraint is binding and it yields:

$$f = v\left(0,0\right) = \left[1 - \lambda^b e^{-x^s}\beta - \lambda^s\xi\left(0,0\right)(1-\beta)\right](1-c).$$

where $\xi(0,0) = 1 - \eta^s(x^s)$. The profit for the market-maker mode is

$$\tilde{\Pi}(0) = S(1 - e^{-\frac{B}{S}})v(0,0). \tag{2.15}$$

Market-making middleman: If the intermediary is a market-making middleman, then $x^m \in (0, B)$ and $x^s \in \left(0, \frac{B}{S}\right)$, satisfying $V^m(x^m) = V^s(x^s)$. Applying (2.8), (2.9), and (2.12), this indifference condition generates an expression for the price $p^m = p^m(x^m)$:

$$p^m = 1 - \lambda^b e^{-x^s} \beta (1 - c) - \frac{x^m e^{-x^s}}{\min\{K, x^m\}} \left(v(x^m, K) - f\right). \tag{2.16}$$

Together with (2.1), this equation defines the relationship between p^m and x^m. Applying this expression, we can see that the condition (2.7) is eventually reduced to (2.13). The profit for the marketmaking middleman mode is

$$\tilde{\Pi}(x^m) = \max_{x^m, f, K} \Pi(x^m, f, K) = S(1 - e^{-x^s})f + \min\{K, x^m\} p^m - Kc$$

subject to (2.13) and $x^m \in (0, B)$. Note that $K > x^m$ cannot be profitable since it is a mere increase in capacity costs. The profit maximization requires the following properties.

Lemma 2.1. *The market-making middleman sets: $K = x^m$ and $f = v(x^m, K)$.*

Proof. See the Appendix. □

The intermediary's capacity should satisfy all the forthcoming demands, and the intermediation fee should be set to extract the full intermediation surplus.

Profit-maximizing intermediation mode: We are now in the position to derive a profit-maximizing intermediation mode. To do so, it is important to observe that relative to the pure middleman mode, an active platform with multiple-market search can undermine the D market by lowering the available supply. This influences the middleman's price in the following way. With $v(\cdot) = f$, the incentive constraint (2.7) is binding, and the middleman's equilibrium price is given by

$$p^m = 1 - \lambda^b e^{-x^s} \beta(1 - c)$$

for any $x^s \geq 0$ (see (2.16)). This shows that p^m decreases with x^m: the outside value of buyers depends positively on the size of the middleman sector, since a larger scale of the middleman crowds out the platform and increases the chance that a buyer can find an active seller in the D market (who was unsuccessful in the platform). Hence, in order to extend the size of the middleman sector, the intermediary must lower the price p^m. In other words, a larger platform allows for a price increase by reducing agents' outside trade opportunities.

Proposition 2.2 (Market-making middleman/Pure Market-maker)**.** *Given multi-market search technologies, there exists a unique directed search equilibrium with active intermediation. The intermediary will open a platform and act as:*

- *a market-making middleman if $\lambda^b\beta \leq \frac{1}{2}$ or if $\lambda^b\beta > \frac{1}{2}$ and $\frac{B}{S} \geq \bar{x}$, some $\bar{x} \in (0, \infty)$;*
- *a pure market-maker if $\lambda^b\beta > \frac{1}{2}$ and $\frac{B}{S} < \bar{x}$.*

Proof. See the Appendix. □

With multiple-market search technologies, there is a cross-market feedback from the D market to the C market, which makes using the platform as part or all of its intermediation activities profitable. Additionally, the intermediary must decide whether it wants to operate as a pure market maker. Our result shows that it depends on parameter values. If $\lambda^b\beta \leq \frac{1}{2}$ then the buyers' outside option value is low. In this case, the middleman sector generates high enough profits for the market-making middleman mode to be adopted for any value of $\frac{B}{S}$. If instead $\lambda^b\beta > \frac{1}{2}$ then the buyers' outside option value is high, and attracting buyers to the middleman sector is costly. In this case, the intermediary will act as a market-making middleman if $\frac{B}{S}$ is high, where the D market is tight for buyers and they expect a low value from it, and as a pure market maker if $\frac{B}{S}$ is low, where buyers expect a high value from the D market. Indeed, the same logic applies to the following comparative statics result.

Corollary 2.1 (Comparative statics)**.** *Consider a parameter space in which the market-making middleman mode is profit-maximizing. Then, an increase in buyer's bargaining power β or buyer's meeting rate λ^b in the D market, or a decrease in the buyer-seller population ratio, $\frac{B}{S}$, leads to a smaller middleman sector x^m and a larger platform x^s.*

Proof. See the Appendix. □

2.4 Extensions

This section considers extensions of the model. As we show below, our main insight, that the profit of using a platform as part or all of the intermediation business is relatively large when agents can search in multiple markets rather than in a single market only, is robust to these extensions.[15]

2.4.1 Matching functions

In this section, we extend the linear matching function in the D market to a more general matching function, $M = M(B^D, S^D)$. As is standard in the literature we assume that the matching function is homogeneous of degree one in B^D and S^D, $M(1, \frac{1}{x^D}) = \frac{M(B^D,S^D)}{B^D}$ and $M(x^D, 1) = \frac{M(B^D,S^D)}{S^D}$, where $x^D = \frac{B^D}{S^D}$ is the buyer-seller ratio in the D market. Then, we allow for the dependence of the individual match probabilities on the relative measure of buyers and sellers,

$$\lambda^b(x^D) = M(1, \frac{1}{x^D}) \text{ and } \lambda^s(x^D) = M(x^D, 1) = x^D\lambda^b(x^D) \tag{2.17}$$

where $\lambda^b(x^D)$ is strictly concave and decreasing in x^D.

With single-market search technologies, the result will not be affected by this extension (for instance, if we use the pessimistic beliefs of agents, then the matching probability in the D market is simply replaced by another constant $\lambda^i(x^D)$, $i = b, s$, with $x^D = \frac{B}{S}$). Therefore, we only consider multi-market search technologies. As mentioned before, we shall let agents exit if they have traded successfully in the C market, because if agents stayed in the D market as in the previous section, then again the analysis would remain essentially unchanged. Then, the population in the D market is given by

$$B^D = B - \min\{x^m, K\} - S(1 - e^{-x^s}) \text{ and } S^D = Se^{-x^s}.$$

With this modification, the buyers' probability to meet an available seller changes from $\lambda^b e^{-x^s}$ to $\lambda^b(x^D)$, and the sellers' probability to meet an available buyer changes from $\lambda^s\xi(x^m, K)$ to $\lambda^s(x^D) = x^D\lambda^b(x^D)$.

In what follows, we derive a condition for a pure middleman mode to be selected under multi-market search technologies. This is the case when, for example, $\lambda^{b\prime}(x^D) = 0$, i.e., when there is no feedback from the D-market to the intermediary's decision in the C market. We proceed with the following steps. First, note that, as before, there is no gain of having an excess capacity $K > x^m$. In addition, a pure middleman wants to

[15] For expositional simplicity, we let $c = 0$ and make the tie-breaking assumption that when the middleman is indifferent between $K = x^m$ and $K > x^m$ we set $K = x^m$.

avoid stockouts $K < x^m$ if

$$\frac{d\tilde{\Pi}(K)}{dK} = \frac{d}{dK}K\left(1 - \lambda^b(x^D)\beta\right) = 1 - \lambda^b(x^D)\beta + \frac{K}{S}\lambda^{b\prime}(x^D)\beta > 0,$$

for any $x^D = \frac{B-K}{S} \geq 0$, which states that the elasticity of the middleman's price $p^m = 1 - \lambda^b(x^D)\beta$ should satisfy

$$z(K) \equiv -\frac{\partial p^m/\partial K}{p^m/K} = -\frac{K\lambda^{b\prime}(x^D)\beta}{S(1 - \lambda^b(x^D)\beta)} \leq 1.$$

This condition guarantees that a pure middleman should satisfy all the forthcoming demand $K = x^m$.

Second, when all buyers are served by the middleman $x^m = K = B$, the marginal gain of allocating buyers to the platform, measured by the intermediation fee,

$$f = 1 - \lambda^b(x^D)\beta - x^D\lambda^b(x^D)(1 - \beta),$$

can not exceed the marginal opportunity cost, measured by the lost revenue in the middleman sector,

$$1 - \lambda^b(0)\beta - K\lambda^{b\prime}(0)\beta\frac{dx^D(K, x^s(K))}{dK}\,|_{x^s(K)=0},$$

where $x^s(K) = \frac{B-K}{S}$ and

$$\frac{dx^D(K, x^s(K))}{dK}\,|_{x^s(K)=0} = \frac{d}{dK}\frac{B - K - S(1 - e^{-x^s(K)})}{Se^{-x^s(K)}}\,|_{x^s(K)=0} = \frac{-S + (B - K - S)}{S^2e^{-x^s(K)}}\,|_{K=B} = 0.$$

Hence, the intermediary can be a pure middleman even with multiple-market search technologies.

Proposition 2.3. *With a non-linear matching function in the D market outlined above, a pure middleman mode can be profitable even with multi-market search technologies only if the middleman's price is inelastic at the full capacity $x^m = K = B$. Otherwise, the intermediary should be a marketmaking middleman or a pure market maker.*

Proof. See the Appendix. □

Figure 2.2 plots the size of the middleman sector, represented by $\frac{x^m}{B}$, and the elasticity of middleman's price with respect to capacity, evaluated at $x^m = K = B$.[16] It shows that when a pure middleman mode is selected $\frac{x^m}{B} = 1$ the price is inelastic $z(B) < 1$,

[16]The figures is drawn with $S = 1$ and $\lambda^b(x^D) = \frac{1-e^{-x^D}}{x^D}$.

whereas when an active platform is used the price is elastic $z(B) > 1$. This confirms that given the appropriate restriction on the meeting rate $\lambda^b(x^D)$, our main conclusion in the baseline model is valid with an alternative assumption that agents exit after successful trade in the C market. It is intuitive that when the middleman's price is elastic, there is strong enough negative feedback from the D market on the price that makes the exclusive use of the middleman sector not profitable.

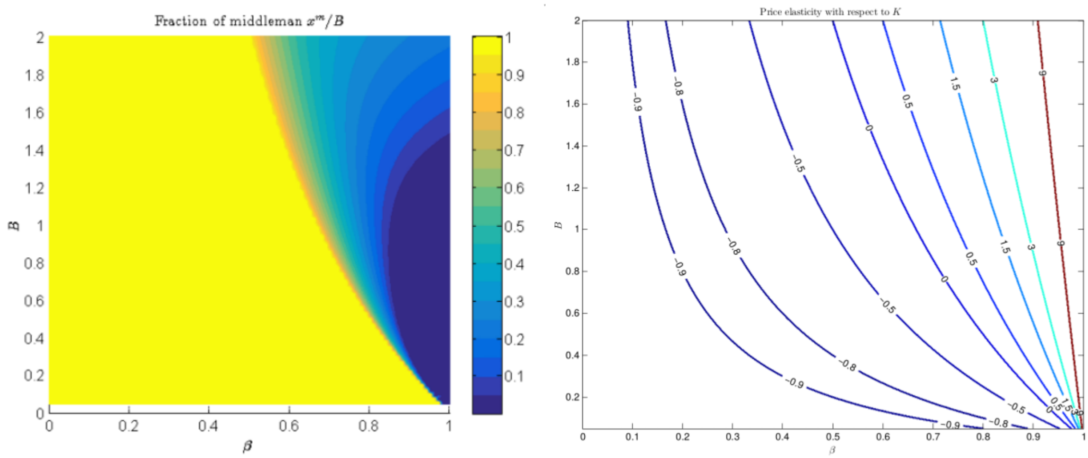

Figure 2.2: Size of middleman sector $\frac{x^m}{B}$ (Left) and Price elasticity $z(B)$ (Right) with Non-linear matching function

2.4.2 Endowment economy

In our baseline model, we simplify the middleman's inventory stocking by assuming that the good is supplied by competitive producers in the frictionless wholesale market. In this section, we study the implication of wholesale-market frictions in an endowment economy (i.e., with no producers). Suppose that each seller owns one unit of endowment. In total, a mass of S commodities are available. In the wholesale market, the middleman can access a fraction α of sellers, where we assume that $\alpha \in (0, 1)$ is an exogenous parameter. Then, the middleman's inventory should satisfy the aggregate resource constraint,

$$K \leq \alpha S. \tag{2.18}$$

In a world with unlimited production capacity, sellers are willing to supply as long as the wholesale price, denoted as p^w, is enough to compensate for the marginal cost; whereas in an endowment economy, sellers are only willing to supply if p^w is high enough to compensate for trading opportunities they lose in other channels. Once contacted by the middleman, sellers choose among selling the endowment to the middleman, or joining the C market platform and/or joining the D market. To simplify the analysis, we abstract away the influence of what sellers can expect from the D market on the determination of the wholesale price, and assume that sellers in the D market receive zero trade share,

$$\beta = 1.$$

Our main conclusion does not depend on this simplification. Then, the middleman's offer to buy from sellers is accepted if and only if

$$p^w \geq W(x^s), \tag{2.19}$$

where $W(x^s)$ is the expected value of sellers to operate in the C market platform.

Single-market search: The determination of the intermediation mode depends on the available resources. If $B \leq \alpha S$, then the middleman can stock the full inventory to cover the entire population of buyers. In this case, by closing the platform, the middleman makes the highest possible profit, $\Pi = B(1 - \lambda^b)$, with the wholesale price $p^w = 0$, just like in the baseline model. If $B > \alpha S$, then the middleman's inventory will not be enough to cover all buyers, and so the intermediary may wish to use a platform even under a single-market search technology. With the wholesale price p^w determined by the binding constraint, (2.19), the fee f and the price p^m determined by the binding participation constraint of buyers, $V = \max\{V^s(x^s), V^m(x^m)\} = \lambda^b$, the intermediary's problem can be written as the choice of the size of its inventory K and the allocation x^m that maximizes

$$\Pi\left(x^m, f, K\right) = (S - K)(1 - e^{-x^s})f + \min\left\{K, x^m\right\} p^m - Kp^w$$

where $x^s = \frac{B - x^m}{S - K}$, subject to the resource constraint (2.18). To guarantee non-negative price/fees/profits, we shall assume sufficiently low values of $\lambda^b > 0$ whenever necessary (see the proof of Proposition 2.4).

As expected, the solution is characterized by the binding resource constraint (2.18) and an active platform $x^s > 0$ when $B > \alpha S$. Note that the intermediary could deactivate the platform since it would lead to the lowest wholesale price of middleman $p^w = 0$. However, it turns out that the benefit of fee revenue from the active platform outweighs the cost savings in the middleman sector. Hence, even with single-market technologies, the aggregate resource constraint can be one reason for the intermediary to open the platform sector in the endowment economy.

Proposition 2.4. *Consider the endowment economy outlined above with single-market search technology, and the zero trade share of sellers in the D market. The intermediation chooses to be:*

- *a pure middleman if $B \leq \alpha S$;*
- *a market-making middleman with $K = \alpha S \leq x^m$ if $B > \alpha S$.*

Proof. See the Appendix. □

The result $x^m \geq K$ occurs because, in line with the previous setup, an excess inventory means extra costs in the middleman sector and lost revenues in the platform. Figure 2.3 demonstrates that when $B > \alpha S$, it is possible that the intermediary attracts an excessive number of buyers to the middleman sector $x^m > K$, resulting in stockouts, in order to lower the wholesale price in the middleman sector.[17] When this occurs, the resource constraint is tight and the outside value of agents is high so that economizing on stocking costs is relatively important.

Multi-market search: With multi-market search technologies, the participation constraint of agents is not the issue but the intermediation fee and the middleman's price should be acceptable relative to the outside value. Hence, the intermediary faces the incentive constraints, (2.5) – (2.7) with an appropriate modification of the match probability in the D market (see the details in the proof of Proposition 2.5). As before, these conditions are reduced to $f \leq v(x^m, K)$.

To be consistent, we maintain the assumption of a zero trade share of the sellers, $\beta = 1$, in the D market. This assumption now implies that sellers are fully exploited in the C market, thus $p^w = W(x^s) = 0$ for any $x^s \geq 0$.

With the multi-market search setup, the buyers' outside option value depends positively on the number of sellers available in the D market. This has the following consequences. First, just as in the baseline setup, a pure middleman mode can never be

[17] The figures in this subsection are drawn with $B = 0.8$ and $S = 1$. We cut out the region where negative profits result, for high values of λ^b.

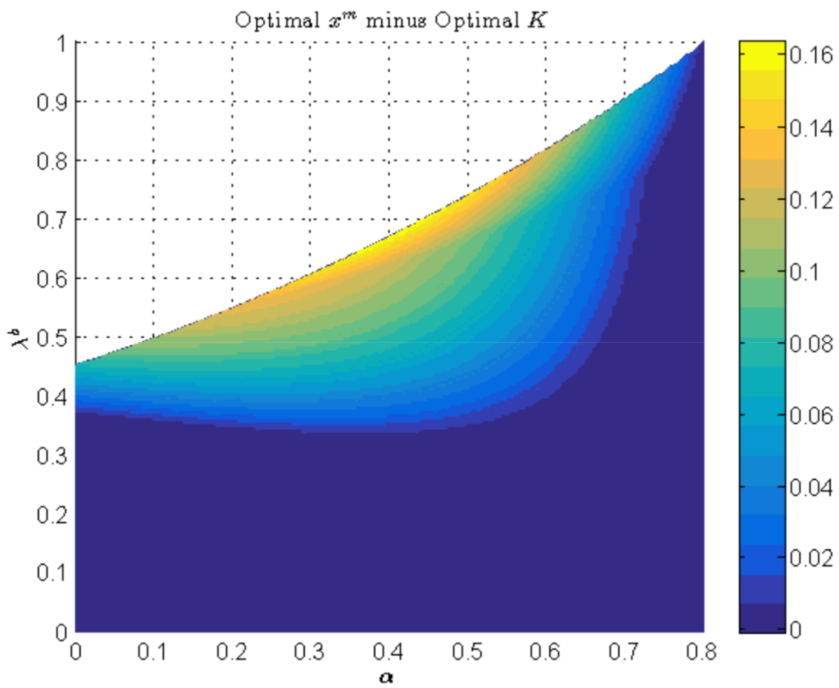

Figure 2.3: Values of $x^m - K$ with single-market search in endowment economy

profit maximizing. Second, in our endowment economy, the intermediary may wish to stock more inventories than the number of buyers visiting the middleman sector. This is because a larger K will crowd out the supply available in the D market, which will eventually lower the outside value of buyers and increase the profit. Therefore, unlike in all the previous setups, the solution here allows for an excess inventory in the middleman sector.

Proposition 2.5. *Consider the endowment economy outlined above with multi-market search technologies, and the zero trade share of sellers in the D market. The intermediation chooses to be a market-making middleman or a pure market-maker with* $x^m \leq K = \alpha S$.

Proof. See the Appendix. □

Figure 4 shows the occurrence of excess inventory holdings in the middleman sector with high values of λ^b and α. This confirms our intuition that the crowding-out effect of excess inventory is stronger when the agents outside value in the D market is higher.

Comparing Proposition 2.4 and 2.5, we can summarize the implication of search frictions in wholesale markets represented by α and the agents' search technologies in retail markets on the choice of intermediation mode in our endowment economy as

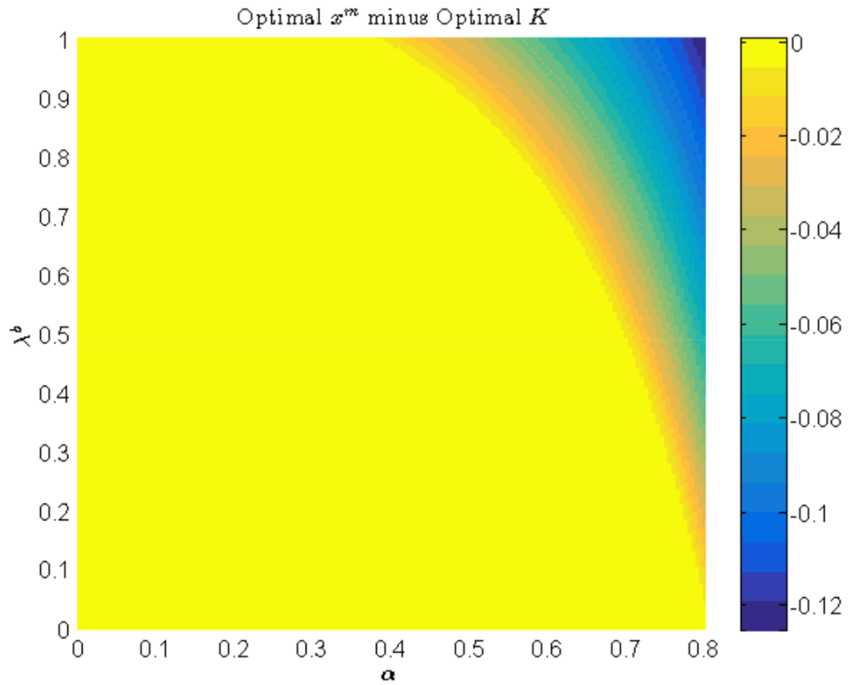

Figure 2.4: Values of $x^m - K$ with multi-market search in endowment economy

follows.

- For $\alpha S \geq B$, the middleman can stock the full inventory that satisfies all the buyers' demand. As in the benchmark setup, the intermediary chooses to be a pure middleman with single-market search, but uses an active platform with multi-market search. Unlike in the previous setups, the middleman holds an excessive amount of inventory.

- For $\alpha S < B$, the full inventory is not possible due to the aggregate resource constraint. The intermediary uses a platform irrespective of whether agents search in a single or in multiple markets. Our main insight is still valid. Namely, *the intermediation mode is further away from the pure middleman mode when agents search in multiple markets, rather than in a single market*. The size of the middleman sector, measured by x^m, is smaller with multi-market search than with single-market search technologies.

2.4.3 Cost functions

Inventory Costs. In the baseline model, we assume zero costs of storing the inventory for the middleman. In this section, we consider the following convex inventory-cost function,

$$C'(K) \geq 0;\;\; C''(K) \geq 0;\;\; C(0) = 0;\;\; C'(K) < 1 - \lambda^b.$$

The last condition guarantees that $C(B) < B(1 - \lambda^b)$. We assume $\beta = 1$ for simplicity. With positive inventory costs, it may be profitable to activate a platform even with single-market search. Still, we show that our main insight is valid.

As in the baseline model, profit maximization requires $K = x^m$. With single-market search, the problem of a market-making middleman can be described as

$$\max_{x^m, f} \Pi(x^m) = S(1 - e^{-x^s}) f(x^s) + x^m p^m - C(x^m). \tag{2.20}$$

subject to $p^m = 1 - \lambda^b$ and $f(x^s) = 1 - \lambda^b e^{x^s}$. The first constraint is given by the buyers' participation constraint in the C market, i.e., $V^m = 1 - p^m \geq \lambda^b$, while the second constraint is given by the buyers' indifference condition, i.e., $V^m = V^s = e^{-x^s}(1 - f)$ (a formal derivation can be found in the Appendix): the platform fee $f = f(x^s)$ is strictly decreasing (increasing) in x^s (x^m). Intuitively, the tighter the platform, the lower the fee that the intermediary needs to offer, in order to make buyers indifferent between the platform and the middleman sector. The negative dependence of the platform fee on the platform size favors the middleman mode. The first order condition becomes

$$\frac{\partial \Pi(x^m)}{\partial x^m} = -e^{-x^s} + \lambda^b e^{x^s} + 1 - \lambda^b - C'(x^m) \equiv \Theta_{Sfoc}(x^m) = 0. \tag{2.21}$$

With multiple-market search, the objective function is the same as in (2.20), but the constraints are $p^m(x^s) = f(x^s) = 1 - \lambda^b e^{-x^s}$ (by (2.16) and $v(\cdot) = f$ as in Lemma 1). As before, the positive dependence of the middleman's price and the platform fee on the platform size favors the market-maker mode. Observe that

$$\begin{aligned} \frac{\partial \Pi(x^m)}{\partial x^m} &= (1 - e^{-x^s})(1 - 2\lambda^b e^{-x^s}) - \frac{\lambda^b x^m e^{-x^s}}{S} - C'(x^m) \\ &< (1 - e^{-x^s})(1 - 2\lambda^b e^{-x^s}) - C'(x^m) \\ &= \Theta_{Sfoc}(x^m) - \lambda^b \left[2e^{-x^s}(1 - e^{-x^s}) + e^{x^s} - 1\right] \\ &< \Theta_{Sfoc}(x^m), \end{aligned} \tag{2.22}$$

implying that the marginal profit of increasing the size of the middleman sector is smaller with multi-market search than with single-market search. The logic behind this is essentially the same as in the baseline model, and is generalized as follows.

Proposition 2.6. *Consider the convex inventory costs of a middleman defined above. Then, a platform is activated even under a single-market search technology. Still, the size of the platform with multi-market search is larger than or equal to that with single-market search.*

Proof. See the Appendix. □

Prior production/purchase before joining the platform. In real-life markets, sellers sometimes need to prepare (produce or purchase) their product for sale prior to market entry. For example, online sellers find it important to display their product's image and keep it ready for delivery before actual transactions occur. A similar issue arises when asset holders are required to commit to their portfolio before trading with their brokers. In these situations, because sellers incur costs irrespective of their success in the platform, attracting sellers to the platform is costly and so the relative profitability of the market-maker mode is reduced. We show, however, that our insight remains valid in such a setting. Interestingly, we also find that a platform can be activated even when the net profit obtained from the platform business is negative.

The only modification now is to introduce a participation constraint for sellers to operate on a platform. With single-market search, this is irrelevant because in any case the pure middleman mode remains profit maximizing. With multiple-market search, the participation constraint is given by

$$W(x^s) - f_p \geq c_E, \tag{2.23}$$

where $c_E \geq 0$ represents entry costs of sellers to the platform, $f_p \geq 0$ (or $f_p \leq 0$), a platform participation fee (or subsidy) to each individual seller, and $W(x^s) = \eta(x^s)(p^m - f)$ the equilibrium value of sellers in the platform. With $\beta = 1$, i.e., zero payoff in the D market for sellers, the intermediary sets $f = p^m = 1 - \lambda^b e^{-x^s}$, satisfying the incentive constraint (6) (note that the participation in the D market does not require prior production/purchase as before), and $f_p = -c_E$. That is, the intermediary should subsidize the entry cost and fully extract the trade surplus in the platform. The profit of a market-making middleman is

$$\tilde{\Pi}(x^m) = S\left[(1 - e^{-x^s})f - c_E\right] + x^m p^m,$$

while the profit of a middleman is

$$\tilde{\Pi}(B) = B(1 - \lambda^b).$$

Comparing these profits, one can find a value of $x^m < B$ (e.g. imagine a neighbourhood of $x^m = B - Sx^s \approx B$) and $c_E > 0$ for which the platform profit is negative but $\tilde{\Pi}(x^m) > \tilde{\Pi}(B)$. This leads to the following result.

Proposition 2.7. *Suppose sellers incur production/purchasing costs prior to platform entry. Then, an active platform can be profit maximizing even when the platform entry cost is higher than the platform fee revenue.*

One benefit of having an active platform in the C market for the intermediary is to reduce competition so that it can set a higher price in the middleman sector. This benefit can be the major source of profits for market-making middlemen even when the platform-entry costs are so high that the net profit from the platform business is negative.

2.5 Examples

Our analysis shows that a marketmaking middleman is more likely to emerge with multi-market search technologies than with single-market technologies. In this section, we offer some real market examples.

Online retailers. The electronic commerce company Amazon.com is traditionally an online retailer, who mainly aims at selling its inventories to customers. In the late 1990s, Amazon was facing fierce competition from local brick and mortar rivals, as well as chain stores such as Walmart, Sears, etc., and especially from eBay. According to the book, *The Everything Store: Jeff Bezos and the Age of Amazon*, Jeff Bezos worried that eBay may become the leading online retailer who attracts the majority of customers. In the summer of 1998, he invited eBay's management team and suggested the possibility of a joint venture or even of buying out their business. This is perhaps Amazon's first attempt to set up an online marketplace. In the end, this trial failed. After several more trials and errors, however, Amazon finally launched their own marketplace in the early 2000s.[18]

[18]The entry version of our model in Chapter 3 captures well Amazon's reaction to the entry by eBay.

Amazon's launch of the platform business influenced significantly the book industry. On the one hand, Amazon attracts many of its competitors to join their platform. Indeed, Amazon drove physical book and record stores out of business, and many bookstore owners re-launched their business on the Amazon-website platform. On the other hand, Amazon lowers the chance of buyers to trade outside. As local bookstores disappeared, it became the habit for most book buyers to start their everyday online-shopping using Amazon as the prime site De los Santos et al. (2012). Overall, these observed phenomena are in line with our theory.[19][20] Not surprisingly, Amazon promoted this shopping pattern to customers in other product categories.

The general picture of the online travel agency industry is similar. Before the rise of Internet, most intermediaries in this industry acted as a pure middleman. In the middleman mode, hotels sell rooms to a middleman in bulk at discounted prices. The middleman then sells them to customers at a markup price. With the online reservation system, a market-making mode became popular, wherein hotels pay a market maker (e.g. Booking.com) commission fees upon successful reservations. The hotels post their services and prices on the platform. Expedia used to be a pure middleman but is nowadays a representative market-making middleman who employs both of these intermediation modes.

Specialist markets. The New York Stock Exchange (NYSE) is a specialist market, which is defined as a hybrid market that includes an auction component (e.g., a floor auction or a limit order book) together with one or more specialists (also called designated market makers). The specialists have some responsibility for the market: as brokers, they pair executable customer orders; and as dealers, they post quotes with

[19]Nowadays, most buyers and sellers use Amazon as the main website (the first one to visit). On the seller side, according to a survey on Amazon sellers conducted in 2016, more than three-quarters of participants sell through multiple channels, online marketplaces, web-stores and bricks-and-mortar stores. The second most popular channel, after Amazon, is eBay, with 73% selling through this marketplace. On the buyer side, according to a recent Reuters/Ipsos poll, 51 percent of consumers plan to do most of their shopping on the Amazon.com.

[20]An alternative (or complementary) to our theory would be a product selection story where Amazon uses the platform for third-party sellers to add new products with the demands too small for Amazon to offer. Once a product is "tested" to be popular enough, Amazon starts to also offer it through the middleman sector. This would be certainly a valid explanation but by far not the exclusive one. First, if this explanation were correct, we should eventually observe that most popular products are listed by Amazon, and most not-so-popular products are listed by independent sellers. In reality, however, many high-demand products are listed by both Amazon and third-party sellers at the same time, and importantly, they are competing with each other. This competition goes against the proposed explanation, but is more in line with our theory. In fact, Amazon could avoid fierce competition with strong competitors operating in the Amazon marketplace, such as GreenCupboards or independent sellers who own 'Buy Boxes', by giving up dealing with such a product in the middleman sector, which should in turn increase their fee revenue.

reasonable depth (Conroy and Winkler, 1986).

As for their role as dealers in the exchanges, our model suggests that, at least for less active securities (represented by smaller outside option values in the model), the specialists' market can provide predictable immediacy and increase the trading volume and liquidity. This is consistent with the trend to adopt hybrid markets in derivative exchanges and stock markets around the world, especially for thinly-traded securities. For example, several European stock exchanges implemented a program which gives less active stocks an option of accompanying a designated dealer in the auction market. These initiatives were effective not only in enhancing the creation of hybrid specialist markets, but also in increasing trade volumes and reducing liquidity risks (Nimalendran and Petrella 2003, Anand et al. 2009, Menkveld and Wang 2013, and Venkataraman and Waisburd 2007.)

Another prediction from our analysis is related to the changing competitive environment faced by securities exchanges. As a broader implication, our result that the increased outside pressure goes hand in hand with more decentralized trades, captures the background trend in general: the market for NYSE-listed stocks was highly centralized in the year of 2007 with the NYSE executing 79% of volume in its listings; in 2009, this share dropped to 25% (Securities and Exchange Commission 2010); today, the order-flow in NYSE-listed stocks is divided among many trading venues – 11 exchanges, more than 40 alternative trading systems, and more than 250 broker-dealers in the U.S. (Tuttle 2014). As a more specific implication, we show that the increased pressure from outside markets will scale up the platform component. This is indeed the case. Starting from 2006, the NYSE adopted the new hybrid trading system featuring an expanded platform sector "NYSE Arca", which allows investors to choose whether to trade electronically or by using traditional floor brokers and specialists. The new system is further supplemented by several dark pools, akin to platforms, owned by the NYSE. These strategies are also adopted by NASDAQ which has been thought of as a typical dealers' market. In addition, the use of fees is widely adopted, as is consistent with our theory. For instance, in 2014, the NYSE offered banks a discount of trading costs by more than 80% conditional on their agreement to stay away from the outside dark pools and other off-exchange venues.[21]

[21]See a report "NYSE Plan Would Revamp Trading" in the Wall Street Journal, 2014. http://www.wsj.com/articles/intercontinental-exchange-proposing-major-stock-market-overhaul-

Real estate agencies. While intermediaries in housing markets are mostly thought of as brokers, i.e., platforms, the business mode employed by the Trump family is a marketmaking middleman. The Trump Organization holds several hundred thousand square feet of prime Manhattan real estate in New York City (NYC) and some more in other big cities. Besides developing and owning residential real estate, the Trump family operates a brokerage company that deals with luxury apartments, the Trump International Realty. Both of these companies target the same market in NYC. Indeed, the Trump's business mode is a marketmaking middleman – both owning his own residential towers, and offering broker services. According to Forbes, the latter portion of Trump's empire becomes by far his largest business with a valuation of 562 million in 2006. Another example is Thor Equities, a large-scale real estate company, which owns and redevelops retail properties in Soho, Madison Avenue, and Fifth Avenue, and also runs brokerage agencies, Thor Retail Advisors and Town Residential.

In the endowment economy version of our model, we show that the marketmaking middleman over-invests in inventory with multi-market search, up to the point where the resource constraint is binding. Perhaps, the real estate market in NYC is an appropriate example of this since it is well known to be competitive and tight for house/apartment hunters. In addition, most new developments in big cities are renovations of old houses, and so we can roughly regard the total supply as fixed. Notably, top real estate firms in NYC attempt to expand their business by being engaged in many new joint projects with developers. Mapped into our model, these efforts are aimed at relaxing their resource constraint and increasing their inventory. For example, Nest Seekers, a real estate brokerage and marketing firm in NYC, works tightly with constructors on new developments. They work together from the very early stage of layout design and fund raising (in some cases Nest Seekers offers their own capital) to the later marketing stage. Nest Seekers provides qualified sales and administrative staff to the sales office, prepares pricing schedules, manages all contracts with the brokerage community, and is eventually in charge of the entire marketing process. This co-development business is one step beyond the middleman mode formulated in our theory, but is considered as an alternative way to secure their inventory.[22] This business mode is adopted in

1418844900.

[22]Strictly speaking, Nest Seekers does not own properties, but becomes the exclusive agent of projects. So far, they have co-developed/marketed more than 30 projects. See https://www.nestseekers.com/NewDevelopments. A report titled "Inside the fight for Manhattans most

many other big real-estate companies in NYC, such as Douglas Elliman, Stribling, and Corcoran.

Finally note that some intermediaries do not help to promote market new developments, but also manage apartment complexes, which constitutes another source of "inventory". For example, Brown Harris Stevens provides the residential management service for its customers since cooperative apartments were first introduced to NYC. These cooperative apartments usually contain hundreds of units in one building, and Brown Harris Stevens is then in charge of listing these properties when they are for rent or on sale.

2.6 Empirical evidence

The model's predictions on the choice of intermediation mode can be empirically tested. We take Amazon as the centralized market and eBay as the decentralized market. Our model predicts that Amazon is more likely to sell the product as a middleman when the chance of buyers to meet a seller in the decentralized market, λ^b, is low, the buyers' bargaining power β is low, and the total demand B is high. That is,

$$\Pr(\text{Amazon's middlemen mode is active}) = f(\underset{-}{\lambda^b}, \underset{-}{\beta}, \underset{+}{B}),$$

where $-$ $(+)$ indicates a negative (positive) correlation.

We collected data from *www.amazon.com* and *www.ebay.com*, focusing on one product category, namely pans. We think this product choice is appropriate not only because there are many observations for both eBay and Amazon but also because pans require some minimum consideration and search before a purchase decision is made. In addition, we analyze the theoretically most relevant case where Amazon acts as a marketmaking middleman for this product category: For 32% of the sample, Amazon acts as a middleman; for the other 68%, Amazon acts as a market maker. Our data matches each product on sale at Amazon to a list of the offers at eBay. Using the information on individual prices and sellers, we construct proxies for key parameters in the model: λ^b, β and B as explained below (see Appendix III for more details). Since our data are not experimental, our evidence should be interpreted as suggestive rather

valuable new development exclusives" by The Real Deal introduces more detailed information on how brokers cooperate with developers, which is available in http://therealdeal.com/2016/03/15/inside-the-fight-for-manhattans-most-valuable-new-development-exclusives/ (visited on July 15, 2016).

than causal.

Table 2.1: Regressions for Amazon's intermediation mode

	(1) Linear	(2) Linear	(3) Probit	(4) Probit
sellersEbayRelative	-0.00630***		-0.00778***	
	(0.000765)		(0.00119)	
sellersEbayRefined		-0.00159**		-0.00156*
		(0.000595)		(0.000698)
sellersAmazon		-0.000552		-0.000669
		(0.000917)		(0.000996)
log(rank)	-0.103***	-0.107***	-0.105***	-0.110***
	(0.00442)	(0.00463)	(0.00500)	(0.00517)
priceDiff	0.105***	0.111***	0.124***	0.133***
	(0.00820)	(0.00830)	(0.0107)	(0.0111)
log(price)	0.0390***	0.0406***	0.0484***	0.0510***
	(0.00608)	(0.00613)	(0.00680)	(0.00688)
listedDays	0.0602***	0.0660***	0.0717***	0.0784***
	(0.00480)	(0.00486)	(0.00599)	(0.00604)
Observations	6457	6457	6457	6457
Adjusted R^2	0.136	0.130		

Note: Columns (1) and (2) use linear probability model, and columns (3) and (4) use Probit models. For Probit models, the marginal effects evaluated at the sample mean are reported. *sellersEbayRefined* is the number of sellers on a refined list of sellers on eBay by matching the title of offers with the Amazon product title and restricting the price of offers between 0.5 and 1.5 times the Amazon price. *sellersAmazon* is the number of third-party sellers on Amazon. Robust standard errors are reported in parentheses. Finally, * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$.

Table 2.1 summarizes various cross-sectional regressions of Amazon's intermediation mode, represented by *sellByAmazon*, which is a dummy variable that takes a value of 1 if the product is sold by Amazon. It is an indicator that Amazon is an active middleman for that product. Other variables we use for the linear regression in Column 1 are discussed below. *sellersEbayRelative* is a proxy for λ^b and is defined as the number of sellers on eBay divided by the number of third-party sellers on Amazon. *rank* is a proxy for total demand B and is defined a the sales rank within the product category. Rankings are negatively correlated with sales (e.g., a product with a rank value of 100 is associated with more sales than a product with a rank value 200). *priceDiff* is a proxy for β and is defined as the log of the median price of eBay offers minus the log of the Amazon price. *listedDays* controls for the number of days since the product was first listed on Amazon. Our theoretical model predicts that *sellByAmazon* is negatively correlated to *sellersEbayRelative*, negatively correlated to *log(rank)* and positively correlated to *priceDiff*. As shown in Table 2.1, all explanatory variables have the expected signs and are statistically significant in all the specifications (including those where we use alternative proxies and probit regressions).[23] To quantify the effect of available options in eBay, we find that the chance that Amazon acts as a middleman decreases by 3.7 percent for a one-standard deviation increase in *sellersEbayRelative* (λ^b), and increases by 0.1 percent for a one percent increase in the median eBay price relative to the Amazon price (proxied by *priceDiff*, β). In Appendix III, we give more detailed information on the data and we experiment with a number of different specifications, but none of them alters our main results.

2.7 Conclusion

This paper developed a model in which market structure is determined endogenously by the choice of intermediation mode. We considered two representative business modes of intermediation that are widely used in real-life markets: a marketmaking mode and a middleman mode. We derived conditions for a mixture of the two modes, a *marketmaking middleman* to emerge.

One implication of our theory is that intermediaries can use a platform to reduce competition with sellers in the decentralized market. However, this is done by inducing consumers to search excessively and so generates inefficiencies. For future research, it would be interesting to examine this issue from the viewpoint of a regulator.

[23] Recent work by Zhu and Liu (2018) also examines empirically the product choice by Amazon. While their approach is very different from ours, it is interesting to note that both share some common evidence. For instance, we find that the number of sellers on Amazon is negatively associated with the likelihood of Amazon to act as a middleman. This may reflect a crowding out effect of Amazon on third-party sellers. Similarly, Zhu and Liu (2018) find that Amazon's entry could discourage third-party sellers and eventually force them to leave the platform. Also, our evidence suggests that Amazon is more likely to sell more established products with higher prices. This is also consistent with Zhu and Liu (2018)'s findings that Amazon may be targeting on successful products to exploit the surplus from third-party sellers.

Appendices

Appendix I: Omitted proofs

Proof of Lemma 2.1

Using $K \leq x^m$ and (2.16), the intermediary's problem can be written as

$$\begin{aligned}
\max_{x^m,f,K} \Pi\left(x^m, f, K\right) &= S(1-e^{-x^s})f + \min\left\{K, x^m\right\} p^m - Kc \\
&= S(1-e^{-\frac{B-x^m}{S}})f + K(1-\lambda^b e^{-\frac{B-x^m}{S}}\beta)(1-c) - x^m e^{-\frac{B-x^m}{S}}\left(v(x^m,K)-f\right)
\end{aligned}$$

subject to (2.13) and

$$0 < K \leq x^m < B.$$

Observe that: $\lim_{x^m \to B} \Pi\left(x^m, f, K\right) = \tilde{\Pi}(B)$ and $\lim_{x^m \to 0} \Pi\left(x^m, f, K\right) = \tilde{\Pi}(0)$, where $\tilde{\Pi}(B) = B(1-\lambda^b\beta)(1-c)$ is the profit for the pure middleman mode (2.14) and $\tilde{\Pi}(0) = S(1-e^{-\frac{B}{S}})f$ is the profit for the pure market-maker mode (2.15). Hence, we can compactify the constraint set and set up a general problem to pin down a profit-maximizing intermediation mode using the following Lagrangian:

$$\mathcal{L} = \Pi\left(x^m, f, K\right) + \mu_k(x^m - K) + \mu_b(B - x^m) + \mu_v\left(v(x^m,K) - f\right) + \mu_0 K,$$

where the μ's ≥ 0 are the Lagrange multiplier of each constraint. In the proof of Proposition 2.2, we show that the following first order conditions are necessary and sufficient:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x^m} &= \frac{\partial \Pi\left(x^m, f, K\right)}{\partial x^m} + \mu_k - \mu_b + \mu_v \frac{\partial v\left(x^m, K\right)}{\partial x^m} = 0, & (24)\\
\frac{\partial \mathcal{L}}{\partial f} &= \frac{\partial \Pi\left(x^m, f, K\right)}{\partial f} - \mu_v = 0, & (25)\\
\frac{\partial \mathcal{L}}{\partial K} &= \frac{\partial \Pi\left(x^m, f, K\right)}{\partial K} - \mu_k + \mu_0 + \mu_v \frac{\partial v\left(x^m, K\right)}{\partial K} = 0. & (26)
\end{aligned}$$

The solution is characterized by these and the complementary slackness conditions of the four constraints.

We now prove the claims in the lemma. First, (25) implies that we must have

$$\mu_v = S(1-e^{-x^s}) + x^m e^{-x^s} > 0,$$

which implies the binding constraint (2.13),

$$f = v\left(x^m, K\right) = \left[1 - \lambda^b e^{-\frac{B-x^m}{S}}\beta - \lambda^s\left\{1 - \frac{K}{B} - \frac{S}{B}(1-e^{-\frac{B-x^m}{S}})\right\}(1-\beta)\right](1-c).$$

Second, applying μ_v from (25) into (26) gives

$$\mu_k = \left[1 - \lambda^b e^{-\frac{B-x^m}{S}}\beta + \lambda^b\left(1 - e^{-\frac{B-x^m}{S}}\right)(1-\beta)\right](1-c) + \mu_0 > 0,$$

which implies that $K = x^m$. This completes the proof of Lemma 2.1. ■

Proof of Proposition 2.2

⊙ Active platform. First of all, we show that the platform will always be active (i.e., $x^m < B$) in equilibrium. Substituting μ_k, μ_v into (24),

$$\begin{aligned}
(1-c)^{-1}(\mu_b - \mu_0) &= -e^{-\frac{B-x^m}{S}}\left[1 - \lambda^b e^{-\frac{B-x^m}{S}}\beta - \lambda^b\left\{\frac{B}{S} - (1-e^{-\frac{B-x^m}{S}})\right\}(1-\beta)\right] \\
&\quad -\lambda^b \frac{x^m}{S} e^{-\frac{B-x^m}{S}} + 1 - \lambda^b\beta + \lambda^b(1-e^{-\frac{B-x^m}{S}})^2 \\
&\equiv \phi(x^m \mid B, S, \beta, \lambda^b). & (27)
\end{aligned}$$

Suppose that the solution is $x^m = B$. Then, (27) yields $\phi(B \mid \cdot) = (1-c)^{-1}\mu_b = -\frac{B}{S}\lambda^b\beta < 0$, which contradicts $\mu_b \geq 0$. Hence, the solution must satisfy $x^m < B$ (which implies $\mu_b = 0$).

⊙ Market-making middleman or pure market-maker. Second, we derive the condition for a pure market-maker $x^m = 0$ or a market-making middleman $x^m > 0$. Since $\phi(B \mid \cdot) < 0$, if $\phi(0 \mid \cdot) > 0$, there exists $x^m \in (0, B)$ that satisfies $\phi(x^m \mid \cdot) = 0$, i.e. a market-making middleman. Further,

$$\frac{\partial\phi(x^m \mid \cdot)}{\partial x^m}\mid_{\phi=0} = -\frac{1}{S}\left[1-\lambda^b\beta+\lambda^b(1-e^{-\frac{B-x^m}{S}})2\lambda^b(1-e^{-\frac{B-x^m}{S}})e^{-\frac{B-x^m}{S}}\right] - \frac{\lambda^b}{S}e^{-\frac{B-x^m}{S}}(1-e^{-\frac{B-x^m}{S}}) < 0.$$

This implies that the allocation of the middleman sector $x^m \in (0, B)$ is unique (if it exists), and that if $\phi(0 \mid \cdot) < 0$ then $\phi(x^m \mid \cdot) < 0$ for all $x^m \in [0, B]$ and the solution must be a pure market maker, $x^m = 0$.

Now, we need to investigate the sign of it:

$$\begin{aligned}\phi(0 \mid B, S, \beta, \lambda^b) &= -e^{-x}\left[1-\lambda^b e^{-x}\beta - \lambda^b\left(x-1+e^{-x}\right)(1-\beta)\right] + 1 - \lambda^b\beta + \lambda^b(1-e^{-x})^2 \\ &\equiv \Theta(x),\end{aligned}$$

where $x \equiv \frac{B}{S}$. Observe that:

$$\Theta(0) = 0 < 1 - \lambda^b\beta + \lambda^b = \Theta(\infty),$$

and

$$\frac{\partial\Theta(x)}{\partial x} = e^{-x}\left[1-\lambda^b x + \lambda^b\beta(x-2) + 4\lambda^b(1-e^{-x})\right].$$

This derivative has the following properties: $\frac{\partial\Theta(x)}{\partial x}\mid_{x=0} = 1 - 2\lambda^b\beta$;

$$\frac{\partial\Theta(x)}{\partial x}\mid_{\Theta(x)=0} = 1 - \lambda^b\beta(1+e^{-x}) + \lambda^b(1-e^{-x})(1+2e^{-x}) \equiv \Upsilon(x).$$

There are two cases.

- When $\lambda^b\beta \leq \frac{1}{2}$, we have $\frac{\partial\Theta(x)}{\partial x}\mid_{x=0} \geq 0$ and $\frac{\partial\Theta(x)}{\partial x}\mid_{\Theta(x)=0} > 0$, implying that no $x \in (0, \infty)$ exists such that $\Theta(x) = 0$. Hence, $\Theta(x) = \phi(0 \mid \cdot) > 0$ for all $x \in (0, \infty)$.
- When $\lambda^b\beta > \frac{1}{2}$, we have $\frac{\partial\Theta(x)}{\partial x}\mid_{x=0} < 0$. Hence, there exists at least one $\bar{x} \in (0, \infty)$ such that $\Theta(x) < 0$ for $x < \bar{x}$ and $\Theta(x) \geq 0$ for $x \geq \bar{x}$. Below we show that such a value has to be unique. For this purpose, observe that:

  $$\Upsilon(0) = 1 - 2\lambda^b\beta < 0 < 1 + \lambda^b(1-\beta) = \Upsilon(\infty), \frac{\partial\Upsilon(x)}{\partial x} = \lambda^b e^{-x}(4e^{-x} - 1 + \beta),$$
  $$\frac{\partial\Upsilon(x)}{\partial x}\mid_{x=0} = \lambda^b(3+\beta) > 0, \frac{\partial^2\Upsilon(x)}{\partial x^2}\mid_{\frac{\partial\Upsilon(x)}{\partial x}=0} = -4e^{-x}\lambda^b e^{-x} < 0.$$

  These properties imply that there exists an $x' \in (0, \infty)$ such that $\Upsilon(x) <$ for all $x < x'$ and $\Upsilon(x) \geq 0$ for all $x \geq x'$. This implies that $\bar{x}$ is unique.

To summarize, we have shown that if $\lambda^b\beta \leq \frac{1}{2}$ then the solution is a market-making middleman $x^m \in (0, B)$ for all $x = \frac{B}{S} \in (0, \infty)$. If $\lambda^b\beta > \frac{1}{2}$ then there exists a unique critical value $\bar{x} \in (0, \infty)$ such that the solution is a market-making middleman for $x \geq \bar{x}$ and is a pure market-maker $x^m = 0$ for $x < \bar{x}$.

⊙ Second order condition. Finally, we verify the second order condition. Define $\mathbf{X} \equiv [x^m, f, K]$ and write the binding constraints as

$$h_1(\mathbf{X}) = v(x^m, K) - f, \; h_2(\mathbf{X}) = x^m - K.$$

The solution characterized above is a maximum if the Hessian of $\mathcal{L}$ with respect to $\mathbf{X}$ at the solution denoted by $(\mathbf{X}^*, \bar{}^*)$ is negative definite on the constraint set $\{\mathbf{w} : \mathbf{Dh}(\mathbf{X}^*)\,\mathbf{w} = 0\}$ with $\mathbf{h} \equiv [h_1(\mathbf{X}), h_2(\mathbf{X})]$. This can be verified by using the bordered Hessian matrix, denoted by H.

$$
\begin{aligned}
H &\equiv \begin{bmatrix} 0 & \mathbf{Dh}(\mathbf{X}^*) \\ \mathbf{Dh}(\mathbf{X}^*)^T & D^2_{\mathbf{X}}\mathcal{L}(\mathbf{X}^*, \bar{}^*) \end{bmatrix} \\
&= \begin{bmatrix}
0 & 0 & \frac{\partial h_1}{\partial x^m} & \frac{\partial h_1}{\partial f} & \frac{\partial h_1}{\partial K} \\
0 & 0 & \frac{\partial h_2}{\partial x^m} & \frac{\partial h_2}{\partial f} & \frac{\partial h_2}{\partial K} \\
\frac{\partial h_1}{\partial x^m} & \frac{\partial h_2}{\partial x^m} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^{m2}} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial f \partial x^m} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial K \partial x^m} \\
\frac{\partial h_1}{\partial f} & \frac{\partial h_2}{\partial f} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^m \partial f} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial f^2} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial K \partial f} \\
\frac{\partial h_1}{\partial K} & \frac{\partial h_2}{\partial K} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^m \partial K} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial f \partial K} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial K^2}
\end{bmatrix} \\
&= \begin{bmatrix}
0 & 0 & -\frac{\lambda^b}{S} e^{-x^s}(1-c) & -1 & \frac{\lambda^s}{B}(1-\beta)(1-c) \\
0 & 0 & 1 & 0 & -1 \\
-\frac{\lambda^b}{S} e^{-x^s}(1-c) & 1 & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^{m2}} & \frac{x^m}{S} e^{-x^s} & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^m \partial K} \\
-1 & 0 & \frac{x^m}{S} e^{-x^s} & 0 & 0 \\
\frac{\lambda^s}{B}(1-\beta)(1-c) & -1 & \frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^m \partial K} & 0 & 0
\end{bmatrix}
\end{aligned}
$$

with

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^{m2}} &= -\frac{1}{S} e^{-x^s} v \\
&+ \left(-\frac{1}{S}\frac{x^m}{S} \lambda^b e^{-x^s} \beta + 2\left(1 + \frac{x^m}{S}\right) e^{-x^s} \frac{\lambda^b}{S} e^{-x^s} - \frac{\lambda^b}{S}\left(1 - e^{-x^s}\right) e^{-x^s} \right)(1-c), \\
\frac{\partial^2 \mathcal{L}(\mathbf{X}^*, \bar{}^*)}{\partial x^m \partial K} &= -\left(\frac{\lambda^b}{S} e^{-x^s} \beta + \left(1 + \frac{x^m}{S}\right) e^{-x^s} \frac{\lambda^s}{B}(1-\beta) \right)(1-c).
\end{aligned}
$$

The determinant is given by

$$
|H| = -\frac{1}{S}\left[e^{-x^s} v(x^m, K^*) + \frac{x^m}{S} \lambda^b e^{-x^s} \beta (1-c) + 3\lambda^b e^{-x^s}\left(1 - e^{-x^s}\right)(1-c) \right] < 0.
$$

Thus, the sufficient condition is satisfied. This completes the proof of Proposition 2.2. ■

Proof of Corollary 2.1

In (27), we have:

$$
\begin{aligned}
\frac{\partial \phi(x^m \mid ., ., \beta, .)}{\partial \beta}\Big|_{(\phi(x^m|\cdot)=0)} &= -\lambda^b(1 - e^{-2x^s}) - \lambda^b e^{-x^s}\left(\frac{B}{S} - 1 + e^{-x^s}\right) < 0, \\
\frac{\partial \phi(x^m \mid B, ., ., .)}{\partial B}\Big|_{(\phi(x^m|\cdot)=0)} &= \frac{1}{S}\left[1 + \lambda^b(1-\beta) - \lambda^b e^{-2x^s} + \lambda^b e^{-x^s}(1 - e^{-x^s})\right] > 0 \\
\frac{\partial \phi(x^m \mid ., S, ., .)}{\partial S}\Big|_{(\phi(x^m|\cdot)=0)} &= -\frac{x^s}{S}\left[1 + \lambda^b(1-\beta) - \lambda^b e^{-2x^s} + \lambda^b e^{-x^s}\left(\frac{B}{x^s} - e^{-x^s}\right)\right] < 0 \\
\frac{\partial \phi(x^m \mid ., ., ., \lambda^b)}{\partial \lambda^b}\Big|_{(\phi(x^m|\cdot)=0)} &= -\frac{1 - e^{-x^s}}{\lambda^b} < 0.
\end{aligned}
$$

Hence, since $\frac{\partial \phi(x^m|\cdot)}{\partial x^m}\big|_{(\phi(x^m|\cdot)=0)} < 0$ (see the proof of Proposition 2.2), it follows that: $\frac{\partial x^m}{\partial \beta} < 0$; $\frac{\partial x^m}{\partial B} < 0$; $\frac{\partial x^m}{\partial S} > 0$; $\frac{\partial x^m}{\partial \lambda^b} < 0$. This completes the proof of Corollary 2.1. ■

Proof of Proposition 2.3

The proof takes steps that are very similar to the ones we made in the proof of Proposition 2.1. With the non-linear matching function, the intermediary's profit function is modified to

$$\begin{aligned}\Pi\left(x^m, f, K\right) &= S(1-e^{-x^s})f + \min\left\{K, x^m\right\} p^m \\ &= S(1-e^{-\frac{B-x^m}{S}})f + K(1-\lambda^b(x^D)\beta) - x^m e^{-\frac{B-x^m}{S}}\left(v(x^m, K) - f\right),\end{aligned}$$

where $x^D = \frac{\max\{B-\min\{x^m,K\}-S(1-e^{-x^s}),0\}}{Se^{-x^s}}$, and the surplus function to

$$v\left(x^m, K\right) = 1 - \lambda^b(x^D)\beta - \lambda^s(x^D)(1-\beta).$$

With these profit and surplus functions, the constraints and the Lagrangian remain unchanged, and the first orders are given by (24) – (26) (the second order conditions are presented below). As before, (25) implies that we must have

$$\mu_v = S(1-e^{-x^s}) + x^m e^{-x^s} > 0,$$

and the binding constraint (2.13). Further, substituting μ_v from (25) into (26) gives

$$\mu_k = \mu_0 + 1 - \lambda^b(x^D)\beta + \frac{K}{Se^{-x^s}}\lambda^{b\prime}(x^D)\beta \tag{28}$$

$$+ \frac{1-e^{-x^S}}{e^{-x^s}}\left(\lambda^{b\prime}(x^D)\beta + (\lambda^b(x^D) + x^D\lambda^{b\prime}(x^D))(1-\beta)\right). \tag{29}$$

Substituting μ_k, μ_v into (24) gives,

$$\begin{aligned}\mu_b &= \mu_0 - e^{-x^s}\left(1 - \lambda^b(x^D)\beta - \lambda^b(x^D)x^D(1-\beta)\right) + 1 - \lambda^b(x^D)\beta + \frac{B-K}{S}\frac{K}{Se^{-x^s}}\lambda^{b\prime}(x^D)\beta \\ &\quad + \frac{B-K}{S}\frac{1-e^{-x^S}}{e^{-x^s}}\left(\lambda^{b\prime}(x^D)\beta + (\lambda^b(x^D) + x^D\lambda^{b\prime}(x^D))(1-\beta)\right).\end{aligned} \tag{30}$$

Suppose now that $x^m = B$ and $K > 0$. Then, $\mu_k > 0$ in (28) if and only if

$$1 - \lambda^b(x^D)\beta + \frac{K}{S}\lambda^{b\prime}(x^D)\beta > 0,$$

and $\mu_b \geq 0$ in (30) if and only if

$$\frac{B-K}{S}\left[(1-\beta)\lambda^b(x^D) + \frac{K}{S}\lambda^{b\prime}(x^D)\beta\right] \geq 0,$$

with $x^D = \frac{B-K}{S}$. Both of these conditions are satisfied only when $K = B$ (which implies $x^D = 0$, satisfying the latter condition) and

$$1 - \lambda^b(0)\beta + \frac{B}{S}\lambda^{b\prime}(0)\beta > 0 \tag{31}$$

(satisfying the former condition with $x^D = 0$). Under this condition, the solution is unique, $K = B = x^m$, $x^s = 0$ and $f = v(B, B)$. Hence, we have shown that the solution can be a pure middleman $x^s = 0$ only if (31) holds and otherwise the solution must be $x^s > 0$ (either a marketmaking middleman or a pure marketmaker).

Finally, we verify the second order condition. With the modified profit and surplus functions, as before, the bordered Hessian matrix is given by

$$\begin{aligned}
H &\equiv \begin{bmatrix} 0 & D\mathbf{h}(\mathbf{X}^*) \\ D\mathbf{h}(\mathbf{X}^*)^T & D^2_{\mathbf{X}} L(\mathbf{X}^*, \bar{}^*) \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & \frac{\partial h_1}{\partial x^m} & \frac{\partial h_1}{\partial f} & \frac{\partial h_1}{\partial K} \\ 0 & 0 & \frac{\partial h_2}{\partial x^m} & \frac{\partial h_2}{\partial f} & \frac{\partial h_2}{\partial K} \\ \frac{\partial h_1}{\partial x^m} & \frac{\partial h_2}{\partial x^m} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial x^{m2}} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial f \partial x^m} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial K \partial x^m} \\ \frac{\partial h_1}{\partial f} & \frac{\partial h_2}{\partial f} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial x^m \partial f} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial f^2} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial K \partial f} \\ \frac{\partial h_1}{\partial K} & \frac{\partial h_2}{\partial K} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial x^m \partial K} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial f \partial K} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial K^2} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & \frac{\partial v(\mathbf{X}^*)}{\partial x^m} & -1 & \frac{\partial v(\mathbf{X}^*)}{\partial K} \\ 0 & 0 & 1 & 0 & -1 \\ \frac{\partial v(\mathbf{X}^*)}{\partial x^m} & 1 & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial x^{m2}} & \frac{B}{S} & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial x^m \partial K} \\ -1 & 0 & \frac{B}{S} & 0 & 0 \\ \frac{\partial v(\mathbf{X}^*)}{\partial K} & -1 & \frac{\partial^2 L(\mathbf{X}^*,\bar{}^*)}{\partial x^m \partial K} & 0 & 0 \end{bmatrix}
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial^2 L(\mathbf{X}^*, \bar{}^*)}{\partial x^{m2}} &= -\frac{1}{S} f - B \frac{\partial^2 \lambda^b}{\partial x^{m2}}(\mathbf{X}^*)\beta - (2 + \frac{B}{S}) \frac{\partial v(\mathbf{X}^*)}{\partial x^m}, \\
\frac{\partial^2 L(\mathbf{X}^*, \bar{}^*)}{\partial x^m \partial K} &= -\frac{\partial \lambda^b}{\partial x^m}\beta - B \frac{\partial^2 \lambda^b}{\partial x^m \partial K}\beta - \frac{B}{S}\frac{\partial v(\mathbf{X}^*)}{\partial K}.
\end{aligned}$$

The determinant is $|H| = -\frac{1}{S}(1 - \lambda^b(0)\beta) - \frac{B}{S^2}\lambda^{b\prime}(0)\beta < 0$. This completes the proof of Proposition 2.3. ■

Proof of Proposition 2.4

As stated in the main text, for $\alpha S \geq B$ the intermediary can achieve the highest possible profit by choosing to be a pure middleman. What remains here is to prove the proposition for $\alpha S < B$. Applying the analysis in the previous section, we derive that the seller value equals, $W(x^s) = (1 - e^{-x^s} - x^s e^{-x^s})(1 - f)$ and the indifferent condition for buyers becomes, $V^m(x^m) = V^s(x^s)$ where $V^m(x^m) = \min\{\frac{K}{x^m}, 1\}(1 - p^m)$ and $V^s(x^s) = e^{-x^s}(1 - p^s)$. The binding participation constraint for buyers implies that $p^m = 1 - \frac{\lambda^b}{\min\{\frac{K}{x^m},1\}}$ and $f = 1 - \frac{\lambda^b}{e^{-x^s}}$, and the binding condition, (2.19), implies that $p^w = (1 - e^{-x^s} - x^s e^{-x^s})\frac{\lambda^b}{e^{-x^s}}$.

To guarantee $f \geq 0$, it is sufficient to assume that

$$\lambda^b \leq e^{-\frac{B - \alpha S}{S}}.$$

This also guarantees $p^m - p^w = 1 - \lambda^b(1 - \frac{1 - e^{-x^s} - x^s e^{-x^s}}{e^{-x^s}}) > 0$ and that profits are non negative.

Using all these expressions of prices and fee, we can write the profit function as

$$\Pi(x^m, K) = (S - K)(1 - e^{-x^s})(1 - \frac{\lambda^b}{e^{-x^s}}) + \min\{K, x^m\} - x^m\lambda^b - K(1 - e^{-x^s} - x^s e^{-x^s})\frac{\lambda^b}{e^{-x^s}},$$

where $x^s = \frac{B - x^m}{S - K}$. Differentiation yields

$$\frac{\partial \Pi(x^m, K)}{\partial x^m} = \frac{S}{S - K}\frac{1 - e^{-x^s}}{e^{-x^s}}\lambda^b + \frac{\partial \min\{K, x^m\}}{\partial x^m} - e^{-x^s}, \tag{32}$$

which is positive if $\min\{K, x^m\} = x^m$. Hence, the solution has to satisfy $x^m \geq K$.

Observe that: $\lim_{x^m \to B} \Pi(x^m, K) = \Pi$ and $\lim_{x^m \to 0} \Pi(x^m, K) = \tilde{\Pi}(0)$, where $\Pi =$

$\alpha S - B\lambda^b$ is the profit for the pure middleman mode and $\tilde{\Pi}(0) = S(1-e^{-\frac{B}{S}})(1-\frac{\lambda^b}{e^{-\frac{B}{S}}})$ is the profit for the pure market-maker mode. Hence, as before, we can find a profit-maximizing intermediation mode using the following Lagrangian:

$$\mathcal{L} = \Pi(x^m, K) + \mu_k(x^m - K) + \mu_b(B - x^m) + \mu_0 K + \mu_s(\alpha S - K).$$

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial x^m} = \frac{\partial \Pi(x^m, K)}{\partial x^m} + \mu_k - \mu_b = 0, \quad (33)$$

$$\frac{\partial \mathcal{L}}{\partial K} = \frac{\partial \Pi(x^m, K)}{\partial K} - \mu_k + \mu_0 - \mu_s = 0, \quad (34)$$

where

$$\frac{\partial \Pi(x^m, K)}{\partial K} = e^{-x^s} + x^s e^{-x^s} - \frac{S}{S-K}\frac{1-e^{-x^s}}{e^{-x^s}}\lambda^b x^s.$$

Suppose $x^m = B$. Then, we must have $\mu_k = 0$ (since $B > \alpha S \geq K$) and so (33) implies we also must have $\frac{\partial \Pi(x^m,K)}{\partial x^m}\mid_{(x^m=B)} = \mu_b \geq 0$. However, $\frac{\partial \Pi(x^m,K)}{\partial x^m}\mid_{(x^m=B)} = -1 < 0$, a contradiction. Hence, the solution must satisfy $x^m < B$ (and $\mu_b = 0$), i.e., an active platform.

Summing up the two first order conditions with $\mu_b = 0$,

$$\begin{aligned}
\mu_s - \mu_0 &= \frac{\partial \Pi(x^m, K)}{\partial K} + \frac{\partial \Pi(x^m, K)}{\partial x^m} \\
&= x^s e^{-x^s} + (1 - x^s)\frac{S}{S-K}\frac{1-e^{-x^s}}{e^{-x^s}}\lambda^b \\
&= -x^s \frac{\partial \Pi(x^m, K)}{\partial x^m} + \frac{S}{S-K}\frac{1-e^{-x^s}}{e^{-x^s}}\lambda^b > 0
\end{aligned}$$

where the last inequality follows from (33) and $\mu_b = 0$ that implies $\frac{\partial \Pi(x^m,K)}{\partial x^m} = -\mu_k \leq 0$. This implies $\mu_s > 0$, i.e., the binding resource constraint (2.18), which implies $K = \alpha S$. This completes the proof of Proposition 2.4. ■

Proof of Proposition 2.5

In our endowment economy, the middleman's inventory purchase influences market tightness not only in the C market platform, but also in the D market. Given all sellers are in the D market, the probability that a buyer meets a seller available for trade in the D market changes from $\lambda^b e^{-x^s}$ to $\lambda^b \frac{S-K}{S} e^{-x^s}$. With this change and using the analysis of multi-market search shown in the previous section, the value of sellers, $W(x^s) = (1 - e^{-x^s} - x^s e^{-x^s})(v(x^m, K) - f)$ and the middleman's price, $p^m = 1 - \lambda^b \frac{S-K}{S} e^{-x^s} - \frac{x^m e^{-x^s}}{\min\{x^m, K\}}(v(x^m, K) - f)$, where $v(x^m, K) = 1 - \lambda^b \frac{S-K}{S} e^{-x^s}$. Substituting these expressions into the profit function, it becomes immediate that the profit is strictly increasing in the fee f. Hence, the incentive constraints are binding, $f = v(x^m, K)$. Using this result, we can write the profit function as

$$\Pi(x^m, K) = (S-K)(1-e^{-x^s})(1-\lambda^b \frac{S-K}{S}e^{-x^s}) + \min\{K, x^m\}(1-\lambda^b \frac{S-K}{S}e^{-x^s}),$$

where $x^s = \frac{B-x^m}{S-K}$. Differentiation yields

$$\begin{aligned}
\frac{\partial \Pi(x^m, K)}{\partial x^m} &= -e^{-x^s}\left(1 - \lambda^b \frac{S-K}{S}e^{-x^s} - \lambda^b \frac{S-K}{S}(1-e^{-x^s})\right) - \frac{\min\{K, x^m\}}{S}\lambda^b e^{-x^s} \\
&\quad + \frac{\partial \min\{K, x^m\}}{\partial x^m}\left(1 - \lambda^b \frac{S-K}{S}e^{-x^s}\right),
\end{aligned}$$

which is negative if $\min\{K, x^m\} = K$. Hence, the solution has to satisfy $x^m \leq K$.

Suppose $x^m = B$. Then,

$$\frac{\partial \Pi(x^m, K)}{\partial x^m} \mid_{x^m = B} = -\frac{B}{S}\lambda^b < 0.$$

Hence, the solution has to be $x^m < B$, i.e., an active platform.

The Lagrangian the becomes,

$$\mathcal{L} = \Pi(x^m, K) + \mu_0 x^m + \mu_k(K - x^m) + \mu_s(\alpha S - K).$$

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial x^m} = \frac{\partial \Pi(x^m, K)}{\partial x^m} + \mu_0 - \mu_k = 0, \tag{35}$$

$$\frac{\partial \mathcal{L}}{\partial K} = \frac{\partial \Pi(x^m, K)}{\partial K} + \mu_k - \mu_s = 0, \tag{36}$$

where

$$\begin{aligned} \frac{\partial \Pi(x^m, K)}{\partial K} &= x^s e^{-x^s}\left(1 - \lambda^b \frac{S-K}{S} e^{-x^s} + x^s \lambda^b \frac{S-K}{S}(1 - e^{-x^s})\right) + \frac{x^s x^m}{S}\lambda^b e^{-x^s} \\ &+ \left(1 - e^{-x^s}\right)\left(1 - 2\lambda^b \frac{S-K}{S} e^{-x^s}\right) + \frac{x^m}{S}\lambda^b e^{-x^s}. \end{aligned}$$

Combining (35) and (36),

$$\frac{\partial \Pi(x^m, K)}{\partial x^m} + \frac{\partial \Pi(x^m, K)}{\partial K} = x^s e^{-x^s}\left[\left(1 - \lambda^b \frac{S-K}{S} e^{-x^s}\right) + \frac{S-K}{S}(1 - e^{-x^s})\lambda^b + \frac{x^m}{S}\lambda^b\right] = \mu_s - \mu_0,$$

which implies $\mu_s > 0$ and $K = \alpha S$. This completes the proof of Proposition 2.5. ■

Proof of Proposition 2.6

⊙ Single-market search. Note $K = x^m$ in equilibrium. With single-market search, the problem of a middleman can be described as

$$\tilde{\Pi}(K) = \max_{p^m, K} \min\{K, B\} p^m - C(K) \text{ s.t. } V^m = 1 - p^m \geq \lambda^b.$$

The solution is $p^m = 1 - \lambda^b$ and $K = B$ (as $C'(K) < 1 - \lambda^b$ for $K < B$), and so the middleman's profit is $\tilde{\Pi}(B) = B(1 - \lambda^b) - C(B)$. When a platform is open, individual sellers solve:

$$\max_x (1 - e^{-x})(p - f^s) \text{ s.t. } V^s = \frac{1 - e^{-x}}{x}(1 - p - f^b)$$

The optimal solution is described as $V^s = e^{-x}(1 - f)$ with $f \equiv f^b + f^s$. In an equilibrium with a market-making middleman, it has to hold that $x = x^s$ and $V^s = V^m = 1 - p^m \geq \lambda^b$, which implies $p^m = 1 - \lambda^b$ and $f = 1 - \lambda^b e^{x^s}$. Hence, the problem of a market-making middleman is given by (2.20), with $\lim_{x^m \to B} \Pi(x^m) = \tilde{\Pi}(B)$. Observe in (2.21) that: $\Theta_{Sfoc}(B) = -C'(B) < 0$ and $\Theta'_{Sfoc}(x^m) = -\frac{1}{S}(e^{-x^s} + \lambda^b e^{x^s}) - C''(x^m) < 0$. Therefore, a market-making middleman is profit-maximizing if

$$\Theta_{Sfoc}(0) = 1 - e^{-\frac{B}{S}} + \lambda^b(e^{\frac{B}{S}} - 1) - C'(0) > 0. \tag{37}$$

Otherwise, the pure market-maker mode is selected.

⊙ Multi-market search. Note $K = x^m$ in equilibrium. With multi-market search, the only modification is to introduce the marginal inventory cost $C'(x^m)$ in the first order condition. Using (2.22), we can write the first order condition as:

$$\Theta_{Mfoc}(x^m) \equiv (1 - e^{-x^s})(1 - 2\lambda^b e^{-x^s}) - \frac{\lambda^b x^m e^{-x^s}}{S} - C'(x^m) = 0.$$

Observe that: $\Theta_{Mfoc}(B) = -\frac{\lambda^b B}{S} - C'(B) < 0$ and

$$\Theta'_{Mfoc}(x^m) = -\frac{e^{-x^s}}{S}\left[1 - \lambda^b e^{-x^s} + 3\lambda^b(1 - e^{-x^s})\right] - \frac{\lambda^b x^m e^{-x^s}}{S^2} - C''(x^m) < 0.$$

Therefore, a market-making middleman is profit-maximizing if

$$\Theta_{Mfoc}(0) = (1 - e^{-\frac{B}{S}})(1 - 2\lambda^b e^{-\frac{B}{S}}) - C'(0) > 0. \tag{38}$$

Otherwise, a pure market-maker is selected.
⊙ Comparison. Comparing (37) and (38),

$$\Theta_{Mfoc}(0) \quad = \quad \Theta_{Sfoc}(0) - \lambda^b \left[2e^{-x^s}(1 - e^{-x^s}) + e^{x^s} - 1\right] < \Theta_{Sfoc}(0),$$

implying that whenever a middleman sector is active, i.e. $x^m > 0$, with multi-market search, it has to be active with single-market search as well. Further, whenever a market-making middleman mode is selected, since $\Theta'_{Mfoc}(x^m) = \frac{\partial^2 \Pi(x^m)}{\partial x^{m2}} < 0$, the comparison given in the main text implies that the platform sector x^s $(= \frac{B - x^m}{S})$ is always larger with multi-market search than single-market search. This completes the proof of Proposition 2.6. ■

Appendix II: Participation fees

In this Appendix, we show that our main result does not change in a version of our model where the middleman's supply is not observable in the participation stage, but instead the intermediary can use participation fees/subsidy. Suppose now that in the first stage the intermediary announces a set of fees $F \equiv \{f^b, f^s, g^b, g^s\}$ for the platform, where $f^b,\ f^s \in [0,1]$ is a transaction fee charged to a buyer or a seller, respectively, and $g^b,\ g^s \in [-1,1]$ is a registration fee charged to a buyer or a seller, respectively.

As is consistent with the main analysis, we follow the literature of two-sided markets and assume that agents hold pessimistic beliefs on the participation decision of agents on the other side of the market (Caillaud and Jullien, 2003). Agents believe that the intermediary would never supply anything at all unless the C market attracts some buyers. This is the worst situation for the intermediary, and (2.3) is not the right participation constraint. A pessimistic belief of sellers means that sellers believe the number of buyers participating in the C market is zero whenever

$$\lambda^b \beta > -g^b,$$

where $\lambda^b\beta$ is the expected payoff of buyers in the D market and $-g^b$ is the payoff buyers receive in the C market (it is a participation subsidy when $g^b < 0$).

Single-market search: To induce the participation of agents under those beliefs, the best the intermediary can do is to use a divide-and-conquer strategy, denoted by h. To divide buyers and conquer sellers, referred to as $h = D_b C_s$, it is required that

$$D_b \quad : \quad -g^b \geq \lambda^b \beta, \tag{39}$$
$$C_s \quad : \quad W - g^s \geq 0. \tag{40}$$

The divide-condition D_b tells us that the intermediary should subsidize the participating buyers so that they receive at least what they would get in the D market, *even if the C market is empty*. This makes sure buyers will participate in the C market whatever happens to the other side of the market. The conquer-condition C_s guarantees the participation of sellers, by giving them a nonnegative payoff – the participation fee $g^s \geq 0$ should be no greater than the expected value of sellers in the C market, $W = W(x^s)$. Observing that the intermediary offers buyers enough to participate, sellers understand that all buyers are in the C market, the D market is empty, and so the expected payoff from the D market is zero. Here, the expected value of sellers in the C market W is defined under the sellers' belief that the intermediary will select the capacity level optimally given the full participation of buyers.

Similarly, a strategy to divide sellers and conquer buyers, referred to as $h = D_s C_b$, requires that

$$D_s \quad : \quad -g^s \geq \lambda^s (1-\beta), \tag{41}$$
$$C_b \quad : \quad V - g^b \geq 0. \tag{42}$$

where $V = \max\{V^s(x^s), V^m(x^m)\}$ is the expected value of buyers in the C market.

Given the participation decision of agents described above, the intermediary's problem of determining the intermediation fees $F = \{f^b, f^s, g^b, g^s\}$ for $h = \{D_b C_s, D_s C_b\}$ is described as

$$\Pi = \max_{F,h} \{B g^b + S g^s + \max_{p^m, K} \Pi(p^m, f, K)\},$$

subject to (39) and (40) if $h = D_b C_s$, or (41) and (42) if $h = D_s C_b$. Here, Bg^b and Sg^s are participation fees from buyers and sellers, respectively, and $\Pi(\cdot)$ is the expected profit in the C market described above. Under either of the divide-and-conquer strategies, the choice of participation fees g^i, $i = b, s$, does not influence anyone's behaviors in the C market. The choice of transaction fees affects the expected value of agents and thus the participation fees and intermediary's profits. However, it does not alter the original solution, a pure middleman, remains optimal.

Proposition .8. *With unobservable capacity and with participation fees, the intermediary sets $f > 1$, $p^m = 1$ and $K = B$. All the buyers buy from the middleman, $x^m = B$, and the platform is inactive, $x^s = 0$. The intermediary makes profits,*

$$\Pi = B - \min\{B\lambda^b\beta, S\lambda^s(1-\beta)\},$$

guaranteeing the participation of agents by $h = D_bC_s$ if $\beta < \frac{1}{2}$ and $h = D_sC_b$ if $\beta > \frac{1}{2}$.

Proof. Consider first $h = D_bC_s$. Then, by (39) and (40), $g^b = -\lambda^b\beta$ and $g^s = W$. For $f > 1$, no buyers go to the platform $x^s = 0$ and all buyers are in the middleman sector $x^m = B$, yielding $g^s = W = 0$. By selecting $K = B$ and $p^m = 1$, the intermediary makes profits,

$$\Pi = -B\lambda^b\beta + \Pi(p^m, 1, B) = (-\lambda^b\beta + 1)B.$$

To show that this is indeed the maximum profit, we have to check two possible cases. Suppose $f = f^b + f^s \le 1$ and $K = 0$. Then, $x^s = \frac{B}{S}$ and $x^m = 0$, and $g^s = W(B/S) \ge 0$, if there is a non-negative surplus in the platform for buyers, $f^b + p^s \le 1$, and for sellers, $f^s \le p^s$. The resulting profit satisfies

$$\begin{aligned}
Bg^b + Sg^s + \Pi(p^m, f, 0) &= -B\lambda^b\beta + S(1-e^{-\frac{B}{S}})(p^s - f^s) + S(1-e^{-\frac{B}{S}})f \\
&= -B\lambda^b\beta + S(1-e^{-\frac{B}{S}})(f^b + p^s) \\
&< -B\lambda^b\beta + B = \Pi
\end{aligned}$$

for all $f^b + p^s \le 1$. Hence, this is not profitable.

Suppose $f = f^b + f^s \le 1$ and $K \in (0, B]$, and both sectors have a non-negative surplus to buyers, i.e., $p^m \le 1$ and $f^b + p^s \le 1$. This leads to $x^m \in (0, B)$ and $x^s \in (0, \frac{B}{S})$ that satisfy the add-up requirement (2.1) and the indifferent condition (2.2). Then, $g^s = W(x^s) \ge 0$, and the resulting profit is

$$\begin{aligned}
&Bg^b + Sg^s + \Pi(p^m, f, K) \\
&\quad = -B\lambda^b\beta + S(1-e^{-x^s})(p^s - f^s) + S(1-e^{-x^s})f + \min\{K, x^m\}p^m \\
&\quad < -B\lambda^b\beta + Sx^s(f^b + p^s) + x^m p^m \\
&\quad \le -B\lambda^b\beta + (Sx^s + x^m)\max\{f^b + p^s, p^m\} \\
&\quad \le -B\lambda^b\beta + B = \Pi
\end{aligned}$$

for all $f^b + p^s \le 1$ and $p^m \le 1$. Hence, this is not profitable either. All in all, no deviation is profitable for $h = D_bC_s$.

Consider next $h = D_sC_b$. Then, by (41) and (42), $g^s = -\lambda^s(1-\beta)$ and $g^b = V$. When $f > 1$, no one go to the platform $x^s = 0$ and all buyers are in the middleman sector $x^m = B$ as long as $p^m \le 1$. This yields $g^b = V = V^m(B) \ge 0$ and $\Pi(p^m, f, B) = Bp^m$ with $K = B$. The profits are

$$\Pi = -S\lambda^s(1-\beta) + B(1-p^m) + \Pi(p^m, f, K) = -S\lambda^s(1-\beta) + B.$$

To show that this is indeed the maximum profit, we have to check two possible cases. Suppose $f = f^b + f^s \le 1$ and $K = 0$. Then, $x^s = \frac{B}{S}$ and $x^m = 0$, and $g^b = V = V^s(B/S) \ge 0$, if there is a non-negative surplus in the platform for buyers, $f^b + p^s \le 1$, and for sellers, $f^s \leq p^s$. This leads to

$$\begin{aligned} Sg^s + Bg^b + \Pi(p^m, f, 0) &= -S\lambda^s(1-\beta) + B\frac{1-e^{-\frac{B}{S}}}{\frac{B}{S}}(1-p^s-f^b) + S(1-e^{-\frac{B}{S}})f \\ &= -S\lambda^s(1-\beta) + S(1-e^{-\frac{B}{S}})(1-p^s+f^s) \\ &< -S\lambda^s(1-\beta) + B = \Pi \end{aligned}$$

for all $f^s \leq p^s$. Hence, this is not profitable.

Suppose $f = f^b + f^s \leq 1$ and $K \in (0, B]$, and both sectors have a non-negative surplus to buyers, i.e., $p^m \leq 1$ and $f^b + p^s \leq 1$. This leads to $x^m \in (0, B)$ and $x^s \in (0, \frac{B}{S})$ that satisfy the add-up constraint (2.1), $Sx^s + x^m = B$, and the indifferent condition (2.2), $V^s(x^s) = V^m(x^m)$. Then, $g^b = V = V^s(x^s)$, and the resulting profit is

$$\begin{aligned} &Sg^s + Bg^b + \Pi(p^m, f, K) \\ &= -S\lambda^s(1-\beta) + B\frac{1-e^{-x^s}}{x^s}(1-p^s-f^b) + S(1-e^{-x^s})f + \min\{K, x^m\}p^m. \end{aligned}$$

There are two cases. Suppose $K \geq x^m$. Then, the indifferent condition (2.2) implies that

$$p^m = 1 - \frac{1-e^{-x^s}}{x^s}(1-p^s-f^b).$$

Applying this expression to the profits, we get

$$\begin{aligned} &Sg^s + Bg^b + \Pi(p^m, f, K) \\ &= -S\lambda^s(1-\beta) + B\frac{1-e^{-x^s}}{x^s}(1-p^s-f^b) + S(1-e^{-x^s})f + x^m\left(1 - \frac{1-e^{-x^s}}{x^s}(1-p^s-f^b)\right) \\ &= -S\lambda^s(1-\beta) + (B-x^m)\frac{1-e^{-x^s}}{x^s}(1-p^s-f^b) + S(1-e^{-x^s})f + x^m \\ &= -S\lambda^s(1-\beta) + S(1-e^{-x^s})(1-p^s+f^s) + x^m \\ &< -S\lambda^s(1-\beta) + B \end{aligned}$$

for all $f^s \leq p^s$. Suppose $K < x^m$. Then, the indifferent condition implies that

$$p^m = 1 - \frac{x^m}{K}\frac{1-e^{-x^s}}{x^s}(1-p^s-f^b).$$

Applying this expression to the profits, we get

$$\begin{aligned} &Sg^s + Bg^b + \Pi(p^m, f, K) \\ &= -S\lambda^s(1-\beta) + B\frac{1-e^{-x^s}}{x^s}(1-p^s-f^b) + S(1-e^{-x^s})f + K\left(1 - \frac{x^m}{K}\frac{1-e^{-x^s}}{x^s}(1-p^s-f^b)\right) \\ &= -S\lambda^s(1-\beta) + (B-x^m)\frac{1-e^{-x^s}}{x^s}(1-p^s-f^b) + S(1-e^{-x^s})f + K \\ &= -S\lambda^s(1-\beta) + S(1-e^{-x^s})(1-p^s+f^s) + K \\ &< -S\lambda^s(1-\beta) + B \end{aligned}$$

for all $f^s \leq p^s$. Hence, any deviation is not profitable for $h = D_sC_b$.

Finally, since the intermediary makes the maximum revenue B for either h, which side should be subsidized is determined by the required costs: noting $B\lambda^b = S\lambda^s$, we have $B\lambda^b\beta \gtreqless S\lambda^s(1-\beta) \iff \beta \gtreqless \frac{1}{2}$. This completes the proof of Proposition .8. ∎

Multi-market search: With multiple-market search, any non-positive registration fee can ensure that agents are in the C market, since the participation to the C market is not exclusive. Hence, attracting one side of the market becomes less costly. By contrast, conquering the other side becomes more costly, since the conquered side still holds the trading opportunity in the D market. The D_sC_b condition with multiple-market search is

$$\begin{aligned} &D_s: \ -g^s \geq 0, \\ &C_b: \ \max\{V^s(x^s), V^m(x^m)\} - g^b \geq \lambda^b e^{-x^s}\beta(1-c). \end{aligned}$$

The divide-condition D_s tells that now a non-positive fee is sufficient to convince one side to participate. The conquer-condition C_b now needs to compensate for the outside option in the D market. Similarly, the D_bC_s condition becomes

$$\begin{aligned} &D_b: \ -g^b \geq 0, \\ &C_s: \ W(x^s) - g^s \geq \lambda^s \xi(x^s, x^m)(1-\beta)(1-c). \end{aligned}$$

Participation fees are designed to induce buyers and sellers' participation. Once agents join the C market, the participation fees become sunk costs, and will not influence their trading decision.

The intermediary's problem of choosing $F = \{f^b, f^s, g^b, g^s\}$ together with $h = \{D_bC_s, D_sC_b\}$ and $p^m, K \in [0, B]$ are described as

$$\Pi = \max_{F,h,K}\left\{Bg^b + Sg^s + \max_{p^m}\Pi(p^m, f, K)\right\}, \tag{43}$$

where $\Pi(p^m, f, K) = S(1-e^{-x^s})f + \min\{K, x^m\}p^m - Kc$. Besides the divide-and-conquer constraints, this maximization problem is also subject to the incentive constraints as described in the main text.

Proposition .9. *In the extended problem described in (43) with unobservable capacity, participation fees and multiple-market search, the determination of the profit-maximizing intermediation mode is identical to the one described in Proposition 2.2, with $g^i = 0$, $i = s, b$.*

Proof. It suffices to prove that the solution is $g^i = 0$, $i = s, b$ for each intermediation mode, since then the problem (43) will become identical to the one we have already solved in the main text. For a pure middleman mode ($x^m = B$), the intermediary sets $g^b = 0$ to divide buyers, with $p^m = 1 - \lambda^b\beta(1-c)$ satisfying (2.7). For a pure market-maker mode ($x^s = 0$), either with D_bC_s or D_sC_b, the intermediary sets the transaction fee to satisfy the binding incentive constraint (2.13), $f = v(0,0) = \left[1 - \lambda^b e^{-B/S} - \lambda^s\xi(0,0)\right](1-c)$, and $g^b = g^s = 0$.

For a hybrid mode, the intermediary's problem is subject to the incentive constraint (2.13), and p^m satisfying (2.16) so that buyers are indifferent between the two modes. We can rewrite the maximization problem (43) as a two-stage problem over a vector $\mathbf{X} \equiv (x^m, f, K) \in \mathbb{X}$, where $\mathbb{X} \equiv [0, B] \times [0, 1] \times [0, K]$:

$$\begin{aligned} &\text{Stage 1: } \max_{(f,K)} Bg^b(\mathbf{X}) + Sg^s(\mathbf{X}) + \Pi(x^m(f,K), f, K) \\ &\qquad\quad s.t.\ 0 \leq f \leq v(x^m(f,K), K),\ 0 \leq K \leq B. \\ &\text{Stage 2: } \max_{x^m} \Pi(x^m, f, K) \\ &\qquad\quad s.t.\ f \leq v(x^m, K),\ 0 \leq x^m \leq B, \end{aligned} \tag{$\mathcal{A}$}$$

where $g^b(\mathbf{X})$ and $g^s(\mathbf{X})$ are given by the binding divide-and-conquer conditions,

$$g^b(\mathbf{X}) = 0, g^s(\mathbf{X}) = \left(1 - e^{-x^s} - x^s e^{-x^s}\right)(v(x^m, K) - f),$$

if $h = D_bC_s$, or

$$g^s(\mathbf{X}) = 0, g^b(\mathbf{X}) = e^{-x^s}(v(x^m, K) - f).$$

if $h = D_s C_b$. As our objective is to prove $g^i(\mathbf{X}) = 0$, $i = s, b$, all that remains here is to show that $f = v(x^m, K)$ at the solution. However, it is immediate that the objective function in ($\mathcal{A}$) is strictly increasing in f and any change in f ($< v(x^m, K)$) does not influence the other constraints. Hence, as in the original problem, we must have $f = v(x^m, K)$. This completes the proof of Proposition .9. ■

Appendix III: Empirical examination

Data and Variables. From *www.amazon.com* and *www.ebay.com*, we retrieve all paginated results listed in the category of Amazon called: *"All Pans"*, which is a subcategory of *"Home & Kitchen: Kitchen & Dining: Cookware"*. This subcategory includes 400 pages of more than 9000 products as of August 2018.[24] For each pan, we obtain the price (*price*), the sales rank in the category *"Home & Kitchen"* (*rank*), the listing days since the first listed date on Amazon by either Amazon or some other sellers on Amazon's market-maker platform (*listedDays*), the number of third-party sellers that sell this product on Amazon (*sellersAmazon*), whether the product is sold by Amazon itself (*sellByAmazon*) and the title of the product.

Sellers could offer various prices for a product on Amazon. We obtain price information from the default page Amazon displays when users search for a product. This gives us the price at which the majority of transactions are processed. Amazon does not publish sales data but does provide a sales ranking for each product. Since ranking information is provided at different levels of categories, in order to make the sales ranking as comparable as possible, we adopt the ranking at the highest possible level *"Home & Kitchen"*. This gives us the variable *rank*.

The title of the product is used to link each product on Amazon to the outside option available at eBay as the theory develops. For each product collected on Amazon, we search its "Amazon product name" on eBay to obtain all related offers. As a proxy for the buyers' matching probability in the decentralized market λ^b, we count the number of all the offers shown in eBay's raw search result. We call this variable *sellersEbayAll*. Admittedly, this is a very noisy measure. EBay tends to provide a long list with offers that are only loosely related to the product. For example, in some cases a pan offered on eBay only matches with some key features of a pan offered on Amazon such as size but it does not match other features such as materials. In this case, we compare the similarity between the eBay product title and the Amazon product title. In some other cases, the titles are similar but the products turn out to be different. For example, searching a pan on eBay only yields an offer of the lid of the same pan on Amazon. To solve this issue, we use the following two-step procedure. We first select offers with product names similar to the Amazon product name.[25] We then refine the list by restricting the offer price between 0.5 and 1.5 times the Amazon price. The rationale for this procedure is that if the offer price is far away from the Amazon offer, the product is likely to vary in quality or could even be a distinct product. Counting the number of sellers in this refined list leads to another proxy for λ^b, *sellersEbayRefined*. This is a more precise measure for the relevant number of sellers on eBay. We will use *sellersEbayRefined* in our main regressions, and use *sellersEbayAll* as a robustness check.

As an alternative proxy for λ^b, we could use the number of sellers on eBay relative to that of Amazon,[26]

$$sellersEbayRelative = \frac{sellersEbayRefined}{sellersAmazon}.$$

This measure proxies the relative success probability of meeting a seller on eBay versus Amazon. It is constructed based on a typical buyer's online shopping experience. When a buyer discovers dozens of sellers on Amazon, it is relatively less likely that he can find even better offers outside Amazon, so the perceived outside value of going to eBay is low. In contrast, for a buyer who observes only few sellers on Amazon, the expected payoff of searching on the outside market is high. *sellersEbayRelative* is therefore

[24] The URL of the list of all pans is *https://www.amazon.com/pans/b?node=3737221* (visited on August 24, 2018).

[25] Here, we use the Fuzzy String Matching Library in Python which computes a score between 0 and 100, with 100 indicating the exact matching. The function `fuzz.token_set_ratio()` computes the score and only selects offers with a score higher than 80. We also tried other criteria scores such as 60 and 90. The results are robust.

likely to be positively correlated with λ^b.

As a proxy for buyers' bargaining power in the outside market, β, we compute the price difference between eBay and Amazon. For each product, we find the median price in the refined eBay offer list and compute the log of this price minus the log of Amazon price. This defines the variable *priceDiff*. We expect this variable to be negatively correlated with β (recall that a higher β implies a larger share of the surplus for the buyer so a lower price in the decentralized market).

Descriptive Analysis. We collected information for 9066 products on Amazon and found matched eBay offers for 7944 of them. Variables may have missing values leading to a smaller sample size. For example, ranking information might be provided not in the aggregate category *"Home & Kitchen"* but in some subcategories with incomplete ranking. We did try different (sub)categories to extract ranking information, and it turned out that *"Home & Kitchen"* gave us the largest valid sample with 7942 observations. In the regressions below, we exclude products without any matched eBay offers to avoid missing *priceDiff*, and exclude products without any third-party sellers on Amazon to avoid missing *sellersEbayRelative*. Finally, we only collect offers for brand new products.

Table 2: Summary Statistics

Variables	Obs.	Mean	Std.	Min.	Max.
sellByAmazon	9066	0.32	0.46	0.00	1.00
listedDays	8168	1759.54	1458.12	8.00	6864.00
price	8856	64.03	107.00	0.01	2118.83
rank	7942	440711	314288	28	2581111
sellersAmazon	9066	3.70	5.36	0.00	77.00
sellersEbayAll	9066	14.69	14.13	0.00	60.00
sellersEbayRefined	9066	6.53	8.05	0.00	43.00
sellersEbayRelative	8487	3.30	5.35	0.00	43.00
priceDiff	7944	0.07	0.74	-5.08	6.86
sellersEbayAll_60	9066	20.88	15.87	0.00	62.00
sellersEbayRefined_60	9066	10.54	10.53	0.00	48.00
sellersEbayRelative_60	8487	5.59	7.69	0.00	44.00
priceDiff_60	8349	0.07	0.72	-5.08	6.66

Note: The table reports summary sample statistics for the merged scraped data from www.amazon.com and www.ebay.com. The last four variables *sellersEbayAll_60*, *sellersEbayRefined_60*, *sellersEbayRelative_60*, *priceDiff_60* are defined on a dataset constructed by searching only the first 60 characters of Amazon product title in eBay's search engine. They are used in robustness checks. Finally, we calculate the statistics of each variable with all valid observations in the dataset.

Table 2 presents summary statistics for our main variables of interest. For 32% of the products in our sample, Amazon acts as a middleman; for the other 68% products, Amazon acts as a pure platform. On average, the products have been on sale at Amazon for almost 5 years, although this varies across products from several days to 18 years. There is a large variation in the price and ranking. The maximum price is as high as \$2118.83. The mean price is \$64, the 25th percentile is \$18.9, the 75th percentile is \$72.9. The number of third-party sellers for a product ranges from 0 to 77 with a mean of 3.7 sellers. The number of sellers on eBay is much larger with a mean of 14.69 for *sellersEbayAll* and a mean of 6.53 for *sellersEbayRefined*. On average, the number of sellers on eBay is more than three times as high as the number of sellers on Amazon. Finally, variables with suffix 60 come from another dataset constructed for robustness checks and will be discussed later.

[26]We use the number of third-party sellers (that is excluding Amazon if Amazon sells) in the denominator.

Table 3: Correlations among proxy variables

	logRank	*priceDiff*	*sellerEbay Refined*	*sellerEbay All*	*sellerEbay Relative*
logRank	1.0000				
priceDiff	-0.1359	1.0000			
sellersEbay Refined	-0.1462	-0.0901	1.0000		
sellersEbay All	-0.0595	-0.0797	0.7063	1.0000	
sellersEbay Relative	0.06839	-0.1203	0.6790	0.4866	1.0000

In table 3, the linear correlations among proxies are very weak. Correlations among proxies for different parameters are around 0.1.

Robustness Checks. We shall pursue a number of robustness checks. A first concern is that our result could be driven by the way that we count the number of eBay offers. To address this issue, instead of refining the list of eBay offers, we use the raw list of eBay offers to calculate the number of sellers, *sellersEbayAll*, and replace *sellersEbayRelative* by *sellersEbayAll*/*sellersAmazon*. Our results are robust to this change as shown in Table 4: although the coefficients of *sellersEbayAll* and *sellersEbayRelative* become smaller, they remain negative. The coefficient of *sellersEbayAll* now becomes non-significant, while *sellersEbayRelative* is still statistically significant. Relative to the result summarized in Table 2.1, the coefficients of the other variables remain almost the same.

A second concern is a bias caused by using the eBay search engine. We find that the number of offers provided by the eBay search engine is negatively correlated with the length of search text. In general, the longer the search text is, the lower the number of results that the eBay search engine can provide. Hence, the longer the product name is, the less likely it can find good matches in its database. This implies that we may ignore good matches if we provide a very long product name with too much information. For example, the same product may have different product titles by different sellers emphasizing different product features, such as size and color of the pan. In some cases, eBay can not give any offer when searching the whole Amazon product title, but does give the right offers when searching part of the Amazon product title. More importantly, there exists anecdotal evidence showing that the product title on Amazon is longer if it is registered by Amazon itself rather than by third-party sellers. If this is true, we may have spurious correlations. To solve this issue, we construct a second dataset by searching all product names using only the first 60 characters on eBay.[27]

The variables *sellersEbayAll_60*, *sellersEbayRefined_60*, *sellersEbayRelative_60* and *priceDiff_60* are constructed in this new dataset. As shown in the last four rows in the summary statistics Table 2, the average number of sellers for each product becomes larger. For example, in terms of the length of the raw search list, the average increases from 14.69 to 20.88. However, the relative prices between eBay and Amazon do not change much. The results using this new dataset are reported in Table 5 and yield similar relationships as our main ones. There are more observations in the regressions because some Amazon product titles which had no eBay offer before can now be matched. As in our previous regressions, the coefficients of other variables remain almost unchanged.

[27] In our data, the median length of product title is 65, the minimum is 9, and the maximum is 245. We also tried to cut the first 50 or 80 characters. The results are similar.

Table 4: Regressions for Amazon's intermediation mode using the raw eBay search results

	(1) Linear	(2) Linear	(3) Probit	(4) Probit
sellersEbayRelative	-0.00354***		-0.00478***	
	(0.000428)		(0.000613)	
sellersEbayAll		-0.000226		-0.000439
		(0.000382)		(0.000426)
sellersAmazon		-0.000691		-0.000765
		(0.000915)		(0.000996)
log(rank)	-0.101***	-0.106***	-0.102***	-0.108***
	(0.00444)	(0.00461)	(0.00500)	(0.00513)
priceDiff	0.100***	0.112***	0.122***	0.136***
	(0.00844)	(0.00847)	(0.0111)	(0.0112)
log(price)	0.0346***	0.0402***	0.0440***	0.0503***
	(0.00614)	(0.00622)	(0.00689)	(0.00698)
listedDays	0.0606***	0.0664***	0.0719***	0.0788***
	(0.00479)	(0.00485)	(0.00596)	(0.00604)
Observations	6457	6457	6457	6457
Adjusted R^2	0.135	0.129		

Note: This table reports the robustness check using the raw eBay search results reflected in *sellersEbayAll* and *sellersEbayRelative*. Except for this change, the specification is the same as before.

Table 5: Regressions for Amazon's intermediation mode using first 60 characters to search eBay offers

	(1) Linear	(2) Linear	(3) Probit	(4) Probit
sellersEbayRelative	-0.00450***		-0.00488***	
	(0.000595)		(0.000816)	
sellersEbayRefined		-0.000352		-0.000170
		(0.000499)		(0.000556)
sellersAmazon		-0.00587***		-0.00708***
		(0.000847)		(0.00112)
log(rank)	-0.101***	-0.0950***	-0.102***	-0.0984***
	(0.00423)	(0.00460)	(0.00468)	(0.00515)
priceDiff	0.114***	0.126***	0.134***	0.146***
	(0.00819)	(0.00849)	(0.0105)	(0.0106)
log(price)	0.0431***	0.0503***	0.0529***	0.0586***
	(0.00602)	(0.00614)	(0.00658)	(0.00667)
listedDays	0.0611***	0.0646***	0.0741***	0.0747***
	(0.00449)	(0.00480)	(0.00566)	(0.00576)
Observations	6822	6822	6822	6822
Adjusted R^2	0.138	0.100		

Note: This table reports the robustness check based on eBay search results using only the first 60 characters of the Amazon product title. Except for using new variable reflecting this change, *sellersEbayAll_60*, *sellersEbayRefined_60*, *sellersEbayRelative_60* and *priceDiff_60*, the specification remains the same as before.

3 Competing Intermediaries

3.1 Introduction

This chapter is based on the working paper co-authored by Pieter Gautier and Makoto Watanabe. We develop a theory of competing intermediaries who can adopt hybrid business modes. We consider two representative business modes of intermediation. In one mode, an intermediary acts as a *middleman*, who holds inventory and resells to buyers, e.g., supermarkets. In the other mode, an intermediary acts as a *marketmaker*, who offers a platform for fees, where the participating buyers and sellers can search and trade with each other, e.g., auction sites and many real estate agencies.

In most real-life markets intermediaries are not one of those extremes but operate both as a middleman and a market-maker at the same time. This is what we call a *market-making middleman.* One well-known example is the electronic intermediary Amazon, who started as a pure middleman, buying and reselling products in its name. After facing the competition from eBay, Amazon moved toward a market-maker mode by allowing third-party sellers to join its marketplace. Today, the market-maker sector has accounted for around half of the gross merchandise volume of Amazon. A similar pattern has been observed in financial markets. For example, the New York Stock Exchange (NYSE) adopted an expanded platform "NYSE Arca" after a serious market share drop around 2008. In housing markets, the Trump Organization established a luxury residential real estate brokerage firm, competing with thousands of housing brokers in New York City. Recently, Ikea is exploring some online sales website that combines a middleman mode (Ikea products) and a market-maker mode (rival products), with Alibaba and Amazon in mind.[1]

A common feature of these marketmaking middlemen is the presence of competing intermediaries. Amazon's marketplace is, at least partially, due to the competition with eBay.[2] NYSE's platform sector is clearly driven by the decentralized equity trading — the order-flow in NYSE-listed stocks today is divided among many trading venues, 11 exchanges, more than 40 alternative trading systems and more than 250 broker-dealers in the U.S. (Tuttle, 2014). In short, the hybrid intermediation mode seems to be a way to keep ahead of competing intermediaries.

In the previous chapter, we have shown that in the presence of outside markets, a marketmaking middlemen mode can be profit-maximizing for a monopolistic intermediary. This is because a middleman with its large inventory holdings can reduce the out-of-stock risk, while a marketmaker with its enrolled third-party sellers can reduce the outside option value of buyers and lower the outside competition pressure. It is not clear, though, whether this intuition holds when the outside market is not passive but is operated by another strategic player. The current paper fills this gap by explaining the emergence of marketmaking middlemen in a duopoly.

In our model, there is a finite mass of buyers and sellers who can only meet by using an intermediary. There are two intermediaries open to agents, call them an incumbent and an entrant. We assume the entrant is restricted to be a pure market-maker, while the incumbent can combine two business modes: as a middleman, he is prepared to serve many buyers at a time by holding inventories; as a marketmaker, he offers a platform and charges transaction fees. Both intermediaries rely on a transaction fee paid ex-post when a transaction takes place between two matched parties. Besides, the incumbent has an additional pricing instrument, the price of its inventory which affects the allocation of the attending buyers among his two business modes.

We formulate the incumbent market as a directed search market to feature the intermediary's technology of spreading price and capacity information efficiently. For example, one can receive instantly all relevant information such as prices, the terms

[1] See a report of Financial Times *https://on.ft.com/2StN4GU* (visited on Feb 25th, 2019).

[2] See details in the book of *The Everything Store: Jeff Bezos and the Age of Amazon* by Stone (2013).

of trade and stocks of individual sellers using the search function in web-based platforms. In this setting, each seller is subject to an inventory capacity of discrete units (normalized to one unit in the model), whereas the middleman has continuous access to production.[3] Naturally, the middleman is more efficient in matching demands with supplies in a directed search equilibrium.

With this setup, we investigate a Bertrand duopoly competition game. A key difference between the two intermediaries is that agents hold an optimistic (pessimistic) belief towards the incumbent (entrant). That is, agents are willing to visit and trade at the incumbent whenever possible. This assumption on pessimistic beliefs about the participation decisions of agents on the other side of the market is consistent with Caillaud and Jullien (2003). It is in this sense that the incumbent is more established. Based on the pessimistic beliefs about the entrant, we analyze the existence and sustainability of the market structure in two situations, *single-market search* versus *multiple-market search*.

Under single-market search, agents have to choose which intermediary to visit in advance. Under the pessimistic beliefs against the entrant and the two-sidedness of a platform, it is a classical result that in the equilibrium all agents only visit and trade at the incumbent. That is, market monopolization is the only equilibrium. Given that the middleman mode is more efficient in realizing transactions, the incumbent uses the middleman-mode exclusively when agents search in a single market. This conclusion is the same as Chapter 2.

In multi-market search, we allow for the possibility that agents search sequentially. As in Chapter 2, we assume the incumbent intermediary opens prior to the entrant. Accordingly, a pessimistic expectation against the entrant means that agents are willing to *first* visit and trade at the incumbent whenever possible. This changes the nature of competitive strategies — the prices/fees charged by the incumbent must be acceptable relative to the available option at the entrant; otherwise, buyers and sellers can easily switch to the entrant. Thus, under multiple-market search, the entrant represents an outside trading option for agents.

Consistent with Chapter 2, the incumbent faces a trade-off between trading quan-

[3] Or equivalently, the middleman is subject to an inventory capacity of a mass K, and assuming a low inventory cost, then the optimal K is chosen to satisfy all demands. Essentially, the out-of-stock risk is zero at the middleman, as is shown in Chapter 2.

tity and marginal trading profit. On the one hand, a larger middleman sector leads to more transactions and consequently to larger profits. On the other hand, sellers are less likely to trade on a smaller-scaled platform, so more sellers are available when a buyer attempts search at the entrant market, a downward restriction on the price/fees that the incumbent can charge. This trade-off determines the middleman scale and eventually the intermediation mode of the incumbent. Beyond the results in Chapter 2, we further show that a pure middleman incumbent is profitable only when the entrant charges a fee at the monopoly level so that a buyer's expected value at the entrant is zero.

An important difference between Chapter 2 and this chapter is that the entrant intermediary in our framework also faces strategic choices.[4] He can either act as a "second source", set a relatively high fee, wait for the leftover agents who are not matched at the incumbent market, or undercut the incumbent fiercely with a transaction fee low enough in order to break the pessimistic beliefs. In the latter case, the entrant, despite the pessimistic beliefs, becomes the sole active source. Therefore, how costly the undercutting determines how profitable a sole source can be. And this, in turn, gives different equilibria.

We first show that a marketmaking middleman incumbent emerges in the equilibria of multi-market search. A pure middleman incumbent can not show up in an equilibrium of pure strategies because the entrant has an incentive to undercut the incumbent, leading to a positive buyer's value at the entrant intermediary. According to the best response of the incumbent, it is profitable for the incumbent to activate its platform trading whenever a buyer's outside value at the entrant is positive. On the other hand, a pure market-making incumbent is also not possible in a pure strategy equilibrium. This follows the logic of Varian (1980): The entrant has an incentive to undercut the incumbent for a discrete jump in total transactions, as long as transaction fee is positive; at the fee level of zero, the entrant would rather increase the fee to the highest level, extract full surplus and work as a second source.

Our main result is that there exists a pure strategy equilibrium when undercutting the incumbent is costly. In the equilibrium, the incumbent works as a market-making middleman and is the first source to implement transactions, and the entrant is a sec-

[4]In the model of Chapter 2, the decentralized market, a counterpart of the entrant intermediary here, is passive.

ond source. Both make positive profits. Moreover, we extend the insights of Chapter 2: As the buyers' outside value becomes larger, the incumbent operates a smaller middleman sector and a larger platform sector in the equilibrium. Finally, when undercutting the incumbent is more profitable, we show there exists a mixed strategy equilibrium where the market-maker sector is activated with positive probability.

Literature review This paper extends Chapter 2 to the case of competing intermediaries. The incumbent (entrant) here corresponds to the centralized market (decentralized market). So the best response analysis of the incumbent in our model corresponds to the profit-maximizing intermediation mode analysis (Proposition 2) in that paper. We show the main conclusion of Chapter 2 can be extended to the case of duopoly competition despite the strategic behavior of the entrant. In particular, when undercutting the incumbent is not profitable, there exists a pure strategy equilibrium that resembles the market structure as is characterized in Chapter 2.

This paper is related to the middleman literature and the two-sided market literature, as has been discussed extensively in Chapter 2. Our model is in particular, closely related to Caillaud and Jullien (2003) who examined a Bertrand competition game between two intermediaries. They considered a rich set of contracting possibilities including the "divide-and-conquer" strategy. Using pessimistic beliefs, they characterized various equilibria assuming both intermediaries are pure market-makers. Compared to them, we abstract from the possibility of using negative participation fees for breaking the pessimistic beliefs and focus on the endogenous intermediation structure out of the market competition. A "divide-and-conquer" strategy would be useful for the entrant to break the disadvantageous beliefs. It is not clear whether the "divide-and-conquer" strategy would reinforce or undermine our results. We leave it for future research.

Our paper is also related to the vast literature of discontinuous games, see Dasgupta and Maskin (1986) for example. In our model, an equilibrium of pure strategies exists simply because cutting the transaction fee slightly is not enough to for the entrant to get a discrete jump in demands. As the entrant's fee decreases, a buyer finds it more appealing to visit the incumbent platform, since even he is unmatched, the outside option value — to trade at the entrant — is now higher. Consequently, more sellers are successfully matched at the incumbent platform, leaving fewer sellers available at the entrant. As a result, buyers' expected value at the entrant does not increase with a decreasing transaction fee there. This would be the case as long as the incumbent is of a mixed mode. This barrier to an undercutting strategy makes a pure strategy equilibrium possible. When indeed undercutting is more profitable for the entrant, we make use of Proposition 5 in Dasgupta et al. (1986) to show the existence of a mixed strategy equilibrium.

The plan of the article is as follows. Section 2 introduces the model. In Section 3, we consider the case of single-market search, which serves as a benchmark for our subsequent analysis. Section 4 contains our main results about multi-market search. Finally, Section 5 concludes. The more technical proofs are relegated to Appendix I. Appendix II contains our extension to the case that the entrant intermediary is a pure middleman.

3.2 Setup

Agents We consider a large economy with two populations, a mass B of buyers and a mass S of sellers. Agents of each type are homogeneous. Each buyer has unit demand for a homogeneous good, and each seller is able to sell one unit of that good. The consumption value for the buyers is normalized to 1. Sellers can purchase the good from a wholesale market. We assume the wholesale market is competitive, the demand of all suppliers is always satisfied with a price equal to the marginal cost which is normalized to zero.

Intermediaries Buyers and sellers can only meet by using intermediaries. There are two intermediaries in the economy, an incumbent I, and an entrant E. Both have some matching technology to facilitate trading. The names "incumbent" and "entrant" are not related to strategic entry deterrence. Instead, they reflect two advantages of I over E: First, the matching technology of I is more advanced in the sense that pricing and capacity information can be spread within the scope of I; Second, I faces favorable beliefs from agents. We will discuss each feature in details, starting with the matching technology.

Trade at intermediary E Intermediary E is a pure *marketmaker*, or equivalently a *platform*, and it makes profits by extracting a transaction fee denoted by $f^e \in [0,1]$. E owns an inferior matching technology. It is "inferior" because E is not able to spread price and stock information, and a buyer only receives such information after meeting a seller. We characterize this matching technology by random matching and bilateral bargaining. In particular, here we employ a simple linear matching function as follows. Suppose all buyers and sellers participate in E, then a buyers finds a seller with probability λ^b and a seller finds a buyer with probability λ^s, satisfying $B\lambda^b = S\lambda^s$. If a subset of buyers $B^E \leq B$ and sellers $S^E \leq S$ participate, then the meeting probabilities become $\lambda^{b\prime} \equiv \lambda^b \frac{S^E}{S}$ and $\lambda^{s\prime} \equiv \lambda^s \frac{B^E}{B}$, respectively.[5] Matched partners follow an efficient bargaining process, which yields a linear sharing of the total surplus, with a share of $\beta \in [0, 1]$ for the buyer and a share of $1 - \beta$ for the seller. In this paper, we focus on the case of $\beta = 1$, i.e., the buyer gets the full trading surplus.

Trade at intermediary I Intermediary I can work as a *marketmaking middleman*. First, as a *marketmaker*, I can extract transaction fees from platform trade. I's platform has a more advanced matching technology — the prices and capacities of all the individual suppliers are publicly observable — and hence buyers can make use of this information to "direct" their search. Still, given that individual buyers cannot coordinate their search activities, the limited selling capacity of individual sellers creates a possibility that some units remain unsold while at the same time some demands are not satisfied. In this sense, I is subject to coordination frictions. Second, other than the platform, I can also work as a *middleman* — he purchases a good from the wholesale market and resells it to buyers. Therefore, I can be both the manager of the platform (called a marketmaker) and a middleman participant to the platform, so called a "marketmaking middleman".

Compared to a seller, the middleman sector of I has an inventory advantage — he is able to hold more inventory, precisely a continuum of inventory, to lower the out-of-stock risk. For simplicity, we consider an extreme of the flexible inventory technology: A continuous access to production. That is, the middleman can produce as he gets an order from some buyer. Effectively, the out-of-stock probability is zero.[6]

[5]To understand $\lambda^{b\prime}$, imagine that all sellers are registered at E, but not all would be available. In our model, a seller is absent from market E if he has sold the unit inventory in other markets. If the seller that a buyer is supposed to meet is not available, then the meeting would fail. Given an mass of $S^E < S$ sells join the platform, the probability a buyer finds an available seller would be $\lambda^b \frac{S^E}{S}$. The same logic applies to $\lambda^{s\prime}$. It is easy to verify that $B^E \lambda^{b\prime} = S^E \lambda^{s\prime}$.

[6]An alternative interpretation of the inventory technologies is that individual sellers have to *produce in advance* (up to the inventory constraint of 1 unit) whereas middlemen can *produce to orders*. Suppose the middleman has to "produce in advance" and holds a mass K of inventory. Given a zero inventory cost, it is a weakly dominant strategy to have K larger or equal to the mass of visiting buyers. This gives a zero

Now let's describe how the directed search works in I. Given a mass of $B^I \in (0, B]$ buyers and a mass of $S^I \in (0, S]$ sellers have decided to participate in I, the matching process in I is specified by a directed search game which consists of the following stages. In the first stage, all the suppliers, i.e., the participating sellers with the unit selling-capacity and the middleman with a continuous access to production, post a price which they are willing to sell at. We denote the posting price of an individual seller by p^s, and that of the middleman by p^m. Observing the prices and the capacities, all buyers simultaneously decide which supplier to visit in the second stage. As is standard in the literature, we assume that each buyer can visit one supplier, one of the sellers or the middleman. Assuming buyers cannot coordinate their actions over which supplier to visit, we investigate a symmetric equilibrium where all buyers use an identical strategy for any configuration of the announced prices. Each individual seller (if any) has an expected queue x^s of buyers and the middleman has an expected queue x^m of buyers. These quantities should satisfy two requirements. The first requirement is a standard accounting identity,

$$S^I x^s + x^m = B^I, \tag{3.1}$$

which states that the number of buyers visiting individual sellers $S^I x^s$ and the middleman x^m should sum up to the total population of participating buyers B^I. The second requirement is that buyers search optimally:

$$x^m = \begin{cases} B^I & \text{if } V^m(B^I) \geq V^s(0) \\ (0, B^I) & \text{if } V^m(x^m) = V^s(x^s) \\ 0 & \text{if } V^m(0) \leq V^s(\frac{B^I}{S^I}), \end{cases} \tag{3.2}$$

where $V^i(x^i)$ is the equilibrium value of buyers in the C market of visiting a seller if $i = s$ and the middleman if $i = m$ (yet to be specified below). Combining (3.1) and (3.2) gives the counterpart for $x^s \in [0, \frac{B^I}{S^I}]$. We define the intermediation mode of I as follows.

Definition 3.1 (Intermediation Mode)**.** *Suppose $B^C \in (0, B]$ buyers and $S^C \in [0, S]$ sellers participate in C market. Then, given the equilibrium search conditions (2.1) and (2.2), we say that the intermediary acts as:*

---

out-of-stock probability.

- *a pure middleman if* $x^m = B^C$;
- *a market-making middleman if* $x^m \in (0, B^C)$;
- *a pure market-maker if* $x^m = 0$.

The Game The situations we consider in the following sections all have the same timing structure.

1. Two intermediaries setting fees/prices. Intermediary I decides whether or not to activate the middleman sector and/or the platform, and announces a transaction fee $f^I \in [0,1]$ charged to a seller, a price $p^m \in [0,1]$ charged to a buyer on his own inventory. Intermediary E announces a transaction fee $f^E \in [0,1]$ charged to a seller. Let's denote these prices (or fees, interchangeably) by $P = \{f^I, p^m, f^E\}$.[7]
2. Observing P, buyers and sellers simultaneously decide whether to participate in the two (or one out of two) intermediaries, yielding a distribution of buyers/sellers across intermediaries.
3. In the I market, the participating buyers, sellers and middleman are engaged in a directed search game as has been specified. In the E market, agents search randomly and follow the efficient sharing rule for the trade surplus.

Let $\mathcal{N} = \{B^I, B^E, S^I, S^E, x^m\}$ denote a distribution of buyers/sellers across intermediaries with B^I (B^E) the mass of buyers visiting I (E), S^I (S^E) the mass of sellers visiting I (E), and x^m the mass of buyers visiting the middlemen if $B^I > 0$. A market allocation is a mapping $\mathcal{N}(\cdot)$ that associates to each price/fee P an equilibrium distribution of buyers/sellers $\mathcal{N}(P)$. Hence, $\mathcal{N}(\cdot)$ generates a reduced-form price-setting game among intermediaries where a Nash equilibrium can be defined.

Definition 3.2. *An equilibrium of the game with a market allocation* $\mathcal{N}(\cdot)$ *is a price vector* $P^* = \{f^{I*}, p^{m*}, f^{E*}\}$ *and an associated distribution of buyers and sellers* $\mathcal{N}(P^*)$ *where neither intermediary* I *nor* E *has an incentive to deviate under* $\mathcal{N}(\cdot)$.

Some regularities on $\mathcal{N}(\cdot)$ are necessary to proceed. First, we impose a tight structure on $\mathcal{N}(\cdot)$ according to pessimistic beliefs about E which is yet another advantage

[7] Here we assume that transaction fees are charged to sellers for simplicity. But as is shown in Chapter 2, the results hold if the fee is charged on buyers or split between a matched pair.

of I relative to E. A pessimistic market allocation differs slightly under single/multi-market search. In the single-market search, we impose $\mathcal{N}(\cdot)$ such that buyers and sellers coordinate on the distribution with zero market share for E as long as it is in their interests to do so.

Definition 3.3 (Pessimistic Expectations in Single-market Search)**.** *Under single-market search, a pessimistic market allocation against E, $\mathcal{N}(\cdot)$, is that $B^I = B, S^I = S, B^E = S^E = 0$ as long as*

$$\max\{V^m(\mathcal{N}, P), V^s(\mathcal{N}, P)\} \geq 0,$$

for some $x^m \in [0, B]$; and $B^E = B, S^E = S, B^I = S^I = 0$ otherwise.

In the single market search case, agents decide which intermediary to join under the expectation that all others visit I. A buyer or a seller only chooses to join E if the fees of I are so high that he would rather visit E. Since he expects no counterparts at E, the expected value at E is 0.

In the multi-market search, intermediary I opens first, and participating in I does not rule out the possibility of visiting E later. A pessimistic market allocation against E is that agents first visit and trade at I and then E as long as it is in their interests to do so. We defer a formal definition to Section 3.4.

3.3 Single-market Search

In this section, we show that, under single-market search, I chooses to be a pure middleman in equilibrium. We start by analyzing the best response of I. For any offers from E, buyers have an expected value $V^E = 0$ under favorable beliefs towards I.[8] Given $f^I, p^m \leq 1$, the expected value of an agent at I is non-negative. Thus, for all price P, we have the agent distribution $\mathcal{N}$ that $B^I = B, S^I = S, B^E = S^E = 0$.

Suppose I is a pure middleman, the best response involves setting $p^m = 1 - V^E = 1$ which gives a profit of B. Since this is the whole trading surplus, it is easy to speculate that the pure middleman mode is the optimal response. This is indeed true: Even I can also get the full surplus for each transaction through the platform, as long as there is some matching frictions that deliver less transactions, using a platform is dominated.

[8] E faces a chicken and egg problem. To get a positive market share despite pessimistic beliefs, E must adopt a divide and conquer strategy. This is analyzed in Caillaud and Jullien (2003).

We now derive the matching function of I's platform. Suppose I has an active platform with S sellers, $B - x^m$ buyers, and an intermediation fee $f^I \leq 1$. Then, the platform generates a non-negative trade surplus $1 - f^I \geq 0$. The number of buyers visiting an individual seller is a random variable, denoted by N, which follows a Poisson distribution, $\text{Prob}[N = n] = \frac{e^{-x}x^n}{n!}$, with an expected queue of buyers $x \geq 0$ (due to coordination frictions — see more explanation in Chapter 2). With limited selling capacity, each seller is able to serve only one buyer. A seller with an expected queue $x^s \geq 0$ has a probability $1 - e^{-x^s}$ $(= \text{Prob}[N \geq 1])$ of successfully selling, while each buyer has a probability $\eta^s(x^s) = \frac{1-e^{-x^s}}{x^s}$ of successfully buying. Hence, the expected value of a seller on the platform with a price p^s and an expected queue x^s is given by

$$W(x^s) = x^s \eta^s(x^s)(p^s - f^I),$$

while the expected value of a buyer who visits the seller is

$$V^s(x^s) = \eta^s(x^s)(1 - p^s).$$

And the platform as a whole generates an expected trading volume of $S(1 - e^{-x^s})$. Hence, the resulting profit of I satisfies

$$S(1 - e^{-x^s})f^I + x^m p^m < (Sx^s + x^m)\max\{f^I, p^m\} \leq B.$$

From the second inequality it follows that $f^I, p^m \leq 1$. And the first inequality follows directly from the matching function of the platform. In a nutshell, the pure middleman mode dominates any other modes with an active platform, and in the equilibrium, I is able to gain a monopoly profit B.

Proposition 3.1 (Pure Middleman)**.** *Given single-market search technologies, in the equilibrium with a pessimistic market allocation against E, I acts as a pure middleman, setting* $f^I = p^m = 1$, *all buyers and sellers join intermediary I, and E is inactive with* $f^E \in [0, 1]$.

3.4 Multi-market Search

This section is devoted to the analysis of competition between two intermediaries who offer non-exclusive services, so that users may engage in multi-market search. We say an intermediary is the *first source* if buyers/sellers would like to first visit it and trade there, and the *second source* if it is visited only when buyers/sellers did not trade at their first source.

In this section, we assume that intermediary I opens prior to intermediary E and I is the first source by default as in Chapter 2. This is the pessimistic belief against E under multi-market search. A formal definition will be given in Definition 3.4. It can be understood as the common belief in the economy and reflects the fact that the incumbent has been well-established in the market.

Under such a belief, when fees/prices of I are comparable to that of E, buyers and sellers will indeed firstly visit I and trade there. It is only when E's fees are *much* lower than that of I that buyers and sellers will forgo trading opportunities of I and trade at E instead. Then, I becomes inactive, and E is the only active intermediary, called the *sole source*.

We start by examining the directed search equilibrium at intermediary I and the best response of I assuming that I is the first source. This analysis is enough to show that equilibrium with pure mode I does not exist. We then turn to the best response analysis of E and the existence of an equilibrium with I of a mixed mode. We show that under certain conditions, there exists an equilibrium of pure strategies that features (1) both intermediaries are active; (2) agents meet and trade first at I and then join E, i.e. I is the first source and E is the second source; (3) I is a marketmaking middleman. Moreover, we show there exists an equilibrium of mixed strategies where the market-maker sector of I is activated with positive probability.

3.4.1 Directed search equilibrium at incumbent intermediary

We work backward and first describe the directed search equilibrium at intermediary I, given that such an equilibrium exists with some $P = \{f^I, p^m, f^E\}$.[9] As under single-market search, any directed search equilibrium has to satisfy (3.1) and (3.2). Given the multiple-market search technology, what is new here is that, when deciding whether or not to accept an offer in the I market, buyers expect a non-negative value at E, defined by

$$V^E(x^m, f^E) = \lambda^b e^{-x^s}(1 - f^E).$$

V^E can be understood as follows: The outside payoff is $1 - f^E$ if the buyer is matched with a seller who has failed to trade in the I market. This happens with probability

[9] This section is a simplified version of the directed search equilibrium assuming seller's bargaining power at E is zero. For a full analysis, see the multi-market search analysis in Chapter 2

$\lambda^b e^{-x^s}$. Hence, the larger the platform size x^s, the higher the chance that a seller trades in the I market, and the lower the chance that a buyer can trade successfully in the E market and the lower his expected payoff at E.

Whenever I's platform is active $x^s > 0$, it must satisfy the incentive constraints:

$$1 - p^s \geq V^E(x^m, f^E), \tag{3.3}$$

$$p^s - f^I \geq 0. \tag{3.4}$$

The constraint of buyers (3.3) states that the offered price/fee in the platform is acceptable only if the offered payoff, $1 - p^s$, is no less than the expected value in the E market $V^E(x^m, f^E)$. The constraint of sellers (3.4) states that the payoff in the I market $p^s - f^I$ should be no less than the expected payoff in the E market which in principle depends on a seller's chance of engaging in a trade in the I market. However, since the bargaining power of sellers is zero, so a seller's expected payoff at E is reduced to zero.

We have a similar condition of buyers for the middleman sector:

$$1 - p^m \geq V^E(x^m, f^E), \tag{3.5}$$

where the middleman's price must be acceptable for buyers relative to the expected payoff in the E market.

Given the outside option of the E market, the equilibrium value of buyers in the I market is $V = \max\{V(x^s), V(x^m)\}$, where

$$V^s(x^s) = \eta^s(x^s)(1 - p^s) + (1 - \eta^s(x^s))\, V^E(x^m, f^E) \tag{3.6}$$

for an active platform $x^s > 0$ and

$$V^m(x^m) = 1 - p^m \tag{3.7}$$

for an active middleman sector $x^m > 0$. Here, if a buyer visits a seller (or a middleman), then he gets served with probability $\eta^s(x^s)$ (or probability 1) and his payoff is $1 - p^s$ (or $1 - p^m$). If not served at intermediary I, then he enters the E market and he can find an available seller with probability $\lambda^b e^{-x^s}$, and gets a payoff $1 - f^E$. Similarly, the equilibrium value of active sellers in the platform is given by

$$W(x^s) = x^s \eta^s(x^s)(p^s - f^s - c) + (1 - x^s \eta^s(x^s)) \times 0. \tag{3.8}$$

A seller trades successfully on I's platform with probability $x^s\eta^s(x^s)$ and receives $p^s - f^I$. If not successful, he enters E's platform where he meets a buyer with some probability and get an expected value of 0 as the seller's bargaining parameter is assumed to be zero.

We now proceed to determined the equilibrium price p^s take as given the first-stage price/fees $P = \{f^I, p^m, f^E\}$. Suppose a seller deviates to a price $p \neq p^s$ that attracts an expected queue $x \neq x^s$ of buyers. Note that given the limited selling-capacity, this deviation has measure zero and does not affect the expected utility in the I market, V. Since buyers must be indifferent between visiting any seller (including the deviating seller), the equilibrium market-utility should satisfy

$$V = \eta^s(x)(1-p) + (1-\eta^s(x))V^E, \tag{3.9}$$

where $\eta^s(x) \equiv \frac{1-e^{-x}}{x}$ is the probability that a buyer is served by this deviating seller. If not served, which occurs with probability $1-\eta^s(x)$, his expected utility at intermediary E is $V^E = \lambda^b e^{-x^s}(1-f^E)$. Given market utility V, (3.9) determines the relationship between x and p, which we denote by $x = x(p|V)$. This yields a downward sloping demand faced by the seller: when the seller raises his price p, the queue length of buyers x becomes smaller, and vice versa.

Given the search behavior of buyers described above and the market utility V, the seller's optimal price should satisfy

$$p^s(V) = \arg\max_p \left(1 - e^{-x(p|V)}\right)\left(p - f^I\right)$$

Substituting out p using (3.9), the sellers' objective function can be written as

$$W(x) = (1-e^{-x})(1 - V^E - f^I) - x(V - V^E),$$

where $x = x(p|V)$ satisfies (3.9) and $1 - V^E$ is the intermediated trade surplus, i.e., the total trading surplus at intermediary I net of the outside options at E. Since choosing a price is isomorphic to choosing a queue, the first order condition is

$$\frac{\partial W(x)}{\partial x} = e^{-x}(1 - V^E - f^I) - (V - V^E) = 0$$

The second order condition can be easily verified. Arranging the first order condition using (3.9) and evaluating it at $x^s = x(p^s|V)$, we obtain the equilibrium price $p^s =$

$p^s(V)$ which can be written as

$$p^s - f^I = \left(1 - \frac{x^s e^{-x^s}}{1 - e^{-x^s}}\right)\left(1 - V^E - f^I\right). \tag{3.10}$$

For the platform to be active, the price and fees must satisfy the incentive constraints (3.3) and (3.4). Substituting in (3.10) yields

$$f^I \leq 1 - V^E(x^m, f^E), \tag{3.11}$$

which states that for the platform to be active $x^s > 0$, the transaction fee f^I should not be greater than the intermediated trade surplus, $1 - V^E(x^m, f^E)$. Whenever (3.3) and (3.4) are satisfied, (3.11) must hold, and whenever (3.11) is satisfied, (3.3) and (3.4) must hold. Hence, we can say that the market maker faces the constraint (3.11) for an active platform.

3.4.2 The best response of the incumbent

For a given fee $f^E \in [0,1]$, we analyze the most profitable choice of intermediation mode and price/fee $\{f^I, p^m\}$ by I. We impose a "bad expectation" market allocation against E, which we formally define as follows.

Definition 3.4 (Pessimistic Expectations in Multi-market Search)**.** *Under multi-market search, a pessimistic market allocation against E, $\mathcal{N}(\cdot)$, is that $B^I = B, S^I = S$ as long as*

$$\max\{V^m(\mathcal{N}, P), V^s(\mathcal{N}, P)\} \geq V^E(\mathcal{N}, P), \tag{3.12}$$

where $B^E = B - x^m - S\left(1 - e^{-x^s}\right)$ and $B^E = S - S\left(1 - e^{-x^s}\right)$ for some $x^m \in [0, B]$. Otherwise, the distribution follows that all users forgo the trading opportunity at I and only visit E, $B^E = B, S^E = S, B^I = S^I = 0$.

Therefore, under the pessimistic expectation allocation, buyers/sellers first visit and trade at I and then E whenever possible. Notice in the definitions above, we do not impose restrictions on x^m. The equilibrium x^m is determined by conditions (3.1) and (3.2). Also note that incentive constraints (3.5) and (3.11) imply the pessimistic expectation condition (3.12). So whenever I optimizes its profits obeying the incentive constraints, buyers/sellers follow the pessimistic expectation distribution.

Our next step is to determine the profit of each intermediation mode, denoted by

$\tilde{\Pi}(x^m)$.

Pure Middleman: If I does not open the platform, then $x^m = B$ and any encountered seller at E is always available for trade. Hence, the middleman serves all buyers at the highest possible price $p^m = 1 - \lambda^b(1 - f^E)$ with a binding incentive constraint (3.5) and makes profits

$$\tilde{\Pi}^I(B) = B(1 - \lambda^b(1 - f^E)). \tag{3.13}$$

Pure Market-maker: When the middleman sector is not open, $x^s = \frac{B}{S}$. Given the equilibrium price p^s in the platform in (3.10), the intermediary charges a fee f^I in order to maximize

$$S\left(1 - e^{-\frac{B}{S}}\right) f^I,$$

subject to the constraint (3.11). The constraint is binding and it yields:

$$f^I = V^E(0, f^E) = 1 - \lambda^b e^{-\frac{B}{S}}(1 - f^E).$$

The profit for the market-maker mode is

$$\tilde{\Pi}^I(0) = S(1 - e^{-\frac{B}{S}})(1 - \lambda^b e^{-\frac{B}{S}}(1 - f^E)). \tag{3.14}$$

Market-making middleman: If the intermediary is a market-making middleman, then $x^m \in (0, B)$ and $x^s \in \left(0, \frac{B}{S}\right)$, satisfying $V^m(x^m) = V^s(x^s)$. Applying (3.6), (3.7), and (3.10), this indifference condition generates an expression for the price $p^m = p^m(x^m)$:

$$p^m(x^m) = 1 - V^E(x^m, f^E) - e^{-x^s}(1 - V^E(x^m, f^E) - f^I), \tag{3.15}$$

Together with (3.1), this equation defines the relationship between p^m and x^m. Applying this expression, we can see that the condition (3.5) is eventually reduced to (3.11). The profit for the marketmaking middleman mode is

$$\tilde{\Pi}(x^m) = \max_{x^m, f^I} \Pi(x^m, f, K) = \max_{x^m, f^I} \left\{ S(1 - e^{-x^s}) f^I + x^m p^m \right\}$$

subject to (3.11) and $x^m \in (0, B)$.

Profit-maximizing intermediation mode To derive a profit-maximizing intermediation mode, it is important to observe that relative to the pure middleman mode, an active platform at I with multiple-market search can undermine intermediary E by lowering the available supply since $S^E = Se^{-x^s}$, which relaxes the constraint on f^I imposed by the incentive constraint (3.11).

This influences the middleman's price similarly. With $f^I = 1 - V^E(x^m, f^E)$, the incentive constraint (3.5) is binding, and the middleman's equilibrium price is given by

$$p^m = 1 - V^E(x^m, f^E)$$

for any $x^s \geq 0$ according to (3.15). This shows that p^m decreases with x^m: the outside value of buyers depends positively on the size of the middleman sector, since a larger scale of the middleman crowds out the platform and increases the chance that a buyer can find an active seller in the E market (who was unsuccessful in I's platform). Hence, in order to extend the size of the middleman sector, I must lower the price p^m. In other words, a larger platform of I allows for a price increase by reducing agents' outside trade opportunities at E. This key insight from Chapter 2 holds for any $f^E < 1$.

Proposition 3.2 (Market-making middleman/Pure Market-maker)**.** *Given multi-market search technologies, there exists a unique directed search equilibrium with active intermediation of I. In particular, I will act as*

- *a pure middleman $x^{m*} = B$ if $f^E = 1$;*
- *a pure market-maker $x^{m*} = 0$ if $\lambda^b e^{-B/S} \geq \frac{1}{2}$ and $f^E \leq 1 - \frac{1}{2\lambda^b e^{-B/S}}$;*
- *a market-making middleman $x^{m*} \in (0, B)$ if $\lambda^b e^{-B/S} < \frac{1}{2}$, or $\lambda^b e^{-B/S} \geq \frac{1}{2}$ and $f^E > 1 - \frac{1}{2\lambda^b e^{-B/S}}$, and x^{m*} is characterized by $f^E = b^I(x^{m*})$, where $b^I(x^m)$ is defined by*

$$b^I(x^m) \equiv 1 - \frac{S(1 - e^{-x^s})}{(2S(1 - e^{-x^s}) + x^m)\lambda^b e^{-x^s}}. \tag{3.16}$$

And in all cases, the optimal prices are set at $p^{m*} = f^{I*} = 1 - \lambda^b e^{-x^{s*}}(1 - f^E)$.

Proof. See the Appendix. □

When $f^E = 1$, we are essentially back to the single-market search scenario where agents do not have outside options. So the pure middleman with $p^{m*} = 1$ is optimal. Whenever $f^E < 1$, with multiple-market search technologies, there is a cross-market feedback from E to I, which makes using the platform as part or all of I's intermediation activities profitable. Additionally, I must decide whether it wants to operate as a pure market maker. Our result shows that it depends on parameter values. If $\lambda^b e^{-B/S} < \frac{1}{2}$, then the buyers' outside option value is low. In this case, the middleman sector generates high enough profits for the market-making middleman mode to be adopted for any value of f^E. If instead $\lambda^b e^{-B/S} \geq \frac{1}{2}$ then the buyers' outside option value is high, and attracting buyers to the middleman sector is costly. In this case, the intermediary will act as a market-making middleman if $f^E > 1 - \frac{1}{2\lambda^b e^{-B/S}}$, where buyers expect a low value from E market, and as a pure market maker if $f^E \leq 1 - \frac{1}{2\lambda^b e^{-B/S}}$, where buyers expect a high value from the E market.

When the mixed mode is activated, $b^I(\cdot)$ in equation (3.16) characterizes the optimal intermediation structure that I is willing to pursue ("b" for best response function). It is essentially the best response of I in facing a transaction fee f^E. It is easy to check that $b^I(x^m)$ is monotonically increasing in x^m, implying that as f^E decreases, I moves towards the pure platform mode. Eventually, as f^E approaches $1 - \frac{1}{2\lambda^b e^{-B/S}}$ (if it is positive), I becomes a pure market-maker.[10] Being different from the optimal intermediation structure shown by Chapter 2, the pure middleman can be the optimal mode. Moreover, it is not clear which intermediation mode will show up in an equilibrium which we now turn to.

3.4.3 Equilibrium candidate with a pure mode incumbent

Armed with the characterization in Proposition 3.2, we can now perform the equilibrium analysis. Doing so, we shall also derive the best responses of E. We start by ruling out any pure strategy equilibrium where I acts a pure middleman or pure market-maker.

A pure middleman intermediary I does not arise in equilibrium because, given Proposition 3.2, I only adopts a pure middleman mode with $p^m = 1$ when $f^E = 1$. But facing $p^m = 1$, E would rather set $f^E = 1 - \varepsilon$, for some $\varepsilon > 0$ to become the only active intermediary and make a profit of $B\lambda^b f^E > 0$.[11]

[10] This is so because $\lim_{x^m \to 0} b^I(x^m) = 1 - \frac{1}{2\lambda^b e^{-B/S}}$, $\lim_{x^m \to B} b^I(x^m) = 1$. And if $1 - \frac{1}{2\lambda^b e^{-B/S}} < 0$, then $\lim_{f^E \to 0} x^m > 0$.

[11] The best response analysis of E when I is a pure middleman is as follows. When I is a pure middleman, E can only make transactions by setting a fee f^E low enough so that buyers' incentive constraint to trade at I, condition (3.5), is violated, i.e., $1 - p^m < \lambda^b(1 - f^E)$. That is, to undercut I by setting $f^E = 1 - \frac{1-p^m}{\lambda^b} - \varepsilon$ for some $\varepsilon > 0$, as long as this leads to a non-negative f^E. In this way, I becomes

Turn to the case when I is a pure market-maker. Under pessimistic expectation against E, when

$$V^E = \lambda^b e^{-B/S}(1 - f^E) \leq 1 - f^I,$$

E is the second source. Otherwise, I becomes inactive and E is the sole active source.

Consider an equilibrium candidate where E sets a fee $f^{Ec} \in [0,1]$ ("c" for candidate) and a pure market-maker I sets a fee $f^{Ic} \equiv 1 - \lambda^b e^{-B/S}(1 - f^{Ec})$.[12] In this proposed equilibrium, E wants to undercut I whenever possible, or deviate to $f^{Ed} = 1$ to get the whole trading surplus from its transactions ("d" for deviation). In the case that $f^{Ec} > 0$, it is profitable to undercut f^{Ic} by setting $f^{Ed} = f^{Ec} - \varepsilon$. As such, E becomes the sole source and makes a profit of $B\lambda^b f^{Ed}$. On the other hand, if $f^{Ec} = 0$, then E would rather take the full surplus of each transaction by $f^{Ed} = 1$ and makes a profit of $B\lambda^b e^{-B/S} > 0$.

We summarize these observations in the following lemma.

Lemma 3.1. *There does no exist a pure strategy equilibrium where I is a pure middleman or a pure market-maker.*

In Chapter 2, the pure middleman mode is suboptimal with a passive outside platform. Lemma 3.1 shows the same conclusion holds in an equilibrium when the outside market is operated by another intermediary E and responses actively. E has an incentive to undercut I, and this gives a positive outside value for buyers. To lower buyers' outside value, I activates the market-maker mode. Lemma 3.1 asserts an even stronger claim: The pure market-maker mode can not be in an equilibrium. This follows the intuition of the classical Bertrand-Edgeworth game. Since the matching probability is less than one at the market-maker sector, it is a profitable deviation for E to set $f^E = 1$ to abstract full trading surplus from agents that are not matched at I.

3.4.4 The best response of the entrant

Now we turn to the existence of an equilibrium with I of a mixed mode. According to Proposition 3.2, as a market-making middleman, I's optimal strategy features $p^m = f^I$. Let's denote the price/fee level by ψ. To construct the equilibrium, we first analyze the most profitable fee choice of E taking ψ as given.

inactive and E makes a profit of $B\lambda^b f^E$.

[12] According to Proposition 3.2, f^{Ic} is required to satisfy the best response of I. Any $f^I \neq f^{Ic}$ would lead to a deviation of I.

Under pessimistic beliefs against E, for any $f^E \in [0,1]$, if there exists an $x^m \in [0, B]$ such that buyers and sellers are willing to first search in I before turning to transaction opportunities in E, as stated in (3.12), then E remains to be the second source. This means E can have two roles in the multi-market search environment: (1) E can work as a second source following pessimistic beliefs; or (2) E can break the beliefs and act as a sole source by undercutting I.

E **works as a sole source** Let's first explore the possibility of E being a sole source. Insert $V^m = 1 - \psi$ and $V^m = e^{-x^s}(1 - V^E - \psi) + V^E$, the condition to maintain a bad market allocation (3.12) becomes $f^E \geq 1 - \frac{1}{\lambda^b e^{-x^s}}(1-\psi)$, for all $x^s \in [0, B/S]$. The right hand side takes the minimum value at $x^s = B/S$. Therefore, as long as

$$f^E \geq 1 - \frac{1}{\lambda^b e^{-B/S}}(1-\psi), \tag{3.17}$$

E is a second source.

To become the sole source, E can set f^E slightly lower than the right hand side of (3.17) and make profits of

$$\Pi^E_{sole}(\psi) = B\lambda^b\Big(1 - \frac{1}{\lambda^b e^{-B/S}}(1-\psi) - \varepsilon\Big). \tag{3.18}$$

Obviously, if $\psi < 1 - \lambda^b e^{B/S}$, then $\Pi^E_{sole} < 0$.

E **works as a second source** To be a second source, E has to choose an $f^E \geq 1 - \frac{1}{\lambda^b e^{-B/S}}(1-\psi)$ to satisfy (3.17). That is, E's profit maximization problem is

$$\Pi^E_{2nd}(\psi) = \max_{f^E \in [1 - \frac{1}{\lambda^b e^{-B/S}}(1-\psi), 1]} \Big(B - x^m - S(1 - e^{-x^s})\Big)\lambda^b e^{-x^s} f^E, \tag{3.19}$$

subject to the equilibrium constraint (3.1) and (3.2). As I, it is in E's interest to have $V^m = V^s$. On the one hand, $V^m > V^s$ leads to a pure middleman incumbent, leaving zero market share to E. On the other hand, $V^m < V^s$ means E needs to set f^E unnecessarily low — E could increase f^E without changing the distribution of agents and make higher profits. This is formally stated in the following lemma.

Lemma 3.2. *Given $p^m = f^I = \psi$, the optimal solution of problem (3.19) features*

$$V^m(p^m) = V^s(x^m, f^I, f^E). \tag{3.20}$$

Proof. Let $\{\hat{f}^E, \hat{x}^m\}$ denote the optimum. Suppose at the optimum $V^m(p^m) > V^s(\hat{x}^m, f^I, \hat{f}^E)$, then $\hat{x}^m = B$. It follows that

$$V^s(\hat{x}^m, f^I, \hat{f}^E) = 1 - f^I = 1 - p^m = V^m(p^m).$$

And we have reached a contradiction.

Suppose at the optimum, $V^m(p^m) < V^s(\hat{x}^m, f^I, \hat{f}^E)$, or

$$V^s = e^{-B/S}(1 - f^I) + (1 - e^{-B/S})\lambda^b e^{-B/S}(1 - f^E) > 1 - p^m.$$

This implies that $\hat{f}^E < 1 - \frac{1-\psi}{\lambda^b e^{-B/S}}$. But then E gains a higher profit by deviating to $\tilde{f}^E = 1 - \frac{1-\psi}{\lambda^b e^{-B/S}}$. At $\tilde{f}^E$, E maintains $\hat{x}^m$ while extract higher fees from each transaction. □

Insert the expression of $V^i, i = m, s, E$, into constraint (3.20) we have

$$\lambda^b e^{-x^s}(1 - f^E) = 1 - \psi. \tag{3.21}$$

Constraint (3.21) states the trade-off that E faces: increasing f^E leads to less favorable outside option for buyers on I's platform, hence more buyers visit I's middleman sector (x^m increases), and there are more unmatched sellers left for E (e^{-x^s} increases). Substitute for f^E from (3.21) and insert into (3.19) yields

$$\Pi^E_{2nd}(\psi) = \max_{x^m \in [0,B]} \left(B - x^m - S(1 - e^{-x^s})\right)\left(\lambda^b e^{-x^s} - (1 - \psi)\right). \tag{3.22}$$

Facing a ψ, by choosing an f^E, E essentially chooses an x^m to balance its demand and supply. A higher x^m implies less buyers join E after trading at I, i.e., $B - x^m - S(1 - e^{-x^s})$ decreases in x^m. At the same time, more sellers join E since the matching probability is now lower for sellers at I's platform, i.e., e^{-x^s} increases in x^m. So the intermediation structure at I affects the supply and demand at intermediary E.

This trade-off crucially depends on the benefit E has by marginally increasing x^m, which is determined by the equilibrium price level ψ. When ψ is high, it is profitable to have more buyers at I's platform by decreasing x^m and ultimately increase participating buyers to E. E can achieve this by decreasing f^E. When ψ is low, E finds it less profitable to have more participating buyers, and the optimal f^E should be higher. So ψ and f^E are strategic substitutes via market structure x^m. From this point of view, this is a quantity competition.

Meanwhile, there is an aspect that ψ and f^E are strategic complements. Holding x^s constant, (3.21) implies a Bertrand type price competition. In a nutshell, the game we characterize has both a price competition element (when holding x^s constant) and a quantity competition element (when tracing the change due to x^s).

Finally, to have these trade-offs, there much be some space for E to manipulate f^E. Notice when $\psi < 1 - \lambda^b$, there does not exist an $x^m \in [0, B]$ that gives positive profits for E either as a second or a sole source. Hence, such a ψ will not exist in a pure strategy equilibrium. For our equilibrium analysis, we focus on $\psi > 1 - \lambda^b$.

Lemma 3.3. *There does not exist an (pure strategy) equilibrium where $\psi \leq 1 - \lambda^b$.*

Proof. Suppose in the equilibrium, $\psi \leq 1 - \lambda^b$, then

$$V^m = 1 - \psi \geq e^{-x^s}(1 - \psi) + (1 - e^{-x^s})\lambda^b(1 - f^E) = V^s,$$

for all $x^m \in [0, B]$ and $f^E \in [0, 1]$, where equality only holds when $x^m = B$, and $f^E = 0$. This means $x^{m*} = B$. According to Proposition 3.2, this is consistent with I's optimal choice only when $f^{E*} = 1$. But given $f^{E*} = 1.$, I would rather set $\psi = 1$, and this contradicts with $\psi \leq 1 - \lambda^b$. $\square$

The following proposition then characterizes the best response of E for $\psi \in (1 - \lambda^b, 1]$.

Proposition 3.3. *Given multi-market search technologies, and $f^I = p^m = \psi \in (1 - \lambda^b, 1]$, intermediary E's optimal strategy has the following property:*

- *For $\psi \in (1 - \lambda^b, 1 - \lambda^b e^{-B/S}]$, E works as the second source, $\Pi^E_{2nd} > 0 \geq \Pi^E_{sole}$, and the optimal $x^{m*} \in (0, B)$ satisfies*

$$1 - \psi = \lambda^b e^{-x^{s*}}\left(1 - \frac{B - x^m - S(1 - e^{-x^s})}{S(1 - e^{-x^s})}\right); \tag{3.23}$$

Define $\phi(B, S, \lambda^b) \equiv \lambda^b e^{-B/S}\left(1 - \frac{B - S(1 - e^{-B/S})}{S(1 - e^{-B/S})}\right)$,

- *if $\phi(B, S, \lambda^b) \geq 0$, we have*

- *for* $\psi \in (1-\lambda^b e^{-B/S}, 1-\phi(B,S,\lambda^b))$*, E may work as a second or a sole source since both deliver positive profits,* $\Pi^E_{2nd}, \Pi^E_{sole} > 0$*, and if E works as the second source,* $x^{m*} \in (0,B)$ *satisfies (3.23),*
- *for* $\psi \in [1-\phi(B,S,\lambda^b), 1]$*, E undercuts I to become the sole source* $\Pi^E_{sole} > \Pi^E_{2nd} > 0$*;*

- *if* $\phi(B,S,\lambda^b) < 0$*, then for* $\psi \in (1-\lambda^b e^{-B/S}, 1]$*, E may work as a second or a sole source since both deliver positive profits,* $\Pi^E_{2nd}, \Pi^E_{sole} > 0$*, and if E works as a second source,* $x^{m*} \in (0,B)$ *satisfies (3.23).*

Equation (3.23) is the best response function of E. It indicates for a given ψ, the optimal market structure, represented by x^m, that E would like to choose. Consider a range of $[\underline{x}^m, B]$ where the right hand side of (3.23) is non-negative. It then follows that

$$\frac{\partial \psi}{\partial x^m} = -\frac{1}{S}\Big(1-\psi+\frac{\lambda^b e^{-x^s}(1-e^{-x^s}-x^s e^{-x^s})}{S^2(1-e^{-x^s})^2}\Big) < 0,$$

for $x^m \in [\underline{x}^m, B]$. This corresponds exactly to the intuition above: as ψ increases, E finds it's more profitable to compete, he lowers f^E to make the middleman sector of I less favorable and x^m decreases.

3.4.5 Equilibrium analysis

The equilibrium should jointly solve the optimal responses of two intermediaries, (3.16) and (3.23), together with the equilibrium conditions (3.1) and (3.21). Inserting the equilibrium condition (3.21) into (3.23) gives an alternative form of the best response function of E that facilitates our analysis:

$$f^E = b^E(x^m) \equiv \frac{B-x^m-S(1-e^{-x^s})}{S(1-e^{-x^s})}. \tag{3.24}$$

The equilibrium $x^{m*} \in (0,B)$ if it exists should solve $b^I(x^m) = b^E(x^m)$. Proposition 3.4 gives a sufficient condition for the existence and uniqueness of the equilibrium.

Proposition 3.4. *There exists a unique equilibrium that features a market-making middleman I as a first source, and a second source market-maker E, if*

$$1-\lambda^b e^{-B/S} \geq \psi > 1-\lambda^b, \tag{3.25}$$

where $\psi^* = 1 - \lambda^b e^{-x^{s*}}\left(1 - \frac{B-x^{m*}-S(1-e^{-x^{s*}})}{S(1-e^{-x^{s*}})}\right)$, $x^{s*} = \frac{B-x^{m*}}{S}$, *and* $x^{m*} \in (0, B)$ *solves*

$$\frac{B - x^m}{S(1-e^{-x^s})} = \frac{S(1-e^{-x^s})}{\lambda^b e^{-x^s}(2S(1-e^{-x^s}+x^m)}. \tag{3.26}$$

The equilibrium is characterized by a distribution of buyers and sellers

$$\mathcal{N}^* = \{B^I = B, B^E = B - x^{m*} - S(1-e^{-x^{s*}}), S^I = S, S^E = Se^{-x^{s*}}, x^{m*}\},$$

and a price vector P^* *that*

$$f^{I*} = p^{m*} = \psi^*, f^{E*} = 1 - \frac{1-\psi^*}{\lambda^b e^{-x^{s*}}}.$$

And both intermediaries make positive profits.

The equilibrium, if it exists, features a first source intermediary I of a mixed mode serving all agents, and a second source intermediary E that serves the rest of agents who are not matched at I. At the equilibrium, the intermediary structure of I, x^{m*}, is derived by $b^I(x^m) = b^E(x^m)$ indicated by (3.26), which together with $\psi^* \leq 1 - \lambda^b e^{-B/S}$ guarantees x^{m*} is the mutually best response.

Figure 3.1 demonstrates an equilibrium by two variables, the price/fee level represented by f^E, and the market structure represented by x^m. It plots the two best response functions $b^I(x^m)$ in (3.16) and $b^E(x^m)$ in (3.24), and we have marked values as x^m approaches 0 and B.[13] The interaction of the two best responses gives the equilibrium $\{f^{E*}, x^{m*}\}$. The equilibrium distribution and price variables can be accordingly derived, as stated in the proposition.

The figure illustrates the comparative statics. Let's consider exogenous changes of the buyer's meeting rate λ^b and the buyer population B. As λ^b increases to $\lambda^{b\prime}$, $b^I(x^m)$ moves upward while $b^E(x^m)$ does not move, leading to a smaller x^{m*}. This is illustrated

[13]We make use of the following observations:

$$\lim_{x^m \to 0} b^I(x^m) = 1 - \frac{1}{2\lambda^b e^{-B/S}}, \quad \lim_{x^m \to B} b^I(x^m) = 1,$$
$$\lim_{x^m \to 0} b^E(x^m) = \frac{B - S(1-e^{-B/S})}{S(1-e^{-B/S})}, \quad \lim_{x^m \to B} b^E(x^m) = 0.$$

We have plotted one particular scenario that $1 - \frac{1}{2\lambda^b e^{-B/S}} > 0$ and $\frac{B-S(1-e^{-B/S})}{S(1-e^{-B/S})} < 1$. But these restrictions are not required for the existence of an equilibrium.

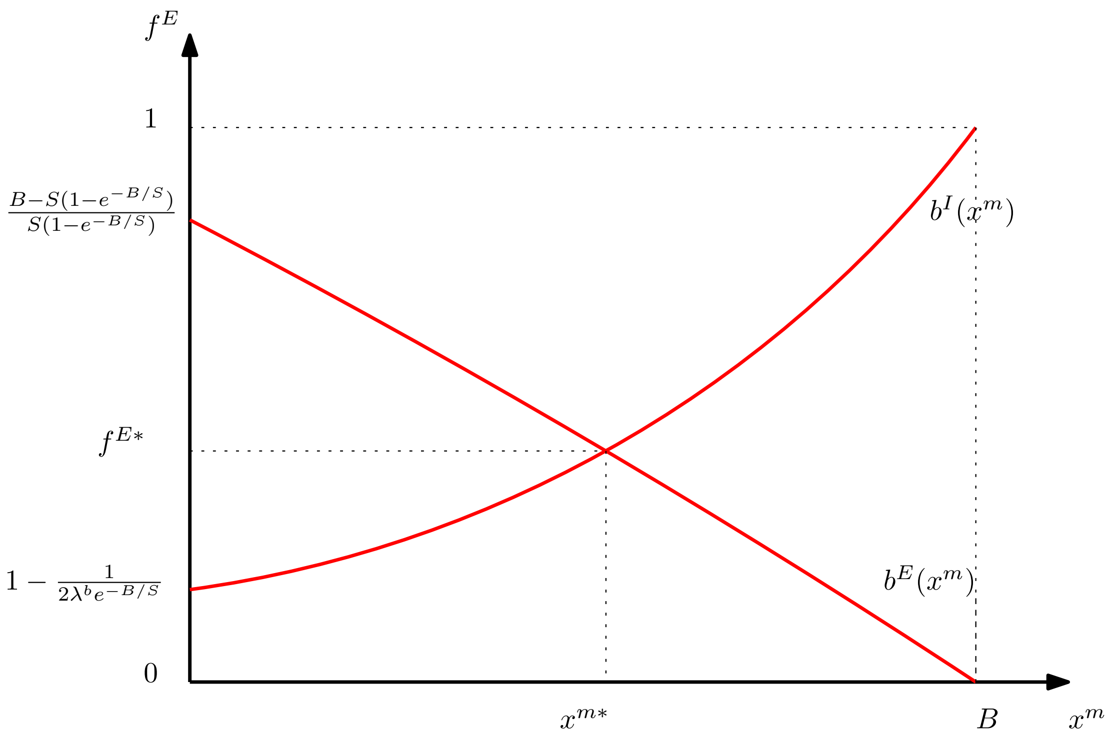

Figure 3.1: Equilibrium under multi-market search

in Figure 3.2. Now consider an exogenous change of B, it follows that

$$\frac{\partial b^E(x^m)}{\partial B} = \frac{1-e^{-x^s}-x^s e^{-x^s}}{S(1-e^{-x^s})} > 0,$$
$$\frac{\partial b^I(x^m)}{\partial B} = -\frac{\lambda^b x^m e^{-2x^s}}{S^2(1-e^{-x^s})^2} < 0.$$

That is, as the population of buyers B increases, $b^I(x^m)$ moves down while $b^E(x^m)$ moves upward, leading to a higher x^{m*}. This is illustrated in Figure 3.3 (the mass of buyers increases from B to B'). Similar comparative statics can be done on the seller population S. We summarize these observations in Corollary 3.1.

Corollary 3.1 (Comparative statics). *Consider a parameter space in which a pure strategy equilibrium exits. Then, an increase in the buyer's meeting rate λ^b in the D market, or a decrease in the buyer-seller population ratio B/S, leads to a smaller middleman sector x^m and a larger platform x^s of I.*

Numerically, it is easy to verify that the sufficient condition (3.25) is satisfied in some parameter space. For example, taking $B = S = 1$, and set a grid of λ^b with two decimals from 0.01 to 0.99, then (3.25) holds for all λ^b grid points between smaller than 0.95. While for λ^b grid points between 0.95 and 0.98 (3.25) is violated, it can be verified that being the second source is more profitable than being a sole source for E, so a pure

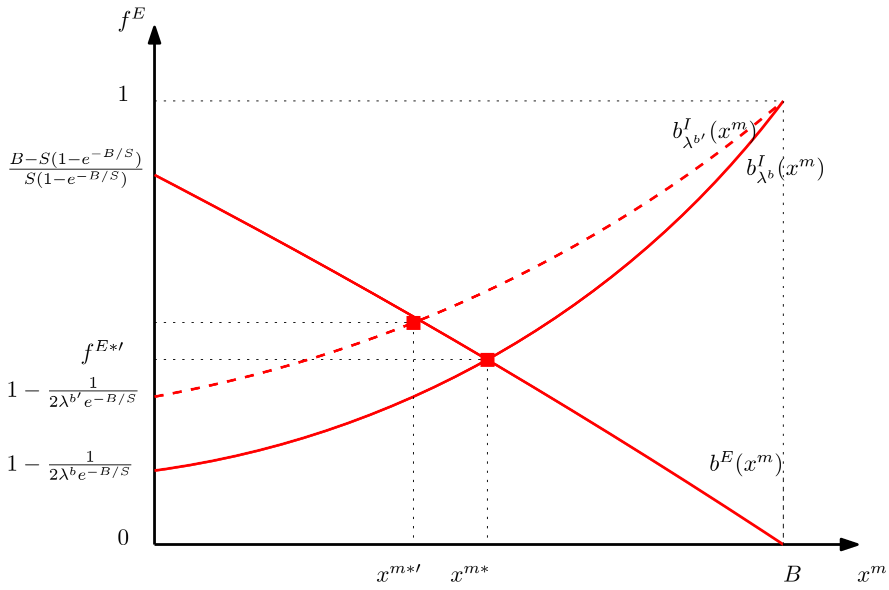

Figure 3.2: Comparative Statics w.r.t. λ^b

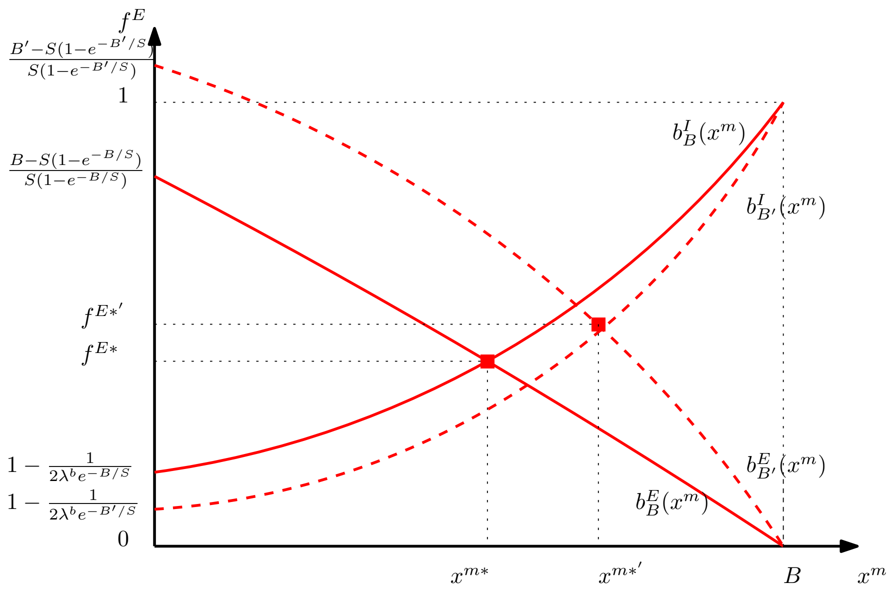

Figure 3.3: Comparative Statics w.r.t. B

strategy equilibrium still exits. For the grid point$\lambda^b = 0.99$, E finds it more profitable to undercut I and become the sole source. Then there does not exist a pure strategy equilibrium.

Equilibrium of mixed strategies When the pure strategies equilibrium does not exist, by applying Theorem 5* of Dasgupta et al. (1986), we show that because Π^I (Π^E) is bounded and weakly lower semi-continuous in f and p^m (in f^E), and $\Pi^I + \Pi^E$ is upper semi-continuous, there exists a mixed strategy equilibrium.

Proposition 3.5. *There exists a mixed strategy equilibrium under multi-market search.*

Given that $f^E \in [0, 1)$ is selected with positive probability, according to Proposition 3.2, I activates its platform with positive probability in equilibrium.

Corollary 3.2. *In the mixed strategies equilibrium, I's platform is activated with positive probability.*

3.5 Conclusion

This paper has proposed a framework to analyze imperfect competition between two intermediaries, with a particular focus on the intermediation mode. We considered two representative business modes of intermediation that the incumbent can adopt: a market-making mode and a middleman mode. We show that the incumbent faces a trade-off between more transactions by using the middleman mode and more profits per transaction by using the market-making mode. This trade-off determines its intermediation model in the competition game with entrant. Therefore, the main insight of Chapter 2 holds with competing intermediaries.

Appendices

Appendix I: Omitted proofs

Proof for proposition 3.2

Using (3.15), the intermediary's problem can be written as

$$\begin{aligned}
\Pi\left(x^m, f^I\right) &= \max_{x^m, f^I} S(1-e^{-x^s}) f^I + x^m p^m \\
&= \max_{x^m, f^I} S(1-e^{-\frac{B-x^m}{S}}) f^I + x^m(1-V^E(x^m, f^E)) - x^m e^{-x^s}(1-V^E(x^m, f^E) - f^I)
\end{aligned}$$

subject to (3.11) and $0 < x^m < B$.

Observe that: $\lim_{x^m \to B} \Pi\left(x^m, f^I\right) = \tilde{\Pi}(B)$ and $\lim_{x^m \to 0} \Pi\left(x^m, f, K\right) = \tilde{\Pi}(0)$, where $\tilde{\Pi}(B)$ is the profit for the pure middleman mode (3.13) and $\tilde{\Pi}(0)$ is the profit for the pure market-maker mode (3.14). Hence, we can compactify the constraint set and set up a general problem to pin down a profit-maximizing intermediation mode using the following Lagrangian:

$$\mathcal{L} = \Pi\left(x^m, f^I\right) + \mu_0 x^m + \mu_b(B - x^m) + \mu_f\left(1 - V^E(x^m, f^E) - f^I\right),$$

where the μ's ≥ 0 are the lagrange multiplier of each constraint. The following first order conditions are necessary:

$$\frac{\partial \mathcal{L}}{\partial x^m} = \frac{\partial \Pi\left(x^m, f^I\right)}{\partial x^m} + \mu_0 - \mu_b - \mu_f \frac{\partial V^E\left(x^m, f^I\right)}{\partial x^m} = 0, \tag{27}$$

$$\frac{\partial \mathcal{L}}{\partial f^I} = \frac{\partial \Pi\left(x^m, f^I\right)}{\partial f^I} - \mu_f = 0, \tag{28}$$

The solution is characterized by these and the complementary slackness conditions of the three constraints.

First, (28) implies that we must have

$$\mu_f = S(1-e^{-x^s}) + x^m e^{-x^s} > 0,$$

which implies the binding constraint (3.11),

$$f^I = V^E\left(x^m, f^E\right). \tag{29}$$

Second, we show that the pure middleman mode is optimal if $f^E = 1$. Given $f^E = 1$, then $V^E(x^m, f^E = 1) = 0$. Using (29) and (3.15), the intermediary's problem can be written as $\Pi\left(x^m\right) = \max_{x^m} S(1-e^{-x^s}) + x^m$, where $\Pi\left(x^m, f^I\right)$ is concave in x^m. The first order condition with respect to x^m is $1 - e^{-x^s} = 0$. Therefore, at the optimal $x^{m*} = B$.

Next, given $f^E < 1$, we show that I's platform is active, i.e., $x^m < B$ at the optimum. Substituting μ_f into (27),

$$\begin{aligned}
\mu_b - \mu_0 &= (1-e^{-x^s})(1-\lambda^b e^{-x^s}(1-f^E)) - (x^m + S(1-e^{-x^s}))\frac{\lambda^b}{S} e^{-x^S}(1-f^E) \\
&\equiv \phi(x^m \mid B, S, \lambda^b, f^E).
\end{aligned} \tag{30}$$

Suppose that the solution is $x^{m*} = B$. Then, (30) yields $\phi(B \mid \cdot) = \mu_b = -\frac{B}{S}\lambda^b(1 - f^E) < 0$, which contradicts $\mu_b \geq 0$. Hence, the solution must satisfy $x^{m*} < B$ (which implies $\mu_b = 0$). Suppose that the solution is $x^{m*} = 0$. Then, (30) yields $\phi(0 \mid \cdot) = -\mu_0 = (1-e^{-B/S})(1-2\lambda^b e^{-B/S}(1-f^E))$, which requires

$$f^E \leq 1 - \frac{1}{2\lambda^b e^{-B/S}}.$$

This leads to the conditions in the proposition. Set $\mu_b = \mu_0 = 0$, we have $\phi(x^m \mid B, S, \lambda^b, f^E) = 0$ according to (30). Since $x^m < B$, we have $1 - e^{-x^s} > 0$. This gives condition (3.16).

Finally, it is straightforward to verify the second order condition using the Hessian of $\mathcal{L}$ with respect to $[f^I, x^m]$.■

Proof for proposition 3.3

First, from (3.18) and (3.22), it is straightforward to see that if $\psi \leq 1 - \lambda^b$, then $\Pi^E_{2nd} \leq 0$; and if $\psi \leq 1 - \lambda^b e^{-B/S}$, then $\Pi^E_{sole} \leq 0$. These observations give the signs of the profits in all three cases of the proposition.

Second, the first order condition of Π^E_{2nd} with respect to x^m is

$$\frac{\partial \Pi^E_{2nd}(x^m, \psi)}{\partial x^m}|_{x^m} = -S(1 - e^{-x^s})(\lambda^b e^{-x^s} - (1 - \psi)) + (B - x^m - S(1 - x^{-x^S}))\lambda^b e^{-x^S} = 0. \tag{31}$$

Since $\Pi^E_{2nd}(x^m)$ is continuously differentiable on $[0, B]$, the maximum point is among $x^m = 0$, $x^m = B$ or an $\hat{x}^m$ such that $\frac{\partial \Pi^E_{2nd}(x^m,\psi)}{\partial x^m}|_{\hat{x}^m} = 0$. Under $\psi > 1 - \lambda^b$, $x^m = B$ is not a maximum point since $\Pi^E_{2nd}(x^m = B) = 0$. When $\frac{\partial \Pi^E_{2nd}}{\partial x^m}|_{x^m=0} > 0$, that is $1 - \lambda^b e^{-B/S}\left(1 - \frac{B - S(1 - e^{-B/S})}{S(1 - e^{-B/S})}\right) > \psi$, $x^m = 0$ is not a maximum point. The following discussion depends on the sign of $\phi(B, S, \lambda^b) \equiv \lambda^b e^{-B/S}\left(1 - \frac{B - S(1 - e^{-B/S})}{S(1 - e^{-B/S})}\right)$.

If $\phi(B, S, \lambda^b) \geq 0$, then for $\psi \in [1 - \phi(B, S, \lambda^b), 1]$, as a second source E chooses $x^{m*} = 0$, which is strictly dominated by being the first source (lowering f^E slightly by some $\varepsilon > 0$)

$$\Pi^E_{2nd}(\psi) = (B - S(1 - e^{-B/S})(\lambda^b e^{-B/S} - (1 - \psi)) < B(\lambda^b e^{-B/S} - (1 - \psi) - \varepsilon) = \Pi^E_{sole}(\psi).$$

For $\psi \in (1 - \lambda^b e^{-B/S}, 1 - \phi(B, S, \lambda^b))$, E compares Π^E_{sole} and Π^E_{2nd} to decide which is more profitable.

If $\phi(B, S, \lambda^b) < 0$, then for $\psi \in [1 - \lambda^b e^{-B/S}, 1]$, one needs to compare the profit of being second source and sole source. These observations give that $0 < x^{m*} < B$ in the first two bullet points of the proposition.

Finally, rearranging (31) gives (3.23). ■

Proof for proposition 3.4

Define

$$g(x^m) \equiv b^I(x^m) - b^E(x^m).$$

First, notice $g(x^m)$ is continuous differentiable with respect to x^m. Taking the limits of x^m to 0 and B, we have

$$\begin{aligned}
\lim_{x^m \to 0} g(x^m) &= \lim_{x^m \to 0} b^I(x^m) - \lim_{x^m \to 0} b^E(x^m) \\
&= 1 - \frac{1}{2\lambda^b e^{-x^s}} - \frac{B - S(1 - e^{-B/S})}{S(1 - e^{-B/S})} \\
&= -\frac{1}{2\lambda^b e^{-x^s}} - \frac{B}{S(1 - e^{-B/S})} < 0; \\
\lim_{x^m \to B} g(x^m) &= \lim_{x^m \to B} b^I(x^m) - \lim_{x^m \to B} b^E(x^m) \\
&= 1 - 0 > 0.
\end{aligned}$$

According to the Intermediate Value Theorem, there exits an $x^{m*} \in (0, B)$ such that $b^I(x^{m*}) = b^E(x^{m*})$.

Second, x^{m*} is unique. This is because g is monotone increasing in x^m on $(0, B)$. Taking the first order derivate with respect to x^m, we have

$$g'(x^m) = b^{I'}(x^m) - b^{E'}(x^m) > 0,$$

where

$$b^{I'}(x^m) = \frac{(1 + 2(1 - e^{-x^s}))S(1 - e^{-x^s}) + x^m}{(2S(1 - e^{-x^s}) + x^m)^2 \lambda^b e^{-x^s}} > 0,$$
$$b^{E'}(x^m) = -\frac{1}{S^3(1 - e^{-x^s})^2}(1 - e^{x^s} - x^s e^{-x^s}) < 0.$$

Thirdly, if $\psi = 1 - \lambda^b e^{-x^{s*}}\left(1 - \frac{B - x^{m*} - S(1 - e^{-x^{s*}})}{S(1 - e^{-x^{s*}})}\right) \in (1 - \lambda^b, 1 - \lambda^b e^{-B/S})$, then according to Proposition 3.3, E has no incentive to deviate to the sole source since $\Pi^E_{sole} < 0$.

Finally, according to Proposition 3.2 and Proposition 3.3, both intermediaries make positive profits. ■

Proof for corollary 3.1

Let's consider a marginal increase in λ^b, B and S in turns.

Consider an increase in λ^b: $\lambda^{b\prime} = \lambda^b + \varepsilon$ with $\varepsilon > 0$. We show the equilibrium market structure under $\lambda^{b\prime}$ follows $x^{m*\prime} < x^{m*}$, where x^{m*} is the in the equilibrium under λ^b and $x^{m*\prime}$ is the in the equilibrium under $\lambda^{b\prime}$. We denote the best response function under λ^b by $b^i_{\lambda^b}(x^m)$ and that under $\lambda^{b\prime}(x^m)$ by $b^i_{\lambda^{b\prime}}(x^m)$, $i = I, E$. Since $\frac{\partial b^E(x^m)}{\partial \lambda^b} = 0$, $\frac{\partial b^I(x^m)}{\partial \lambda^b} > 0$, for $x^m \in (0, B)$, we have $b^E_{\lambda^{b\prime}}(x^m) = b^E_{\lambda^b}(x^m)$, and $b^I_{\lambda^{b\prime}}(x^m) > b^I_{\lambda^b}(x^m)$.

Suppose $x^{m*\prime} = x^{m*}$, then

$$b^I_{\lambda^{b\prime}}(x^{m*\prime}) > b^I_{\lambda^b}(x^{m*\prime}) = b^I_{\lambda^b}(x^{m*}) = b^E_{\lambda^b}(x^{m*}) = b^E_{\lambda^{b\prime}}(x^{m*}) = b^E_{\lambda^{b\prime}}(x^{m*\prime}).$$

But $b^I_{\lambda^{b\prime}}(x^{m*\prime}) > b^E_{\lambda^{b\prime}}(x^{m*\prime})$ implies that $x^{m*\prime} = x^{m*}$ is not in an equilibrium.

Suppose $x^{m*\prime} > x^{m*}$, then

$$b^I_{\lambda^{b\prime}}(x^{m*\prime}) > b^I_{\lambda^b}(x^{m*\prime}) > b^I_{\lambda^b}(x^{m*}) = b^E_{\lambda^b}(x^{m*}) = b^E_{\lambda^{b\prime}}(x^{m*}) > b^E_{\lambda^{b\prime}}(x^{m*\prime}).$$

Again, $b^I_{\lambda^{b\prime}}(x^{m*\prime}) > b^E_{\lambda^{b\prime}}(x^{m*\prime})$ implies that $x^{m*\prime} > x^{m*}$ can not be in an equilibrium.

Consider an increase in B: $B' = B + \varepsilon$ with $\varepsilon > 0$. We show the equilibrium market structure under B' follows $x^{m*\prime} > x^{m*}$. Since

$$\frac{\partial b^E(x^m)}{\partial B} = \frac{1 - e^{-x^s} - x^s e^{-x^s}}{S(1 - e^{-x^s})} > 0,$$
$$\frac{\partial b^I(x^m)}{\partial B} = -\frac{2S(1 - e^{-x^s})^2 + x^m}{(2S(1 - e^{-x^s}) + x^m)^2 \lambda^b e^{-x^s}} < 0,$$

for $x^m \in (0, B)$, we have $b^E_{B'}(x^m) > b^E_B(x^m)$, and $b^I_{B'}(x^m) < b^I_B(x^m)$.

Suppose $x^{m*\prime} = x^{m*}$, then

$$b^I_{B'}(x^{m*\prime}) < b^I_B(x^{m*\prime}) = b^I_B(x^{m*}) = b^E_B(x^{m*}) < b^E_{B'}(x^{m*}) = b^E_{B'}(x^{m*\prime}).$$

But $b^I_{B'}(x^{m*\prime}) < b^E_{B'}(x^{m*\prime})$ implies that $x^{m*\prime} = x^{m*}$ can not be in an equilibrium.

Suppose $x^{m*\prime} < x^{m*}$, then

$$b^I_{B'}(x^{m*\prime}) < b^I_B(x^{m*\prime}) < b^I_B(x^{m*}) = b^E_B(x^{m*}) < b^E_{B'}(x^{m*}) < b^E_{B'}(x^{m*\prime}).$$

Again, $b^I_{B'}(x^{m*\prime}) < b^E_{B'}(x^{m*\prime})$ implies that $x^{m*\prime} < x^{m*}$ can not be in an equilibrium.

Consider an increase in S: $S' = S + \varepsilon$ with $\varepsilon > 0$. We show the equilibrium market structure under S' follows $x^{m*\prime} < x^{m*}$. Since

$$\frac{\partial b^E(x^m)}{\partial S} = -\frac{(1-e^{-x^s} - x^s e^{-x^s})x^s}{S(1-e^{-x^s})^2} < 0,$$
$$\frac{\partial b^I(x^m)}{\partial S} = \frac{x^s\left[2(1-e^{-x^s})^2 + \frac{x^m}{S}(1-\frac{1-e^{-x^s}}{x^s})\right]}{S(2(1-e^{-x^s}) + x^m/S)^2\lambda^b e^{-x^s}} > 0,$$

for $x^m \in (0, B)$, we have $b^E_{S'}(x^m) < b^E_S(x^m)$, and $b^I_{S'}(x^m) > b^I_S(x^m)$.

Suppose $x^{m*\prime} = x^{m*}$, then

$$b^I_{S'}(x^{m*\prime}) > b^I_S(x^{m*\prime}) = b^I_S(x^{m*}) = b^E_S(x^{m*}) > b^E_{S'}(x^{m*}) = b^E_{S'}(x^{m*\prime}).$$

Sut $b^I_{S'}(x^{m*\prime}) > b^E_{S'}(x^{m*\prime})$ implies that $x^{m*\prime} = x^{m*}$ can not be in an equilibrium.

Suppose $x^{m*\prime} > x^{m*}$, then

$$b^I_{S'}(x^{m*\prime}) > b^I_S(x^{m*\prime}) > b^I_S(x^{m*}) = b^E_S(x^{m*}) > b^E_{S'}(x^{m*}) > b^E_{S'}(x^{m*\prime}).$$

Again, $b^I_{S'}(x^{m*\prime}) > b^E_{S'}(x^{m*\prime})$ implies that $x^{m*\prime} > x^{m*}$ can not be in an equilibrium. This completes the proof of corollary 3.1.■

Proof for proposition 3.5

Consider a game between I, who selects $(f, p^m) \in [0, \bar{f}] \times [0, \bar{p}]$ with a payoff $\Pi^I = \Pi^I(f, p^m \mid f^E)$, and E, who selects $f^E \in [0, 1]$ with a payoff $\Pi^E = \Pi^E(f^E \mid f, p^m)$. Here, we set $\bar{f}, \bar{p} > 1$ and $f > 1$ ($\bar{p} > 1$) leads to an inactive platform (middleman sector). We apply Theorem 5 of Dasgupta and Maskin (1986) to show there exists a mixed strategy equilibrium.

Given Theorem 5 of Dasgupta et al. (1986), it is sufficient to show that Π^I (Π^E) is bounded and weakly lower semi-continuous in f and p^m (in f^E), and $\Pi^I + \Pi^E$ is upper semi-continuous. Clearly, Π^I (Π^E) is bounded in $(f, p^m) \in [0, \bar{f}] \times [0, \bar{p}]$ (in $f^E \in [0, 1]$).

Both of the profit functions are continuous except at

$$\min\{f, p^m\} = 1 - V^E(f^E), \tag{32}$$

where $V^E(f^E)$ is evaluated at $x^m = 0$. So we shall pay attention to this discontinuity point.

First, we show that $\Pi^I(f, p^m \mid f^E)$ is weakly lower semi-continuous in (f, p^m). Give the discontinuous point in (32), we have

$$\Pi^I(f, p^m \mid f^E) = \begin{cases} S(1-e^{-x^s})f + x^m p^m, & \text{if } \min\{f, p^m\} \le 1 - V^E(f^E) \\ 0 & \text{otherwise,} \end{cases}$$

where in the second situation, the price/fee of I is not competitive to the fee of E, hence agents will trade via E, rather than I, and so I will become inactive. Consider some $f_\varepsilon \in [0, 1]$, and some $f, p^m > 0$ such that $\min\{f, p^m\} = 1 - V^E(f^E)$. For any sequence $\{(f^{(j)}, p^{m(j)})\}$ converging to (f, p^m) such that no two $f^{(j)}$'s, and no two $p^{m(j)}$'s are the same, and $f^{(j)} \le f, p^{m(j)} \le p^m$, we must have $\min\{f^{(j)}, p^{m(j)}\} \le 1 - V^E(f^E)$. Hence,

$$\lim_{j\to\infty} \Pi^I(f^{(j)}, p^{m(j)} \mid f^E) = \Pi^I(f, p^m \mid f^E),$$

satisfying the definition of weakly lower semi-continuity (see Definition 6 in page 13 of Dasgupta and Maskin, 1986, or condition (9) in page 384 of Maskin, 1986).

Second, we shall show that $\Pi^E(f^E \mid f, p^m)$ is lower semi-continuous in f^E. Consider a potential discontinuity point $f_0 \in (0,1)$ satisfying (32) such that

$$\Pi^E(f^E \mid f, p^m) = \begin{cases} B\lambda^b f^E, & \text{if } f^E < f_0 \\ (B - x^m - S(1-e^{-x^s}))\lambda^b f^E & \text{if } f^E \geq f_0. \end{cases}$$

Clearly, this function is lower semi-continuous, since for every $\epsilon > 0$ there exists a neighborhood U of f_0 such that $\Pi^E(f^E \mid \cdot) \geq \Pi^E(f_0 \mid \cdot) - \epsilon$ for all $f^E \in U$.

Finally, we prove the upper semi-continuity of $\Pi^I + \Pi^E$. For this purpose, consider all sequences of $\{f^{(j)}, p^{m(j)}, f^{E(j)}\}$ that converges to $\{\hat{f}, \hat{p}^m, \hat{f}^E\}$ that satisfies $\min\{\hat{f}, \hat{p}^m\} = 1 - V^E(\hat{f}^E)$.

Consider first an extreme in which case $\min\{f^{(j)}, p^{m(j)}\} \leq 1 - V^E(f^{E(j)})$ for all j. As the equilibrium is that I is visited prior to E, we must have

$$\lim_{j\to\infty} \Pi^I(f^{(j)}, p^{m(j)} \mid f^{E(j)}) + \Pi^E(f^{E(j)} \mid f^{(j)}, p^{m(j)}) = \Pi^I(\hat{f}, \hat{p}^m \mid \hat{f}^E) + \Pi^E(\hat{f}^E \mid \hat{f}, \hat{p}^m).$$

Consider next the other extreme in which $\min\{f^{(j)}, p^{m(j)}\} > 1 - V^E(f^{E(j)})$ for all j. Then, in the equilibrium only E is active and we must have

$$\lim_{j\to\infty} \Pi^I(f^{(j)}, p^{m(j)} \mid f^{E(j)}) + \Pi^E(f^{E(j)} \mid f^{(j)}, p^{m(j)}) = B\lambda^b \hat{f}^E.$$

If $\hat{f} \geq \hat{p^m}$, then

$$\Pi^I(\hat{f}, \hat{p}^m \mid \hat{f}^E) + \Pi^E(\hat{f}^E \mid \hat{f}, \hat{p}^m) = B\hat{p}^m = B(1 - \lambda^b(1-\hat{f}^E)) > B\lambda^b \hat{f}^E.$$

If $\hat{f} < \hat{p^m}$, then

$$\Pi^I(\hat{f}, \hat{p}^m \mid \hat{f}^E) + \Pi^E(\hat{f}^E \mid \hat{f}, \hat{p}^m) = B(1-e^{-\frac{B}{S}})\hat{f} + B\lambda^b e^{-\frac{B}{S}}\hat{f}^E > B[(1-e^{-\frac{B}{S}}) + \lambda^b e^{-\frac{B}{S}}]\hat{f}^E > B\lambda^b \hat{f}^E.$$

Thus,

$$\lim_{j\to\infty} \Pi^I(f^{(j)}, p^{m(j)} \mid f^{E(j)}) + \Pi^E(f^{E(j)} \mid f^{(j)}, p^{m(j)}) < \Pi^I(\hat{f}, \hat{p}^m \mid \hat{f}^E) + \Pi^E(\hat{f}^E \mid \hat{f}, \hat{p}^m).$$

As these two extreme cases give the upper and lower bounds respectively, all the other sequences give some limits in between. Therefore,

$$\lim_{j\to\infty} \Pi^I(f^{(j)}, p^{m(j)} \mid f^{E(j)}) + \Pi^E(f^{E(j)} \mid f^{(j)}, p^{m(j)}) \leq \Pi^I(\hat{f}, \hat{p}^m \mid \hat{f}^E) + \Pi^E(\hat{f}^E \mid \hat{f}, \hat{p}^m),$$

for any of the sequences converging to $\{\hat{f}, \hat{p}^m, \hat{f}^E\}$, and so $\Pi^I + \Pi^E$ is upper semi-continuous. This completes the proof of Proposition 3.5. ■

Appendix II : The game with a pure middleman entrant

⊙ Single-market search. First of all, suppose E is a middleman with a price p^E, then $V^E = 1 - p^E$. If I is a pure middleman, then it makes a profit of Bp^m with price $p^m = 1 - V^E$. If I activates a platform, it must satisfy the participation constraints,

$$\eta^s(x^s)(1 - p^s) \geq V^E,$$
$$p^s - f^I \geq 0.$$

Under these conditions, it holds that

$$f^I \leq 1 - V^E.$$

Hence, the resulting profit of I satisfies $S(1 - e^{-x^s})f^I + x^m p^m < (Sx^s + x^m)\max\{f^I, p^m\} \leq B(1 - V^E)$. That is, the pure middleman mode dominates any other modes with an active platform. Therefore, under single-market search, I must be a pure middleman in all possible equilibria.

⊙ Multi-market search: E is a pure middleman. With multi-market search, when E is a pure middleman, an active platform of I has to satisfy the incentive constraints,

$$\begin{aligned} 1 - p^s &\geq 1 - p^E \\ p^s - f^I &\geq 0. \end{aligned}$$

These constraints imply: $f^I \leq p^E$. Similarly, an active middleman sector has to satisfy $p^m \leq p^E$. Then, if $\max\{p^m, f^I\} \leq p^E$, then I can be a market-making middleman, and if

$$\min\{p^m, f\} \leq p^E,$$

then trade can occur in either one of the sectors, and so I can be an active intermediary. The profit of I is

$$S(1 - e^{-x^s})f + x^m p^m.$$

Noting $x^s = \frac{B - x^m}{S}$, we see from this expression that the profit maximization requires that $x^m = B$ with $p^m = f = p^E$. Hence an active platform is not profitable. Then, since the two intermediaries compete with price, any equilibrium must be subject to the Bertrand undercutting, leading to $p^m = p^E = 0$ and zero profits.

4

Managerial Labor Market Competition and Incentive Contracts

4.1 Introduction

Executives are incentivized by having their compensation closely tied to firm performance in the form of bonuses, stocks, options, etc. Traditionally, it is believed that incentive contracts are designed to align the interests of executives with those of shareholders. In recent decades, however, we have seen that competition for executives is increasingly influential in shaping incentive contracts. For example, in the "battle for talent", IBM targets the 50th percentile of both cash and equity compensation among a large group of benchmark companies. The contract of individual executives is further adjusted according to "the skills and experience of senior executives that are highly sought after by other companies and, in particular, by our (IBM's, added) competitors." Similarly, Johnson & Johnson compare "salaries, annual performance bonuses, long-term incentives, and total direct compensation to the Executive Peer Group companies" who compete with Johnson & Johnson "for executive talent".[1]

Despite its relevance for the industry, a characterization of how heterogeneous firms compete for executives is still missing in the literature, and the consequences for executive contracts have remained unclear. For example, in the assignment models

[1]See detailed compensation policy in IBM and Johnson & Johnson proxy statements, which are accessible at `https://www.ibm.com/annualreport/2017/assets/downloads/IBM_Proxy_2018.pdf` and `http://www.investor.jnj.com/gov/annualmeetingmaterials.cfm` (visited on Oct 27, 2018).

(e.g., Gabaix and Landier 2008, Edmans et al. 2009), equilibria are static and dynamic features such as career concerns or job ladder effects are absent. In the multiple-period models (e.g., Holmström 1999, Oyer 2004, Giannetti 2011), it is usually assumed that all companies compete with the same spot market wage, and executives cannot transit to potentially more productive companies. Other dynamic models concentrate more on the firm/rank choice of executives rather than competition between firms (e.g., Gayle et al. 2015).

This chapter focuses on the competition between heterogeneous firms in a tractable framework that combines dynamic moral hazard and equilibrium labor search. In particular, I allow executives to *search on-the-job* along a *hierarchical job ladder towards larger firms* as in Postel-Vinay and Robin (2002). This feature, which is missing in the existing studies on managerial labor markets, drives the key results.[2] The model considers two types of agents: executives and firms. Executives are heterogeneous in the *general* managerial productivity, which evolves stochastically depending on their current and past effort. Firms are heterogeneous in time-invariant asset size.[3] As in Gabaix and Landier (2008), the marginal impact of an executive's productivity increases with the value of the firm under his or her control. While output is observable, the effort is not. Thus, a moral hazard problem arises. To resolve the problem, the firm and the executive sign a long-term incentive contract. Moreover, the executive has limited commitment to the relationship and may encounter outside poaching offers from an external labor market. By making use of poaching offers, the executive can renegotiate with the current firm or transit to a larger poaching firm, where the compensation contract is determined in a sequential auction. Essentially, the current and the poaching firms are engaged in a Bertrand competition for the executive.

The competition from poaching offers impacts executive incentive contracts via two channels. *First*, as in Postel-Vinay and Robin (2002), competition from outside offers increases total compensation. When the poaching firm is smaller than the current firm, the executive may use the offer to negotiate with the current firm for a higher pay.

[2]The executive job ladder exists in the real world. The career path of Richard C. Notebaert is a good example as is described by Giannetti (2011): "Notebaert led the regional phone company Ameritech Corporation before its 1999 acquisition by SBC Communication Inc.; after, he held the top job at Tellabs Inc., a telecom-equipment maker; finally, in 2002, he became CEO of Qwest Communications International Inc." This anecdotal description is consistent with the data evidence in the literature. Huson et al. (2001) report that the fraction of outsider CEOs increased from 15.3% in the 1970s to 30.0% at the beginning of the 1990s. A similar pattern is reported by Murphy and Zabojnik (2007).

[3]I measure firm size by market capitalization (value of debt plus equity).

When the poaching firm is larger, it can always outbid the current firm since firm size contributes to the production. Thus, the executive uses the current firm as a threat point to negotiate with the poaching firm and transits to the poaching firm. In either case, the executive climbs up the job ladder towards a higher compensation level and (or) a larger firm size. *Second*, poaching offers generate a new source of incentives and consequently reduce the need for performance-based incentives. Poaching firms are willing to bid more for more productive executives. Meanwhile, the productivity of an executive is stochastically determined by his or her past effort. Together, these factors imply that effort today will lead to a more favorable offer from the same poaching firm in the future. This potential gain from labor market competition becomes what I call *labor market incentives* in this paper. Firms can take advantage of these labor market incentives and give fewer performance-based incentives to executives, but still resolve the moral hazard issue.

These two channels enable the model to shed light on two puzzling facts: the *firm-size pay-growth premium* and the *firm-size incentive premium*, both of which are firstly documented in this paper, complementing other stylized facts in the literature (see, e.g., Edmans et al. 2017). The *firm-size pay-growth premium* refers to the empirical finding that starting with the same total compensation, the executive of a larger firm experiences a higher compensation growth. Based on the data for U.S. listed firms, I find that for a 1% increase in firm size, the annual pay-growth rate increases by 15.4%. This big gap in pay-growth rate significantly contributes to the pay differentials between small and large firms. My explanation for this premium is as follows. Executive compensation grows because firms desire to retain executives in response to poaching offers. Due to a firm-size effect in production, larger firms are more capable of countering poaching offers; hence, their executive compensation tends to grow faster.

The *firm-size incentive premium* refers to the empirical fact that performance-based incentives embedded in bonuses, stocks, options, etc. increase with firm size after controlling for total compensation. As will be elaborated in Section 3, a 1% increase in firm size leads to a 0.35% increase in performance-based incentives.[4] This incentive

[4]Performance-based incentives are measured by the dollar change in firm-related wealth per percentage change in firm value. It is an ex ante measure of incentives in compensation package, before the realization of firm performance. Consistent with the literature (Jensen and Murphy 1990, Hall and Liebman 1998), I use all firm-related wealth instead of only current compensation to calculate performance-based incentives. See Section 3 for more details on motivating facts.

premium has excluded the size differentials in pay levels, and it reflects that larger firms tend to allocate a higher fraction of the compensation package to performance-related pay. My explanation for this premium is based on labor market incentives. In the model, an executive is motivated by two sources of incentives which substitute for each other: performance-based incentives and labor market incentives. I show that labor market incentives decrease with firm size. To motivate executives in larger firms, the performance-based incentives are required to be higher.

There are two reasons why executives in larger firms receive less labor market incentives. The first reason lies in the job ladder. Executives from larger firms are located "higher" on the job ladder. Consequently, the chance of receiving an outside offer that beats the current value is lower. Thus, labor market incentives for larger firms are smaller. Indeed, for individuals that are at the bottom of the job ladder, labor market incentives can be large enough that no performance-related pay is required, which goes back to the original model of Postel-Vinay and Robin (2002). The second reason is based on a wealth effect. Executives of larger firms are expected to receive higher compensation in the future, i.e., the size premium in pay-growth; thus, the certainty equivalents of their future expected utilities are higher. Given a diminishing marginal utility, at a higher certainty equivalent, the utility gain from a more favorable poaching offer is smaller. As a result, labor market incentives are smaller as firm size increases.

To provide empirical evidence and structurally estimate the model, I assembled a new dataset on executive job turnovers by merging the ExecuComp and BoardEX databases. ExecuComp is the standard data source for executive compensation studies. It contains annual records on top executives' compensation in firms comprising the S&P 500, MidCap, and SmallCap indices. BoardEX contains detailed executive employment history, and it helps to identify the employment status after executives leave the S&P firms. For executives that are not identified in BoardEX, I further searched for executive profiles and biographies using LinkedIn and Bloomberg. In the final data sample, there are 35,088 executives and a total of 218,168 executive-year observations spanning the period 1992 to 2016.

I first provide reduced-form evidence to support the model set-up and implications. Using the merged data, I document a job-to-job transition rate of around 5%, which is stable over the years and across industries. Moreover, there is a job ladder in the firm-size dimension: about 65% of job-to-job transitions are towards larger firms. This justifies the hierarchical job ladder featured in the model. Second, I test whether the job ladder "position" of an executive matters for his/her chance of job-to-job transitions. Specifically, using a Cox model, I find that executives in larger firms are less likely to experience job-to-job transitions, which is in line with the model's prediction. Finally, using the variation across industries, I find that firm-size premiums in both pay-growth and incentives are higher in industries where the managerial labor market is more active. I proxy the activeness of an executive labor market by the job-to-job transition rate, the fraction of outside CEOs and the average of the general ability index (Custódio et al. 2013).

It is difficult to numerically solve for the optimal contract in the presence of an incentive compatibility constraint, limited-commitment constraints, and shocks of large support, as one needs to solve for the promised value in each state of the world. I address this issue by using the recursive Lagrangian approach (Marcet and Marimon 2017), under which I only need to solve for one Lagrangian multiplier to find the optimal contract. This multiplier represents the weight of the executive in a constructed Pareto problem, and it keeps track of various constraints and job-to-job transitions. Based on this multiplier, the optimal incentive pay and promised values can be solved.

Using the simulated method of moments (SMM), I estimate the model by targeting data moments on executive compensation, incentives, and turnovers, as well as on firm size and profitability. Importantly, I do not explicitly target the firm-size premiums in compensation growth and performance-based incentives. Yet, the estimated model quantitatively captures both. The predictions of the estimated model are very close to the premium estimates from the data, which corroborates that the model mechanism plays an essential role in explaining both premiums. A counter-factual decomposition shows that labor market incentives account for more than 40% of total incentives among small-firm executives, around 15% for medium-firm executives and less than 5% for large-firm executives.

Based on the structural estimation, I use a counterfactual exercise to quantitatively account for the sharp increases in executive compensation since the mid-1970s. I show that with an exogenous rise in the arrival rate of poaching offers, the model generates increases in total compensation and performance-based incentives, more inequality of compensation across executives, and a higher correlation between compensation and firm size. Quantitatively, the above changes in model-simulated data match well with the data facts documented by Frydman and Saks (2010). The intuition is that the managerial labor market was much thinner before the 1970s, which is supported by the evidence presented in Frydman (2005) and Murphy and Zabojnik (2007).

Finally, there is a clear policy implication of the model regarding how to regulate the compensation of highly paid executives, especially in large firms. Rather than only focusing on large firms, it is important to lower the bids for executives from small and medium firms. This could be achieved via various reforms such as more independent compensation committees, greater mandatory pay (or pay ratio) disclosure or say-on-pay legislation. In this way, the competitive pressure in the overall managerial labor market will decrease. In the model, there is a spillover effect whereby higher bids from a set of firms boost executive pay not only in those firms but also in all firms that are higher on the job ladder. In a comparative static analysis, I show that, compared to an increase in the bids from large firms, the same increment in those of small and medium firms has a similar effect on the compensation of large firms, and has a more substantial impact on the compensation of the whole managerial market.

The rest of the paper is organized as follows. In Section 2, I provide a detailed literature review. In Section 3, I present the motivating facts of the firm-size pay-growth premium and incentive premium. I further show that both premiums significantly increase when the executive labor market is more active. I then set up the model in Section 4, where I characterize the optimal contract and explain the premiums. Section 5 presents reduced-form evidence and Section 6 estimates the model. Section 7 explains the sharp increase in executive pay since the mid-1970s. Section 8 discusses the policy implications of the research. Finally, Section 9 concludes.

4.2 Literature review

This paper contributes to two strands of literature in understanding pay differentials between small and large firms. The first strand explains the differentials using assignment models. Gabaix and Landier (2008), Tervio (2008) and Eisfeldt and Kuhnen (2013) present competitive assignment models to explain why total compensation increases with firm size. Consistent with these studies, I use a multiplicative production function to characterize the contribution of executives. My model provides a similar prediction of the relationship between total compensation and firm size. Since my model is dynamic, it also captures the growth of total compensation, which is absent in the existing literature. More importantly, it afford a different view on the pay differentials between small and large firms. In this paper, executives are paid much more in larger firms not because they are more talented (e.g., Gabaix and Landier 2008) but because they are lucky to be matched with a firm whose size makes it better able to counter outside offers. Further along this strand of research, Edmans et al. (2009) and Edmans and Gabaix (2011) add a moral hazard problem to the assignment framework and explain why performance-based incentives increase with firm size. Their explanation is based on the notation that total compensation increases with firm size. Yet, these models do not explain why, after controlling for total compensation, a firm-size incentive premium still exists. My model is a dynamic and search-frictional version of their framework and highlights a hierarchical job ladder. Besides the explanation given in Edmans et al. (2009), the job ladder in my model gives rise to labor market incentives, which contributes to the understanding of the firm-size incentive premium.

The second strand of literature explains the pay differentials using agency problems. Margiotta and Miller (2000) derive and estimate a multi-period principal-agent model with moral hazard. Based on this model, Gayle and Miller (2009) show that large firms face a more severe moral hazard problem, so that higher equity incentives are needed to satisfy the incentive compatibility condition. Gayle et al. (2015) embed the model developed by Margiotta and Miller (2000) into a generalized Roy model. They find that the quality of the signal is unambiguously poorer in larger firms, and this explains most pay differentials between small and larger firms. In contrast to my focus on managerial labor market competition, Gayle et al. (2015) find that the career concern channel does not explain the size premium in their estimation. The critical difference between the model used by Gayle et al. (2015) and my model is that in their model job-to-job transitions are based on a Roy model and are in general not directed towards larger firms, whereas the driving force of my explanation is a hierarchical job ladder where executives move from small to large firms. The different approach to modeling job-to-job transitions explains why labor market incentives contribute much less in the framework of Gayle et al. (2015). Using executives' job-to-job transition data, I show that the hierarchical job ladder does exist.

This paper also contributes to the literature explaining the rise of executive compensation in recent decades. My paper is closest in spirit to the explanation based on executive mobility. The literature shows that the increases in compensation coincide with the increased occupational mobility of executives, which is brought about by an increased importance of executives general managerial skills in comparison to firm-specific knowledge (Frydman 2005, Frydman and Saks 2010, Murphy and Zabojnik 2007). Giannetti (2011) develops a model to show that job-hopping opportunities can help explain not only the increase in total pay, but also the structure of managerial contracts. In Section 7, I provide a counterfactual analysis showing that with an exogenous increase of poaching offer arrival rate from 5% to 40% per year, my model can account for the sharp increase in total compensation and performance-based incentives, as well as a much higher correlation between firm size and total compensation.

The fourth stream of the literature to which I contribute is the one on executive turnovers, see e.g., Taylor (2010), Jenter and Kanaan (2015), Kaplan and Minton (2012) and Peters and Wagner (2014). In particular, on the incentive effect of turnovers, Remesal et al. (2018) estimate a dynamic moral hazard model allowing for endogenous compensation and dismissals. Their estimation shows that dismissal threats play a small role in CEO incentives, whereas the bulk of CEO incentives comes from the flow and deferred compensation. These results justify my focus on performance-related compensation. Wang and Yang (2016) study the optimal termination in a dynamic contract with moral hazard and stochastic market value shocks. The model generates rich insights on voluntary and involuntary dismissals, and termination plays distinct roles depending on the level of market values. Their analysis, however, is abstract from firm size, and the market value shocks are homogeneous to all executives. In contrast, poaching offers in my model are heterogeneous across executives.

Finally, this paper is closely related to the work of Abrahám et al. (2016), who aim to explain wage inequality in the general labor market by combining repeated moral hazard and on-the-job search. Besides the differences in topics, there is a critical difference that distinguishes the two papers: The productivity of agents is independent over time in the model of Abrahám et al. (2016), while it is persistent in my model. Therefore, in my model, working hard today rewards the agent in the future. It is this feature that gives rise to labor market incentives and explains the firm-size incentive premium. This feature is absent in their model.

In terms of modeling, this paper links two strands of literature. One strand is the extensive literature on optimal long-term contracts with private information and commitment frictions, e.g., Townsend (1982), Rogerson (1985), Spear and Srivastava (1987), Phelan and Townsend (1991), Harris and Holmstrom (1982), Thomas and Worrall (1990) and Phelan (1995). I build on this literature by embedding an optimal contracting problem with moral hazard and two-sided limited commitment into an equilibrium search model. In doing so, the outside environment is endogenized, which significantly changes the optimal contract. Another important strand of literature uses structural search models to evaluate wage dispersions. Postel-Vinay and Robin (2002), Cahuc et al. (2006) and Lise et al. (2016) among others estimate models with job ladders and sequential auctions. Compared to this literature, I add a dynamic moral hazard problem, which allows me to understand how search frictions influence a long-term contract. The model of Postel-Vinay and Robin (2002) is a special case of my model when the moral hazard problem is absent. In addition, the managerial labor market is an appropriate environment for their framework. In real life, it happens very often that executives are contacted and ”auctioned” by competing firms for promotion, as is described by Khurana (2004).

4.3 Motivating facts

This paper is motivated by two firm-size premiums: the pay-growth premium and the incentive premium. As far as I am aware, these facts are firstly documented. Moreover, I show that both premiums are larger in industries where the managerial labor market is more active, where labor market thickness is measured by job-to-job transition rates, the general ability index and the fraction of inside CEOs at the industry or industry-year level. The primary data source for the analyses is Standard & Poor's ExecuComp database. Variables about executive labor markets come from a newly assembled dataset on executive turnovers and the other two datasets provided by Custódio et al. (2013) and Martijn Cremers and Grinstein (2013). All nominal quantities are converted into constant 2016 dollars using GDP deflater. Section 5 presents a statistical description of the data.

4.3.1 Size pay-growth premium

I measure firm size by market capitalization, defined by the common shares outstanding times the fiscal year close price. The executive annual compensation-growth rate is measured by the first-order difference of $\log(tdc1)$ where $tdc1$ is the total compensation including the sum of salary and bonus, the value of restricted stocks and options granted, and the value of retirement and long-term compensation schemes. Column (1) in table 4.1 presents a regression of $\Delta \log(tdc1)$ on firm size, controlling for lagged $\log(tdc1)$ and dummies on tenure, age and year times industry.[5] The estimation indicates that starting from the same level of total compensation, for a 1% increase in firm size, the annual compensation-growth rate increases by 11.2%. The premium is slightly larger with an estimate of 15.4% in column (2) after further controlling for *operating profitability*, *market-book ratio*, *annualized stock return*, title dummies such as *director*, *CEO*, *CFO*, etc. Definitions of these variables are provided in the table note.

To link the size pay-growth premium with managerial labor markets, I explore the variation across industries. An industry is an appropriate sub-labor market since more than 60% of executive job-to-job transitions are within the industry (see details in Section 5). As a direct test of whether size pay-growth premium is related to a more active managerial labor market, I use four proxies to measure the labor market thickness and test if the interactions between these proxies and firm size are significant. The first two proxies are job-to-job transition rates on the industry-year level (Fama-French 48 industries and fiscal years). *EE90* is the job-to-job transition rate where a job-to-job transition is defined by an executive leaves the current firm and starts to work in another firm within *90* days. Similarly, *EE190* is the job-to-job transition rate where a job-to-job transition is defined by a gap of no more than 190 days. The third proxy *gai* is the mean of the general ability index of CEOs at the industry-year level. The general ability index itself is the first principal component of five proxies that measure the generality of a CEO's human capital based on his or her lifetime work experience (Custódio et al. 2013).[6] The last proxy, *inside CEO*, is the industry-level percentage of the CEOs who are promoted inside the firm (Martijn Cremers and Grinstein 2013). It accounts for all

[5]The result is robust by controlling for lagged total compensation using 100 or 200 group dummies that equally divide the sample according to the value of $\log(tdc1)$.

[6]The five proxies to measure the general ability of CEO's are: the number of positions that CEO performed during his/her career, the number of firms where a CEO worked, the number of industries at the four-digit SIC level where a CEO worked, a dummy variable that equals one if a CEO held a CEO position at another firm, and a dummy variable that equals one if a CEO worked for a multi-division firm.

Table 4.1: Pay-growth increases with firm size

	$\Delta \log(tdc1)$					
	(1)	(2)	(3)	(4)	(5)	(6)
log(firm size)$_{-1}$	0.112***	0.154***	0.108***	0.107***	0.141***	0.127***
	(0.00903)	(0.0129)	(0.00183)	(0.00189)	(0.00177)	(0.00489)
log(firm size)$_{-1}$ $\times$ *EE90*			0.0711*			
			(0.0403)			
log(firm size)$_{-1}$ $\times$ *EE190*				0.0759**		
				(0.0353)		
log(firm size)$_{-1}$ $\times$ *gai*					0.0233***	
					(0.00546)	
log(firm size)$_{-1}$ $\times$ *inside CEO*						-0.0232***
						(0.00696)
log(tdc1)$_{-1}$	-0.290***	-0.390***	-0.251***	-0.251***	-0.304***	-0.253***
	(0.0200)	(0.0262)	(0.00173)	(0.00173)	(0.00267)	(0.00173)
other controls		X	X	X	X	X
tenure dummies	X	X	X	X	X	X
age dummies	X	X	X	X	X	X
year dummies	X	X	X	X	X	X
industry	X	X				
year $\times$ *industry*	X	X				
Observations	129,068	106,819	106,820	106,820	58,188	106,820
adj. R^2	0.157	0.216	0.260	0.260	0.233	0.262

Note: This table presents firm-size pay-growth premium and its correlation with the activeness of executive labor markets. The dependent variable is the first-order difference of $\log(tdc1)$ where *tdc1* is the total compensation including the sum of salary and bonus, the value of restricted stocks and options granted, and the value of retirement and long-term compensation schemes. Firm size is measured by the market capitalization, defined by the common shares outstanding times the fiscal year close price. I control for lagged *tdc1*. Whenever possible, I also control for age, tenure, and year times industry dummies. Other controls include *operating profitability*, *market-book ratio*, *annualized stock return*, *director*, *CEO*, *CFO* and *interlock*. *director* is a dummy which equals to 1 if the executive served as a director during the fiscal year. *CEO* and *CFO* are dummies defined by whether the executive served as a CEO (and CFO) during the fiscal year. *interlock* is a dummy which equals to 1 if the executive is involved in an interlock relationship. An interlock relationship generally involves one of the following situations: (1) The executive serves on the board committee that makes his or her compensation decisions; (2) the executive serves on the board of another company that has an executive executive serving on the compensation committee of the indicated executive's company; (3) the executive serves on the compensation committee of another company that has an executive executive serving on the board of the indicated executive's company. I use four variables to measure the activeness of the executive labor market at the industry or industry-year level. *EE90* is the industry-year level job-to-job transition rate where a job-to-job transition is defined by an executive leaves the current firm and starts to work in another firm within *90* days. *EE190* defines a job-to-job transition with a gap of no more than 190 days. *gai* is the mean of general ability index of CEOs at the industry-year level. The original data is provided by Custódio et al. (2013). *insider CEO* is the industry level percentage of internally promoted new CEOs between 1993 and 2005. The original data on this variable is provided by Martijn Cremers and Grinstein (2013). For all variables, an industry is based on Fama-French 48 categories, a year is based on fiscal years. Standard errors clustered on the firm $\times$ fiscal-year level are shown in parentheses, and I denote symbols of significance by * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

new CEOs between 1993 and 2005 using Fama-French 48-industry categories.

The picture that emerges in the last four columns of table 4.1 is not ambiguous: All four interaction terms are statistically and economically significant, and the signs confirm that the size growth premium is larger in industries/years where the executive labor market is more active. Specifically, the coefficient of 0.0759 on the interaction with *EE190* and a standard deviation of 0.0224 for *EE190*, imply that a one-standard-deviation increase in job-to-job transition measured by *EE190* gives an increase in pay-growth premium from 7.3% to 7.47%. Similarly, a one-standard-deviation increase in job-to-job transition measured by *EE90* implies that pay-growth premium increases from 7.3% to 7.43%. Given a standard deviation of 0.253 for *gai* and a coefficient of 0.0233, a one-standard deviation increase in *gai* leads to an increase in pay-growth premium from 7.3% to 7.89%. Given a standard deviation of 0.122 for *inside CEO* and a coefficient of 0.0233, a one-standard deviation increase in *inside CEO* leads to an increase of pay-growth premium from 7.3% to 7.58%.

4.3.2 Size incentive premium

I measure performance-based incentives in executive contracts by "*delta*". By definition, *delta* equals the dollar increase in executives' firm-related wealth for a percentage increase in firm value. It measures incentives before firm performance is realized. Thus, it is ex ante.[7] As has been documented in Edmans et al. (2009) and is replicated in table 4.2 column (1), *delta* is positively correlated with firm size: For a 1% increase in firm size, measured by market capitalization, performance-based incentives increase by 0.59%. Edmans et al. (2009) argued that because executives in larger firms are paid higher, they require more incentives to induce effort.

However, the level of total compensation does not explain the entire size incentive premium. The positive correlation between performance-based incentives and firm size remains after controlling for total compensation, $\log(tdc1)$, in table 4.2 column (2): For a 1% increase in firm size, *delta* increases by 0.36%, which accounts for more

[7] *delta* is also known as "the value of equity at stake" or "dollar-percentage incentives". Empirical studies of pay-to-performance have used a wide range of specifications to measure this relationship. Two common alternatives are the dollar change in executive wealth per dollar change in firm value (the Jensen-Murphy statistic) and the dollar amount of wealth that an executive has at risk for a 1% change in the firms value (the value of equity at stake or *delta*). The Jensen-Murphy statistic is the correct measure of incentives for activities whose dollar impact is the same regardless of firm size, and the value of equity at stake is appropriate for actions whose value scales with firm size. The latter is the modeling approach of this paper.

Table 4.2: Performance-based incentives increase with firm size

	log(*delta*)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(firm size)	0.585***	0.360***	0.315***	0.316***	0.325***	0.316***
	(0.0141)	(0.0247)	(0.0029)	(0.0029)	(0.0036)	(0.0029)
log(firm size) $\times$ *EE90*			0.772*			
			(0.1228)			
log(firm size) $\times$ *EE190*				0.716**		
				(0.1054)		
log(firm size) $\times$ *gai*					0.055***	
					(0.0112)	
log(firm size) $\times$ *inside CEO*						-0.087***
						(0.0196)
log(tdc1)		0.609***	0.693***	0.692***	0.687***	0.684***
		(0.035)	(0.046)	(0.046)	(0.056)	(0.046)
other controls	X	X	X	X	X	X
tenure dummies	X	X	X	X	X	X
age dummies	X	X	X	X	X	X
year dummies	X	X	X	X	X	X
industry	X	X				
year $\times$ *industry*	X	X				
Observations	146,747	128,006	128,006	128,006	79,476	128,006
adj. R^2	0.442	0.482	0.486	0.487	0.482	0.485

Note: This table presents firm-size incentive premium and its correlation with the activeness of executive labor markets. The dependent variable is log($delta$) where *delta* is the dollar change in firm related wealth for a percentage change in firm value. Firm size is measured by the market capitalization, defined by the common shares outstanding times the fiscal year close price. *tdc1* is the total compensation including the sum of salary and bonus, the value of restricted stocks and options granted, and the value of retirement and long-term compensation schemes. Whenever possible, I control for age, tenure, and year times industry dummies. Other controls include *operating profitability*, *market-book ratio*, *annualized stock return*, *director*, *CEO*, *CFO* and *interlock*. *director* is a dummy which equals to 1 if the executive served as a director during the fiscal year. *CEO* and *CFO* are dummies defined by whether the executive served as a CEO (and CFO) during the fiscal year. *interlock* is a dummy which equals to 1 if the executive is involved in an interlock relationship. Please refer to the footnote of table 4.1 for a definition of *interlock*. I use four variables to measure the activeness of the executive labor market at the industry or industry-year level. *EE90* is the industry-year level job-to-job transition rate where a job-to-job transition is defined by an executive leaves the current firm and starts to work in another firm within *90* days. *EE190* defines a job-to-job transition with a gap of no more than 190 days. *gai* is the mean of general ability index of CEOs at the industry-year level. The original data is provided by Custódio et al. (2013). *insider CEO* is the industry level percentage of internally promoted new CEOs. The original data on this variable is provided by Martijn Cremers and Grinstein (2013). For all variables, an industry is based on Fama-French 48 categories, a year is based on fiscal years. Standard errors clustered on the firm $\times$ fiscal-year level are shown in parentheses, and I denote symbols of significance by * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

than half of the size premium estimated in column (1). The estimated elasticity 0.36 of incentives to size in column (2) is the *size incentive premium* that I aim to explain. It excludes the pay-level effects and only reflects the proportion of incentive-related pay. As I will show in Section 6, the estimates of size incentive premium in both columns (1) and (2) can be quantitatively captured by the model.

I further test if size incentive premium is related to managerial labor market using the same four proxies as in the last subsection: *EE90*, *EE190*, *gai* and *inside CEO*. As presented in columns (3) to (6) in table 4.2, all interaction terms are statistically and economically significant, and the signs indicate that the size incentive premium is larger in industries/years where the executive labor market is more active.[8]

Finally, I show that size incentive premium decreases as executives approach retirement age. Starting from Gibbons and Murphy (1992), *age* has been used as an indicator for career concerns: The older the executive is, the less influential that managerial labor market is on incentive contract design. If size incentive premium is at least partly caused by managerial labor markets, we would expect the premium to decrease with age. This is indeed the case, as is shown in figure 4.1. The size incentive premium starts with 0.652 at age 35, and gradually decreases to around 0.35 after age 50. This pattern holds with or without controls.

4.4 The model

In this section, I construct an equilibrium model of the managerial labor market. The model features on-the-job search, poaching offers and contract renegotiation. I embed a bilateral moral hazard problem into the labor market equilibrium. Poaching offers are used to renegotiate with the current firm, leading to compensation growth. Thus, the size growth premium is linked to a firm's capability of overbidding poaching offers. Poaching offers also generate new incentives, called labor market incentives in the model, which constitute a wedge between the total incentives required to motivate executives and performance-based incentives provided by firms. That is, the size

[8] Specifically, given a coefficient of 0.415 on the interaction with *EE190* and a standard deviation of 0.0224 for *EE190*, a one-standard-deviation increase in *EE190* leads to an increase in the elasticity from 0.525 to 0.534. Similarly, a one-standard-deviation increase in job-to-job transition measured by *EE90* implies a higher elasticity of 0.532. A standard deviation of 0.253 for *gai*, together with the coefficient of 0.0648, means that with a one-standard deviation increase in *gai*, the elasticity increases by 0.016. A standard deviation of 0.122 for *inside CEO*, together with the coefficient of 0.0458, means that a one-standard deviation increase in *inside CEO* leads to an increase of 0.0056 of the elasticity.

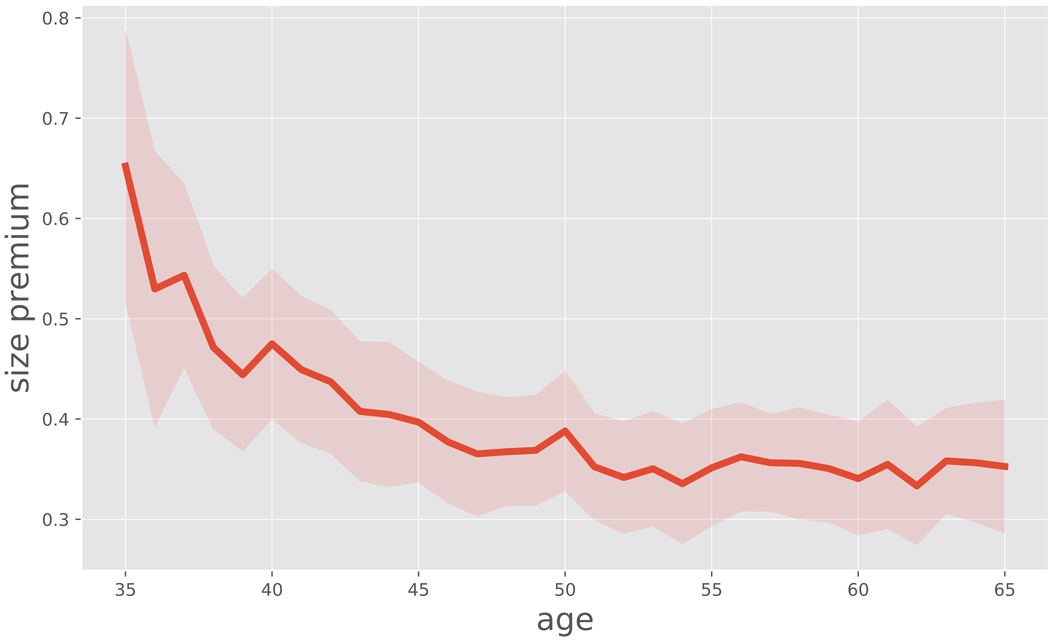

Figure 4.1: Size premium in performance-based incentives decreases with age

Note: The figure depicts the size premium in performance-based incentives at each age from 35 to 65. Each point is one estimated coefficient of the interaction term between one *age dummy* and *log(firm size)* in the following regression,

$$log(delta)_{it} = \Phi' age\ dummies_{it} \times log(firm\ size)_{it} + \Psi' X_{it} + \epsilon_{it},$$

where i denotes an executive, t denotes the fiscal year, *age dummies* is a set of 31 dummies for each age from 35 to 65, *firm size* is measured by market capitalization by the end of the fiscal year, calculated by a firm's common shares outstanding times the close price by fiscal year, X denotes a vector of control variables including a constant term. I control for total compensation *log(tdc1)* and dummies of executive tenure, age, and fiscal year time industry. A 95% confidence interval is plotted using the standard error clustered on firm times fiscal year. The full regression result is provided in Appendix B.

premium in performance-based incentives is linked to labor market incentives. These mechanisms are used to explain the size premiums in both annual pay-growth and performance-based incentives. I now formally introduce the model.

4.4.1 Ingredients

Agents There is a fixed measure of individuals: They are either employed as executives or not hired as executives but looking for executive jobs. I call the latter "executive candidates". Individuals die with some probability. Once an individual dies, a newborn enters the economy.

Individuals want to maximize expected lifetime utility,

$$\mathbb{E}_0 \Sigma_{t=0}^{\infty} (\beta \times (1-\eta))^t (u(w_t) - c(e_t)),$$

where $\beta \in (0,1)$ is the discount factor, $\eta \in (0,1)$ is the death probability, utility of consumption $u : \mathbb{R} \to \mathbb{R}$ is increasing and concave and $c(\cdot)$ is the dis-utility of effort. The effort e_t takes two values, $e_t \in \{0, 1\}$, and cost of $e_t = 0$ is normalized to zero. I denote $c(1)$ by c.

Executives are heterogeneous in *general* managerial skills, or productivity, denoted by $z \in \mathbb{Z} = \{z^{(1)}, z^{(2)}, ..., z^{(n_z)}\}$. z is observable to the executive himself or herself and to firms that he or she meets, and can be carried with the individual through job-to-job transitions.[9]

Individual productivity z changes over time according to a Markov process. Denote z_t as the beginning of period t productivity. Given z_t and effort e_t, the end of period t productivity z_{t+1} follows $\Gamma_z(z_{t+1}|z_t, e_t)$. I denote the process by $\Gamma_z(z_{t+1}|z_t)$ for $e_t = 1$, and $\Gamma^s_z(z_{t+1}|z_t)$ for $e_t = 0$ (s is for shirking). To start the process, I assume all unmatched executive candidates have the same starting productivity, $z = z_o$. In the following, whenever it is not confusing, I will denote z_t by z and z_{t+1} by z'.

While z and z' are observable to firms, effort e is not. Hence, there is *moral hazard.* To impose some structure on the moral hazard problem, I define the likelihood ratio as follows:

$$g(z'|z) \equiv \frac{\Gamma^s(z'|z)}{\Gamma(z'|z)}.$$

As a likelihood ratio, its expectation is one, $E[g(z'|z)] = 1$. I further assume that taking effort delivers a higher expected productivity, $E_\Gamma[z'g(z'|z)] < E_\Gamma[z']$, and that taking effort is more likely to deliver a higher productivity, i.e., $g(z'|z)$ is non-increasing in z'[10].

On the other side of the managerial labor market are firms characterized by the scale of assets, called firm size, denoted by $s \in \mathbb{S} = [\underline{s}, \bar{s}]$. Firm size is permanent and exogenous.[11] A match between a worker of productivity z and a firm of size s produces a flow of output (or a cash flow) $y(s, z) = \alpha_0 s^{\alpha_1} z$, $\alpha_0 \in (0, 1), \alpha_1 \in (0, 1]$. This function form entails that executive effort "roll out" across the entire firm up to a scale of α_0. It has constant return to scale if $\alpha_1 = 1$ and decreasing return to scales if $\alpha_1 < 1$.[12]

[9]Here I treat productivity as general management skills rather than firm-specific skills. However, firm-specific skills could be included using a productivity discount upon a job-to-job transition. This is left as a future extension.

[10]This is the monotone likelihood ratio property (MLRP).

[11]From the point of view of the labor economics literature, one could interpret firm size here as "the productivity of the job" or "firm type". Instead of using the total number of employees, I use total asset value as a proxy for firm size since the performance of the firm is usually measured by return on assets.

[12]There has been some discussion in this literature on the appropriate production function of execu-

Managerial labor market The managerial labor market is search-frictional. Executives and firms are imperfectly informed about executive types and location of firms. The search friction precludes the optimal assignments assumed in Gabaix and Landier (2008). Agents are only informed about each other's types when they meet. Search is random; executives and executive candidates all sample from the same, exogenous job offer distribution $F(s)$. Unmatched candidates meet firms with probability λ_0, while on-the-job executives meet firms with probability λ. I treat these parameters as exogenous.[13]

When a candidate meets a firm, they bargain on a contract. Suppose the continuation value of an unmatched executive candidate is W^0. Then, the firm ultimately offers a contract with a continuation value W^0, for there is no other credible threat. The individual then enters the next period as an employed executive.

When an on-the-job executive meets an outside firm, a compensation renegotiation is triggered. Otherwise, the executive has an interest in transiting to the outside firm. I allow the incumbent firm to respond to outside offers: A sequential auction is played between the executive and both firms as in Postel-Vinay and Robin (2002). If the poaching firm is larger, the executive moves to the alternative firm, for the poaching firm can always pay more than the current one can match. Alternatively, if the poaching firm is smaller, then the executive may use the outside offer to negotiate up his/her compensation. This sequential auction mechanism characterizes labor market competition in this paper.

Timing Time is discrete, indexed by t, and continues forever. The period of an executive candidate is simple — he or she is matched to a firm with some probability and starts with a contract of continuation value W^0. An on-the-job executive enters a period with a beginning-of-period productivity z and current firm of size s. The timing is shown in figure 4.2.

1. **Compensation:** The firm s firstly pays compensation w for this period, in accor-

tives, see e.g., Edmans et al. (2017). Take s as firm size and z as the executive's per unit contribution to shareholder values. An additive production function such as $y(s,z) = s + z$ implies that the effect of executives on firm value is independent of firm size. This specification is appropriate for a perk consumption. A multiplicative production function such as $y(s,z) = sz$ is appropriate for executives' actions that can roll out across the entire firm and thus have a greater effect in a larger company. The latter is the function form adopted here.

[13] So we are in a "partial" equilibrium, in contrast to the "general" equilibrium where the labor market tightness is determined in the equilibrium.

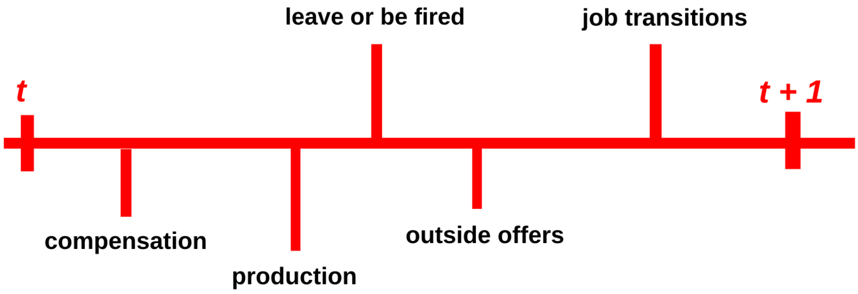

Figure 4.2: Timing

dance with the contract.

2. **Production:** Then, the executive enters the production phase. He or she chooses an effort level, $e \in \{0, 1\}$. His or her productivity z' is then realized according to $\Gamma(z'|z, e)$. The firm only observes the output $y(z, s)$, not the effort e. This is the moral hazard problem.

3. **Labor market:** With probability η the executive dies; otherwise, with probability λ_1, a job offer of firm size $\tilde{s} \sim F(s)$ arrives. The renegotiation game is triggered. The executive may stay in the current firm and receive higher compensation, or transit to the poaching firm. The value of the contract to the executive is determined by a sequential auction between the current and poaching firms.

The compensation w, effort choice e and job-to-job transitions in future periods are stipulated in the contract between the firm and the executive, defined on a proper state of the world, which we now turn to.

Contractual environment A contract defines the transfers and actions of the executive and the firm in a matched pair for all future histories, where a history summarizes the past states of the world. I define a history as follows. Call $h_t = (z'_t, \tilde{s}_t)$ the state of the world by the end of period t, where z'_t is the realized productivity by the end of t and $\tilde{s}_t$ is the size of a poaching firm. Let $\tilde{s} = s_o$ if there is no poaching firm. The history of productivity and the poaching firm up to period t, denoted by $h^t = (h_1, h_2, .., h_t)$, is common knowledge to the executive and the firm, and is fully contractible.

The two elements included in the history — productivity and poaching firm — correspond to the frictions I have in the contracting problem, namely moral hazard and search frictions. First, while productivity is included in the history and is contractible, executive's effort is not and needs to be induced by incentives. Hence, an incentive compatibility constraint is required. Second, by including poaching firms in the history, I allow a contract to stipulate whether and how to counter poaching offers. That is, competing for executives is included in the contracting problem.

Countering outside offers is optimal (or subgame perfect in game terminology); it is, therefore, necessary to allow limited commitment for both sides — to terminate the contract when the surplus is negative. Firms cannot commit to the relationship if the profits are negative. When the outside offer comes from a larger firm, the firm's participation constraint binds, and the match separates. Likewise, executives cannot commit to the match if the current firm cannot provide more than the outside value, be the unmatched value W^0 or the offer value of a poaching firm. In the former case, the executive is fired by the board. In the latter case, the executive transits to the poaching firm.

Given the information structure, I define a feasible contract as a plan that defines compensation $w_t(h^{t-1})$, a recommended effort choice $e_t(h^{t-1})$ and whether to terminate the contract $I_t(h^t)$ in every future history, represented by

$$\{e_t(h^{t-1}), w_t(h^{t-1}), I_t(h^t)\}_{t=0}^{\infty},$$

that satisfies the participation constraints of both sides and an incentive compatibility constraint.

To further simplify, I impose two assumptions. First, I assume taking effort, $e = 1$, is optimal. This assumption is consistent with Gayle et al. (2015) and in accordance with the fact that almost all executives in my data are provided with an incentive package. Secondly, I assume a reasonable support of productivity z such that the profits of a firm are always non-negative at the the unmatched value W^0. As a result, firing is excluded, and $I_t(h^t) = 1$ is equivalent to a job-to-job transition.[14]

A simplified contract state space To recursively write up the contracting problem, I use the executive's beginning-of-period expected utility, denoted by V, as a co-state variable to summarize the history of productivities and outside offers. A dynamic

[14] If I allow a large domain of z such that for some z the profit is negative at promised continuation value W^0, then firing happens. This is left as an extension in the future.

contract, defined recursively, is

$$\sigma \equiv \{e(V), w(V), W(z', \tilde{s}, V) | z' \in \mathbb{Z}, s' \in \mathbb{S} \text{ and } V \in \Phi\},$$

where e is the effort level suggested by the contract (optimal level is assumed to be 1), w is the compensation, W is the promised value given for a given state $(z', \tilde{s})$, and Φ is the set of feasible and incentive compatible expected utilities that can be derived following Abreu et al. (1990).[15]

4.4.2 Optimal contracting problem

In this section, I first characterize the participation constraints derived from the sequential auction, then I describe the contracting problem.

Sequential auction Here I illustrate the sequential auction using value functions.[16] Let $\Pi(z, s, V)$ denote the discounted profit of a firm with size s, executive of beginning-of-period productivity z and a promised value to the executive V. The maximum bidding values $\overline{W}(z, s)$ are defined by

$$\Pi(z, s, \overline{W}) = 0.$$

The firm would rather fire the executive (normalizing the vacancy value to 0) if he or she demands a value higher than $\overline{W}$. I let $\overline{W}(z, s^o) \equiv W^0$, meaning that when there is no outside offer, the executive's outside value is simply W^0. I call $\overline{W}(z, s)$ the *bidding frontier* to highlight that it is a function (frontier) in terms of z and s.

The sequential auction works as follows. When the executive from a firm of size s (hereafter firm s) meets a poaching firm of size $\tilde{s}$ (hereafter firm $\tilde{s}$), both firms enter a Bertrand competition won by the larger one. Since it is willing to extract a positive marginal profit out of every match, the best firm s can do is to provide a promised utility $\overline{W}(z', s)$. When $\tilde{s} > s$, the executive moves to a potentially better match with firm $\tilde{s}$, if the latter offers at least the $\overline{W}(z', s)$. Any less generous offer on the part of firm $\tilde{s}$ is successfully countered by firm s. If $\tilde{s}$ is smaller than s, then $\overline{W}(z', s) > \overline{W}(z', \tilde{s})$, in which case firm $\tilde{s}$ will never raise its offer up to this level. Rather, the executive will

[15]Promised utilities as co-states have been used extensively in models with incentive or participation constraints. Among others, Phelan and Townsend (1991) studied a model of risk-sharing with incentive constraints, Kocherlakota (1996) analyzed the risk-sharing model with two-sided limited commitment, Hopenhayn and Nicolini (1997) studied a model of unemployment insurance, and Alvarez and Jermann (2000) studied a decentralized version of the above risk-sharing model with debt constraints.

[16]What distinguishes this model from the original sequential auction framework is that here the wage is not flat. Firms compete on a stream of wages contingent on all possible future histories. At each period, the contract defines a wage on each state $(z, \tilde{s})$.

stay at his or her current firm, and be promoted to the continuation value $\overline{W}(z', \tilde{s})$ that makes him/her indifferent between staying and joining firm $\tilde{s}$.

The above argument defines outside values of the executive contingent on the state $(z', \tilde{s})$,

$$W(z', \tilde{s}) \geq \min\{\overline{W}(z', \tilde{s}), \overline{W}(z', s)\}.$$

This is the participation constraint of the executive in a contracting problem.

The contracting problem In designing the contract, the firm chooses a wage w and a set of promised values $W(z', \tilde{s})$ depending on the state $(z', \tilde{s})$. For ease of notation, I denote an effective discount factor, $\tilde{\beta} = \beta(1 - \eta)$, and write the mixture distribution of outside offers as follows:

$$\tilde{F}(s) = \mathbb{I}(s = s^o)(1 - \lambda_1) + \mathbb{I}(s \neq s^o)\lambda_1 F(s).$$

The expected profit of the firm can be expressed recursively as

$$\Pi(z, s, V) = \max_{w, W(z', \tilde{s})} \sum_{z' \in \mathbb{Z}} \left[y(s, z') - w + \tilde{\beta} \sum_{\tilde{s} \leq s} \Pi(z', s, W(z', \tilde{s}))\tilde{F}(\tilde{s}) \right] \Gamma(z'|z). \tag{BE-F}$$

subject to the promise-keeping constraint,

$$V = u(w) - c + \tilde{\beta} \sum_{z' \in \mathbb{Z}} \sum_{\tilde{s} \in \mathrm{S}} W(z', \tilde{s})\tilde{F}(\tilde{s})\Gamma(z'|z), \tag{PKC}$$

the incentive compatibility constraint,

$$\tilde{\beta} \sum_{z' \in \mathbb{Z}} \sum_{\tilde{s} \in \mathrm{S}} W(z', \tilde{s})\tilde{F}(\tilde{s})(1 - g(z, z'))\Gamma(z'|z) \geq c. \tag{IC}$$

and the participation constraints of the executive and the firm,

$$W(z', \tilde{s}) \geq \min\{\overline{W}(z', \tilde{s}), \overline{W}(z', s)\} \tag{PC-E}$$

$$W(z', \tilde{s}) \leq \overline{W}(z', s). \tag{PC-F}$$

The objective function (Bellman Equation of the Firm, BE-F) includes a flow profit of $y(s, z') - w$, taking into account that the match may separate either because the executive dies, which happens with probability η, or transits to another firm, which happens with probability $\sum_{\tilde{s} > s} \tilde{F}(s)$.

The promise-keeping constraint (PKC) makes sure that the choices of the firm honor the promise made in previous periods to deliver a value V to the executive, and the promised value V contains all the relevant information in the history. The right-hand side of the constraint is the lifetime utility of the executive given the choices made by the firm. (PKC) is also the Bellman equation of an executive with state (z, s, V).

The incentive compatibility constraint (IC) differentiates itself from the promise-keeping constraint by the term $1 - g(z'|z)$. It asserts that the continuation value of effort is higher than no effort. This creates incentives for the executive to pursue the shareholders' interests rather than his or her own.

Finally, the participation constraints are stated in (PC-E) and (PC-F). The firm commits to the relationship as long as the promised value is no more than $\overline{W}(z', s)$. The sequential auction pins down the outside value of the executive, which is the minimum of bidding frontier of the poaching firm, $\overline{W}(z', \tilde{s})$, and of the current firm, $\overline{W}(z', s)$.

4.4.3 Equilibrium definition

Before turning to the characterization of the optimal contract, I define the equilibrium. An equilibrium is an executive unemployment value W^0, a value function of employed executives W that satisfies (PKC), a profit function of firms Π and an optimal contract policy $\sigma = \{w, e, W(z', \tilde{s})\}$ for $z' \in \mathbb{Z}$ and $\tilde{s} \in \mathrm{S}$ that solves the contracting problem (BE-F) with associated constraints (PKC), (IC), (PC-E) and (PC-F), a stochastic process of executive productivity Γ that follows the optimal effort choice, and a distribution of executives across employment states evolving according to flow equations.

Proposition 4.1. *The equilibrium exists.*[17]

4.4.4 Contract characterization

In this section, I derive a characterization of the optimal contract. The characterization builds on and extends the dynamic limited commitment literature, pioneered by Thomas and Worrall (1988) and Kocherlakota (1996), the dynamic moral hazard literature, pioneered by Spear and Srivastava (1987), and related literature in labor search such as Lentz (2014).

Proposition 4.2. $\Pi(z, s, V)$ *is continuous differentiable, decreasing and concave in* V*, and increasing in* z *and* s*. An optimal contract evolves according to the following updating rule.*

[17] The proof of the existence of the equilibrium is an exercise applying Schauder's fixed point theorem, as shown by Menzio and Shi (2010).

Given the beginning-of-period state (z, s, V), the current period compensation is given by w^*,

$$\frac{\partial \Pi(z, s, V)}{\partial V} = -\frac{1}{u'(w^*)}, \tag{4.1}$$

and the continuation value $W^*(z', \tilde{s})$ follows

$$W^*(z', \tilde{s}) = \begin{cases} \overline{W}(z', s) & \text{if } \overline{W}(z', \tilde{s}) \geq \overline{W}(z', s) \text{ or } W(z') > \overline{W}(z', s) \\ \overline{W}(z', \tilde{s}) & \text{if } \overline{W}(z', s) > \overline{W}(z', \tilde{s}) > W^i(z') \\ W(z') & \text{if } \overline{W}(z', s) \geq W^i(z') \geq \overline{W}(z', \tilde{s}) \end{cases} \tag{4.2}$$

where $W(z')$ satisfies

$$\frac{\partial \Pi(z', s, W(z'))}{\partial W(z')} - \frac{\partial \Pi(z, s, V)}{\partial V} = -\mu(1 - g(z, z')). \tag{4.3}$$

Proof. The properties of $\Pi(z, s, V)$ follow immediately from the proof of proposition 4.1. To characterize the optimal contract, I assign Lagrangian multipliers λ to (PKC), μ to (IC), $\tilde{\beta}\mu_0(z', \tilde{s})$ to (PC-E) and $\tilde{\beta}\mu_1(z', \tilde{s})$ to (PC-F). The first order condition w.r.t w gives

$$u'(w) = \frac{1}{\lambda},$$

and the envelop theorem gives

$$-\frac{\partial \Pi(z, s, V)}{\partial V} = \lambda.$$

They together give (4.1). Participation constraints (PC-E) and (PC-F) can be simplified. If $\overline{W}(z', \tilde{s}) \geq \overline{W}(z', s)$, we have $W(z;, \tilde{s}) = \overline{W}(z', s)$. This is the first case in line 1 of (4.2). If $\overline{W}(z', \tilde{s}) \geq \overline{W}(z', s)$, the participation constraints become $\overline{W}(z', \tilde{s}) \leq W(z;, \tilde{s}) \leq \overline{W}(z', s)$. Use this to derive the first order condition w.r.t $W(z', \tilde{s})$:

$$-\frac{\partial \Pi(z', s, W(z', s))}{\partial W(z', s)} = \lambda + \mu(1 - g(z, z')) + \mu_0(z', \tilde{s}) - \mu_1(z', \tilde{s}).$$

If $\mu_0(z', \tilde{s}) = \mu_1(z', \tilde{s}) = 0$, $W(z', \tilde{s}) = W(z')$ defined by (4.3). This is the case in line 3 of (4.2). If $\mu_0(z', \tilde{s}) > \mu_1(z', \tilde{s}) = 0$, $W(z', \tilde{s}) = \overline{W}(z', \tilde{s})$. This is the case in line 2 of (4.2). Finally, if $\mu_1(z', \tilde{s}) > \mu_0(z', \tilde{s}) = 0$, $W(z', \tilde{s}) = \overline{W}(z', s)$. This is the second condition in line 1 of (4.2). □

Proposition 4.2 states that, if one abstracts from the participation constraints, an optimal contract inherits the essential properties of the classical infinite repeated moral hazard model (Spear and Srivastava 1987). Equation (4.1) states that the current period compensation w^* is directly linked to the promised continuation utility V, by equating the principal's and agent's marginal rates of substitution between the present and future compensation. Equation (4.3) says, abstract from participation constraints, the continuation utility $W(z')$ only changes to induce the executive effort. In the extreme case that the IC constraint is not binding ($\mu = 0$, μ is the multiplier of the IC constraint), $W(z') = V$ remains constant. Thus, the pay is also constant over time. Generally speaking, a higher V induces a higher $W(z')$. That is, an optimal dynamic contract has some memory.

When outside offers are realized such that the participation constraint is binding, the contract is no longer dependent on history, and the continuation value depends only on the current state. This is what Kocherlakota (1996) calls *amnesia*. More precisely, when the outside firm is larger $\tilde{s} \geq s$, the continuation value is equal to the bidding frontier of the current firm $W(z', \tilde{s}) = \overline{W}(z', s)$; when the outside firm is smaller, $\tilde{s} < s$, the continuation value depends on whether the bidding frontier of the outside firm $\overline{W}(z', \tilde{s})$ can improve upon $W(z')$.

Even when the participation constraint is binding, amnesia of the optimal contract is not "complete" — although $\overline{W}$ does not depend on the previously promised utility V, it does depend on the executive's productivity z', which is stochastically determined by past effort. Therefore, the boundaries of participation constraints carry the memory of the prior effort choice. This is where the incentives from the labor market come into effect.

4.4.5 Explaining the size pay-growth premium

With the characterization of the optimal contract, we are ready to explain the size premium in pay-growth and incentives. I start by defining two sets of poaching firms $\tilde{s}$: larger or smaller than the current firm.

$$\mathcal{M}_1(s) \equiv \{\tilde{s} \in \mathbb{S} | \tilde{s} > s\},$$
$$\mathcal{M}_2(z, s, W) \equiv \{\tilde{s} \in \mathbb{S} | \overline{W}(z, s) > \overline{W}(z, \tilde{s}), W < \overline{W}(z, \tilde{s})\}.$$

Given a poaching firm that belongs to the set $\mathcal{M}_1$, the executive will transit to such a firm and receive the full surplus of his or her previous job $\overline{W}(z, s)$. Given a poaching firm in $\mathcal{M}_2$, the executive will stay in the current firm but use the outside offer to renegotiate up to $\overline{W}(z, \tilde{s})$. Any poaching firm that is not in $\mathcal{M}_1$ or $\mathcal{M}_2$ is not competitive in the sense that it cannot be used to negotiate compensation with the incumbent firm.

Accordingly, the Bellman equation of executives can be written as:[18]

$$V = u(w) - c + \tilde{\beta} \sum_{z'} \Bigg[\lambda_1 \sum_{s' \in \mathcal{M}_1} F(s') \overline{W}(z', s) + \lambda_1 \sum_{s' \in \mathcal{M}_2} F(s') \overline{W}(z', s')$$
$$+ \Big(1 - \lambda_1 \sum_{s' \in \mathcal{M}_1 \cup \mathcal{M}_2} F(s') \Big) W(z') \Bigg] \Gamma(z'|z), \tag{PKC'}$$

(PKC′) shows that compensation grows mainly in two cases: i) There is a poaching firm from set $\mathcal{M}_2$ and total compensation increases without a job turnover; ii) There is a poaching firm from set $\mathcal{M}_1$ and total compensation grows upon a job-to-job transition. The firm-size pay-growth premium observed in the data refers to the growth in the former case. In the latter case, compensation may also decrease if an executive is willing to make sacrifices on his or her current pay for the sake of higher pay in the future.[19]

To understand the firm-size pay-growth premium, consider two executives from a small firm s_1 and a large firm s_2, $s_2 > s_1$. For simplicity, suppose they have the same continuation value $W(z')$. Since the firm s_2 has a higher output and is more capable of overbidding outside offers, the corresponding set $\mathcal{M}_2$ is larger. That is, there exist poaching firms with a size between s_1 and s_2 such that the firm s_2 can overbid and retain the executive with compensation growth while the firm s_1 cannot overbid and consequently lose the executive. Therefore, the total pay increases faster in the larger firm s_2.

4.4.6 Explaining the size incentive premium

To explain the firm-size incentive premium, I define "performance-based incentives" and "labor market incentives" in the model. Using these definitions to rewrite the IC constraint, I then show that the two sources of incentives substitute for each other given a constant effort cost. Finally, I explain that labor market incentives de-

[18]We can similarly rewrite the Bellman equations of firms using the optimal continuation value, and this equation is consistent with Postel-Vinay and Robin (2002):

$$\Pi(z, s, V) = \max_{w, W(z')} \sum_{z'} \Bigg[y(s) z' - w + \tilde{\beta} \Bigg(\lambda_1 \sum_{s' \in \mathcal{M}_2} F(s') \Pi_1(z', s, \overline{W}(z', s'))$$
$$+ \Big(1 - \lambda_1 \sum_{s' \in \mathcal{M}_1 \cup \mathcal{M}_2} F(s') \Big) \Pi_1(z', s, W(z')) \Bigg) \Bigg] \Gamma(z'|z). \tag{BE-F'}$$

[19]Where compensation information is available in both the original and target firms, it would be interesting to examine whether there is also a firm-size compensation-growth premium in job-to-job transitions. This is, however, not possible with the current dataset.

crease with firm size. Thus, performance-based incentives increase with firm size.

I first define an "incentive operator", $\mathcal{I}(\cdot)$, which calculates the incentives an executive receives from a continuation utility scheme:

$$\mathcal{I}\Big(W(z')\Big) \equiv \int_{z'} W(z')(1-g(z,z'))\Gamma(z'|z).$$

I then rewrite the IC constraint using the incentive operator:

$$\begin{aligned}
&\lambda_1 \int_{\tilde{s}\in\mathcal{M}_1} dF(\tilde{s})\mathcal{I}\Big(\overline{W}(z',s)\Big) + \lambda_1 \int_{\tilde{s}\in\mathcal{M}_2} \mathcal{I}\Big(\overline{W}(z',\tilde{s})\Big)F(\tilde{s}) \\
&+ \Big(1-\lambda_1 \sum_{\tilde{s}\in\mathcal{M}_1\cup\mathcal{M}_2} F(\tilde{s})\Big)\mathcal{I}\Big(W(z')\Big) \geq c/\tilde{\beta}, \qquad \text{(IC}'\text{)}
\end{aligned}$$

The incentives comprise: i) incentives brought by larger firms in $\mathcal{M}_1$; ii) incentives brought by smaller firms in $\mathcal{M}_2$; iii) incentives in performance-related pay when there are no poaching firms from $\mathcal{M}_1$ or $\mathcal{M}_2$.

The incentives when there are no competitive poaching offers are the *performance-based incentives*, denoted by Ξ_p:

$$\Xi_p\Big(W(z')\Big) \equiv \Big(1-\lambda_1 \sum_{\mathcal{M}_1\cup\mathcal{M}_2} F(s')\Big)\mathcal{I}\Big(W(z')\Big), \qquad (4.4)$$

and the incentives from poaching offers are the *labor market incentives*, denoted by Ξ_m:

$$\Xi_m\Big(s, W(z')\Big) \equiv \lambda_1 \int_{\tilde{s}\in\mathcal{M}_1} dF(\tilde{s})\mathcal{I}\Big(\overline{W}(z',s)\Big) + \lambda_1 \int_{\tilde{s}\in\mathcal{M}_2} \mathcal{I}\Big(\overline{W}(z',\tilde{s})\Big)F(\tilde{s}). \qquad (4.5)$$

Ξ_m would be zero if there were no poaching offers.

Because of labor market incentives, the need for performance-based incentives is less. Intuitively, firms appreciate higher productivities and are willing to bid higher for a more productive executive. The sequential auction in the model therefore begets labor market incentives for executive effort: If working hard today is not only an input into current production but also an investment in the (inalienable and transferable) human capital, then it is intuitive that the objectives of the firm and of the executive become better aligned and the need for short-term compensation incentives decreases.

Mathematically, Ξ_m is an expectation of incentives from all possible poaching offers. When the poaching firm is larger than the current firm, the incentives are from the bidding frontier of the current firm. When the poaching firm is smaller than the current firm, the incentives are from the bidding frontier of the poaching firm.

The magnitude of Ξ_m is determined by current firm size s and the promised continuation value $W(z')$. In particular, firm size s enters Ξ_m via bidding frontiers. That is, Ξ_m depends on s even though the moral hazard problem fundamentally does not. On the other hand, $W(z')$ determines the lower bound of set $\mathcal{M}_2$. The larger the promised continuation value $W(z')$, the less likely a poaching firm can be used to renegotiate with the current firm, and the lower the labor market incentives.

Based on this, there is a simple "job ladder" explanation for the size premium when comparing executives of different pay levels. Such an incentive premium is reported in column (1) in table 4.2. Since executives of larger firms tend to have higher total compensation, the corresponding continuation values are larger; they are thus higher on the job ladder. Accordingly, the chance of encountering a competitive poaching offer that beats the current value is smaller. Hence, labor market incentives are lower. As a result, executives in larger firms require more incentives in performance-related pay.[20]As we will see in the following section, this "job ladder" argument also applies to explaining the size incentive premium among executives with the same total compensation.

Labor market incentives decrease with firm size Now I compare labor market incentives for executives who have the same total compensation but come from firms of different size. I show that when there is enough concavity in the utility function, labor market incentives decrease with firm size. Therefore, larger firms need to provide more performance-based incentives. This explains the firm-size incentive premium.

Consider two executives from firms s_1 and s_2, $s_1 < s_2$. The executives have the same total compensation. Figure 4.3 illustrates the possible poaching firm sizes for the two executives and the associated incentives. The poaching firm size ranges from $\underline{s}$ to $\bar{s}$. I denote the lower bound of $\mathcal{M}_2$ for firms s_1 and s_2 by s_1^{lb} and s_2^{lb}, respectively. Notice that $s_2^{lb} > s_1^{lb}$ because they are determined by lifetime utilities rather than current period compensation. Although the two executives have the same total compensation, the one in s_2 has higher lifetime utility. The left side of the axis depicts sets $\mathcal{M}_1$, $\mathcal{M}_2$ and corresponding labor market incentives in the two sets for the executive in s_1. The right side of the axis depicts the counterparts for s_2. Taking the difference between

[20]This is an alternative explanation in addition to the current explanations based on moral hazard (Gayle and Miller, 2009) and on multiplicative utility (Edmans et al., 2009).

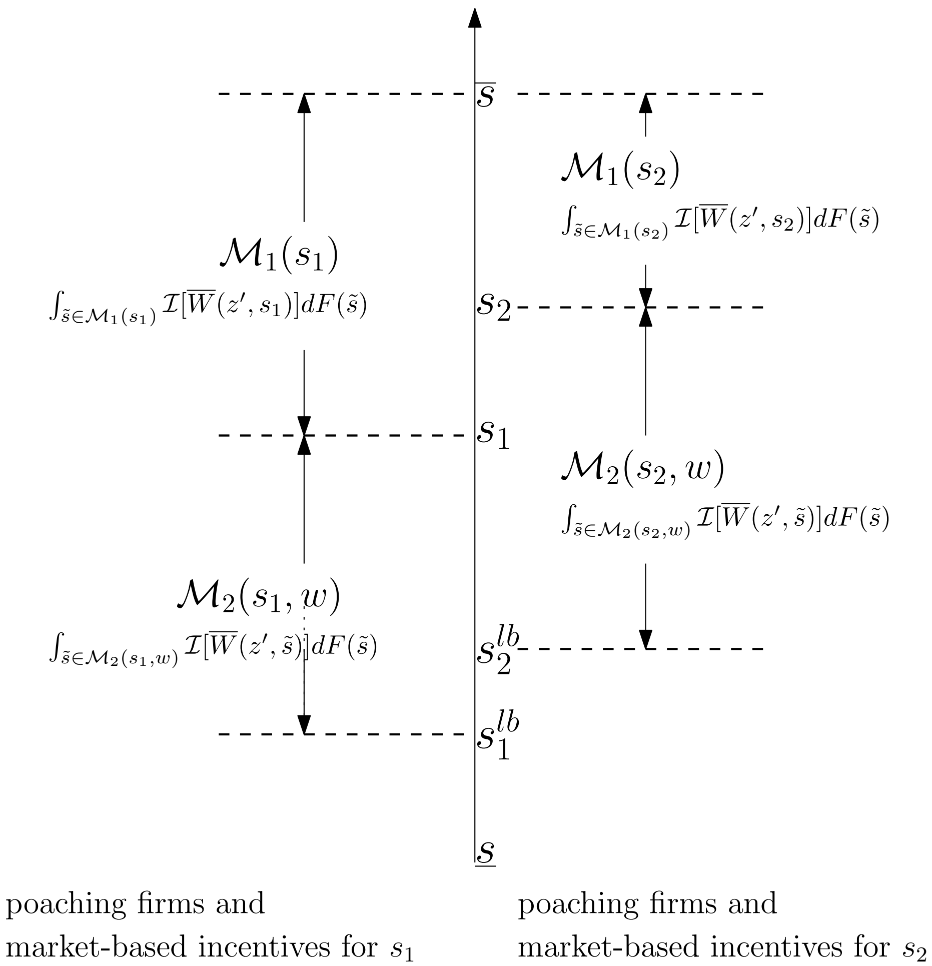

Figure 4.3: Compare labor market incentives

Note: The figure illustrates labor market incentives for executives with the same compensation w from firms of size s_1 and s_2. The vertical axis labels the size of poaching firms $[\underline{s}, \overline{s}]$. s_1^{lb} is the lower bound of set $\mathcal{M}_2(s_1, w)$ and s_2^{lb} is the lower bound of set $\mathcal{M}_2(s_2, w)$. The labor market incentives of s_1 and s_2 are on the left and right of the vertical axis, respectively. The notation for each interval is followed by the value of incentives from poaching firms of that interval.

$\Xi_m(s_2)$ and $\Xi_m(s_1)$, we have

$$\begin{aligned}\Xi_m(s_2) - \Xi_m(s_1) = &- \int_{s_1^{lb}}^{s_2^{lb}} d\tilde{F}(\tilde{s})\mathcal{I}(\overline{W}(z', \tilde{s})) \\ &+ \int_{s_2}^{\overline{s}} d\tilde{F}(\tilde{s})\Big(\mathcal{I}(\overline{W}(z', s_2) - \mathcal{I}(\overline{W}(z', s_1)\Big) \\ &+ \int_{s_1}^{s_2} \Big(\mathcal{I}(\overline{W}(z', \tilde{s}) - \mathcal{I}(\overline{W}(z', s_1)\Big)d\tilde{F}(\tilde{s}). \qquad (4.6)\end{aligned}$$

Their labor market incentives are different in two respects. First, with poaching firms in $[s_1^{lb}, s_2^{lb}]$, the executive in s_1 receives an incentive of $\int_{s_1^{lb}}^{s_2^{lb}} d\tilde{F}(\tilde{s})\mathcal{I}(\overline{W}(z', \tilde{s}))$, while the executive in s_2 has no incentive from the labor market. This is the first item in (4.6), and it corresponds to the job ladder argument previously mentioned — since $s_2^{lb} > s_1^{lb}$, the executive in s_2 is less likely to receive a competitive outside offer, and labor market incentives are lower.

Second, for poaching firms in the range of $[s_1, \bar{s}]$, labor market incentives for firms s_1 and s_2 are drawn on different bidding frontiers, which correspond to the second and third items in (4.6). With poaching firms in this range, the bidding frontier for the executive of firm s_1 is always $\overline{W}(z', s_1)$, since any poaching firm larger than s_1 can bid just $\overline{W}(z', s_1)$ to attract the executive. In contrast, the bidding frontiers for the executive in firm s_2 are either $\overline{W}(z', s_2)$ or $\overline{W}(z', \tilde{s})$ with $\tilde{s} > s_1$, both of which are larger than $\overline{W}(z', s_1)$.[21] Consequently, the certainty equivalent of the executive in s_2 is higher. By diminishing marginal utility, the incentives from these higher bidding frontiers are lower:

$$\mathcal{I}\Big(\overline{W}(z', s_1)\Big) > \mathcal{I}\Big(\overline{W}(z', \tilde{s})\Big) \text{ for } \tilde{s} > s_1.$$

This is a wealth effect of poaching offers — a wealthier executive is harder to incentivize. This wealth effect holds as long as the utility function is sufficiently concave. In the following, I give a sufficient condition under the restriction that the utility function is of the CRRA form and effort cost c is equal to a particular value.

Proposition 4.3 (Labor market incentives and firm size)**.** *Suppose the executives' utility is of the CRRA form, and the cost of effort $c = \bar{c}(s)$, then $\mathcal{I}\Big(\overline{W}(z', s)\Big)$ decreases in s if*

$$\sigma > 1 + \frac{s^{1-\alpha_1}}{\alpha_1}\psi'(s), \tag{4.7}$$

where $\psi(s)$ is a function of s that is positive and increasing in s and

$$\bar{c}(s) \equiv \tilde{\beta} \sum_{z' \in \mathbb{Z}} \overline{W}(z', s)(1 - g(z'|z))\Gamma(z'|z).$$

Proof. See the Appendix. □

To understand the proposition, first notice that $\mathcal{I}\Big(\overline{W}(z', s)\Big)$ is simply a weight sum of $\frac{\Delta \overline{W}(z', s)}{\Delta z'}$ over the domain of z' — the steeper $\overline{W}(z', s)$ with respect to z', the higher the incentives to induce effort. So it would be sufficient to show that $\frac{\Delta \overline{W}(z', s)}{\Delta z'}$ decreases in s

[21] $\overline{W}(z', s)$ is strictly increasing in s.

under the stated condition. To proceed, it follows that

$$\frac{\Delta\overline{W}(z,s)}{\Delta z} = -\frac{\Delta\Pi(z,s,\overline{W})/\Delta z}{\Delta\Pi(z,s,\overline{W})/\overline{W}} = \frac{\tilde{\alpha}\times s}{1/u'(\overline{w})}, \tag{4.8}$$

where $\overline{w}$ is the optimal compensation for the current period, corresponding to a promising continuation value $\overline{W}$. The first equality follows from an implicit differentiation. In the second equality,

$$\Delta\Pi(z,s,\overline{W})/\Delta z = \tilde{\alpha}\times s$$

because keeping the promised value, all increasing output is accrued to the company. In particular, $\tilde{\alpha}$ is α multiplied by a factor that adjusts for the possibility that the executive will leave the firm and the job is destructed. On the denominator,

$$\Delta\Pi(z,s,\overline{W})/\overline{W} = -1/u'(\overline{w})$$

follows directly from condition (4.1) in Proposition 4.2.

There are two opposing effects of s involved in (4.8). On the one hand, the maximum value that larger firms are able to bid changes more with respect to z due to the multiplicative production function. This will generate more labor market incentives, and it is reflected in the numerator of (4.8). On the other hand, the incentives in terms of utilities can actually be lower because the marginal utility for extra returns from the executive labor market is lower now ($\overline{w}$ increases in s making $u'(\overline{w})$ lower). This is reflected in the denominator of (4.8). The second force dominates when the utility function has enough concavity, as stated in the proposition.

The requirement stated in (4.7) is consistent with the literature in this context. The existing studies usually estimate or calibrate a higher σ value. For example, a careful calibration study on CEO incentive pay by Hall and Murphy (2000) uses σ between 2 and 3. Calibration exercises on CEO incentive compensation convexity starting from Dittmann and Maug (2007) are based on $\sigma > 1$. Using an employer-employee matched data from Sweden for the general labor market, Lamadon (2016) estimates that $\sigma = 1.68$. Numerically, I find the right-hand side of (4.7) is approximately equal to one in the parameter space explored in my estimation.

Back to the firm-size incentive premium. If (4.7) is satisfied, and given (4.6), the labor market incentives Ξ_m are lower for the executive in firm s_2. Since the effort cost is the same for both executives, the executive in the larger firm s_2 demands more incentives from the performance-related pay. This explain the firm-size incentive premium.

4.5 Empirical evidence

To quantitatively evaluate the model, I use data on executives employed in U.S. publicly listed firms. Close scrutiny of the managerial labor market allows me to put together a rich array of data from various sources. Specifically, I assemble a new dataset on job turnovers from BoardEX, and merge the job turnover data with two sets of standard data, the executive compensation from ExecuComp, and firm-level information from CompuStat. In the following, I provide a brief description of the relevant data features. In particular, I examine executives' job-to-job transitions, and whether they climb the job ladder towards larger firms. These are the key features of the managerial labor market in the model. Additionally, I examine whether the job-to-job transition rate decreases with firm size as predicted by the model.

4.5.1 Data

The empirical analysis and estimation mainly rely on the ExecuComp database, which provides rich information on executive compensation of the top five to eight executives in companies included in the $S\&P$ 500, MidCap and SmallCap indices for the period of 1992 to 2016. The accounting information from CompuStat and stock returns from CRSP are merged with ExecuComp. The dataset provided by Coles et al. (2006) and Coles et al. (2013) contains performance-based incentives *delta* calculated based on ExecuComp. To collect job turnover information, I extract the full employment histories of executives from the BoardEX database, and supplement them with the information from executives' LinkedIn and Bloomberg pages.

My final sample comprises $35,088$ executives with age between 30 and 65.[22] Of these, $26,972$ episodes cover the full tenure of the executive from beginning to end. The total number of executive-fiscal year observations in my sample is $218,168$. The minimum number of firms covered in a given year is $1,556$ in 1992, and the maximum is $2,235$ in 2007. All nominal quantities are converted into constant 2016 dollars using a GDP deflater from the Bureau of Economic Analysis.

Here I describe the variables that are used in my analysis. Using information from ExecuComp, I identify the *gender* and *age* for each executive, the *tenure* in the current

[22] I select this age range because the managerial labor market is more relevant than for those passing the retirement age.

Table 4.3: Summary statistics

Variable	N	mean	sd	p25	p50	p75
age	218,168	51.04	6.96	46	51	56
male	218,168	0.936	0.244	1	1	1
CEO	218,168	0.184	0.387	0	0	0
CFO	218,168	0.096	0.295	0	0	0
director	218,168	0.339	0.473	0	0	1
interlock	218,168	0.013	0.112	0	0	0
tenure	218,168	4.71	3.793	2	4	6
tdc1	198,673	2,555.527	5,454.153	632.164	1,270.806	2,690.385
delta	146,790	322.518	4,736.982	16.966	50.634	154.411
mkcap	212,271	7,997.377	25,810.758	598.919	1,622.236	5,169.379
at	216,384	15,594.888	98,653.077	542.863	1,796.467	6,570.342
sales	216,276	5,472.709	17,387.175	428.2	1,217.738	3,917.269
profit	209,639	0.119	0.359	0.069	0.121	0.176
annual return	211,067	0.181	0.802	-0.127	0.106	0.356
mbr	183,565	1.669	2.21	0.811	1.198	1.913

Note: The table reports summary sample statistics for my dataset, which covers named executive officers reported in ExecuComp over the period of 1992 to 2016. *age* is the executive's age by the end of the fiscal year. Sample episodes with age lower than 35 or higher than 70 are dropped. Dummy variables *CEO*, *CFO*, *director* and *interlock* indicate whether the executive served as a CEO, or a CFO, or a director, or is involved in the interlock relationship during the fiscal year, respectively. An interlock relationship is described in the note of table 4.1. *tenure* (in years) counts the number of fiscal years that the executive works as a named officer. *tdc1* is the total compensation, composed of the following: Salary, Bonus, Other Annual, Total Value of Restricted Stock Granted, Total Value of Stock Options Granted (using BlackScholes), Long-Term Incentive Payouts, and All Other Total. *delta* is the dollar change in wealth associated with a 1% change in the firms stock price (in $000*s*). *mkcap* (in millions) is the market capitalization of the company, calculated by *csho* (Common Shares Outstanding, in millions of shares) multiplied by *prcc_f* (fiscal year end price). *prcc_f* and *csho* are reported in CompuStat Fundamentals Annual file. *at* (in millions) is the Total Book Assets as reported by the company. *sales* (in millions) is the Net Annual Sales as reported by the company. *profit* is the profitability, calculated by EBITDA/Assets. *annual return* is the annualized stock return which is compounded based on CRSP MSF (Monthly) returns. MSF returns have been adjusted for splits, etc. *mbr* is the Market-to-Book Ratio, calculated by Market Value of Assets divided by Total Book Assets. Market Value of Assets is calculated according to $\textit{Market Value of Assets (MVA)} = prcc\_f * cshpri + dlc + dltt + pstkl - txditc$. Variable definitions are provided in the main text.

executive episode, whether he or she is a *CEO*, or *CFO*, or *director* of the board or is involved in an *interlock* relationship during the fiscal year. Table 4.3 reports summary statistics for my sample. Ninety-three percent of the executives are males and the average age is 51. The average length of an episode is 6.21 years. Among all executive-year observations, 18.4% are CEO spells and 9.6% are CFO spells.

In terms of the compensation information, *tdc1* is the total compensation including salary, bonus, values of stock and options granted, etc. The total compensation has an average of $2,555,000$ dollars, with a 25th percentile of $632,000$ dollars and a 75th percentile of $2,690,000$ dollars. In terms of means, only 16.5% of the total compensation is fixed base salary; the rest is all incentive-related. Performance-based incentives come not only from the total compensation each year, but also from the stocks and options that are granted in previous years. The variable *delta* measures how strong performance-based incentives are in firm-related wealth. It is defined by the dollar change in wealth (in \$000*s* in table 4.3) associated with a 1% change in the firms stock price. The distribution of $delta$ is right-skewed, with a mean of $323,000$ dollars, even larger than its 75th percentile of $154,000$ dollars.

For the firm-side information, I use market capitalization *mkcap*, the market value of a company's outstanding shares, to measure the firm size. In some robustness checks (not shown in the main text), I also use book value of assets *at* and *sales* to measure firm size. They are in millions of dollars. I use operating profitability, denoted by *profit*, to measure firm performance. Two alternative measures for firm performance are stock market annualized return, denoted by *annual return*, and market-to-book ratio, denoted by *mbr*.

The job turnover information comes from the BoardEX database.[23]BoardEX contains details of each executive's employment history, including start and end dates, firm names and positions. It also has extra information on educational background, social networks, etc. I merge the two databases using three sources of information: the executive's first, middle and last names, date of birth, and working experiences, i.e. in which years the executive worked in which firms. If all three aspects are consistent, the executive is identified. For executives that cannot be identified in BoardEX, I search for the respective LinkedIn and Bloomberg pages and manually collect the available employment information. In this way, I am able to identify more than 93% of executives in ExecuComp, 32, 864 in total.

4.5.2 Job-to-job transitions

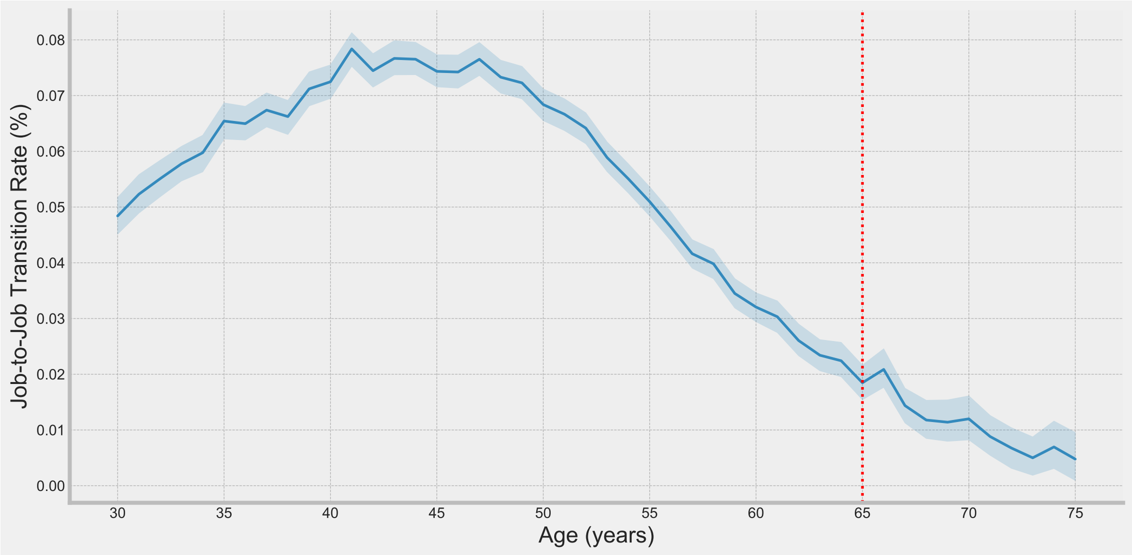

Figure 4.4: Job-to-job transition rate over age

Note: The figure depicts estimated job-to-job transition rates over age with the 95% confidence interval. A job-to-job transition is defined as an executive leaving the current firm and starting to work in another firm within 190 days.

I define a job-to-job transition as the executive leaving their current firm and starting to work in another within 190 days. Otherwise, the event is defined as an exit from the managerial labor market. In the data, the job-to-job transition rate is 4.98% each year over the period of 1992 to 2015, while the job exit rate is slightly higher, at 6.91%. Figure 4.4 illustrates how job-to-job transition changes with age, and figure 4.5 shows how job exit changes with age. To illustrate the trend, the figures also include those who did not retire after age 65. As shown in the figure, the job-to-job transition rate increases gradually before 40 and peaks at around age 45 before decreasing after 50. In contrast, the job exit rate is lower before 55 and peaks sharply at age 65 as expected.

Most job-to-job transitions are within the industry. Among transitions for which industry information is available, 1,717 out of 2,567 transitions are within the industry as defined by the Fama-French 12 industry classification, and 1,407 out of the 2,567 cases as defined by the Fama-French 48 industry classification.

[23]What is missing in the ExecuComp database is the information on executives' employment history. For example, there is no information to identify whether the executive transits to another firm after the current position in an *S&P* firm or whether they simply retire. Moreover, the start and end dates of current employment are not known.

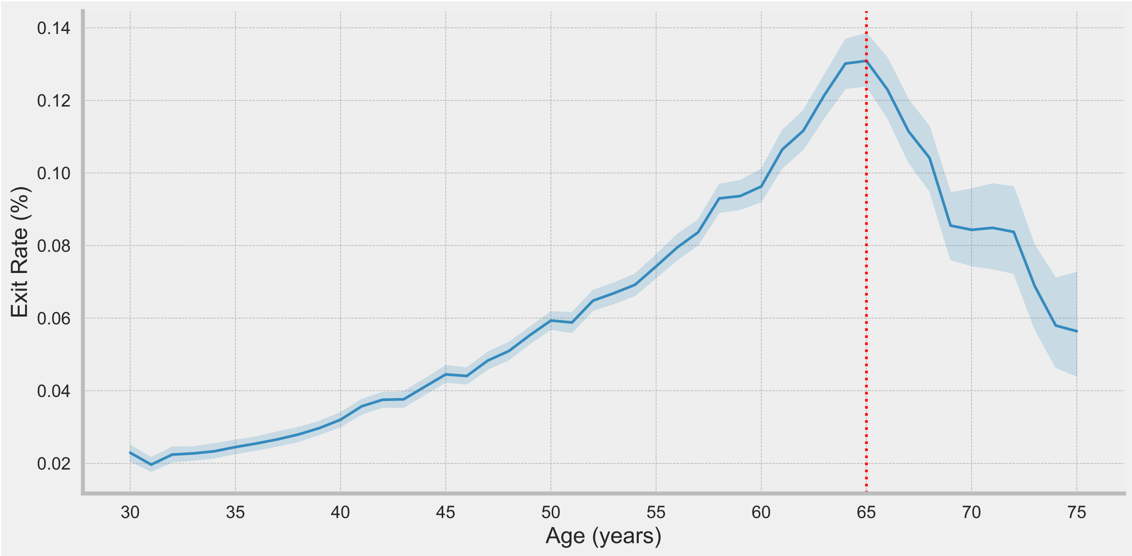

Figure 4.5: Exit rate over age

Note: The figure depicts the estimated exit rates over age with the 95% confidence interval. A job exit is defined as an executive leaving the current firm and not working in another firm within 190 days.

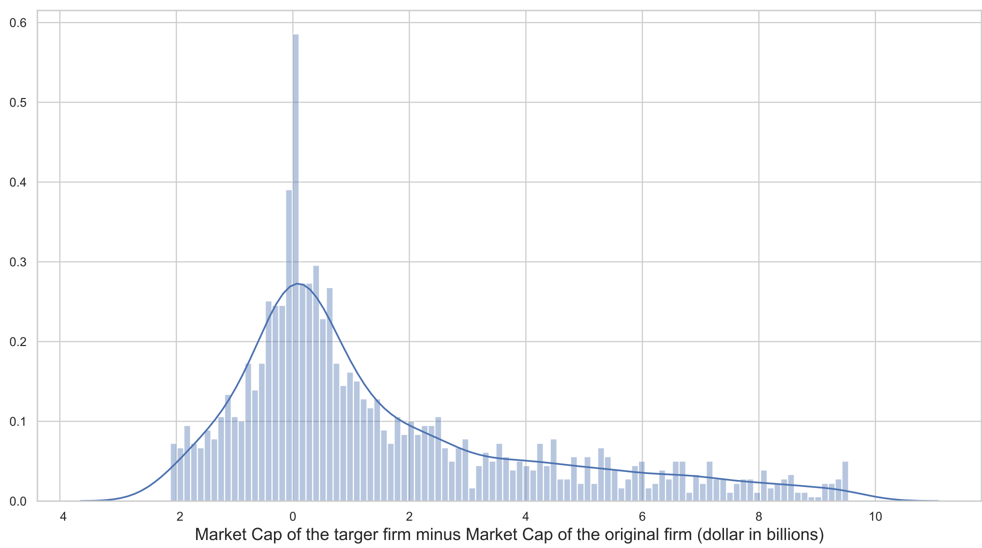

Figure 4.6: Distribution of the change of firm size upon job-to-job transitions

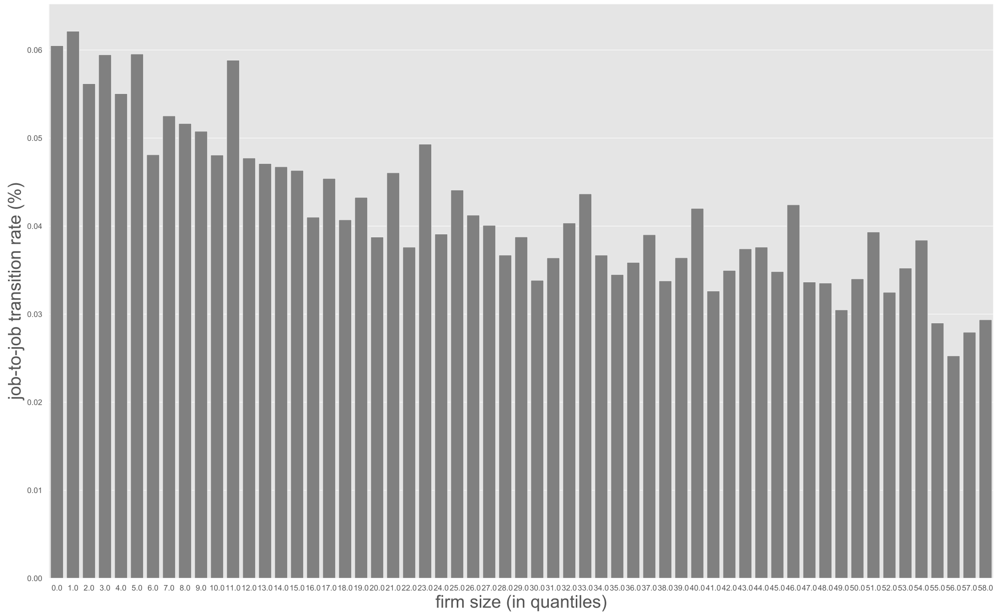

Figure 4.7: Job-to-job transition rates across firm size

Executives transit to larger firms

In my sample, there are $9,138$ job-to-job transitions from a CompuStat firm; only $2,567$ have firm size information on both the original and target firms. The rest are private firms whose size information is not disclosed. Based on the selected sample where size information is available, I find that approximately 60% of job-to-job transitions are associated with a firm size increase. The pattern is stable across age-groups and industries, as shown in table 4.4. I further check the transitions towards smaller firms. It turns out that 20% of these cases are due to a title change from a non-CEO title to a CEO title, while this fraction is only 3.3% in transitions towards larger firms.

Figure 4.6 portrays the distribution of the change of firm size upon a transition. While many transitions are between firms with similar size, there are a lot of "leap" transitions where the target firm is much larger. This lends support to my modeling of the managerial labor market, where executives engage in *random* on-the-job search.

Job-to-job transitions decrease with firm size

Next, I check whether executives in larger firms have fewer transitions, which is predicted by the model. As a first pass, figure 4.7 depicts the transition rates across firm size quantiles. The transition rate decreases from more than 6% at the 5th percentile

Table 4.4: Change of firm size upon job-to-job transitions

Panel A: All executives

Firm size proxy	Total obs.	Firm size decrease obs. (%)	Firm size increase obs. (%)
Market Cap	2,567	985 (39%)	1,582 (61%)
Sales	2,617	1,051 (40%)	1,566 (60%)
Book Assets	2,616	1,038 (40%)	1,578 (60%)

Panel B: Across age groups

Age groups	Total obs.	Firm size decrease obs. (%)	Firm size increase obs. (%)
$\leq$ 40	100	34 (34%)	66 (66%)
[40, 45)	381	135 (35%)	246 (65%)
[45, 50)	701	262 (37%)	439 (63%)
[50, 55)	766	304 (40%)	462 (60%)
[55, 60)	261	179 (43%)	82 (67%)
[60, 65)	73	52 (39%)	21 (61%)
[65, 70)	30	7 (25%)	23 (75%)
$\geq$ 70	6	1 (16%)	5 (84%)

Panel C: Across industries

Fama-French industries (12)	Total obs.	Firm size decrease obs. (%)	Firm size increase obs. (%)
1	119	39 (33%)	80 (67%)
2	88	33 (38%)	55 (61%)
3	281	98 (35%)	183 (65%)
4	120	58 (48%)	62 (52%)
5	71	30 (42%)	41 (58%)
6	609	229 (38%)	380 (62%)
7	60	20 (33%)	40 (67%)
8	96	48 (50%)	48 (50%)
9	381	142 (37%)	239 (63%)
10	197	89 (45%)	108 (65%)
11	314	115 (37%)	199 (63%)
12	231	84 (36%)	147 (64%)

of firm size to around 3% at the 95th percentile of firm size. To further investigate how job-to-job transitions vary with firm size, I estimate a Cox model on how firm size affects the time to job-to-job transitions, controlling for executive age, firm performance indicators, year and industry dummies. For a 1% increase in the firm scale, the hazard rate decreases by 8.3% without controlling for total compensation, and by 2.8% after controlling for total compensation. That is, larger firms have significantly lower job-to-job transition rate.[24]

Table 4.5: Job-to-job transitions and firm size

	Job-to-Job transition	
	(1)	(2)
log(firm size)	0.917****	0.972*
	(0.0109)	(0.0139)
age	0.985****	0.967***
	(0.00273)	(0.0112)
log(tdc1)		0.830****
		(0.0150)
other controls	X	X
year x *industry*	X	X
N	154635	118119
chi2	496.1	491.4

Note: I estimate a Cox proportional hazards model with the event of a job-to-job transition. A job-to-job transition is defined as the executive leaving the current firm (and not returning to the current firm within one year), and starting to work in another firm within 190 days. All variables have the same definition as in table 4.1. All dollar-related variables are adjusted by a GDP deflater. The standard errors are shown in parentheses, and I denote symbols of significance by * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

4.6 Estimation

I estimate the model parameters using Simulated Methods of Moments. That is, I use a set of moments that are informative for the parameters and minimize the distance between data moments and model-generated moments. My moments are partly coefficients from auxiliary regressions, so the approach could alternatively be presented as Indirect Inference. I first introduce the numerical method that I employ to solve the dynamic contracting problem. Then I describe the model specifications and moments

[24]Contradicting the model's prediction that job-to-job transition is not related to compensation level, in the data, when the total compensation rises by 1%, the hazard rate drops by 27% One possible explanation is that the compensation level contains information on ranks which are related to the production function parameters α_0 and α_1. Perhaps a more accurate way to measure production is by "effective firm size", which combines both firm asset scales and executive rank information.

used for identification. Specifically, I do not explicitly target the firm-size pay-growth and incentive premiums. After reporting the parameter estimates, I compare the estimates of the premiums in the data and in the model simulated data. I show that the model quantitatively captures both premiums.

4.6.1 Numerical method

To solve the contracting problem, one needs to find the optimal promised values in each state of the world for the next period. This becomes infeasible as soon as reasonable supports are considered for $\mathbb{Z}$ and $\mathbb{S}$. Instead of solving for promised values directly, I use the recursive Lagrangian techniques developed in Marcet and Marimon (2017) and extended by Mele (2014). Under this framework, the optimal contract can be characterized by maximizing a weighted sum of the lifetime utilities of the firm and the executive, where in each period the social planner optimally updates the Pareto weight of the executive to enforce an incentive compatible allocation. This Pareto weight becomes a new state variable that recursifies the dynamic agency problem. In particular, this endogenously evolving weight summarizes the contract's promises according to which the executive is rewarded or punished based on the performance and outside offers. Ultimately, solving an optimal contract is to find the sequence of Pareto weights that implements an incentive-compatible allocation. Once these weights are solved, the corresponding utilities can be recovered. This technique improves the speed of computation and makes the estimation feasible. I leave more details to Appendix C.

4.6.2 Model specification and parameters

I estimate the model fully parametrically and make several parametric assumptions. Being consistent with the analysis above, I use the constant relative risk aversion utility function:

$$u(w) = \frac{w^{1-\sigma}}{1-\sigma},$$

and a production function:

$$y(z,s) = e^{\alpha_0} s^{\alpha_1} z.$$

I model the process of productivity by an $AR(1)$ process:

$$z_t = \rho_0(e) + \rho_z z_{t-1} + \epsilon_t,$$

where ϵ follows a normal distribution $N(0, \sigma_\epsilon)$, and the mean for no effort, $\rho_0(0)$, is normalized to zero. The process is transformed into a discrete Markov Chain using Tauchen (1986) on a grid of 6 points.[25] Furthermore, I set the sampling distribution of firm size $F(s)$ as a truncated log-normal distribution with expectation of μ_s and standard deviation of σ_s.[26] Finally, the discount rate β is set to be 0.9 for the model is solved annually. I set the number of grid points for the Pareto weight to be 50 and for firm size s to be 20. Table 4.6 lists the complete set of parameters that I estimate.

Table 4.6: Parameters

Parameters	Description
η	the death probability
λ_1	the offer arrival probability
ρ_z	the $AR(1)$ coefficient of productivity shocks
μ_z	the mean of productivity shocks for $e = 1$
σ_z	the standard deviation of productivity shocks
μ_s	the mean of $F(s)$
σ_s	the standard deviation of $F(s)$
c	cost of efforts
σ	relative risk aversion
α_0, α_1	production function parameter

4.6.3 Moments and identifications

I next make a heuristic identification argument that justifies the choice of moments used in the estimation. Firstly, for the identification of the productivity process, the exit rate, and offer arrival rate, there are direct links between the model and the data. The exit rate directly informs η. Likewise, the incidence of job-to-job transitions is monotonically related to λ_1. The parameters of the productivity process, namely ρ_z, μ_z and σ_ϵ, are informed directly by the estimates of an $AR(1)$ process relating to the profitability of each firm-executive match:

$$profit_{it} = \beta_0 + \rho_z profit_{it-1} + \epsilon_{it,0},$$

where i represents the executive-firm match and t represents the year.

Secondly, the two parameters governing the job offer distribution, μ_s and σ_s, are

[25] The choice of grid points is for speed of estimation. The simulated moments are very robust in this choice.

[26] The upper and lower bounds of the truncated normal distribution are calibrated to be the 0.99 and 0.01 quantiles of market capitalization in the data.

disciplined by the mean and variance of firm size. Given $\lambda_1 > 0$, the higher μ_s, the more likely executives can transit to larger firms and the larger the mean of $\log(size)$. Similarly, the higher σ_s, the more heterogeneous the outside firms, and both mean and variance of $\log(size)$ increase.

Thirdly, regarding the production function, α_0 is mainly determined by the level of total compensation, and α_1 is determined by the relationship between firm size and total compensation. Therefore, α_0 and α_1 are identified by the mean and variance of $\log(tdc1)$ and $\beta_{tdc1-size}$ in the following regression of $\log(tdc1)$ on $\log(size)$:

$$\log(tdc1_{it}) = \beta_1 + \beta_{tdc1-size} \log(size_{it}) + \epsilon_{it,1}.$$

The final part of the identification concerns the parameters σ and c. These parameters govern the level of incentives and how these incentives change with compensation level. To be consistent with the incentive variable delta in the data, I construct in the simulated data a "delta" variable defined by the dollar change in pay for a percentage change in productivity. I use the mean and variance of the $\log(delta)$ to inform the effort cost c. To discipline σ, I run the following regression:

$$\log(delta_{it}) = \beta_2 + \beta_{delta-tdc1} \log(tdc1_{it}) + \epsilon_{it,2},$$

and use $\beta_{delta-tdc1}$ to inform σ. Numerical exercises show that $\beta_{delta-tdc1}$ is closely related to σ. The higher σ, the larger $\beta_{delta-tdc1}$.

Firm-size premiums I intentionally leave the firm-size pay-growth premium and incentive premium untargeted in the estimation. Instead, in the real data and the simulated data by the estimated model, I separately estimate these premiums using the same regression specification in order to examine whether the model mechanism can match up with the real world. In both the data and model-generated data, the premiums are estimated as follows. The firm-size pay-growth premium is the coefficient $\beta_{\Delta tdc1-size}$ in the following regression:

$$\Delta \log(tdc1_{it}) = \beta_3 + \beta_{\Delta tdc1-size} \log(size_{it}) + \beta_4 \log(tdc1_{it}) + \epsilon_{it,3}; \qquad (4.9)$$

and the firm-size incentive premium is the coefficient $\beta_{delta-size}$ in the following regression:

$$\log(delta_{it}) = \beta_5 + \beta_{delta-size}\log(size_{it}) + \beta_6\log(tdc1_{it}) + \epsilon_{it,4}. \tag{4.10}$$

The estimates of both premiums in the data are shown in column (2) of table 4.1 and table 4.2 in the section of motivating facts, respectively.

4.6.4 Estimates

Table 4.7: Moments and estimates

Moments	Data	Model	Estimates	Standard Error
Exit rate	0.0691	0.0691	$\eta = 0.0695$	0.0127
J-J transition rate	0.0498	0.0473	$\lambda_1 = 0.3164$	0.0325
$\hat{\rho}_{profit}$	0.7683	0.6299	$\rho_z = 0.8004$	0.0366
$Mean(profit)$	0.1260	0.1144	$\mu_z = 0.0279$	0.0014
$Var(profit)$	0.0144	0.0160	$\sigma_z^2 = 0.1198$	0.0044
$Mean(\log(size))$	7.4515	7.4806	$\mu_s = 1.2356$	0.0365
$Var(\log(size))$	2.3060	2.1610	$\sigma_s = 2.5795$	0.1211
$Mean(\log(tdc1))$	7.2408	7.2665	$\alpha_0 = -1.5534$	0.0147
$Var(log(tdc1))$	1.1846	0.8960	$\alpha_1 = 0.5270$	0.0217
$\beta_{tdc1-size}$	0.3830	0.2822		
$\beta_{delta-tdc1}$	1.1063	1.1997	$\sigma = 1.1038$	0.0030
$Mean(\log(delta))$	8.4994	8.478	$c = 0.0814$	0.0259
$Var(\log(delta))$	3.4438	3.35872		

Table 4.7 reports the targeted values of moments in the data and the corresponding values in the estimated model. The last two columns list the parameter estimates and standard errors. While I arrange moments and parameters along the identification argument made in the previous subsection, all parameters are estimated jointly. Overall, the model provides a decent fit to the data.

Looking into the estimates, a job arrival rate $\lambda_1 = 31.64\%$ is required to match the job-to-job transition rate 4.98% in the data. The magnitude of λ_1 indicates that, on average, the executive will receive an outside offer every three years. Most job offers (about 84%) are from poaching firms that are smaller than the current firm and are used to negotiate compensation with the current firm. This is confirmed by a small mean of poaching firm size. The magnitude of μ_s indicates that most offers are provided by relatively small firms, though the magnitude of σ_s implies the variation of poaching firm size is high. Comparing the data and the model-simulated mean and variance of $\log(size)$, it seems using a log-normal distribution is sufficient to match the firm size distribution in the data.

The process of productivity is matched reasonably well, given I use only 6 grid points. The mean of $\log(tdc1)$ is matched well, but the variance of $log(tdc1)$ and $\beta_{tdc1-size}$ is not. In particular, the variance of $log(tdc1)$ is much lower in the model-generated data. This indicates that the model may miss out some heterogeneous features of firms and executives. Finally, the optimal dynamic contracting employed by the model provides good matches on the mean and variance of $\log(delta)$ and the correlation of delta with total compensation, $\beta_{delta-tdc1}$.

4.6.5 Predicting firm-size premiums

Table 4.8 reports the size-premium estimates in the data and the model simulated data. There are three premiums. The first row is the size pay-growth premium estimated in regression (4.9). The second row and the third row are both the size incentive premiums estimated in regression (4.10) except that the total compensation is not controlled in estimating *incentive premium* **(w/o tdc1)** in the last row. Therefore, it includes premiums for both level and compositional reasons, while the second row is the incentive premium that cannot be attributed to pay levels of total pay, which is the focus of my explanation. Nevertheless, I show all the premiums can be replicated by my model.

Table 4.8: Predictions on size premiums

	Benchmark		Model Variants		
	Data	Model	w/o mkt inc	More offers	Less offers
Size premiums	(1)	(2)	(3)	(4)	(5)
pay-growth premium	0.1542	0.1450	0.1481	0.1624	0.0411
incentive premium	0.3473	0.3122	-0.0444	0.4299	0.1964
incentive premium **(w/o tdc1)**	0.6044	0.6507	0.4202	0.7093	0.4076

Column (1) shows the premium estimates in the data, as reported in table 4.1 and table 4.2. Column (2) shows the estimates in the benchmark model using the estimated parameters. Comparing columns (1) and (2), I find that even without targeting these premiums, the model can quantitatively capture all three premiums. In the model, the size pay-growth premium is driven by the renegotiation, and the size incentive premium is driven by labor market incentives. There is nothing mechanical that forces these estimates to coincide between the data and the model. The fact that the predicted premiums match up so closely with the estimates in the data is reassuring for the ability of the model mechanism to play an important role in explaining the firm size premium. In particular, since my model carries the insights of Edmans et al. (2009), I am able to predict the size incentive premiums with or without controlling for total compensation.

To further clarify the mechanisms behind the premium predictions, in columns (3) to (5), I report the premium estimates in several model variants. In column (3), I simulate a counterfactual scenario where firms ignore labor market incentives when designing incentive contracts. In column (4), I simulate the model using a higher job arrival probability $\lambda_1 = 0.6$. In column (5), I simulate the model with a lower job arrival probability $\lambda_1 = 0.1$.

Column (3) shows that once labor market incentives are ignored, while the pay-growth premium remains almost the same as in column (2), the incentive premium (after controlling for total compensation) in the second row essentially becomes zero. Therefore, the incentive premium of columns(2) is solely driven by labor market incentives. The incentive premium estimated at 0.4202 without controlling for total compensation reflects the notion that total compensation is higher in larger firms, which is the channel proposed by Edmans et al. (2009). Columns (4) and (5) show that when there are more (less) job offers, both the pay-growth and incentive premiums are higher (lower). These exercises illuminate that it is indeed the poaching offers that drive the predicted premiums.

4.6.6 Decomposition

To further evaluate the contribution of labor market incentives, in the data generated by a model where labor market incentives are ignored (column (3) in table 4.8), I cut the firm size into ten groups. The upper panel of figure 4.8 shows the box plots of $\log(delta)$ across ten firm size groups. Clearly, smaller firms are likely to suffer more by ignoring labor market incentives, in consistent with the job ladder mechanism. Indeed, firms that are lower on the job ladder benefit more from executives' concerns of climbing the ladder. I further calculate the ratio of *delta* with and without labor market

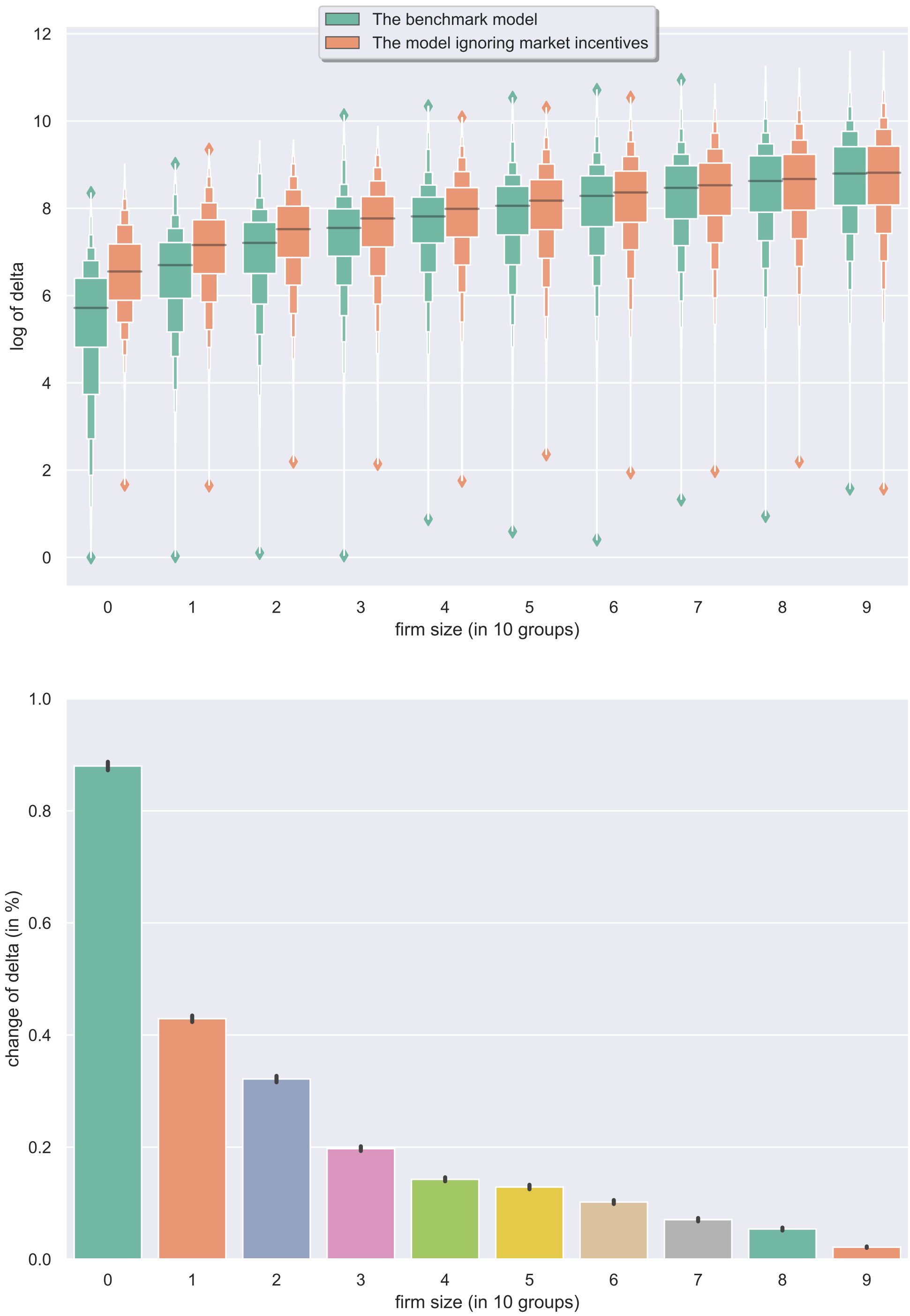

Figure 4.8: Fraction of market incentives is higher in smaller firms

incentives in the lower panel of figure 4.8. The fraction of market incentives is very high for the smallest firm group: The *delta* will be 80% higher when the job ladder is absent. The fraction quickly decreases to around 15% in the medium-size firms, and almost vanishes for top-size firms.

4.7 The long-run trend in executive compensation

Based on the structural estimation, I use a counterfactual exercise to quantitatively explain the sharp increases in executive total pay and performance-based incentives, more inequality across executives, and a higher correlation between executive compensation and firm size since the mid-1970s, as documented by Frydman and Saks (2010). In table 4.9, I select two representative periods 1970 - 1979 and 1990 - 1999 and replicate the data moments from Frydman and Saks (2010). The average total compensation rises from $1,090,000$ dollars before 1979 to $4,350,000$ dollars after 1990, and the average performance-based incentives increase almost six-fold from the 1970s to the 1990s. The interquartile range of third and first quartiles increases from $670,000$ dollars to $3,080,000$ dollars. While firm size is closely related to executive pay in the data after 1992, it was weaker in previous decades. The coefficient increases from 0.199 to 0.264 from the 1970s to the 1990s.

All these changes since the 1970s can be accounted for in my model by an exogenous change of the external executive labor market, measured by the job arrival rate λ_1. While the model does not provide an endogenous mechanism for the increase in λ_1, there is abundant evidence of more active executive labor markets since the mid-1970s. Murphy and Zabojnik (2007) document that an increasing number of CEO openings have been filled through external hires. Huson et al. (2001) document that the fraction of outsider CEOs increased from 15.3% in the 1970s to 30.0% at the beginning of the 1990s. One explanation for the trend is that executive jobs have increasingly placed greater emphasis on general rather than firm-specific skills (Frydman 2005). This is also the view taken by this paper. The executive productivity in the model is "general" and "transferable" between firms.

For the exercise, I calibrate λ_1 to be 5% for 1970 - 1979 and 40% for 1990 - 1999. These values are chosen to match the data moments under the constraint that all other parameters are equal to the estimated values in the structural estimation. Since most firms in the sample of Frydman and Saks (2010) are within the top 500, I keep the largest 500 firms in the simulated data as well. The moments calculated by model-simulated datasets are reported in the last two columns of table 4.9.

Table 4.9: Long-run trend in executive compensation

Moments	Data		Model	
(dollar value in year 2000)	1970 - 1979	1990 - 1999	$\lambda_1 = 0.05$	$\lambda_1 = 0.4$
Mean tdc1 (thousand)	1090	4350	985	4296
Mean size (million)	-	-	2426	5710
Mean delta (thousand)	21.743	120.342	24.972	125.310
$\beta_{tdc1-size}$	0.199	0.264	0.175	0.240
Percentiles of tdc1 (thousand)				
25th percentile	640	1350	109	1217
50th percentile	930	2360	478	2957
75th percentile	1310	4430	1596	5860

The results are consistent with the model intuition and are quantitatively matched with the counterparts in data. As λ_1 increases, executives are more likely to use poaching offers to renegotiate contracts, which leads to higher total compensation $tdc1$ (from $985,000$ dollars to $4,296,000$ dollars) and higher incentives $delta$ (from $24,972$ dollars to $125,310$ dollars for a 1%increase in fir's rate of return). Moreover, as firms bid for executives, the correlation between pay and firm size becomes larger (from 0.175 to 0.240). Finally, since the executive labor market in the model is search-frictional, inequality is amplified with more poaching offers: Lucky executives receive many poaching offers, while unlucky ones get few job-hopping opportunities.[27]

My model also entails predictions for moments that are not disclosed in Frydman and Saks (2010). A more active labor market also induces a larger average firm size. The mean of firm size doubles as λ_1 increases (from 2,425 million to 5,710 million). The predictions for firm-size pay-growth and incentive premiums (not shown in table 4.9) follow a similar pattern as in the last two columns of table 4.8. These predictions require further examination in the data.

[27] While the simulated moments are mostly very close to the data, there are some exceptions. In particular, the model generates much lower *tdc1* in the first two percentiles when $\lambda_1 = 0.05$. This may indicate that the poaching offer distributions of the 1970s and 1990s are different. Thus, separate estimations are required for different periods.

4.8 The spillover effect and policy implications

In this section, I discuss the spillover effect of firms' willingness to bid for executives using comparative statics. The parameter α_0 in the production function of the model represents the firm's (or the board's) willingness to pay for executives. The "spillover" refers to the effect where higher bids from some firms not only raises executive pay in those firms but also increases pay in all firms that are higher on the job ladder. This is because executives who are higher on the job ladder can make use of these bids to negotiate with their present firms. Consequently, the renegotiation leads to higher pay and higher performance-based incentives.

From the perspective of a regulator, executive pay is an essential part of corporate governance and is often determined by a company's board of directors. When compensation is inefficient, it is usually a symptom of an underlying governance problem brought on by conflicted boards and dispersed shareholders. For this reason, I assume that α_0 is negatively correlated with the quality of corporate governance. For example, an entrenched executive tends to have higher bargaining power and face a higher α_0, while a more independent board may impose a lower α_0 on executives. A caveat of this assumption must be emphasized: It is by no means that α_0 should always be negatively correlated with the quality of governance. This assumption should be valid only in the range where α_0 is too big.

Quantitatively, I use counterfactuals of higher α_0 values in firms of different size to evaluate how sizable such spillover effect can be. I consider two counterfactual scenarios. In one scenario, α_0 doubles for firms that are smaller than the size median, the "small/medium firms". I denote this higher bid of small/medium firms as "worse governance in small firms". And it supposes to create a spillover effect on the pay of large firms. To compare this spillover effect, I use the second counterfactual that α_0 doubles for firms that are larger than the median. And this case is denoted as "worse governance in large firms". Figure 4.9 plots the distribution of delta (in the upper panel) and total compensation (in the lower panel) across ten equally divided firm size groups. There are three box plots for each size group, i.e., worse governance in small firms, the benchmark model and worse governance in large firm,s, and the medians are marked as a horizontal line in the middle of each box.

Not surprisingly, the boosts in bids increase total compensation and incentives in

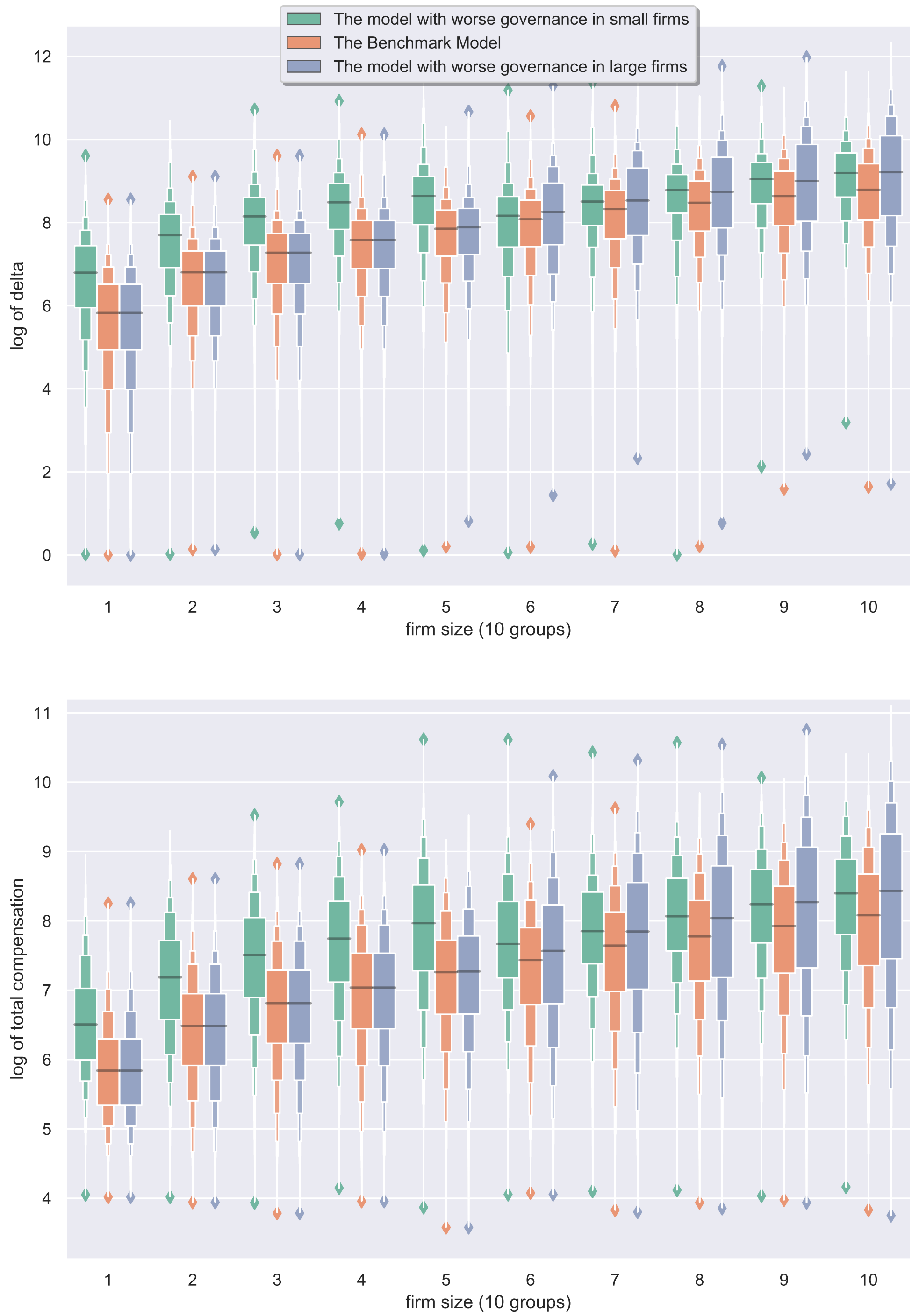

Figure 4.9: Compare higher bids from small/medium firms and from large firms

each separate type of firms. A higher bidding willingness in small/medium firms (in green) raises pay and incentives in firms of the first five groups, while a higher α_0 in large firms (in blue) increases pay and incentives in the largest five groups. Importantly, the rise in α_0 in small/medium firms spillovers to large firms as well. In terms of median, this spillover effect is as large as the effect of higher bids from large firms themselves. As shown in figure 4.9, there are 40% to 50% increases in pay and incentives of the largest two groups of firms, in both cases of higher bids from small/medium firms and higher bids from large firms.

The policy implications of this exercise are as follow. To regulate the compensation of highly paid executives, rather than only focusing on large firms, it is important to lower the bids in small/ medium firms. Thus, large firms will face less competitive pressure. As for particular regulation policies, reforms that have been proposed or implemented including more independent compensation committee, greater mandatory pay (or pay ratio) disclosure, say-on-pay legislation should work in small and medium firms as well.

4.9 Conclusions

This paper has studied the impact of labor market competition on managerial incentive contracts. I developed a dynamic contracting model where executives use poaching offers to renegotiate with the current firm, and showed that poaching offers have both a level and an incentive effect on compensation. The model explains the firm-size pay-growth premium and incentive premium. Empirical evidence from a new dataset on job turnovers supports the job ladder mechanism.

I structurally estimated the model without explicitly targeting firm-size pay-growth and incentive premiums, yet the predicted premiums of the model match up very closely with the estimates in the data. A counterfactual analysis based on the structural estimation showed that with an exogenous increase of poaching offer arrival rate, my model can account for the sharp increase in total pay, performance-based incentives, and the correlation between firm size and pay levels since the mid-1970s.

Quantitative analysis showed that there is a spillover effect from the deterioration of corporate governance in small and medium firms to the compensation growth of the overall executive labor market. The policy implication is that to regulate the compensation of highly paid executives especially in large firms, it is important to improve the corporate governance of small and medium firms and reduce their bids. This will lower the competitive pressure faced by board members of large firms.

Appendices

Appendix I: Model appendices

Proof for proposition 4.3

I start with a lemma showing that $\mathcal{I}\Big(\overline{W}(z',s)\Big)$ is a weighted sum of $\frac{\Delta\overline{W}(z',s)}{\Delta z'}$ over the domain of z'. And then show $\frac{\Delta\overline{W}(z',s)}{\Delta z'}$ decreases in s.

Step 1: Showing that $\mathcal{I}\Big(\overline{W}(z',s)\Big)$ is a weighted sum of $\frac{\Delta\overline{W}(z',s)}{\Delta z'}$

Lemma .1. *Consider a productivity set $\mathbb{Z} = \{z^{(1)}, z^{(2)}, ..., z^{(n_z)}\}$. Suppose there is a distribution of productivity when the executive takes the effort Γ, a distribution when the executive shirks Γ^s, a likelihood ratio $g = \Gamma/\Gamma^s$ and a value function W. All functions are defined on $\mathbb{Z}$, then the incentive the executive receives from W is*

$$\mathcal{I}\Big(W(z)\Big) = \sum_{i=1}^{n_z-1} \omega_i \frac{\Delta W(z^{(i)})}{\Delta z^{(i)}},$$

where $\Delta z^{(i)} = z^{(i+1)} - z^{(i)}$ and $\omega_i \geq 0$.

Proof. Without lose of generality, I assume $g(z) \geq 1$ for $z \in \{z^{(1)}, z^{(2)}, ..., z^{(m)}\}$ and $g(z) < 1$ for $z \in \{z^{(m+1)}, ..., z^{(n_z)}\}$ where $m < n_z$, and define $\gamma(z) \equiv |1 - g(z)| \times \Gamma(z)$. I further denote $W(z^{(i)})$ by W_i and $\gamma(z^{(i)})$ by γ_i. Moreover, $\sum_{z\in\mathbb{Z}}(1-g(z))\Gamma(z) = 0$ implies that

$$\gamma_1 + ... + \gamma_m - \gamma_{m+1} - ... - \gamma_{n_z-1} - \gamma_{n_z} = 0. \tag{11}$$

It follows that

$$\begin{aligned}
\mathcal{I}\Big(W\Big) =& \sum_{z\in\mathbb{Z}} \Big(W(z)(1-g(z))\Gamma(z)\Big) \\
=& -\gamma_1 W_1 - \gamma_2 W_2 - ... - \gamma W_m + \gamma_{m+1} W_{m+1} + \gamma_{n_z} W_{n_z} \\
=& \gamma_1(W_2 - W_1) + (\gamma_1 + \gamma_2)(W_3 - W_2) + ... \\
&+ (\gamma_1 + ... + \gamma_m)(W_{m+1} - W_m) + (\gamma_1 + ... + \gamma_m - \gamma_{m+1})(W_{m+2} - W_{m+1}) + ... \\
&+ (\gamma_1 + ... + \gamma_m - \gamma_{m+1} - ... - \gamma_{n_z-1})(W_{n_z} - W_{n_z-1}) \\
&+ (\gamma_1 + ... + \gamma_m - \gamma_{m+1} - ... - \gamma_{n_z-1} - \gamma_{n_z})W_{n_z} \\
=& \gamma_1 \Delta z_1 \frac{W_2 - W_1}{\Delta z_1} + (\gamma_1 + \gamma_2)\Delta z_2 \frac{(W_3 - W_2)}{\Delta z_2} + ... \\
&+ (\gamma_1 + ... + \gamma_m)\Delta z_m \frac{(W_{m+1} - W_m)}{\Delta z_m} \\
&+ (\gamma_1 + ... + \gamma_m - \gamma_{m+1})\Delta z_{m+1} \frac{W_{m+2} - W_{m+1}}{\Delta z_{m+1}} + ... \\
&+ (\gamma_1 + ... + \gamma_m - \gamma_{m+1} - ... - \gamma_{n_z-1} - \gamma_{n_z-1})\Delta z_{n_z-1} \frac{W_{n_z} - W_{n_z-1}}{\Delta z_{n_z-1}} \\
=& \omega_1 \frac{W_2 - W_1}{\Delta z_1} + \omega_2 \frac{(W_3 - W_2)}{\Delta z_2} + ... \\
&+ \omega_m \frac{(W_{m+1} - W_m)}{\Delta z_m} + \omega_{m+1} \frac{W_{m+2} - W_{m+1}}{\Delta z_{m+1}} + ... + \omega_{n_z-1} \frac{W_{n_z} - W_{n_z-1}}{\Delta z_{n_z-1}} \\
=& \sum_{i=1}^{n_z-1} \omega_i \frac{\Delta W(z^{(i)})}{\Delta z^{(i)}}.
\end{aligned}$$

The first equality follows from the definition of the incentive operator $\mathcal{I}$, the rest steps are simple algebraic transformations, where I have applied condition (11). By construction, ω_i is positive.

□

Step 2: Expressing $\frac{\Delta \overline{W}(z,s)}{\Delta z}$ in terms of s.

Given lemma 1, it is sufficient to show that $\frac{\Delta \overline{W}(z,s)}{\Delta z}$ decreases in s for all $z \in \mathbb{Z}$. Notice that

$$\frac{\Delta \overline{W}(z,s)}{\Delta z} = -\frac{\Delta \Pi(z,s,\overline{W})/\Delta z}{\Delta \Pi(z,s,\overline{W})/\Delta \overline{W}} = u'(\overline{w}(s))\frac{\Delta \Pi(z,s,\overline{W})}{\Delta z}, \tag{12}$$

where $\overline{w}(z,s)$ is the compensation corresponding to $\overline{W}(z,s)$ and satisfies (4.1).

To derive $\overline{w}$, suppose the effort cost is

$$c = \overline{c}(s) \equiv \tilde{\beta} \sum_{z' \in \mathbb{Z}} \overline{W}(z',s)(1-g(z'|z))\Gamma(z'|z),$$

such that the optimal contract indicates that the promised value equals to the bidding frontier

$$W(z',\tilde{s}) = \overline{W}(z',s).$$

Under the optimal contract, the continuation value (profit) of the firm is zero.

According to the Bellman equation of the firm,

$$\begin{aligned}
\Pi(z,s,\overline{W}(z,s)) &= \sum_{z' \in \mathbb{Z}} \Big(\alpha_0 s^{\alpha_1} z' - \overline{w} + \tilde{\beta} \int_{\tilde{s}} \Pi(z',s,W(z',\tilde{s}) d\tilde{F}(\tilde{s})\Big)\Gamma(z'|z) \\
&= \sum_{z' \in \mathbb{Z}} \Big(\alpha_0 s^{\alpha_1} - \overline{w} + \tilde{\beta} \int_{\tilde{s}} \Pi(z',s,\overline{W}(z',s) d\tilde{F}(\tilde{s})\Big)\Gamma(z'|z) \\
&= \sum_{z' \in \mathbb{Z}} \Big(\alpha_0 s^{\alpha_1} - \overline{w}\Big)\Gamma(z'|z) = 0.
\end{aligned}$$

Therefore,

$$\overline{w}(z,s) = \alpha_0 s^{\alpha_1} \sum_{z' \in \mathbb{Z}} z' \Gamma(z'|z)$$

To derive $\frac{\Delta \Pi(z,s,\overline{W})}{\Delta z}$, I use envelop theorem. It follows that

$$\begin{aligned}
\frac{\Delta \Pi(z,s,\overline{W})}{\Delta z} &= \sum_{z' \in \mathbb{Z}} \Big(\alpha_0 s^{\alpha_1} z' + \tilde{\beta} \int_{\tilde{s} \leq s} \Pi(z',s,\overline{W}(z',s) d\tilde{F}(\tilde{s})\Big) \frac{\Delta \Gamma(z'|z)}{\Delta z} \\
&+ \lambda \tilde{\beta} \sum_{z' \in \mathbb{Z}} \Big(\int_{\tilde{s}} \overline{W}(z',s) d\tilde{F}(\tilde{s})\Big) \frac{\Delta \Gamma(z'|z)}{\Delta z} \\
&+ \mu \tilde{\beta} \sum_{z' \in \mathbb{Z}} \Big(\int_{\tilde{s}} \overline{W}(z',s) d\tilde{F}(\tilde{s})\Big) \frac{\Delta \Big((1-g(z'|z)\Gamma(z'|z)\Big)}{\Delta z} \\
=& \alpha_0 s^{\alpha_1} \sum_{z' \in \mathbb{Z}} z' \frac{\Delta \Gamma(z'|z)}{\Delta z} + \tilde{\beta} \sum_{z' \in \mathbb{Z}} \int_{\tilde{s}} \overline{W}(z',s) d\tilde{F}(\tilde{s}) \Big(\lambda \frac{\Delta \Gamma(z'|z)}{\Delta z} + \mu \frac{\Delta \Big((1-g(z'|z)\Gamma(z'|z)\Big)}{\Delta z}\Big).
\end{aligned} \tag{13}$$

Divide both sides by $\alpha_0 \sum_{z' \in \mathbb{Z}} z' \frac{\Delta \Gamma(z'|z)}{\Delta z}$, we have

$$\frac{\frac{\Delta \Pi(z,s,\overline{W})}{\Delta z}}{\alpha_0 \sum_{z' \in \mathbb{Z}} z' \frac{\Delta \Gamma(z'|z)}{\Delta z}} = s^{\alpha_1} + \psi(s), \tag{14}$$

where $\psi(s) \equiv \frac{\tilde{\beta} \sum_{z' \in \mathbb{Z}} \int_{\tilde{s}} \overline{W}(z',s) d\tilde{F}(\tilde{s}) \Big(\lambda \frac{\Delta \Gamma(z'|z)}{\Delta z} + \mu \frac{\Delta \big((1-g(z'|z)\Gamma(z'|z)\big)}{\Delta z}\Big)}{\Delta z} / \alpha \sum_{z' \in \mathbb{Z}} z' \frac{\Delta \Gamma(z'|z)}{\Delta z}$.

Since all items of $\psi(s)$ are positive, $\psi(s) > 0$. Because $\psi(s)$ only depends on s via $\overline{W}$ which is increasing in s, $\psi(s)$ is also increasing in s.

Insert (13) and (14) into (12), we have

$$\frac{\Delta\overline{W}(z,s)}{\Delta z} = u'(\overline{w}(s))\frac{\Delta\Pi(z,s,\overline{W})}{\Delta z} = u'\Big(\alpha_0 s^{\alpha_1}\sum_{z'\in\mathbb{Z}} z'\Gamma(z'|z)\Big)\Big(s^{\alpha_1}+\psi(s)\Big)\alpha_0\sum_{z'\in\mathbb{Z}} z'\frac{\Delta\Gamma(z'|z)}{\Delta z}. \tag{15}$$

Step 3: Showing that $\frac{\Delta\overline{W}(z,s)}{\Delta z}$ decreases in s under the stated condition.

To have

$$\lim_{\Delta s\to 0}\frac{\Delta\overline{W}(z,s+\Delta s)}{\Delta z} - \frac{\Delta\overline{W}(z,s)}{\Delta z} > 0,$$

using (15)

$$\frac{u'\Big((s+\Delta s)^{\alpha_1}\alpha_0\sum_{z'\in\mathbb{Z}} z'\Gamma(z'|z)\Big)}{u'\Big(s^{\alpha_1}\alpha_0\sum_{z'\in\mathbb{Z}} z'\Gamma(z'|z)\Big)} < \frac{s^{\alpha_1}+\psi(s)}{(s+\Delta s)^{\alpha_1}+\psi(s+\Delta s)}.$$

Applying $u'(w) = w^{-\sigma}$, we have

$$\Big(\frac{s}{s+\Delta s}\Big)^{-\alpha_1\sigma} < \frac{s^{\alpha_1}+\psi(s)}{(s+\Delta s)^{\alpha_1}+\psi(s+\Delta s)},$$

or

$$\sigma > \frac{\log\frac{s^{\alpha_1}+\psi(s)}{(s+\Delta s)^{\alpha_1}+\psi(s+\Delta s)}}{\frac{s}{s+\Delta s}}.$$

Take $\Delta s \to 0$ using L'Hopital's rule,

$$\sigma > 1 + \frac{s^{1-\alpha_1}}{\alpha_1}\psi'(s).$$

Appendix II: Empirical appendices

This appendix contains some extra regression results on firm-size incentive premium. Figure 10 is a heat map of performance-based incentives *log(delta)* on total compensation and firm size. It shows that among executives with similar total compensation, those in larger firms get higher performance-based incentives.

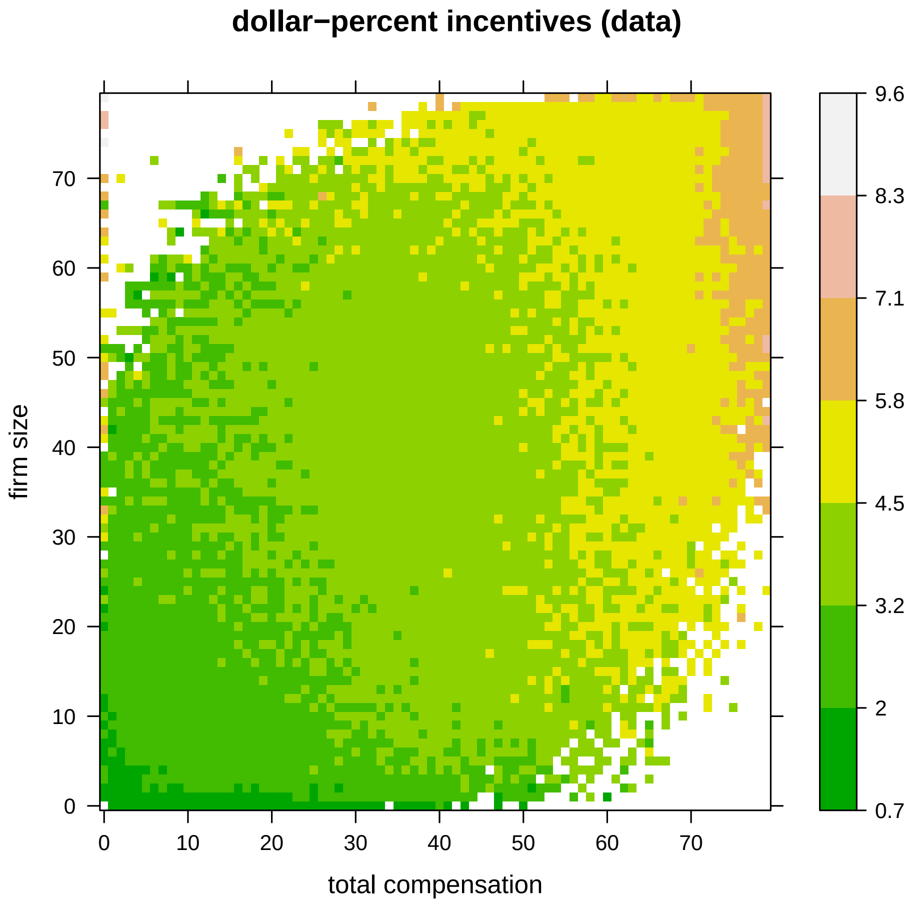

Figure 10: $\log(delta)$ over firm size and total compensation

Note: delta is the wealth-performance sensitivity defined as the dollar change in firm-related wealth for a percentage change in firm value. The total compensation is the sum of salary and bonus, the value of restricted stocks and options granted, and the value of retirement and long-term compensation schemes. The firm size is the market capitalization by the end of the fisical year, calculated by `csho` × `prcc_f` where `csho` is the common shares outstanding and `prcc_f` is the close price by fiscal year. I divide the whole sample into 80 × 80 cells according to the total compensation and firm size, and compute the mean of $\log(delta)$ within each cell.

Table 10 shows more robustness check on firm-size incentive premium. Table 11 contains the full results on the interaction of firm size and proxies of labor market competition. Table 12 contains the full result of the size incentive premium for each age.

Table 10: Performance-based incentives increase with firm size

	log($delta$)				
	(1)	(2)	(3)	(4)	(5)
log(firm size)	0.585***	0.360***	0.331***	0.330***	0.440***
	(0.0141)	(0.0247)	(0.0237)	(0.0236)	(0.0236)
log(tdc1)		0.609***			0.334***
		(0.0350)			(0.0323)
tdc1 Dummies (50)			Yes		
tdc1 Dummies (100)				Yes	
Other contorls					Yes
tenure dummies	Yes	Yes	Yes	Yes	Yes
age dummies	Yes	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes	Yes
industry dummies	Yes	Yes	Yes	Yes	Yes
year $\times$ industry dummies	Yes	Yes	Yes	Yes	Yes
Observations	146,747	128,006	128,006	128,006	109,730
adj. R^2	0.442	0.514	0.523	0.524	0.595

Note: This table reports evidence on firm size premium in performance-based incentives. The dependent variable is the log of *delta* where *delta* is the dollar change in firm related wealth for a percentage change in firm value. Firm size is measured by the market capitalization defined by the common shares outstanding times the fiscal year close price. *tdc1* is the total compensation, including the sum of salary and bonus, the value of restricted stocks and options granted, and value of retirement and long-term compensation schemes. It is the variable *tdc1* in ExecuComp dataset. In all regressions, I have controlled for age dummies, executive tenure dummies, year $\times$ industry dummies. Column (1) is a regression of *log(delta)* on *log(firm size)*, which replicates the cross-sectional regression in the literature. From column (2) to column (4), I add *log(tdc1)*, *tdc1 dummies 50* and *tdc1 dummies 100* (tdc1 values are evenly divided into 50 or 100 groups and then transformed into dummies), respectively. In column (5), I add other controls including *operating profitability*, *market-book ratio*, *annualized stock return*, *director*, *CEO* and *CFO*, *interlock*. Standard errors clustered at the firm $\times$ fiscal year level are shown in parentheses, and I denote symbols of significance by * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Table 11: Size incentive premium increases with managerial labor market competition

	log(*delta*)			
	(1)	(2)	(3)	(4)
log(firm size)	0.525***	0.529***	0.561***	0.571***
	(0.00512)	(0.00499)	(0.00310)	(0.0139)
EE190	1.919*			
	(0.776)			
log(firm size) × *EE190*	0.415***			
	(0.101)			
EE90		2.611**		
		(0.903)		
log(firm size) × *EE90*		0.359**		
		(0.118)		
gai			-1.211***	
			(0.0941)	
log(firm size) × *gai*			0.0648***	
			(0.0118)	
inside CEO				-0.00566***
				(0.00156)
log(firm size) × *inside CEO*				-0.000458*
				(0.000202)
Controls	Yes	Yes	Yes	Yes
Observations	125858	125858	75747	125858
adj. R-sq	0.521	0.521	0.531	0.521

Note: This table reports evidence that the firm size incentive premium increases as the managerial labor market competition becomes thicker. The dependent variable is the log of *delta* where *delta* is the dollar change in firm related wealth for a percentage change in firm value. The independent variables include the log of firm size, several variables that measure the how active the competition in managerial labor markets, and the interaction terms between firm size and labor market competition. In column (1), labor market competition is measured by job-to-job transition rate in each (Fama-French 48) industries and fiscal years. A job-to-job transition is defined as an executive leaving the current firm and starting to work in another firm within *190* days. The same measure is used in column (2) except the gap between jobs is changed to *90* days. Column (3) measures labor market activeness by the average of the general ability index at the industry-year level. The original index is provided by Custódio et al. (2013). Column (4) uses the industry level percentage of new CEOs who are promoted inside the company. The data is provided by Martijn Cremers and Grinstein (2013). The control variables include executive tenure dummies, age dummies, fiscal year dummies, *operating profitability*, *market-book ratio*, *annualized stock return*, whether the executive served as a director, CEO or CFO during the fiscal year, whether the executive is involved in the interlock relationship. For regression including *inside CEO*, I use data from year 1992 to year 2006. For the rest, I use data from year 1992 to year 2015. Standard errors clustered at the firm times fiscal year level are shown in parentheses, and I denote symbols of significance by * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Table 12: Size incentive premium decreases with executive age

	log($delta$)				
	(1)	(2)	(3)	(4)	(5)
age35 × log(firm size)	0.849***	0.652***	0.580***	0.541***	0.539***
	(0.0534)	(0.0649)	(0.0620)	(0.0614)	(0.0611)
age36 × log(firm size)	0.753***	0.530***	0.516***	0.484***	0.481***
	(0.0487)	(0.0658)	(0.0519)	(0.0538)	(0.0529)
age37 × log(firm size)	0.746***	0.543***	0.540***	0.508***	0.506***
	(0.0366)	(0.0440)	(0.0359)	(0.0365)	(0.0365)
age38 × log(firm size)	0.689***	0.471***	0.471***	0.438***	0.436***
	(0.0328)	(0.0390)	(0.0339)	(0.0341)	(0.0332)
age39 × log(firm size)	0.667***	0.444***	0.443***	0.410***	0.410***
	(0.0276)	(0.0365)	(0.0297)	(0.0297)	(0.0296)
age40 × log(firm size)	0.664***	0.475***	0.493***	0.462***	0.461***
	(0.0296)	(0.0358)	(0.0319)	(0.0341)	(0.0338)
age41 × log(firm size)	0.653***	0.449***	0.489***	0.459***	0.457***
	(0.0264)	(0.0350)	(0.0320)	(0.0325)	(0.0324)
age42 × log(firm size)	0.640***	0.437***	0.478***	0.453***	0.450***
	(0.0285)	(0.0342)	(0.0330)	(0.0338)	(0.0335)
age43 × log(firm size)	0.630***	0.408***	0.469***	0.445***	0.444***
	(0.0248)	(0.0333)	(0.0303)	(0.0311)	(0.0309)
age44 × log(firm size)	0.622***	0.405***	0.473***	0.449***	0.447***
	(0.0230)	(0.0345)	(0.0313)	(0.0314)	(0.0314)
age45 × log(firm size)	0.608***	0.397***	0.468***	0.447***	0.446***
	(0.0220)	(0.0287)	(0.0280)	(0.0280)	(0.0278)
age46 × log(firm size)	0.592***	0.377***	0.443***	0.424***	0.422***
	(0.0210)	(0.0293)	(0.0283)	(0.0286)	(0.0284)
age47 × log(firm size)	0.594***	0.365***	0.445***	0.428***	0.426***
	(0.0207)	(0.0297)	(0.0289)	(0.0296)	(0.0295)
age48 × log(firm size)	0.598***	0.367***	0.454***	0.435***	0.434***
	(0.0163)	(0.0259)	(0.0252)	(0.0256)	(0.0257)
age49 × log(firm size)	0.594***	0.369***	0.437***	0.419***	0.417***
	(0.0180)	(0.0264)	(0.0284)	(0.0276)	(0.0277)
age50 × log(firm size)	0.589***	0.388***	0.457***	0.439***	0.438***
	(0.0210)	(0.0287)	(0.0301)	(0.0316)	(0.0317)
age51 × log(firm size)	0.563***	0.352***	0.426***	0.410***	0.409***
	(0.0173)	(0.0254)	(0.0270)	(0.0273)	(0.0275)
age52 × log(firm size)	0.560***	0.342***	0.414***	0.399***	0.398***
	(0.0191)	(0.0268)	(0.0280)	(0.0282)	(0.0281)
age53 × log(firm size)	0.577***	0.350***	0.425***	0.409***	0.408***
	(0.0192)	(0.0274)	(0.0278)	(0.0286)	(0.0287)
age54 × log(firm size)	0.570***	0.335***	0.423***	0.409***	0.409***
	(0.0209)	(0.0288)	(0.0286)	(0.0290)	(0.0293)
age55 × log(firm size)	0.569***	0.351***	0.435***	0.423***	0.423***
	(0.0184)	(0.0279)	(0.0271)	(0.0273)	(0.0273)

Table 12: Size incentive premium decreases with executive age (continue)

	(1)	(2)	(3)	(4)	(5)
age56 × log(firm size)	0.592***	0.362***	0.454***	0.442***	0.441***
	(0.0157)	(0.0260)	(0.0271)	(0.0272)	(0.0270)
age57 × log(firm size)	0.593***	0.356***	0.440***	0.429***	0.428***
	(0.0141)	(0.0233)	(0.0237)	(0.0232)	(0.0230)
age58 × log(firm size)	0.592***	0.356***	0.442***	0.430***	0.429***
	(0.0175)	(0.0266)	(0.0261)	(0.0260)	(0.0261)
age59 × log(firm size)	0.593***	0.351***	0.435***	0.423***	0.422***
	(0.0172)	(0.0256)	(0.0258)	(0.0253)	(0.0254)
age60 × log(firm size)	0.579***	0.341***	0.424***	0.412***	0.412***
	(0.0175)	(0.0271)	(0.0259)	(0.0258)	(0.0259)
age61 × log(firm size)	0.600***	0.355***	0.438***	0.428***	0.427***
	(0.0216)	(0.0307)	(0.0302)	(0.0311)	(0.0309)
age62 × log(firm size)	0.587***	0.333***	0.420***	0.409***	0.408***
	(0.0192)	(0.0282)	(0.0268)	(0.0272)	(0.0272)
age63 × log(firm size)	0.605***	0.358***	0.448***	0.436***	0.435***
	(0.0196)	(0.0252)	(0.0256)	(0.0253)	(0.0255)
age64 × log(firm size)	0.593***	0.356***	0.440***	0.429***	0.429***
	(0.0242)	(0.0285)	(0.0296)	(0.0292)	(0.0289)
age65 × log(firm size)	0.596***	0.353***	0.435***	0.423***	0.423***
	(0.0246)	(0.0318)	(0.0339)	(0.0334)	(0.0332)
logtdc1		0.611***	0.345***		
		(0.0352)	(0.0339)		
tdc1 Dummies (50)				Yes	
tdc1 Dummies (100)					Yes
profit			0.619***	0.598***	0.602***
			(0.117)	(0.116)	(0.116)
annual return			0.102*	0.0999	0.0998
			(0.0488)	(0.0485)	(0.0485)
mbr			0.116***	0.120***	0.120***
			(0.0209)	(0.0213)	(0.0213)
director			0.754***	0.739***	0.737***
			(0.0326)	(0.0307)	(0.0306)
interlock			0.517***	0.529***	0.527***
			(0.0953)	(0.0948)	(0.0947)
CEO			0.593***	0.576***	0.574***
			(0.0387)	(0.0395)	(0.0397)
CFO			0.0837***	0.0711***	0.0711***
			(0.0130)	(0.0131)	(0.0130)
N	146750	128008	109732	109732	109732
adj. R^2	0.432	0.506	0.586	0.590	0.590

Note: This table reports the evidence that firm size incentive premium decreases in executive age. The dependent variable is the log of *delta* where *delta* is the dollar change in firm related wealth for a percentage change in firm value. The key independent variable is the log of firm size where firm size is measured by the market capitalization defined by the common shares outstanding times the fiscal year close price. I allow a different coefficients of firm size across ages from 35 to 65. Control variables include total compensation (*tdc1*), age dummies, executive tenure dummies, year times industry dummies, *profit*, the operating profitability, *mbr*, the market-book ratio, *annual return*, the annualized stock return, *director*, whether the executive served as a director during the fiscal year, *CEO* and *CFO*, whether the executive served as a CEO (and CFO) during the fiscal year, *interlock*, whether the executive is involved in the interlock relationship. Standard error (clustered at the firm $\times$ fiscal year level) are shown in parentheses, and I denote symbols of significance by * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Appendix III: Estimation

Recursive multiplier method

To further characterize the optimal solution, I resort to the tools developed by Marcet and Marimon (2017).[28]In dynamic contracting problems with forward looking constraints such as the IC constraint here, the solution does not satisfy the Bellman equation. MM suggest to study a recursive Lagrangian. Under standard general conditions there is a recursive saddle-point functional equation (analogous to a Bellman equation) that characterizes a recursive solution to the planner's problem. The recursive formulation is obtained after adding a co-state variable λ_t summarizing previous commitments reflected in past Lagrange multipliers. The time-consistent continuation solution is obtained by using the endogenous λ_t as the vector of weights in the objective function. I summarize this method in the following proposition.

Proposition .4 (Marcet and Marimon)**.** *Define Pareto Frontier by*

$$P(z,s,\lambda) = \sup_W \Pi(z,s,W) + \lambda W,$$

where Π and W are defined as in (BE-F) and (PKC), and $\lambda > 0$ is a Pareto weight assigned to the executive. Then there exist positive multipliers of $\{\mu, \mu_0(z'), \mu_1(z')\}$ that solve the following problem

$$P(z,s,\lambda) = \inf_{\mu,\mu_0(z',\tilde{s}),\mu_1(z',\tilde{s})} \sup_w h(z,s,\lambda,w) + \hat{\beta}\sum_{z'} P(z',s,\lambda')\Gamma(z'|z),$$

where multiplier μ corresponds to the incentive compatibility constraint, multipliers $\mu_0(z',\tilde{s}), \mu_1(z',\tilde{s})$ correspond to participation constraints,

$$h(z,s,\lambda,w) = y(s)z' - w + \lambda u(w) - (\lambda+\mu)c,$$

Pareto weight evolves according to

$$\lambda' = \lambda + \mu(1-g(z,z')) + \mu_0(z',\tilde{s}) + \mu_1(z',\tilde{s}),$$

and

$$\hat{\beta} = \tilde{\beta}(1-\lambda_1 \sum_{\mathcal{M}_1\cup\mathcal{M}_2} F(s')).$$

The optimal contract $\{w, W(z',\tilde{s})\}$ follows that

$$u'(w) = \frac{1}{\lambda}, \tag{16}$$

$$W(z',\tilde{s}) = W(z',\tilde{s},\lambda'). \tag{17}$$

Proposition .4 can be illustrated intuitively using the Pareto weight of the executive λ and the multiplier μ of the incentive constraint. Suppose the match starts with a $\lambda^{(0)}$, and assume the participation constraints are not binding so that $\mu_0 = \mu_1 = 0$. $\lambda^{(0)}$ has to satisfy $W(z_O, s, \lambda^{(0)}) = W^0$. To deal with the moral hazard, the optimal contract indicates a $\mu^{(0)} > 0$. Then depending on the realization of z', the weight of the executive will be updated to

$$\lambda^{(i)} = \lambda^{(i-1)} + \mu^{i-1}(1-g(z,z')) \text{ for } i \text{ in } 1,2... \tag{18}$$

The evolve of λ continues as such till the match breaks.

[28]This approach has been used in many applications. A few examples are: growth and business cycles with possible default (Marcet and Marimon (1992), Kehoe and Perri (2002), Cooley, et al. (2004)); social insurance (Attanasio and Rios-Rull (2000)); optimal fiscal and monetary policy design with incomplete markets (Aiyagari, Marcet, Sargent and Seppala (2002), Svensson and Williams (2008)); and political-economy models (Acemoglu, Golosov and Tsyvinskii (2011)).

When there is an outside offer such that the executive moves from his or her current firm to the outside firm, the new match starts with a weight denoted by $\lambda^{(n)}$ such that

$$W(z, \tilde{s}, \lambda^{(n)}) = \overline{W}(z, s),$$

were I have denoted the current productivity by z, current firm by s, and the outside firm by $\tilde{s}$. It means the new match will assign a new weight to the executive so that he or she gets the continuation value $\overline{W}(z, s)$. Then the new Pareto weight will evolve again as illustrated in (18). In a nutshell, proposition .4 allows me to solve the optimal contract in the space of Pareto weight λ instead of in the space of the promised utility. At any moment, I can transit from the metrics of λ back to the metrics of utilities using (16) and (17).

The advantage of this method is I do not need to find the promised utilities $W(z', \tilde{s})$ in each state of the world for the next period. Instead, λ and μ are enough to trace all $W(z', \tilde{s})$. Moreover, λ corresponds to the total compensation level (wage level), while μ corresponds to how much contract incentive is provided in the optimal contract. The two multipliers are enough to understand both theoretically and numerically why keeping the same wage level (the same λ), incentive pays increase with firm size (μ increases with firm size).

Summary

This thesis consists of three studies investigating topics in the fields of intermediaries and executive compensation through the lens of search frictions. The three studies all points to the power of the search frictions in explaining real world phenomena. This summary provides an overview of the research questions addressed, the methodologies used to answer those questions, and the conclusions from this research.

Chapter 2 develops a directed search framework to understand two widely used market structures of intermediaries: the middleman mode and the marketmaker (or platform) mode. In the middleman mode, the intermediary is specialized in buying and selling for its own account and typically operates with inventory holdings (e.g., dealers in financial and steel markets). In the marketmaker mode, the intermediary offers a marketplace for fees, where the participating buyers and sellers can search and trade with each other (e.g., eBay, brokers in real estate and financial markets). Due to the search frictions, the intermediary has a trade-off between a larger transaction volume by operating as a middleman and a higher price/fee by acting as market-maker. This trade-off determines the optimal intermediation mode, and eventually a *marketmaking middleman* such as Amazon, who adopts a mixture of these two intermediation modes, can be optimal in the equilibrium. Our main insight holds in various market environments, including an endowment economy and a market with nonlinear matching functions. The theoretical predictions are examined using scraped data from Amazon and eBay.

Chapter 3 joins in the growing academic interest in understanding the competition among e-commerce giants. We extended the framework of Chapter 2 to a Bertrand competition game between an incumbent intermediary who can mix a middleman and a marketmaker mode, and an entrant intermediary who is restricted to be a marketmaker. Meeting technologies in both intermediaries are subjected to search frictions.

We find that the entrant faces the choice of being a second-source (for intermediation service) with high price/fees and a sole active source by undercutting the incumbent. Despite this strategic complication, we show that the intuition of chapter 2 holds in a duopoly. In particular, for a reasonable set of parameters, there exists an equilibrium of pure strategies where a marketmaking middleman incumbent emerges. Furthermore, in the mixed strategy equilibrium, the incumbent activates its market-maker mode with positive probability.

Chapter 4 assesses the impact of job search ladders on executives' incentive contracts. In the theoretical framework, an executive is poached by outside firms while a dynamic incentive contract is designed to deal with both the moral hazard problem inside the firm and competition from the external labor market. I show that competition for executives increases total compensation, and generates a new source of incentives, called *labor market incentives*, which substitutes for performance-based incentives embedded in bonus, stocks, options, etc. The model is estimated using a newly assembled dataset on job turnovers for executives in U.S. publicly listed firms. Simulations based on the structural estimates show that the model is capable of explaining and replicating the puzzling facts that executives of larger firms experience higher compensation growth and receive higher performance-based incentives.

Samenvatting (Summary in Dutch)

Bibliography

Abrahám, Arpád, Fernando Alvarez-Parra, and Susanne Forstner (2016), "The effects of moral hazard on wage inequality in a frictional labor market."

Abreu, Dilip, David Pearce, and Ennio Stacchetti (1990), "Toward a theory of discounted repeated games with imperfect monitoring." *Econometrica: Journal of the Econometric Society*, 1041–1063.

Acemoglu, Daron and Robert Shimer (1999), "Holdups and efficiency with search frictions." *International Economic Review*, 40, 827–849.

Agrawal, Ashwini and Alexander Ljungqvist (2014), "Managerial labor market frictions and corporate investment."

Alvarez, Fernando and Urban J Jermann (2000), "Efficiency, equilibrium, and asset pricing with risk of default." *Econometrica*, 68, 775–797.

Anand, Amber, Carsten Tanggaard, and Daniel G Weaver (2009), "Paying for market quality." *Journal of Financial and Quantitative Analysis*, 44, 1427–1457.

Armstrong, Mark (2006), "Competition in two-sided markets." *The RAND Journal of Economics*, 37, 668–691.

Armstrong, Mark and Jidong Zhou (2015), "Search deterrence." *The Review of Economic Studies*, 83, 26–57.

Awaya, K. Iwahashi, Y. and M. Watanabe (2018a), "Speculative bubbles in an economy with flippers." *mimeo*.

Awaya, Z. Chen, Y. and M. Watanabe (2018b), "Intermediation and reputation." *mimeo*.

Baker, George P and Brian J Hall (2004), "Ceo incentives and firm size." *Journal of Labor Economics*, 22, 767–798.

Baye, Michael R and John Morgan (2001), "Information gatekeepers on the internet and the competitiveness of homogeneous product markets." *American Economic Review*, 91, 454–474.

Biglaiser, Gary (1993), "Middlemen as experts." *The RAND journal of Economics*, 212–223.

Bloch, Francis and Harl Ryder (2000), "Two-sided search, marriages, and matchmakers." *International Economic Review*, 41, 93–116.

Böhm, Michael J, Daniel Metzger, and Per Strömberg (2015), "'since you're so rich, you must be really smart': Talent and the finance wage premium."

Burdett, Kenneth, Shouyong Shi, and Randall Wright (2001), "Pricing and matching with frictions." *Journal of Political Economy*, 109, 1060–1085.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin (2006), "Wage bargaining with on-the-job search: Theory and evidence." *Econometrica*, 74, 323–364.

Caillaud, Bernard and Bruno Jullien (2001), "Competing cybermediaries." *European Economic Review*, 45, 797–808.

Caillaud, Bernard and Bruno Jullien (2003), "Chicken & egg: Competition among intermediation service providers." *RAND journal of Economics*, 309–328.

Coles, Jeffrey L, Naveen D Daniel, and Lalitha Naveen (2006), "Managerial incentives and risk-taking." *Journal of financial Economics*, 79, 431–468.

Coles, Jeffrey L, Naveen D Daniel, and Lalitha Naveen (2013), "Calculation of compensation incentives and firm-related wealth using execucomp: Data, program, and explanation."

Condorelli, Daniele, Andrea Galeotti, and Vasiliki Skreta (2018), "Selling through referrals." *Journal of Economics & Management Strategy*, 27, 669–685.

Conroy, Robert M and Robert L Winkler (1986), "Market structure: The specialist as dealer and broker." *Journal of Banking & Finance*, 10, 21–36.

Cornelli, Francesca, Zbigniew Kominek, and Alexander Ljungqvist (2013), "Monitoring managers: Does it matter?" *The Journal of Finance*, 68, 431–481.

Custódio, Cláudia, Miguel A Ferreira, and Pedro Matos (2013), "Generalists versus specialists: Lifetime work experience and chief executive officer pay." *Journal of Financial Economics*, 108, 471–492.

Damiano, Ettore and Hao Li (2008), "Competing matchmaking." *Journal of the European Economic Association*, 6, 789–818.

Dasgupta, Partha and Eric Maskin (1986), "The existence of equilibrium in discontinuous economic games, ii: Applications." *Review of Economic Studies*, 53, 27–41.

Dasgupta, Partha, Eric Maskin, et al. (1986), "The existence of equilibrium in discontinuous economic games, i: Theory." *Review of Economic Studies*, 53, 1–26.

De Fraja, Gianni and József Sákovics (2012), "Exclusive nightclubs and lonely hearts columns: non-monotone participation in optional intermediation." *Journal of Economic Behavior & Organization*, 84, 618–632.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R Wildenbeest (2012), "Testing models of consumer search using data on web browsing and purchasing behavior." *American Economic Review*, 102, 2955–80.

Dittmann, Ingolf and Ernst Maug (2007), "Lower salaries and no options? on the optimal structure of executive pay." *The Journal of Finance*, 62, 303–343.

Dittmann, Ingolf, Ernst Maug, and Oliver Spalt (2010), "Sticks or carrots? optimal ceo compensation when managers are loss averse." *The Journal of Finance*, 65, 2015–2050.

Dittmann, Ingolf, Ko-Chia Yu, and Dan Zhang (2017), "How important are risk-taking incentives in executive compensation?" *Review of Finance*, 21, 1805–1846.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen (2005), "Over-the-counter markets." *Econometrica*, 73, 1815–1847.

Edelman, Benjamin, Julian Wright, et al. (2015), "Price restrictions in multi-sided platforms: Practices and responses." *Harvard Business School NOM Unit Working Paper*.

Edmans, Alex and Xavier Gabaix (2011), "The effect of risk on the ceo market." *The Review of Financial Studies*, 24, 2822–2863.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter (2017), "Executive compensation: A survey of theory and evidence." Technical report, National Bureau of Economic Research.

Edmans, Alex, Xavier Gabaix, and Augustin Landier (2009), "A multiplicative model of optimal ceo incentives in market equilibrium." *The Review of Financial Studies*.

Eisfeldt, Andrea L and Camelia M Kuhnen (2013), "Ceo turnover in a competitive assignment framework." *Journal of Financial Economics*, 109, 351–372.

Fama, Eugene F (1980), “Agency problems and the theory of the firm.” *Journal of political economy*, 88, 288–307.

Farhi, Emmanuel and Iván Werning (2013), “Insurance and taxation over the life cycle.” *Review of Economic Studies*, 80, 596–635.

Fingleton, John (1997), “Competition among middlemen when buyers and sellers can trade directly.” *The journal of industrial economics*, 45, 405–427.

Frydman, Carola (2005), “Rising through the ranks. the evolution of the market for corporate executives, 1936-2003.” *Working Paper*.

Frydman, Carola and Raven E Saks (2010), “Executive compensation: A new view from a long-term perspective, 1936–2005.” *The Review of Financial Studies*, 23, 2099–2138.

Gabaix, Xavier and Augustin Landier (2008), “Why has ceo pay increased so much?” *The Quarterly Journal of Economics*, 123, 49–100.

Galeotti, Andrea and José Luis Moraga-González (2009), “Platform intermediation in a market for differentiated products.” *European Economic Review*, 53, 417–428.

Gautier, Pieter A, Bo Hu, and Makoto Watanabe (2018), “Marketmaking middlemen.” *CEPR Discussion Paper*.

Gautier, Pieter A and Coen N Teulings (2015), “Sorting and the output loss due to search frictions.” *Journal of the European Economic Association*, 13, 1136–1166.

Gautier, Pieter A, Coen N Teulings, and Aico Van Vuuren (2010), “On-the-job search, mismatch and efficiency.” *The Review of Economic Studies*, 77, 245–272.

Gayle, George-Levi, Limor Golan, and Robert A Miller (2015), “Promotion, turnover, and compensation in the executive labor market.” *Econometrica*, 83, 2293–2369.

Gayle, George-Levi and Robert A Miller (2009), “Has moral hazard become a more important factor in managerial compensation?” *American Economic Review*, 99, 1740–69.

Gehrig, Thomas (1993), “Intermediation in search markets.” *Journal of Economics & Management Strategy*, 2, 97–120.

Geromichalos, Athanasios and Kuk Mo Jung (2018), “An over-the-counter approach to the forex market.” *International Economic Review*, 59, 859–905.

Giannetti, Mariassunta (2011), “Serial ceo incentives and the structure of managerial contracts.” *Journal of Financial Intermediation*, 20, 633–662.

Gibbons, Robert and Kevin J Murphy (1992), “Optimal incentive contracts in the presence of career concerns: Theory and evidence.” *Journal of political Economy*, 100, 468–505.

Grochulski, Borys and Yuzhe Zhang (2017), “Market-based incentives.” *International Economic Review*, 58, 331–382.

Guerrieri, Veronica, Robert Shimer, and Randall Wright (2010), “Adverse selection in competitive search equilibrium.” *Econometrica*, 78, 1823–1862.

Hagiu, Andrei (2006), “Pricing and commitment by two-sided platforms.” *The RAND Journal of Economics*, 37, 720–737.

Hagiu, Andrei and Bruno Jullien (2011), “Why do intermediaries divert search?” *The RAND Journal of Economics*, 42, 337–362.

Hagiu, Andrei and Julian Wright (2013), “Do you really want to be an ebay?” *Harvard Business Review*, 91, 102–108.

Hagiu, Andrei and Julian Wright (2014), "Marketplace or reseller?" *Management Science*, 61, 184–203.

Hall, Brian J and Jeffrey B Liebman (1998), "Are ceos really paid like bureaucrats?" *The Quarterly Journal of Economics*, 113, 653–691.

Hall, Brian J and Kevin J Murphy (2000), "Optimal exercise prices for executive stock options." *American Economic Review*, 90, 209–214.

Harris, Milton and Bengt Holmstrom (1982), "A theory of wage dynamics." *The Review of Economic Studies*, 49, 315–333.

Hendershott, Terrence and Albert J Menkveld (2014), "Price pressures." *Journal of Financial Economics*, 114, 405–423.

Hermalin, Benjamin E (2005), "Trends in corporate governance." *The Journal of Finance*, 60, 2351–2384.

Holmström, Bengt (1999), "Managerial incentive problems: A dynamic perspective." *The review of Economic studies*, 66, 169–182.

Holzner, Christian and Makoto Watanabe (2016), "Understanding the role of the public employment agency." *Tinbergen Institute Discussion Paper 16-041/VII*.

Hopenhayn, Hugo A and Juan Pablo Nicolini (1997), "Optimal unemployment insurance." *Journal of political economy*, 105, 412–438.

Huson, Mark R, Robert Parrino, and Laura T Starks (2001), "Internal monitoring mechanisms and ceo turnover: A long-term perspective." *The Journal of Finance*, 56, 2265–2297.

Jensen, Michael C and Kevin J Murphy (1990), "Performance pay and top-management incentives." *Journal of political economy*, 98, 225–264.

Jenter, Dirk and Fadi Kanaan (2015), "Ceo turnover and relative performance evaluation." *The Journal of Finance*, 70, 2155–2184.

Johri, Alok and John Leach (2002), "Middlemen and the allocation of heterogeneous goods." *International Economic Review*, 43, 347–361.

Ju, Jiandong, Scott C Linn, and Zhen Zhu (2010), "Middlemen and oligopolistic market makers." *Journal of Economics & Management Strategy*, 19, 1–23.

Kaplan, Steven N and Bernadette A Minton (2012), "How has ceo turnover changed?" *International review of Finance*, 12, 57–87.

Khurana, Rakesh (2004), *Searching for a corporate savior: The irrational quest for charismatic CEOs*. Princeton University Press.

Kocherlakota, Narayana R (1996), "Implications of efficient risk sharing without commitment." *The Review of Economic Studies*, 63, 595–609.

Lagos, Ricardo and Guillaume Rocheteau (2007), "Search in asset markets: Market structure, liquidity, and welfare." *American Economic Review*, 97, 198–202.

Lagos, Ricardo and Guillaume Rocheteau (2009), "Liquidity in asset markets with search frictions." *Econometrica*, 77, 403–426.

Lagos, Ricardo, Guillaume Rocheteau, and Pierre-Olivier Weill (2011), "Crashes and recoveries in illiquid markets." *Journal of Economic Theory*, 146, 2169–2205.

Lagos, Ricardo and Shengxing Zhang (2016), "Monetary exchange in over-the-counter markets: A theory of speculative bubbles, the fed model, and self-fulfilling liquidity crises." *NBER working paper*.

Lamadon, Thibaut (2016), "Productivity shocks, long-term contracts and earnings dynamics." *manuscript, University of Chicago*.

Lentz, Rasmus (2014), "Optimal employment contracts with hidden search." Technical report, National Bureau of Economic Research.

Li, Yiting (1998), "Middlemen and private information." *Journal of Monetary Economics*, 42, 131–159.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin (2016), "Matching, sorting and wages." *Review of Economic Dynamics*, 19, 63–87.

Loertscher, Simon and Andras Niedermayer (2008), "Fee setting intermediaries: on real estate agents, stock brokers, and auction houses." Technical report.

Marcet, Albert and Ramon Marimon (2017), "Recursive contracts."

Margiotta, Mary M and Robert A Miller (2000), "Managerial compensation and the cost of moral hazard." *International Economic Review*, 41, 669–719.

Martijn Cremers, KJ and Yaniv Grinstein (2013), "Does the market for ceo talent explain controversial ceo pay practices?" *Review of Finance*, 18, 921–960.

Maskin, Eric (1986), "The existence of equilibrium with price-setting firms." *The American Economic Review*, 76, 382–386.

Masters, Adrian (2007), "Middlemen in search equilibrium." *International Economic Review*, 48, 343–362.

Mele, Antonio (2014), "Repeated moral hazard and recursive lagrangeans." *Journal of Economic Dynamics and Control*, 42, 69–85.

Menkveld, Albert J and Ting Wang (2013), "How do designated market makers create value for small-caps?" *Journal of Financial Markets*, 16, 571–603.

Menzio, Guido and Shouyong Shi (2010), "Block recursive equilibria for stochastic models of search on the job." *Journal of Economic Theory*, 145, 1453–1494.

Mittal, Hitesh (2008), "Are you playing in a toxic dark pool?: A guide to preventing information leakage." *The Journal of Trading*, 3, 20–33.

Moen, Espen R (1997), "Competitive search equilibrium." *Journal of political Economy*, 105, 385–411.

Murphy, Kevin J and Jan Zabojnik (2004), "Ceo pay and appointments: A market-based explanation for recent trends." *American economic review*, 94, 192–196.

Murphy, Kevin J and Jan Zabojnik (2007), "Managerial capital and the market for ceos."

Nimalendran, Mahendrarajah and Giovanni Petrella (2003), "Do thinly-tradedstocks benefit from specialist intervention?" *Journal of banking & finance*, 27, 1823–1854.

Nocke, Volker, Martin Peitz, and Konrad Stahl (2007), "Platform ownership." *Journal of the European Economic Association*, 5, 1130–1160.

Nosal, Ed, Yuet-yee Wong, and Randall Wright (2015), "More on middlemen: Equilibrium entry and efficiency in intermediated markets." *Journal of Money, Credit and Banking*, 47, 7–37.

Oyer, Paul (2004), "Why do firms use incentives that have no incentive effects?" *The Journal of Finance*, 59, 1619–1650.

Parker, Geoffrey G and Marshall W Van Alstyne (2005), "Two-sided network effects: A theory of information product design." *Management science*, 51, 1494–1504.

Peters, Florian S and Alexander F Wagner (2014), “The executive turnover risk premium.” *The Journal of Finance*, 69, 1529–1563.

Peters, Michael (1991), “Ex ante price offers in matching games non-steady states.” *Econometrica (1986-1998)*, 59, 1425.

Peters, Michael (2000), “Limits of exact equilibria for capacity constrained sellers with costly search.” *Journal of Economic Theory*, 95, 139–168.

Phelan, Christopher (1995), “Repeated moral hazard and one-sided commitment.” *Journal of Economic Theory*, 66, 488–506.

Phelan, Christopher and Robert M Townsend (1991), “Computing multi-period, information-constrained optima.” *The Review of Economic Studies*, 58, 853–881.

Pissarides, Christopher A (2000), *Equilibrium unemployment theory*. MIT press.

Postel-Vinay, Fabien and Jean-Marc Robin (2002), “Equilibrium wage dispersion with worker and employer heterogeneity.” *Econometrica*, 70, 2295–2350.

Remesal, Alvaro et al. (2018), “How important are dismissals in ceo incentives? evidence from a dynamic agency model.”

Rhodes, Andrew, Makoto Watanabe, Jidong Zhou, et al. (2017), “Multiproduct intermediaries and optimal product range.” Technical report, University Library of Munich, Germany.

Rochet, Jean-Charles and Jean Tirole (2003), “Platform competition in two-sided markets.” *Journal of the european economic association*, 1, 990–1029.

Rochet, Jean-Charles and Jean Tirole (2006), “Two-sided markets: a progress report.” *The RAND journal of economics*, 37, 645–667.

Rogerson, William P (1985), “Repeated moral hazard.” *Econometrica: Journal of the Econometric Society*, 69–76.

Rubinstein, Ariel and Asher Wolinsky (1987), “Middlemen.” *The Quarterly Journal of Economics*, 102, 581–593.

Rust, John and George Hall (2003), “Middlemen versus market makers: A theory of competitive exchange.” *Journal of Political Economy*, 111, 353–403.

Rysman, Marc (2009), “The economics of two-sided markets.” *Journal of economic perspectives*, 23, 125–43.

Securities and Exchange Commission (2010), “Securities exchange act release no. 34-61358, 75 fr 3594, concept release.”

Securities and Exchange Commission (2013), “Equity market structure literature review part i: Market fragmentation.” *US Securities and Exchange Commission*.

Securities and Exchange Commission (2014), “Equity market structure literature review part ii: High frequency trading.” *US Securities and Exchange Commission*.

Shevchenko, Andrei (2004), “Middlemen.” *International Economic Review*, 45, 1–24.

Spear, Stephen E and Sanjay Srivastava (1987), “On repeated moral hazard with discounting.” *The Review of Economic Studies*, 54, 599–617.

Spulber, Daniel F (1996), “Market making by price-setting firms.” *The Review of Economic Studies*, 63, 559–580.

Spulber, Daniel F (1999), *Market microstructure: intermediaries and the theory of the firm*. Cambridge University Press.

Stahl, Dale O (1988), "Bertrand competition for inputs and walrasian outcomes." *The American Economic Review*, 189–201.

Stone, Brad (2013), *The everything store: Jeff Bezos and the age of Amazon*. Random House.

Tauchen, George (1986), "Finite state markov-chain approximations to univariate and vector autoregressions." *Economics letters*, 20, 177–181.

Taylor, Lucian A (2010), "Why are ceos rarely fired? evidence from structural estimation." *The Journal of Finance*, 65, 2051–2087.

Tervio, Marko (2008), "The difference that ceos make: An assignment model approach." *American Economic Review*, 98, 642–68.

Teulings, Coen N and Pieter A Gautier (2004), "The right man for the job." *The Review of Economic Studies*, 71, 553–580.

Thomas, Jonathan and Tim Worrall (1988), "Self-enforcing wage contracts." *The Review of Economic Studies*, 55, 541–554.

Thomas, Jonathan and Tim Worrall (1990), "Income fluctuation and asymmetric information: An example of a repeated principal-agent problem." *Journal of Economic Theory*, 51, 367–390.

Townsend, Robert M (1982), "Optimal multiperiod contracts and the gain from enduring relationships under private information." *Journal of political Economy*, 90, 1166–1186.

Tsuyuhara, Kunio (2016), "Dynamic contracts with worker mobility via directed on-the-job search." *International Economic Review*, 57, 1405–1424.

Tuttle, Laura A (2014), "OTC trading: Description of non-ats otc trading in national market system stocks." *SSRN Working Paper 2412100*.

Varian, Hal R (1980), "A model of sales." *The American Economic Review*, 70, 651–659.

Venkataraman, Kumar and Andrew C Waisburd (2007), "The value of the designated market maker." *Journal of Financial and Quantitative Analysis*, 42, 735–758.

Wang, Cheng and Youzhi Yang (2011), "Termination of dynamic contracts in an equilibrium labor market model." *Journal of Economic Theory*, 146, 74.

Wang, Cheng and Youzhi Yang (2015a), "Equilibrium matching and termination." *Journal of Monetary Economics*, 76, 208–229.

Wang, Cheng and Youzhi Yang (2015b), "Outside opportunities and termination." *Games and Economic Behavior*, 91, 207–228.

Wang, Cheng and Youzhi Yang (2016), "Optimal ceo contracting with moral hazard and external shocks." *Working paper*.

Watanabe, Makoto (2010), "A model of merchants." *Journal of Economic Theory*, 145, 1865–1889.

Watanabe, Makoto (2018a), "Middlemen: A directed search equilibrium approach." *Tinbergen Institute working paper*.

Watanabe, Makoto (2018b), "Middlemen: The visible market-makers." *The Japanese Economic Review*, 69, 156–170.

Weill, Pierre-Olivier (2007), "Leaning against the wind." *The Review of Economic Studies*, 74, 1329–1354.

Weyl, E Glen (2010), "A price theory of multi-sided platforms." *American Economic Review*, 100, 1642–72.

Wright, Randall and Yuet-Yee Wong (2014), "Buyers, sellers, and middlemen: Variations on search-theoretic themes." *International Economic Review*, 55, 375–397.

Yavaş, Abdullah (1994), "Middlemen in bilateral search markets." *Journal of Labor Economics*, 12, 406–429.

Yavaş, Abdullah (1996), "Search and trading in intermediated markets." *Journal of Economics & Management Strategy*, 5, 195–216.

Zhu, Feng and Qihong Liu (2018), "Competing with complementors: An empirical look at amazon. com." *Strategic Management Journal*, 39, 2618–2642.