Middlemen: A Directed Search Equilibrium Approach

Makoto Watanabe∗
Universidad Carlos III de Madrid†

July 16, 2011

Abstract

This paper studies an intermediated market for a homogeneous good with middlemen who hold inventories of the good. Using a directed search approach, I investigate a steady state equilibrium. Under market frictions and competition, the ask price of middlemen includes a retail premium for the immediacy service to buyers and the bid price includes a wholesale premium charged to sellers for guaranteed sale. It is shown that the middlemen enhance the economy-wide efficiency, and that the bid-ask spread can be non-monotone in the middlemen’s inventory capacity. The simple model provided here has wide applicability and offers economic insights into many empirically relevant forms of middlemen, especially in the context of financial intermediaries, international trade and retail stores.

Keywords: Directed Search, Intermediation, Inventory holdings

JEL Classification Number: D4, F1, G2, L1, L8, R1

∗I am indebted to Melvyn G. Coles for his many insightful comments and suggestions. Thanks are also due to V. Bhaskar, Gary Biglaiser, Kenneth Burdett, Pieter Gautier, Bruno Jullien, Nobu Kiyotaki, John Moore, Georg Nöldeke, Larry Samuelson, Randy Wright, an editor of the journal, George-Marios Angeletos, three anonymous referees, and participants at various seminars and conferences for useful comments and discussions on earlier drafts. Any remaining errors are my own. Financial support from the Spanish government in the form of research grant, ECO2009-10531, and research fellowship, Ramon y Cajal, is gratefully acknowledged. An earlier version of this article circulated under the title “Middlemen: the bid-ask spread”.

†Correspondence: Makoto Watanabe, Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, Getafe (Madrid) 28903, SPAIN. E-mail: mwatanab@eco.uc3m.es
1 Introduction

Despite recent emergence of electronic commerce and Internet technologies, which have substantially reduced the costs required to finding a buyer or a seller, middlemen still prevail in most markets (e.g. retailers, wholesalers, trading entrepreneurs, dealers or brokers of services and durable goods and assets). This suggests the existence of a premium for middlemen. For instance, while individual buyers of a used-car, durable good or financial asset could make a direct contact with a potential seller, or advertise on a website, the stocks maintained by middlemen are valuable to those who wish to save time and efforts that would have to be spent on searching in private. Similarly, while importers of agricultural products could search local producers for available supply, the middlemen, who are specialized in aggregating individual supply of peasants, managing inventories and breaking bulk shipments, have a relative advantage over such a transaction, which allows them to charge margins for intermediating trade.

The essential feature captured by the above observation is that middlemen can mitigate market frictions and, at the same time, exercise market power. While the former role of middlemen should help improve the resource allocation, it is not obvious how it is related to the latter behavior of middlemen. Indeed, it is important to understand such a connection for accessing the functioning of many real-life markets with middlemen. In financial markets, the liquidity provided by middlemen is measured by the bid-ask spread as well as by the trading delay and the trade volume. In other markets, the excessive level of middlemen’s price markups, as well as the facilitation of intermediation process,

1According to the “Used Car Market Report in 2007,” produced by Manheim Consulting, about 65.7% of the 42.6 million U.S. used car sales in 2006 were from dealers. While the private party sales has increased by 22.4% during the past 11 survey years, the net departmental profit of average dealer has increased at least by 10%. The average transacted price of used vehicle was $10,875 for franchised dealers, $8,675 for independent dealers, and $4,450 for private sellers. In 9 of the 11 years, the net return on used vehicle sales has exceeded that of new vehicles for franchised dealers (e.g., the average net profit per unit was $14 for new cars and $230 for used cars in 2006). They noted that a major reason for these differences was that dealers had greater control over their used vehicle inventory as a result of an active wholesale market.

2On-shelf unavailability (or stockout) is a major cause of customer dissatisfaction (Andersen 1996). For electronics, toys, or video rentals, customers typically spend time and effort in traveling to the store and locating their items on the shelf. Even for online purchases, navigating through webpages and submitting orders could require substantial time and effort (Hann and Terwiesch, 2003). Customers even report stockouts as their chief complaint in the case of mail-order catalog companies (Fitzsimons, 2000).

3According to Feenstra and Hanson (2004), markups on re-exporters of Chinese goods in Hong Kong were 12% of its GDP, while manufacturing accounted for only 6%. The average markup rate accruing to Hong Kong intermediaries on re-exported Chinese goods was 24%. Importantly, as for the patterns of individual trade, it is rarely observed that small exporters are matched with small importers, indicating that small-scaled buyers and sellers find it difficult to identify each other. See further discussion in Section 6.

4When Stigler (1964) and Demsetz (1968) analyzed the role of dealers in securities markets, they viewed dealers as the suppliers of immediacy to the market. To determine the costs to buyers and sellers of using the NYSE to contract with each other, Demsetz captured the compensation to the dealer for the provision of the service by the bid-ask spread. He defined it as “the markup that is paid for predictable immediacy of exchange in organized markets; in other markets, it is the inventory markup of retailer or wholesaler” (p. 36).
continue to be a major policy concern, typically in developing countries, regarding the malfunctioning of the distribution system.\textsuperscript{5}

This paper proposes a new framework based on a standard directed search equilibrium\textsuperscript{6} that allows us to study explicitly the relationship between the matching capability of middlemen and their market power. The key ingredient is the inventory holdings of middlemen. In my framework, middlemen are specialized in buying and selling, and their inventory capacity keeps them less likely to experience a stockout. The gap between the buying price and the selling price of middlemen determines a price markup called the bid-ask spread that can be influenced by their inventory.

To be specific, I consider an infinite horizon model in which each period consists of two sub-periods. In the first sub-period, retail markets are open where buyers can search for a homogeneous good. In the second sub-period, wholesale markets are open where middlemen can acquire their goods from sellers. Retail markets are operated by both middlemen and sellers, while wholesale markets are operated only by sellers. To ensure analytical tractability, I assume that the retail markets are frictional but the wholesale markets are competitive. This assumption guarantees the middlemen’s inventory to be deterministic and identical to individuals across all the periods.\textsuperscript{7} With a constant stock of buyers and sellers, I focus my attention on a steady state.

In the frictional retail markets, middlemen have an advantage over sellers in the inventory capacity. In a directed search equilibrium, this implies that buyers are attracted to middlemen, who can provide them with a lower stockout probability for which a premium can be charged. In the wholesale markets, however, middlemen are always on the short side given a sufficiently large population of sellers. Thus, middlemen can restock their inventories from individual sellers at the sellers’ continuation value of holding onto the good. At such an equilibrium, the ask (retail) price of middlemen includes a premium for immediacy service to buyers, and the bid (wholesale) price includes a premium charged to sellers for guaranteed sale. The bid-ask spread decreases with the number of middlemen because it leads to

\textsuperscript{5}Middlemen are often blamed for siphoning gains from trade away from less developed countries towards developed countries (see, for example, Oxfam (2002)). McMillan, Welch and Rodrik (2004) report that cashew growers in Mozambique only receive 40 – 50\% of the border price. They observe that farmers’ incomes are depressed not only by transport and marketing costs, but also by the market power exercised by middlemen.


\textsuperscript{7}This assumption is related to the one adopted in a monetary model of Lagos and Wright (2005) who establish a monetary equilibrium with a degenerate distribution of divisible money holdings in the presence of market frictions that could potentially lead to a complicated stochastic evolution of individual money balances.
more competitive retail markets (i.e., tighter markets) as is standard in the directed/competitive search literature.

An increase in inventories of middlemen generates two non-trivial effects on the bid-ask spread. On the one hand, a larger inventory maintained by middlemen induces more buyers to search in the middlemen’s market rather than in the sellers’ market. This effect implies a larger retail premium and a smaller wholesale premium for middlemen. On the other hand, as the inventory of middlemen grows it is less likely that an individual middleman will run out of stock. This effect implies a downward pressure on the retail premium (and hence on the bid-ask spread), since buyers know that the unsold inventories yield a lower value to the middleman. These conflicting effects cause a non-monotonic response of the bid-ask spread to changes in the inventory: the spread increases with relatively low inventories and decreases with relatively high inventories. The stockout probability is initially high (low) in the former (latter) situation, thereby buyers are willing to pay a higher premium for a larger inventory of middlemen when the initial capacity is reduced.

As an extension of the analysis, I allow for the endogenous determination of the number of middlemen by free entry. The number of middlemen can be non-monotone in the inventory, and the extensive margin can matter for the determination of the bid-ask spread. In particular, the spread can be non-monotonic when the inventory holding costs are convex: it decreases in relatively low inventories but increases in relatively high inventories reflecting the rapidly changing number of active middlemen. Finally, the middlemen in this economy are efficiency enhancing, and the middlemen’s inventory capacity can improve the matching efficiency and the total welfare of buyers and sellers.

The economic insights obtained from the overall analysis can be summarized as follows. The middlemen studied here are engaged in the activity that people view with low regard – buying from sellers at their reservation price and selling to buyers at a premium – even though they can enhance the economy-wide efficiency with their inventory capacity. When the initial capacity is relatively high, a larger inventory capacity of middlemen leads to a higher total matching rate; a higher total welfare; more competitive markets and a reduction in the bid-ask spread. As it holds true with fixed supply in the middlemen’s market, this result implies that few middlemen, each with many inventories lead to a higher total matching rate, a higher total welfare, and a smaller bid-ask spread compared to having many middlemen, each with few inventories. However, the latter consequence of the bid-ask spread can be reversed when the initial capacity of middlemen is relatively low, in which case buyers would be prepared to pay a large premium for the increased probability of getting served.
The simple model presented here is admittedly stylized and does not aspire to capture the precise workings of any particular market. Nevertheless, it offers wide applicability and sheds light on many empirically relevant forms of middlemen, as will be discussed in detail in Section 6, especially in the context of financial intermediaries, international trade and retail stores.

The rest of the paper is organized as follows. Section 2 presents the basic setup and shows the existence and uniqueness of a steady state equilibrium. Section 3 provides a characterization of the bid-ask spread of middlemen. Section 4 extends the analysis to allow for the free entry of middlemen. Section 5 investigates the matching efficiency and the welfare. Section 6 discusses the related literature and further implications of the model for individual markets. Section 7 concludes. All the proofs are in the Appendix.

2 Model

Consider an economy that has a continuum of buyers, sellers and middlemen, indexed \( b, s \) and \( m \), respectively. Time is discrete and lasts forever. Each period is divided into two subperiods. During the first subperiod, a retail market is open for a homogeneous good to buyers. This retail market is subject to search frictions which I describe in detail below. Each buyer wishes to obtain one unit of the good while each seller holds \( k_s = 1 \) unit, which is her endowment (or produced with zero cost), and each middleman holds \( k_m \geq 1 \) units of the good. The capacity of suppliers \( k_i \) is exogenously given, for both \( i = s, m \). If a buyer successfully purchases the good at a price \( p \), then he obtains a period utility of \( 1 - p \) and exits the market. Otherwise, he receives zero utility in that period and enters the next period. A seller or middleman who sells \( z \) units at a price \( p \) makes a profit of \( zp \) per period during the first subperiod.

Once the retail market is closed, another market opens during the second subperiod. This market is a wholesale market, where middlemen can restock their units to sell in the future retail markets and the sellers who still hold the good can sell to one of the middlemen. In contrast to the retail market, there are no search frictions in the wholesale market. The period is then repeated infinitely. While buyers and sellers leave the market once they complete the trade, middlemen are active in all the periods. Agents discount future payoffs at a rate \( \beta \in [0, 1) \) across periods, but there is no discounting between the two sub-periods.

The environment in each retail market is the same as in the standard competitive/directed search models (see, for example, Accemoglu and Shimer (1999), Burdett, Shi, and Wright (2001), Montgomery
It can be described as a simple two-stage game. In the first stage, sellers and middlemen simultaneously post a price which they are willing to sell at. Observing the prices, buyers simultaneously decide which seller or middleman to visit in the second stage. If more buyers visit a seller or middleman than its capacity (i.e., demand greater than supply), then the good or goods are allocated randomly. Assuming buyers cannot coordinate their actions over which seller or middleman to visit, I study a symmetric equilibrium where all buyers use the identical mixed strategy for any configuration of the announced prices. Further, I focus my attention on a steady-state equilibrium where entry of buyers and sellers are exogenous, and the population of agents and the capacity of middlemen $k_m$ are constant over time – if a buyer (or seller) leaves the market, then it is assumed that another buyer (or seller) enters the market so that the total entry of a buyer (or seller) equals to the total exit of a buyer (or seller), leading to the constant stock of buyers and sellers over time.\footnote{With a homogeneous agents setup, this is the simplest way to describe a market equilibrium. With heterogeneous agents setups, the other specifications of exogenous inflows of agents are considered in the marriage matching and labor force mobility literature. See Burdett and Coles (1999).} In such an equilibrium, all sellers post the identical price $p_s$ and all middlemen post the identical price $p_m$ every period.

In any given period each seller or middleman is characterized by an expected queue of buyers, denoted by $x$. The number of buyers visiting a given seller or middleman who has expected queue $x$ is a random variable, denoted by $n$, which has the Poisson distribution $\text{Prob}(n=k) = \frac{e^{-x}x^k}{k!}$. In a symmetric equilibrium where $x_i$ is the expected queue of buyers at $i$, each buyer visits some seller (and some middleman) with probability $Sx_s$ (and $Mx_m$), where $S$ ($M$) denotes the measure of sellers (middlemen). The measure of buyers is normalized to one so that they should satisfy the adding-up restriction,

$$Mx_m + Sx_s = 1,$$

requiring that the number of buyers visiting individual sellers and middlemen be summed up to the total population of buyers. Each period, each buyer visiting a supplier with capacity $k_i$ get served with probability $\eta(x_i, k_i)$.

In steady state, middlemen should restock the identical units from the sellers for all the periods so that they hold $k_m \geq 1$ units at the beginning of every period. It is sustainable only when there exist a sufficiently large number of sellers, relative to the total demand by middlemen, in the wholesale
Throughout the paper, I assume 

\[ S \geq 1, \]

so that the total demand is below the total supply in the wholesale market each period for all \( k_m \geq 1. \)

Given that the wholesale market is Walrasian, this implies that the wholesale price is set equal to the reservation value of sellers, who are always on the long side, at which the sellers are indifferent between operating in a private market and in a wholesale market. The latter property further implies that the timing of events is irrelevant here – the entire analysis remains unchanged with an alternative setting in which the wholesale market occurs before the retail market in any given period.

**Buyers’ directed search** Assuming for the moment the existence of a symmetric equilibrium, the following lemma gives the buyer’s probability of being served by a supplier who has capacity \( k_i \), denoted by \( \eta(x_i, k_i) \). The derivation is given in Watanabe (2006) (see also Watanabe (2010) for the finite agents version).

**Lemma 1** Given \( x_i > 0 \) and \( k_i \geq 1 \), the buyer’s probability of obtaining a good from a supplier \( i \) that has \( k_i \) units of the goods, \( \eta(x_i, k_i) \), is given by the following closed form expression.

\[
\eta(x_i, k_i) = \frac{\Gamma (k_i, x_i)}{\Gamma (k_i)} + \frac{k_i}{x_i} \left( 1 - \frac{\Gamma (k_i + 1, x_i)}{\Gamma (k_i + 1)} \right)
\]

where \( \Gamma (k) = \int_0^\infty t^{k-1}e^{-t}dt \) and \( \Gamma (k, x) = \int_x^\infty t^{k-1}e^{-t}dt \). \( \eta(\cdot) \) is strictly decreasing (increasing) in \( x_i \) (in \( k_i \)) and satisfies \( \eta(x_s, 1) = (1 - e^{-x_s})/x_s \).

Given \( \eta(\cdot) \) described above, I now characterize the expected queue of buyers. In any equilibrium where \( V_b \) is the value of being a buyer, should a seller or a middleman deviate by setting price \( p \) in a

\[ Mx_m \eta(x_m, k_m) < S(1 - x_s \eta(x_s, 1)), \]

where the L.H.S. is the total restocking units of middlemen (total demand) and the R.H.S. is the total available units by sellers (total supply). Notice that the total number of realized sales, given by \( Mx_m \eta(x_m, k_m) + Sx_s \eta(x_s, 1) \), is strictly below the population of buyers, which is normalized to one. This implies that the above inequality holds true when \( S \geq 1 \). The steady state equilibrium established below can exist even for low values of \( S \) but with restricted values of \( k_m \).

Watanabe (2010) studies the case where demand and supply in the wholesale market are balanced and the wholesale price equals to zero (with the infinite discounting of agents).

---

\(^9\) Below, it is assumed that the capacity constraint of suppliers is binding for a middleman, so that he accumulates inventories up to the limit \( k_m \). In financial markets, it can be justified by the reserve/capacity requirement. In other markets, one can assume a significantly high costs of maintaining an unfilled capacity or displaying unfilled shelves due to reputation concerns. To endogenize \( k_m \) is technically more involved and will be left for future research.

\(^10\) Implicit here is that middlemen are born with \( k_m \) units of initial endowment at period 0, but this assumption is not very important and is just to simplify the presentation. Let \( x_i \eta(x_i, k_i) \) be the realized number of sales by a supplier with capacity \( k_i \). If the following inequality holds true then middleman are on the short side of the wholesale market each period:

\[ Mx_m \eta(x_m, k_m) < S(1 - x_s \eta(x_s, 1)), \]

where the L.H.S. is the total restocking units of middlemen (total demand) and the R.H.S. is the total available units by sellers (total supply). Notice that the total number of realized sales, given by \( Mx_m \eta(x_m, k_m) + Sx_s \eta(x_s, 1) \), is strictly below the population of buyers, which is normalized to one. This implies that the above inequality holds true when \( S \geq 1 \). The steady state equilibrium established below can exist even for low values of \( S \) but with restricted values of \( k_m \). The proof of this statement is in Theorem 2 in Watanabe (2010) for \( \beta = 0 \), and is available upon request for \( \beta \in [0, 1) \). Watanabe (2010) studies the case where demand and supply in the wholesale market are balanced and the wholesale price equals to zero (with the infinite discounting of agents).
period, the expected queue of buyers denoted by $x$ satisfies

$$V^b = \eta(x, k_i) (1 - p) + (1 - \eta(x, k_i)) \beta V^b.$$  

(2)

A buyer choosing $p$ is served with probability $\eta(x, k_i)$ in which case he obtains per-period utility $1 - p$ (and exits the market). If not served by the seller or middleman, the buyer enters the next period and obtains the discounted value $\beta V^b$. The situation is the same for all the other buyers. (2) is an implicit equation that determines $x = x(p, k_i | V^b) \in (0, \infty)$ as a strictly decreasing function of $p$ given $\beta, k_i$ and $V^b$.

**Optimal pricing** Given the directed search of buyers described above, the next step is to characterize the equilibrium retail prices. Denote by $V^s$ the value of being a seller. As middlemen restock at price $\beta V^s$ in the wholesale market, where the sellers are just indifferent between selling (leaving the market) and not selling (remaining in the market), in any equilibrium where $V^b$ and $V^s$ are the value of a buyer and a seller, respectively, the optimal price of a seller who has a capacity $k_s = 1$, denoted by $p_s(V^b, V^s)$, satisfies

$$p_s(V^b, V^s) = \arg\max_p [x(p, 1 | V^b) \eta(x(p, 1 | V^b), 1)p + (1 - x(p, 1 | V^b)) \eta(x(p, 1 | V^b), 1)) \beta V^s]$$

as the seller sells its good at price $p$ with probability $x(p, 1) \eta(x(p, 1), 1)$, and is otherwise guaranteed $\beta V^s$ in the wholesale market (and exits the market), or in the subsequent retail markets. Similarly, the optimal price of a middleman who has capacity $k_m \geq 1$ is given by

$$p_m(V^b, V^s) = \arg\max_p [(p - \beta V^s)x(p, k_m | V^b) \eta(x(p, k_m | V^b), k_m)]$$

where $x(p, k_m) \eta(x(p, k_m), k_m)$ represents the expected number of sales, and the middleman restocks at price $\beta V^s$ in the wholesale market.

Substituting out $p$ using (2), $p = 1 - \frac{1 - \eta(x, 1)}{\eta(x, 1)} V^b$, the objective function of a seller or a middleman denoted by $\Pi_s(x)$ or $\Pi_m(x)$ can be written as

$$\Pi_s(x) = x \eta(x, 1) - x(1 - \beta(1 - \eta(x, 1))) V^b + (1 - x \eta(x, 1)) \beta V^s$$

$$\Pi_m(x) = x \eta(x, k_m) - x(1 - \beta(1 - \eta(x, k_m))) V^b - x \eta(x, k_m) \beta V^s$$

where $x = x(p, k_i | V^b)$ satisfies (2). The first-order condition is

$$\frac{\partial \Pi_s(x)}{\partial x} = \eta(x, k_i) + x \frac{\partial \eta(x, k_i)}{\partial x} (1 - \beta(V^b + V^s)) - (1 - \beta(1 - \eta(x, 1))) V^b - \beta V^s = 0$$

8
for both \( i = s, m \). Denoting by

\[
N \equiv 1 - \beta (V^b + V^s) < 1
\]

the net trading surplus of the supplier \( i \) (against buyers), and rearranging the first order condition above using (2), or

\[
V^b = \frac{\eta_i(1-p)}{1-\beta(1-\eta_i)}, \text{ and}
\]

one can obtain the optimal price of the seller (if \( i = s \)) or the middleman (if \( i = m \)),

\[
p_i(V^b, V^s) = \beta V^s + \varphi^i(x, k_i) N
\]

where

\[
\varphi^i(x, k_i) \equiv -\frac{\partial \eta_i(x, k_i)}{\eta_i(x, k_i)} = \frac{k_i}{x^2} \left( 1 - \frac{\Gamma(k_i + 1, x)}{\Gamma(k_i + 1)} \right),
\]

is the elasticity of the matching rate of buyers. For the analysis that follows below, it is worth mentioning here that the term

\[
\text{Prob.}(n > k_i) = \sum_{n=k_i+1}^{\infty} \frac{e^{-x_i} x_i^n}{n!} = 1 - \frac{\Gamma(k_i + 1, x_i)}{\Gamma(k_i + 1)}
\]

represents the stock-out probability of an individual supplier \( i \), which is the probability that the number of buyers visiting the seller or middleman \( n \) is strictly greater than its capacity \( k_i \). The stockout probability is decreasing in the capacity \( k_i \) and is increasing in the number of buyers to the supplier \( i, x_i \). As it turns out, the relative size of these effects shapes critically the behavior of the stockout probability and the equilibrium price (and the bid-ask spread).

Existence and uniqueness of steady-state equilibrium

**Definition 1** Given the population parameters \( S, M \), the initial endowments \( k_i, i = s, m \), and the discount factor \( \beta \), a steady state equilibrium is a set of expected values \( V^j \) for \( j = b, s, m \), and market outcomes \( x_i, p_i \) for \( i = s, m \) such that:

1. Buyers’ directed search satisfies (1) and (2):

\[
\frac{\partial^2 \Pi_i(x)}{\partial x^2} = -\frac{x_i^{k_i-1} e^{-x_i}}{\Gamma(k_i)} (1 - \beta (V^b + V^s)) < 0.
\]

12The second equation follows from the series definition of cumulative gamma function, \( \sum_{n=0}^{k} \frac{e^{-x_i} x_i^n}{n!} = \frac{\Gamma(k+1, x_i)}{\Gamma(k+1)} \).
2. Sellers’ and middlemen’s retail prices satisfy the first-order conditions (3) for \(i = s, m\);

3. Middlemen restock their inventories from sellers in the wholesale market at price \(\beta V^s\), and hold \(k_m \geq 1\) units in the retail market;

4. Agents have rational expectations.

The analysis above has established the equilibrium prices \(p_i(V^b, V^s)\) given \(V^b\) and \(V^s\). Equilibrium implies buyers are indifferent between any of the individual suppliers \(i = s, m\), leading to

\[
V^b = \eta(x_s, 1)(1 - p_s) + (1 - \eta(x_s, 1))\beta V^b
\]

(4)

\[
= \eta(x_m, k_m)(1 - p_m) + (1 - \eta(x_m, k_m))\beta V^b
\]

(5)

where \(x_i = x(p_i, k_i \mid V^b)\) is the equilibrium queue of buyers at \(i = s, m\). Buyers then successfully purchase the good from the seller or middleman with probability \(\eta(x_i, k_i)\) each period. The value of sellers and middlemen are given by

\[
V^s = x_s\eta(x_s, 1)p_s + (1 - x_s\eta(x_s, 1))\beta V^s
\]

(6)

\[
V^m = x_m\eta(x_m, k_m)(p_m - \beta V^s)/(1 - \beta),
\]

(7)

respectively. Middlemen restock at wholesale price \(\beta V^s\) each period and sellers are indifferent between selling and not selling at that price. To solve for the equilibrium, it is important to observe that indifference conditions (4) and (5) can be reduced to the following simple form: applying (3) for \(i = s\) to (4) with a rearrangement,

\[
\frac{V^b}{1 - \beta V^s} = \frac{\eta(x_s, 1)(1 - \varphi^s(x_s, 1))}{1 - \beta(1 - \eta(x_s, 1)) - \beta\eta(x_s, 1)\varphi^s(x_s, 1)} = \frac{e^{-x_s}}{1 - \beta(1 - e^{-x_s})};
\]

similarly, applying (3) for \(i = m\) to (5) with a rearrangement,

\[
\frac{V^b}{1 - \beta V^s} = \frac{\eta(x_m, k_m)(1 - \varphi^m(x_m, k_m))}{1 - \beta(1 - \eta(x_m, k_m)) - \beta\eta(x_m, k_m)\varphi^m(x_m, k_m)} = \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \frac{1}{1 - \beta \left(1 - \Gamma(k_m, x_m) \right)};
\]

these two equations imply

\[
\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} = e^{-x_s}.
\]

(8)

The adding-up restriction (1) and the indifference condition (8) identify an equilibrium allocation \(x_s, x_m > 0\).
Theorem 1 (Steady state equilibrium) Given $S \in [1, \infty)$, a steady state equilibrium exists and is unique for all $\beta \in [0, 1)$ and $M \in (0, \infty)$, satisfying $V^b \in (0, 1)$, $x_s \in (0, 1)$, $x_m \in (0, \infty)$, $p_i \in (0, 1)$, and $V^i \in (0, k_i)$, $i = s, m$.

The equilibrium allocation of buyers $x_s, x_m > 0$ is determined irrespective of the discount factor $\beta$ each period by (1) and (8). Therefore, the results obtained in Watanabe (2006, 2010), where the case $\beta = 0$ is investigated, are applicable here for all $\beta \in [0, 1)$:

1. For $k_m = 1$ all sellers and middlemen receive the identical number of buyers $x_s = x_m$ and post the identical price $p_s = p_m$;

2. An increase in the capacity of middlemen $k_m$ induces more buyers to visit middlemen and fewer buyers to visit sellers, resulting in an increase in $x_m$ and a decrease in $x_s$;

3. An increase in the proportion of sellers $S$ or middlemen $M$ decreases $x_s, x_m$.

As a lower $x_s$ implies a lower value of sellers $V^s$ and thus a lower wholesale price $\beta V^s$, the above results further lead to Corollary 1.

Corollary 1 (Wholesale (bid) price) For all $\beta \in [0, 1)$, an increase in the population of sellers $S$ or middlemen $M$, or in the capacity of middlemen $k_m$ leads to a lower wholesale (bid) price $\beta V^s$.

3 Bid-ask spread

In this section, I characterize the behaviors of the bid-ask spread of middlemen, i.e., the difference between the ask price (retail price) and the bid price (wholesale price). It is given by

$$p_m - \beta V^s = \varphi^m(x_m, k_m)N.$$ 

Here, $\varphi^m(x_m, k_m)$ represents the middleman’s share of the net trading surplus,

$$N \equiv 1 - \beta(V^b + V^s) = \frac{1 - \beta}{1 - \beta x_s e^{-x_s}},$$

where $V^b = \frac{e^{-x_s}}{1 - \beta x_s e^{-x_s}}$ and $V^s = \frac{1 - e^{-x_s} - x_se^{-x_s}}{1 - \beta x_s e^{-x_s}}$ (see the proof of Theorem 1). It is worth noting that in Rubinstein and Wolinsky (1987), the trading surplus is divided via Nash bargaining and the inventory holdings of middlemen is restricted to one unit. In their model, the surplus share is constant (and equals to $\frac{1}{2}$) and the spread is given by $\frac{1}{2}N$, whereas in my model the surplus share $\varphi^m(x_m, k_m)$ is an endogenous object and the inventory $k_m$ can influence both $\varphi^m(x_m, k_m)$ and $N$. 

11
I begin by showing that the usual market-tightness effect leads to a lower price, as is standard in the directed/competitive search literature—see, for example, Moen (1997) and Acemoglu and Shimer (1999).

**Proposition 1 (Market tightness effects)** An increase in the population of sellers $S$ or middlemen $M$ (relative to that of buyers) leads to a lower retail market price $p_i, i = s, m$, and a lower bid-ask spread for all $k_m \geq 1$ and $\beta \in [0,1)$.

As is consistent with the standard framework, the market-tightness effect implies an intensified competition in the retail-markets, leading to a lower retail price and further to a lower bid-ask spread of middlemen. The discount factor does not affect the equilibrium allocations, thus a higher $\beta$ implies a higher wholesale price and a lower spread.

The bid-ask spread of middlemen can be decomposed into two parts:

$$p_m - \beta V^s = (p_m - p_s) + (p_s - \beta V^s).$$

In this expression, the first term $p_m - p_s \geq 0$ represents the premium a middleman charges to buyers for its high matching rate in the retail market, which is non-negative (see the proof of Proposition 2 and 4)—in the retail market, middlemen offer buyers a high matching rate with a high price and sellers offer buyers a low matching rate with a low price. All buyer are indifferent between visiting any middleman and any seller. The second term $p_s - \beta V^s > 0$ represents the premium a middleman charges to sellers for guaranteed sale in the wholesale market. In what follows, I present the comparative statistics results of $k_m$ on the bid-ask spread, with (i) $\beta = 0$ and (ii) $\beta \in [0,1)$ in separation. As it turns out, the former (special) case identifies important effects that help understand the behavior of the bid-ask spread in the latter (general) case. To simplify the analysis, I normalize the population parameter of sellers to one, $S = 1$.

### 3.1 Special Case $\beta = 0$

In this case, all the agents have an infinite discounting and zero continuation value. Thus, with $\beta = 0$, the wholesale price is zero and the bid-ask spread of middlemen is identical to their retail market price, i.e.,

$$p_m = \varphi^m(x_m, k_m) = \frac{1 - \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)}}{x_m \eta(x_m, k_m)/k_m}.$$
where, as the continuation value is zero for all the agents, the net trading surplus is fixed and equals to the consumption value, $N = 1$. Below, it is shown that the behavior of $p_m$ shapes critically that of the retail market premium, which, in this case, is given by

$$p_m - p_s = \varphi^m(x_m, k_m) - \varphi^s(x_s) \geq 0.$$ 

The wholesale premium with zero continuation value is simply equal to the sellers’ retail market price, $p_s$.

In addition to the usual (market-tightness) effect, there are two important effects of an increase in the inventory of middlemen $k_m$ on their price $p_m$. On the one hand, an increase in the middlemen’s capacity implies an increase in the number of buyers to visit middlemen, rather than sellers. This effect pushes up $p_m$ and pushes down $p_s$. On the other hand, a larger capacity of a middleman implies it is less likely that excess demand occurs and decreases the stockout probability. Because buyers know that the middleman receives zero payoff when $\beta = 0$ (or a lower expected payoff in general when $\beta > 0$) from unsold units, the middleman can extract only a smaller fraction of trading surplus per unit when the capacity $k_m$ is larger. This effect decreases the price $p_m$. The final outcome of these combined effects on the retail price of middlemen $p_m$ is, in general, ambiguous. Denote by

$$X = \frac{1}{Mk_m + 1} < 1$$

the per-period ratio of the total demand to the total supply in the retail market.

**Proposition 2 (Retail market prices/premium)** Suppose $\beta = 0$.

1. The retail price of middlemen $p_m$ is increasing in sufficiently low $k_m$, if and only if $X > X^* \in (0, 1)$, and is decreasing in sufficiently large $k_m$ for any given $X \in (0, 1)$.

2. The retail price of sellers $p_s$ is decreasing in all $k_m$, for any given $X \in (0, 1)$.

3. The retail market premium $p_m - p_s \geq 0$ is increasing in sufficiently low $k_m$ and is decreasing in sufficiently large $k_m$ for any given $X \in (0, 1)$.

The proposition shows that the middlemen’s retail price $p_m$ can be non-monotone in their inventory $k_m$, when the total demand is relatively large. Figure 1 plots the behaviors of the price $p_m$ in response to changes in $k_m$ for given values of $M$ (and hence $X$). The non-monotonicity reflects the dominance of the effect of their capacity to increasing the number of buyers to middlemen, without which the
price increase is impossible. For relatively small $k_m$, an increase in the capacity of middlemen implies a relatively large increase in their probability of serving buyers. When the total demand is sufficiently large, this creates a sufficiently large increase in the number of buyers visiting middlemen so that the effect of increasing the number of buyers can be dominant for the determination of $p_m$. For relatively large $k_m$, the probability of serving buyers is already high and so the number of buyers visiting middlemen does not increase enough to make this effect dominant. The same effect implies a decrease in the number of buyers visiting sellers. Thus, the sellers’ market gets tighter as the capacity of middlemen increases, pushing down the retail price of sellers, $p_s$. Therefore, the non-monotonicity of the retail market premium, $p_m - p_s$, is driven by that of the middlemen’s price: buyers are more (less) willing to pay a premium for a larger capacity of middlemen when the capacity is initially low (high).

Notice that the price decrease of middlemen can occur even without the market tightness effect, but solely due to the effect of decreasing the stockout probability. To confirm this point, I should abstract it from the (market-tightness) effect caused by changes in total supply. For this purpose, I examine the same comparative statistics exercise but, this time, fixing the middlemen’s total supply denoted by $G = Mk_m$. 

Figure 1: Retail price of middlemen
Proposition 3 (Fixed supply in middlemen’s market) Suppose $\beta = 0$ and fix the total supply by middlemen $G = Mk_m$.

1. The retail price of middlemen $p_m$ is increasing in sufficiently low $k_m$, if $X > X^* \in (0, 1)$, and is decreasing in sufficiently large $k_m$ for any given $X \in (0, 1)$.

2. The retail price of sellers $p_s$ is decreasing in both low $k_m$ and high $k_m$ for any given $X \in (0, 1)$.

3. The retail market premium $p_m - p_s$ is increasing in sufficiently low $k_m$ and is decreasing in sufficiently large $k_m$ for any given $X \in (0, 1)$.

The fixed total supply $G = Mk_m$ generates two margins: one is the intensive margin (as already seen above) and the other is the extensive margin, where $M$ decreases in response to an increase in $k_m$. As the extensive margin implies a price increase, the non-monotonicity described above holds with the fixed total supply, and can be generalized as follows: a price increase occurs for relatively low inventory $k_m$ and relatively high total demand, due to the effect of increasing number to middlemen (rather than sellers), whereas a price drop occurs for relatively high $k_m$, due to the effect of decreasing the stockout probability. If the relative total demand $X$ is not large enough, then the former effect cannot be dominant even at relatively low $k_m$. Therefore, all the insights obtained before remain valid and are confirmed to hold true irrespective of changes in the total supply.

3.2 General Case $\beta \in [0, 1)$

In general, the premium in the retail market can be written as

$$p_m - p_s = (\varphi^m(x_m, k_m) - \varphi^s(x_s))N.$$

It satisfies $p_m - p_s = 0$ when $k_m = 1$, $p_m - p_s > 0$ for all $k_m \in (1, \infty)$ and $p_m - p_s \to 0$ as $k_m \to \infty$ (see the proof of Proposition 4), implying the retail premium is non-monotone in the inventory of middlemen $k_m$ – the retail premium is increasing in relatively low $k_m$ and decreasing in relatively high $k_m$. On the other hand, the premium in the wholesale market can be written as

$$p_s - \beta V^* = \varphi^s(x_s)N,$$

where both the share $\varphi^s(\cdot)$ and the surplus $N$ are monotone increasing in $x_s$. Hence, since sellers receive fewer buyers and get lower profits in the retail market as $k_m$ increases, the wholesale premium decreases monotonically in the inventory of middlemen $k_m$. The combined effects of $k_m$ on the bid-ask spread of middlemen are, in general, ambiguous.
Proposition 4 (Bid-ask spread) 1. The bid-ask spread of middlemen $p_m - \beta V_s$ is increasing in low $k_m$, if and only if $X > X^*(\beta) \in (0, 1)$, where $X^*(\beta) \geq X^*$ with equality only when $\beta = 0$, and is decreasing in large $k_m$ for any given $X \in (0, 1)$.

2. Given the total supply by middlemen $G = Mk_m$ fixed, the bid-ask spread of middlemen is increasing in low $k_m$, if $X > X^*(\beta) \in (0, 1)$, and is decreasing in large $k_m$ for any given $X \in (0, 1)$.

The bid-ask spread of middlemen can be non-monotone in their inventory $k_m$ if the relative total demand $X$ is sufficiently large. Figure 2 depicts the response of the bid-ask spread to changes in $k_m$ for given values of $M$ (and hence $X$). For relatively low $k_m$, an increase in the retail premium is larger in magnitude than a decrease in the wholesale premium, so long as $X$ is large enough, so that the bid-ask spread is increasing in $k_m$. The responses of the retail and wholesale premia are plotted in Figure 3. Notice that the net surplus in the retail market decreases with $k_m$. Therefore, the critical value $X^*(\beta)$, above which a price increase can occur, should be larger for $\beta > 0$ than for $\beta = 0$. For relatively high $k_m$, the premia both in the retail market and in the wholesale market decrease with $k_m$, so does the spread. If $X$ is not large enough, then the effect to increase the retail premium cannot be dominant even at sufficiently low $k_m$. Finally, fixing the total supply of middlemen implies, the extensive margin
(of lowering M) works to increase the spread, thus the result holds true as long as $X > X^*(\beta)$. Hence, all the insights provided so far are valid with fixed total supply.

$\begin{align*}
\frac{k_m}{1 + \beta} + V_m &> 0, \\
\text{given values of } k_m \geq 1 \text{ and } M > 0.
\end{align*}$

A symmetric free entry equilibrium is a steady state equilibrium described in Theorem 1 where entry and exit occur until the middlemen operating in the markets earn zero expected net profits, just to cover the cost.

The equilibrium number of middlemen $M > 0$ is determined by the free entry condition, $V_m = \frac{ck_m^{\alpha}}{1 + \beta}$, or

$$N = c^{\alpha - 1}m,$$

where the L.H.S. represents the per-unit net profit of middlemen and the R.H.S. the per-unit cost. Define the upper bound of the cost parameter $\bar{c} \equiv \lim_{M \to 0}(1 - \beta)V_m/k^\alpha < 1$.

**Figure 3:** Retail market premium (left) and Wholesale market premium (right)

### 4 Free entry equilibrium

In this section, I allow for the number of middlemen to be determined endogenously by free entry. Suppose now that the inventory technology of middlemen can be acquired by paying a cost each period, and that the per-period cost of holding inventory $k_m$ is given by $ck_m^{\alpha} > 0$, where $\alpha \geq 0$ stands for the elasticity of inventory cost and $c > 0$ the scale parameter. These technologies enable one to operate as a middleman, so that he can buy multiple units from different sellers in the wholesale market and to serve more than one buyers in the retail market. An agent chooses to be a middlemen if the value of being a middlemen is non-negative, $-\frac{ck_m^{\alpha}}{1 + \beta} + V_m \geq 0$, given values of $k_m \geq 1$ and $M > 0$. A symmetric free entry equilibrium is a steady state equilibrium described in Theorem 1 where entry and exit occur until the middlemen operating in the markets earn zero expected net profits, just to cover the cost.

The equilibrium number of middlemen $M > 0$ is determined by the free entry condition, $V_m = \frac{ck_m^{\alpha}}{1 + \beta}$, or

$$N = c^{\alpha - 1}m,$$

where the L.H.S. represents the per-unit net profit of middlemen and the R.H.S. the per-unit cost. Define the upper bound of the cost parameter $\bar{c} \equiv \lim_{M \to 0}(1 - \beta)V_m/k^\alpha < 1$. 

17
Proposition 5 (Free entry equilibrium) Given values of \( c \in (0, \bar{c}) \) and \( \alpha \in [0, \infty) \), a free entry equilibrium exists and is unique. The equilibrium number of middlemen \( M \in (0, \infty) \) is:

1. decreasing in the cost parameters \( c, \alpha \);
2. increasing in low \( k_m \), if the scale parameter \( c \) is sufficiently high and the elasticity \( \alpha \) is low;
3. decreasing in low \( k_m \) if \( c \) is sufficiently low or \( \alpha \) is high;
4. decreasing in high \( k_m \) for any given \( c \in (0, \bar{c}) \) and \( \alpha \in [0, \infty) \).

The pre-unit profit (i.e., L.H.S. of (9)) is decreasing in the number of middlemen, thereby a larger cost leads to fewer middlemen given values of \( k_m \) (result 1). The next three results 2–4 in the proposition show that \( M \) can be non-monotone in the inventory of middlemen \( k_m \) – see the left figure of Figure 4. For relatively low \( k_m \), if the scale parameter \( c \) of the inventory holding cost is high, then there are few operating middlemen, thus there is a relatively high total demand. In this situation, a larger \( k_m \) creates a sufficiently large increase in the number of buyers to middlemen. This effect leads to an increase in the profit of middlemen and stimulates entry, if the cost elasticity \( \alpha \) is low, but to a decrease in their profit and induces exit, if \( \alpha \) is high. If \( c \) is not that high, there are many operating middlemen (implying a relatively low total demand), so that an increase in the inventory leads to a lower profit and induces exit. For relatively high \( k_m \), an increase in the inventory reduces the profit of middlemen, resulting in fewer middlemen, for any given values of \( c \in (0, \bar{c}) \) and \( \alpha \in [0, \infty) \).

In the free entry equilibrium, changes in the inventory create not only the intensive margin (as described in the previous section), but also the extensive margin for the determination of the bid-ask spread.

Proposition 6 (Bid-ask spread with free entry) In the free entry equilibrium described in Proposition 5, the bid-ask spread of middlemen is, for any given values of \( c \in (0, \bar{c}) \):

1. increasing (decreasing) in low \( k_m \) for relatively high (low) \( \alpha \);
2. increasing (decreasing) in large \( k_m \) if \( \alpha > 1 \) (\( \alpha \leq 1 \));

For relatively low \( k_m \), the bid-ask spread increases with the inventory \( k_m \), due to the effect of attracting more buyers in the intensive margin and the effect of decreasing the market tightness in the extensive margin, if the cost elasticity \( \alpha \) is relatively high. If \( \alpha \) is relatively low, then the spread can be decreasing even with low \( k_m \). While the condition of the spread’s increase is stated in terms of the cost parameter
α here, the essential remains the same as before: with high (low) elasticity of the inventory cost α, the number of middlemen does not increase much or even decreases (increases) in response to an increase in the inventory, as shown in Proposition 5. Thus, the extensive margin of a smaller (larger) number of middlemen creates the situation of higher (lower) relative total demand, just like in Proposition 4, that leads to the higher (lower) spread. For relatively high $k_m$, the spread is decreasing in $k_m$ via the intensive margin, if the inventory cost is not elastic $\alpha \leq 1$, but is increasing via the extensive margin if the cost is elastic $\alpha > 1$. Combining these results, the bid-ask spread can be non-monotone in $k_m$ for not too high $\alpha > 1$—it decreases with low $k_m$ but increases with high $k_m$ (see the right figure of Figure 4).

![Figure 4: Free entry equilibrium](image)

5 Matching efficiency and welfare

I now study the implications of middlemen’s inventory holdings on the matching efficiency and the net welfare. The per-period total matching rate in this economy, denoted by $T$, is given by

$$T = M x_m \eta(x_m, k_m) + x_s \eta(x_s, 1).$$

To show first that the middlemen in my framework are efficiency enhancing, consider an alternative economy in which there are no middlemen, and buyers and sellers can trade only in a private market
each period. Let $S = Mk_m + 1$ be the population of sellers. Then, the total matching rate in this alternative economy is given by

$$T = S x_s \eta(x_s, 1)$$

in each period, where $x_s = 1/S$ is the queue of buyers at individual sellers. Comparing the per-period total matching rates in these two economies, which have the same total supply ($= Mk_m + 1$) and total demand (normalized to one), the following proposition shows $T \geq T_0$ with strict inequality for $k_m > 1$.

**Proposition 7 (Matching efficiency)** The middlemen in this economy are efficiency enhancing in terms of the total matching rate.

Clearly, the total matching rate $T$ is increasing in both the capacity of middlemen $k_m$ and the number of middlemen $M$. When the total supply in the middlemen’s market $G = Mk_m$ is fixed, however, there is a negative relationship between $k_m$ and $M$. The following proposition shows that the total matching rate is increasing in $k_m$ via the intensive margin, given fixed total supply, for both relatively small and large values of $k_m$.

**Proposition 8 (Matching efficiency and inventory)** The total matching rate $T$ is increasing in the inventory capacity of middlemen $k_m$ and the number of middlemen $M$. For any fixed supply in the middlemen’s market $G = Mk_m$, $T$ is increasing in both small and large values of $k_m$.

Figure 5 plots the behaviors of the total matching rate $T$ in response to changes in the inventory capacity of middlemen $k_m$, for given fixed values of $G = Mk_m$. It shows that $T$ is increasing in any values of $k_m$ via the intensive margin. The result implies that few middlemen, each with many inventories, lead to a higher total matching rate than many middlemen, each with few inventories.

Similarly, one can examine the welfare implication of middlemen. In this economy, the total welfare net of middlemen’s profits, denoted by $W$, is given by

$$W = V^b + V^s = \frac{e^{-x_s}}{1 - \beta x_s e^{-x_s}} + \frac{1 - e^{x_s} - x_s e^{-x_s}}{1 - \beta x_s e^{-x_s}} = \frac{1 - x_s e^{-x_s}}{1 - \beta x_s e^{-x_s}},$$

which is monotone decreasing in $x_s \in (0, 1/(M + 1))$, while the net welfare in the economy with no middlemen is given by

$$W = \frac{1 - x_s e^{-x_s}}{1 - \beta x_s e^{-x_s}}.$$
where $\bar{z}_s = 1/(Mk_m + 1)$ is the queue of buyers at individual sellers. Define $\Delta W \equiv W - \overline{W}$ as the welfare difference between the economies with and without middlemen. Clearly, it satisfies $\Delta_w = 0$ when $k_m = 1$. When comparing the economies with and without middlemen, notice that, as before, these two economies have the identical per-period ratio of total demand to total supply, $X = 1/(G + 1) \in (0,1)$.

Proposition 9 (Welfare implications of middlemen) The total net welfare $W$ is increasing in the inventory capacity of middlemen $k_m$ and the number of middlemen $M$. For any fixed supply in the middlemen’s market $G = Mk_m$, the welfare difference $\Delta W$ is increasing in both low and high values of $k_m$.

The welfare results are consistent with the previous analysis on the matching efficiency: A larger supply implies a higher matching rate of individual buyers and a lower matching rate of individual sellers, leading to a higher welfare of buyers $V^b$ and a lower welfare of sellers $V^s$, and to a higher total welfare of buyers’ and sellers $W$. Further, the welfare difference $\Delta W$ increases with $k_m$ via the intensive margin, given fixed supply in the middlemen’s market (which implies $\overline{W}$ is constant and so the behavior of $\Delta W$ is dictated by that of $W$). Figure 6 plots the response of $\Delta W$ to changes in $k_m$ given values of $G = Mk_m$. Therefore, the efficiency implication of middlemen’s inventory capacity can
be generalized as follows: few middlemen, each with many inventories, lead to a higher total matching rate and a higher total welfare than many middlemen, each with few inventories.

6 Discussion

This study contributes to the literature of middlemen, initiated by Rubinstein and Wolinsky (1987), that emphasizes the middlemen’s advantage over the original suppliers in the rate at which they meet buyers. In their model, it is assumed that: (i) the matching rates of agents are exogenous; (ii) the terms of trades are determined by Nash bargaining; (iii) middlemen can hold only one unit of a good as inventory. In contrast, my approach is based on a standard directed search equilibrium and allows me to study both the matching advantage of middlemen and its influence on their market power, because it incorporates: (i) buyers’ choice of where to search so that the matching rate between buyers and suppliers is determined endogenously; (ii) competition among suppliers so that individual suppliers can influence the search-purchase behaviors of buyers through prices; (iii) middlemen’s inventory holdings.
of more than one unit so that the dependence of both the buyers’ search decision and the extent of competition on their inventory can be made explicit.

There are two papers in this literature that are related to the current study. Shevichenko (2004) provides a random meeting model that allows for middlemen to hold a variety of goods as inventory with ex ante heterogeneity of products and preferences. The middlemen studied in his model are agents who can mitigate the severity of double-coincidence of wants problem. In his model, the price is determined by Nash bargaining, and the equilibrium price is dispersed with respect to the type of goods held by middlemen. By holding a larger variety of goods as inventory, middlemen can increases the chance that a random buyer finds his preferred good on their shelves, which may or may not improve their terms of trade. However, the inventory capacity does not affect the rate at which a random buyer meets with a middleman. In contrast to Shevichenko (2004), even though agents and goods are both homogeneous in my model, the inventory capacity of middlemen can affect their meeting rate with buyers, generating the buyers’ tradeoff between the price and the matching rate. It is exactly this simple tradeoff that yields the differential equilibrium prices across sellers with different capacities and the differential degree of retail market competition depending on the capacity level.

In a companion paper, Watanabe (2010) presents a special case of the current model and studies the turnover behaviors of sellers to become middlemen under a simplifying assumption of infinite discounting. It is shown that turnover equilibrium can be multiple - one is stable and has many middlemen, each with fewer units and a high price, and the other is unstable and has few middlemen, each with many units and a low price. Assuming away the turnover issue, the current paper investigates the effect of middlemen’s inventory capacities on the market outcomes, which is perhaps more relevant for the directed search approach to study the functioning of markets with middlemen. While it is true that much complication can arise once the myopic-agents assumption is relaxed, the current paper develops a simple methodology that allows me to establish a steady state equilibrium with forward looking agents and to derive the analytical characterization, both with and without free entry.

In the directed/competitive search literature, Burdett, Shi, and Wright (2001) and Shi (2001a)

---

13 See also Biglaiser (1993) and Li (1998) where middlemen are modeled as a guarantor of product qualities, who help to ameliorate lemons problems, and Masters (2007) and Watanabe (2010) where the decision of agents to become a middleman is endogenized. In another approach used in Gehrig (1993); Spulber (1996), Rust and Hall (2003), Caillaud and Jullien (2003), Hendershott and Zhang (2006) and Loertscher (2007) (see also the book by Spulber (1999)), price setting is emphasized as the middlemen’s main role of market-makings, but the meeting rate is exogenous.

14 Coles and Eeckhout (2003) study a setup in which sellers can post a more general trading mechanism for a finite number of agents. They show that a continuum of equilibria exist including an equilibrium with a simple form of price posting (i.e., the one studied in Burdett, Shi and Wright), while sellers prefer an equilibrium with auction. With a continuum of agents, auction and price posting are practically equivalent, with sellers achieving the same revenue and guaranteeing buyers the same utility. A usual argument applies: relatively high transaction costs associated with
are the first to study the buyers’ tradeoff between the price and the service probability for sellers with different capacities, assuming that the capacity level of high-capacity sellers is $k_m = 2$. Generalizing $k_m \geq 1$, this paper identifies an equilibrium in which all agents are willing to pay a premium for the middlemen’s capacity: (i) buyers are indifferent between visiting a producer offering a low price and a low service rate, and a middleman offering a high price and a high service rate; (ii) sellers are indifferent between selling to buyers in the retail market at a higher price with a risk of not clearing out their stocks, and selling to middlemen in the wholesale markets at a lower price with no risk of unsold goods. It turns out that the price differential in the retail market (i.e., retail premium) is not necessarily monotone in $k_m$. Another implication obtained here is that an economy with few large firms is more efficient than an economy with many small firms, given fixed total supply. The essence is neither cost considerations, nor increasing returns to scale in production technologies, but the way in which firms should mitigate market frictions. This is a new implication of firms’ capacity and in general the operative form of firms as an economic entity for search frictions. The latter implication is interesting especially when it is compared to the Mortensen-Pissarides type of models, which has no concept of firms’ capacity and where with constant returns to the (exogenous) matching function, the composition of market, in terms of the size and the number of firms, does not play any role.

A further related issue to the directed search literature is imperfect information on the capacity of middlemen. Menzio (2007) provides a model of the labor market with asymmetric information on the firms’ productivity and shows that cheap talk can sometimes credibly convey information when wages are determined through bilateral bargaining. In his model, the cost of advertising a high-quality job is to receive tougher wage demands at the bargaining stage, whereas the benefit is to achieve a higher probability of filling the vacancy. In my context, the legal system can prevent middlemen from misreporting prices, but not necessarily how much inventory they hold. Thus, the true inventory level of a middleman can differ from what he communicates to buyers. In this situation, it would be interesting to ask whether the price would be a good enough signal about the true inventory.

The simple model presented here is admittedly stylized and does not aspire to capture the precise

establishing and implementing auction can make sellers prefer price posting. This makes sense in particular for the economy considered here where retail technologies are made explicit and play an important economic role for middlemen’s profits.

15Moen (1997), Mortensen and Wright (2002), and Sattinger (2003) consider fictitious market makers (not middlemen), who replace the Walrasian auctioneer, to interpret the efficiency property of competitive search equilibrium: Given the market makers create different submarkets and announce a pair of price and service probability in each submarket, buyers and sellers are willing to trade in a submarket where the Hosios condition holds endogenously. The latter property holds true in the equilibrium with middlemen constructed in my model, but achieving higher efficiency than the one with their market makers.
workings of any particular market. Nevertheless, it offers wide applicability and sheds light on many empirically relevant forms of middlemen.

A recent literature on financial intermediaries pioneered by Duffie, Garleanu, and Pedersen (2005) uses a bargaining-based search model, together with time varying preference shocks, to formulate the trading frictions that are characteristic of over-the-counter markets. Although the details of the modeling setup are different, the current study is consistent with recent progress in this literature. In particular, Lagos and Rocheteau (2007, 2009) generalize the framework by allowing market participants to hold an unrestricted amount of assets and show that the (average) bid-ask spread of dealers can be non-monotonic in their bargaining power or in the contact rate of investors. In their model, a higher bargaining power or a lower contact rate has a positive effect on the spread, and changes the distribution of trade sizes, which can have an adverse effect on the spread. The latter effect can be considered as a general equilibrium effect that occurs through changes in the investors’ hedging behaviors against the future preference shocks. My model shows there is non-monotonicity in the capacity of dealers. This result arises due to a rather simple tradeoff faced by traders between the price and the (endogenous) matching probability: traders are more willing to pay a higher premium for a larger inventory of dealers when the capacity is initially low than when the capacity is initially high. Another difference is that in my model, agents are allowed to trade directly with each other and the premium for dealers is defined over this option, whereas the dealers are the only avenue of exchange in their model and the premium traders are willing to pay for dealers is defined with respect to no trade.

The current study is also related to the recent literature on middlemen for international trade. This literature studies the role of middlemen to reduce trading costs. One of the insights obtained in this paper is that while middlemen enhance the economy-wide efficiency, they can make the original producers worse off and enjoy a relatively high intermediation margin simultaneously. Further, the following evidence supports the idea pursued here that small exporters are rarely matched with small importers and that the scale is important in intermediation: Bernardo, Blum, Claro, and Horstmann (2010) find that small exporters typically match with one large importer-intermediary; Ahn, Khandelwal and Wei (2010) find that intermediaries tend to make a deal with relatively small Chinese firms who

---

16Other papers in this literature include Afonso (2010), Garleanu (2009), Lagos, Rocheteau, and Weill (2010), Vayanos and Wang (2007), Vayanos and Weill (2008), and Weill (2007, 2008). In the finance literature, the formal models (without search frictions) emphasizing the relationship between the dealers’ inventory holdings and the bid-ask spread are developed, for example, Stoll (1978), Amihud and Mendelson (1980), Ho and Stoll (1981), and Hendershott and Menkveld (2009).

17See, for instance, Ahn, Khandelwal and Wei (2010), Akerman (2010), Blume, Claro and Horstmann (2009), and Rauch and Watson (2004).
find it difficult to penetrate export markets on their own.

Antràs and Costinot (2010) provide a model that studies explicitly the role of search frictions in intermediated trade using a random meeting/bilateral bargaining approach. As in my model with free entry, they identify an ambiguous effect of middlemen’s entry on the intermediation margin (which causes the deterioration of the domestic farmers’ welfare): On the one hand, entry of foreign middlemen makes the domestic market tighter, which increases the surplus share of farmers; On the other hand, the foreign middlemen have exogenously higher bargaining power than the local middlemen, which they justify by arguing that larger middlemen tend to have higher bargaining power, resulting in the reduction of the farmers’ surplus share. Thus, in their model, the total effect depends on the bargaining power of middlemen and the entry cost. In my model, an ambiguous effect of middlemen’s entry on the intermediation margin (and efficiency)\(^{18}\) occurs depending on the inventory capacity of middlemen. When the capacity cost is convex, the intermediation margin decreases in relatively low capacities but increases in relatively high capacities via the extensive margin of the rapidly changing number of active middlemen. In my model, however, the matching rate and the surplus share of agents can be influenced by the capacity of middlemen, thereby the source of middlemen’s market power is made explicit by the tangible economic object. It should be noted that their model fits well with international trade in that if frictions are absent then it is essentially a Ricardian model with two homogeneous goods and two countries, whereas in my model the equilibrium approaches to the competitive equilibrium if frictions disappear (or if the capacity of middlemen becomes unboundedly high).

Finally, the implications of the model regarding the retail market behaviors feature various important evidence recently provided in the empirical industrial organization literature. The demand effect of capacity studied in this paper captures a familiar observation that retailers attract more customers when they increase their inventories. The co-movement of inventory with final sales is a well-established empirical regularity (see the survey by Ramey and West (1999)). As for the effect on the retail price, Dana and Spier (2001) offer interesting evidence from the video rental industry. They surveyed the retail price and the availability of four new releases at 20 video outlets within a 4-mile radius of Northwestern University. They found that Blockbuster Video (the leading intermediary in the rental video industry) charged $3.81 and had 86% availability on average, whereas the outlets from other national chains charged $3.32 and had 60% availability, and the independent stores charged $2.62 and had only 48% availability. These price differences reflect the fact that Blockbuster has increased its new-release

\(^{18}\)The result on the efficiency with free entry was presented in the working paper version, but is omitted here to save space. It is available upon request.
prices since adopting the policy to increase product availability, which allowed them to be more likely to have particular new releases in stock than other major chains. This evidence is consistent with the demand effect of capacity analyzed in the my model. A further insight obtained from my model is that the effect of a higher inventory to increase the retail price may become prevalent when the inventory is low relative to the total demand, as in the case of popular newly released titles, so that individual customers are prepared to pay a large premium for the increased probability of getting served.

My model captures the fact that when intermediaries hold excessive amount of inventories, they often have a clearance sale with large price reduction. First of all, Aguirregabiria (2005) finds that intermediaries' inventory is a critical variable to explain their pricing patterns, especially when customers trade-off the price against the service rate. Further, Aguirregabiria (1999) identifies a negative effect of inventory ordering on the markups in a supermarket chain. In connection with the latter evidence, retailers' price increases following supermarket leveraged buy-outs (LBOs) are observed in Chevalier (1995) and Matsa (2011). All these empirical findings are consistent with the effect of capacities on retail prices illustrated in my model, given that high leverage may lead firms to be cash-constrained and thus may reduce their ability to maintain the available stock for future sales. A related evidence provided by Matsa (2009) shows that stockouts are negatively correlated with competition: supermarkets that face significantly high competition offer on average 5 percent lower stockout rates than otherwise similar stores. This means that, as is consistent with the case of relatively high capacities in my model, a higher competition in the presence of large-scaled retailers, like Wal-Mart, is associated with a lower stockout probability offered to their customers.

7 Conclusion

This paper proposed a simple theory of middlemen using a standard directed search approach. It offers wide applicability and economic insights into many empirically relevant forms of middlemen. The middlemen hold inventories of a good and are specialized in buying and selling. Middlemen’s inventories can provide buyers with immediacy service and sellers with guaranteed sales under market frictions, thereby the ask price of middlemen includes a premium to buyers and the bid price includes a premium charged to sellers. The model generates two important effects of middlemen’s inventories that serve as the critical determinant of the bid-ask spread. On the one hand, it allows middlemen to enjoy a simultaneous increase in both their buying and selling power. On the other hand, it puts downward pressure on their retail price. These conflicting effects cause non-monotonic responses of the
bid-ask spread to changes in their inventories. When free entry of middlemen is allowed, the number of active middlemen can be non-monotone in the inventory, and the extensive margin can matter for the determination of the bid-ask spread. The middlemen in this economy are efficiency enhancing, and few middlemen, each with many inventories lead to a higher total matching rate and a higher total welfare than many middlemen, each with few inventories.
Appendix

Proof of Theorem 1

The analysis in the main text has established that (1), (3), (4), (5), (6) and (7) describe necessary and sufficient conditions for an equilibrium given the stationary inventory restocking of middlemen. All that remains here is to establish a solution to these conditions, \(x_i, x_m, p_s, p_m, V^b, V^s, V^m > 0\), exists and is unique. The proof takes 2 steps. Step 1 establishes a unique solution \(x_s, x_m > 0\) for all \(k_m \geq 1\), \(S \in [1, \infty)\) and \(M \in (0, \infty)\), using (1), (3), (4) and (5). With a slight abuse of notation, let \(x_i(k_m, S, M)\) denote this solution for \(i = s, m\). Given this solution, Step 2 then identifies a unique solution \(V^j \in (0, 1)\) to (4), (5) and (6) for \(j = b, s\). The rest of the equilibrium values are identified immediately: given \(V^b, V^s\) and \(x_i, (3)\) determines a unique \(p_i \in (0, 1)\) for \(i = s, m\); given \(V^s, x_m\) and \(p_m\), (7) determines a unique \(V^m \in (0, k_m)\). For all \(\beta \in [0, 1), k_m \geq 1, S \in [1, \infty), M \in (0, \infty)\), this solution then satisfies (1), (3), (4), (5), (6), and (7) so describes equilibrium.

Step 1 For any \(k_m \geq 1, S \in [1, \infty)\) and \(M \in (0, \infty)\), a solution \(x_i = x_i(k_m, S, M)\) to (1), (3), (4) and (5) exists and is unique for \(i = s, m\) that is: continuous in \(S, M, k_m \in \mathbb{R}_+\); strictly decreasing in \(S, M\) for all \(k_m \geq 1\); strictly increasing (or decreasing) in \(k_m\) for all \(S \in [1, \infty)\) and \(M \in (0, \infty)\) if \(i = m\) (or if \(i = s\)) satisfying \(x_s(1, \cdot) = x_m(1, \cdot) = 1/(S + M)\), \(x_s(k_m, \cdot) \to 0\) and \(x_m(k_m, \cdot) \to 1/M\) as \(k_m \to \infty\).

Proof of Step 1. In the main text, it has been shown that (3), (4) and (5) imply (8). Substituting out \(x_m\) in (8) by using (1),

\[
\Gamma\left(k_m, \frac{1-Sx}{M}\right) = e^{-x_s}
\]

where \(\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt\) and \(\Gamma(k, x) = \int_x^\infty t^{k-1}e^{-t}dt\). The L.H.S. of this equation, denoted by \(\Phi(x_s, k_m, S, M)\), is continuous and strictly increasing in \(x_s\) and \(k_m \in \mathbb{R}_+\), satisfying for any \(S \in [1, \infty)\) and \(M \in (0, \infty)\):

\[
\Phi(x_s, \cdot) \to \frac{\Gamma(k_m, \frac{1}{S+M})}{\Gamma(k_m)} < 1 \text{ as } x_s \to \cdot; \quad \Phi\left(\frac{1}{S+M}, \cdot\right) = \frac{\Gamma\left(k_m, \frac{1}{S+M}\right)}{\Gamma(k_m)} \geq e^{-\frac{1}{S+M}}
\]

with equality only when \(k_m = 1\);

\[
\Phi(x_s, 1, \cdot) = e^{-\frac{1-Sx}{M}}; \quad \Phi(x_m, k_m, \cdot) \to 1 \text{ as } k_m \to \infty.
\]

Similarly, \(\Phi(\cdot)\) is continuous and strictly increasing in \(S, M\) for any \(x_s \in (0, \frac{1}{S+M})\) and \(k_m \geq 1\). It follows therefore that a unique solution \(x_s = x_s(k_m, S, M) \in [0, \frac{1}{S+M}]\) exists that is: continuous and strictly decreasing in \(k_m \in [1, \infty) \subseteq \mathbb{R}_+\) satisfying \(x_s(1, \cdot) = \frac{1}{S+M}\) and \(x_s(k_m, \cdot) \to 0\) as \(k_m \to \infty\) for any \(S, M\); continuous and strictly decreasing in \(S, M\) for all \(k_m \geq 1\).

Applying this solution to (1), one can obtain a unique solution \(x_m = x_m(k_m, S, M) \in [\frac{1}{S+M}, \frac{1}{M}]\) that is: continuous and strictly decreasing in \(S\) and \(M\); continuous and strictly increasing in \(k_m \in [1, \infty) \subseteq \mathbb{R}_+\) satisfying \(x_m(1, \cdot) = \frac{1}{S+M}\) and \(x_m(k_m, \cdot) \to \frac{1}{M}\) as \(k_m \to \infty\). This completes the proof of Step 1.

Step 2 Given \(x_s \in (0, 1/(S + M)]\) established in Step 1, there exists a unique solution \(V^j \in (0, 1), j = b, s\), to (3), (4), and (6).

Proof of Step 2. (3), (4), and (6) imply \(V^b\) satisfies

\[
V^b = \frac{e^{-x_s}}{1 - \gamma x_s e^{-x_s}}.
\]
The R.H.S of this equation, denoted by $\Upsilon_b(x_s)$, is strictly decreasing in $x^s \in (0, \infty)$ and satisfies: $\Upsilon_b() \to 1$ as $x_s \to 0$; $\Upsilon() \to 0$ as $x_s \to \infty$. As equilibrium implies $x_s \in (0, 1/(S + M)]$, there exists a unique $V^b \in (0, 1)$ that satisfies $V^b = \Upsilon_b()$. (3), (4), and (6) also imply

$$V^s = \frac{1 - e^{-x^s} - x_0 e^{-x^s}}{1 - \beta x_s e^{-x^s}}$$

and this time, the R.H.S. of this equation, denoted by $\Upsilon_s(x_s)$, is strictly increasing in $x^s \in (0, \infty)$ and satisfies: $\Upsilon_s() \to 0$ as $x_s \to 0$; $\Upsilon_s() \to 1$ as $x_s \to \infty$, thereby there exists a unique solution $V^s \in (0, 1)$. This completes the proof of Step 2. ■

**Proof of Corollary 1**

Step 1 in the proof of Theorem 1 has shown that $x_s = x_s(k_m, S, M)$ is strictly decreasing in $k_m, S, M$ while Step 2 in the proof of Theorem 1 has shown that $V^s$ is strictly increasing in $x_s$, implying $V^s$ is strictly decreasing in $k_m, S, M$. ■

**Proof of Proposition 1**

○ Retail prices $p_i$, $i = s, m$:

Differentiation yields

$$\frac{dp_i}{dS} = \frac{d}{dS} (\beta V^s + \varphi^i(x_i, k_i) N) = \frac{d\beta V^s}{dx_s} \frac{dx_s}{dS} + \frac{d\varphi^i(x_i, \cdot)}{dx_i} \frac{dx_i}{dS} N + \varphi^i() \frac{dN}{dx_s} \frac{dx_s}{dS}.$$  \hspace{1cm} (11)

for $i = s, m$. Remember that $\frac{dx_i}{dS} < 0$, $i = s, m$, and the first term in (11) is negative (by Corollary 1).

Observe that

$$\frac{\partial \varphi^i(x_m, \cdot)}{\partial x_i} = -\frac{k_i}{x_i} \left(1 - \frac{\Gamma(k_i + 1, x_i)}{\Gamma(k_i + 1)}\right) \frac{\eta(\cdot)}{\partial x_i} + \frac{\partial}{\partial x_i} \left[\frac{k_i}{x_i} \left(1 - \frac{\Gamma(k_i + 1, x_i)}{\Gamma(k_i + 1)}\right)\right].$$  \hspace{1cm} (12)

The first term in the R.H.S. of (12) is positive, and the second term is

$$\frac{\partial}{\partial x_i} \left[\frac{k_i}{x_i} \left(1 - \frac{\Gamma(k_i + 1, x_i)}{\Gamma(k_i + 1)}\right)\right] = \frac{\partial}{\partial x_i} \left[\sum_{j=k_i}^{\infty} \frac{x^j e^{-x_i}}{j!} \frac{k_i}{j + 1}\right] = \sum_{j=k_i}^{\infty} \frac{x^{j-1} e^{-x_i} (j - x_i)}{j!} \frac{k_i}{j + 1} > 0$$

if $x_i < k_i$, which is always the case with $i = s$ for $S \geq 1$. To examine the case $x_m \geq k_m$, rewrite (12) as

$$x_m \eta(\cdot)^2 \frac{\partial \varphi^m(x_m, \cdot)}{\partial x_m} = \frac{x^m k_m e^{-x_m}}{\Gamma(k_m)} \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} - \frac{k_m}{x_m} \left(1 - \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)}\right) \left(\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} - \frac{x^m k_m e^{-x_m}}{\Gamma(k_m)}\right).$$

If $\Gamma(k_m, x_m) \leq \frac{x^m k_m e^{-x_m}}{\Gamma(k_m)}$, then $\frac{\partial \varphi^m(x_m, \cdot)}{\partial x_m} > 0$. Otherwise,

$$x_m \eta(\cdot)^2 \frac{\partial \varphi^m(x_m, \cdot)}{\partial x_m} > \frac{x^m k_m e^{-x_m}}{\Gamma(k_m)} - \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \left(1 - \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)}\right),$$

which holds true with $x_m \geq k_m$ and $\Gamma(k_m, x_m) > \frac{x^m k_m e^{-x_m}}{\Gamma(k_m)}$, since $\Gamma(k_m + 1, x_m) > \frac{x^m k_m e^{-x_m}}{\Gamma(k_m + 1)} - \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} = \frac{x^m k_m e^{-x_m}}{\Gamma(k_m + 1)} > 0$. Now, to examine the R.H.S. of the above inequality, define

$$\Phi_{\eta}(x, k) \equiv \frac{x e^{-x}}{\Gamma(k)} \frac{\Gamma(k, x)}{\Gamma(k)} \left(1 - \frac{\Gamma(k, x)}{\Gamma(k)}\right)$$

30
for $x \geq k \in [1, \infty] \subseteq \mathbb{R}_+$. Observe that $\lim_{x \to \infty} \Phi_g(x, k) = 0$, and

$$\frac{\partial \Phi_g(x, k)}{\partial x} = \frac{x^{-k}e^{-x}}{\Gamma(x, k)} \left( k + 1 - x - 2\frac{\Gamma(k, x)}{\Gamma(k)} \right) > 0 \iff x \lesssim x^+$$

where $x^+ \in (k, k+1)$ is a unique solution to $x^+ = k + 1 - 2\frac{\Gamma(k, x^+)}{\Gamma(k)}$, hence $\frac{\partial \Phi_g(x, k)}{\partial x} > 0$ at $x = k$. Therefore, if $\Phi_g(k, k) > 0$ then $\Phi_g(x, k) > 0$ for all $x \in [k, \infty)$. To show this corner condition $\Phi_g(k, k) > 0$ holds true, notice first that

$$\Phi_g(k, k) > \frac{k^k e^{-k}}{\Gamma(k)} - \frac{1}{4}$$

holds true for any $k \in [1, \infty)$. Now, observe that

$$\frac{d}{dk} \ln \left( \frac{k^k e^{-k}}{\Gamma(k)} \right) = \ln(k) - \psi(k),$$

where $\psi(k) = \frac{d\ln(\Gamma(k))}{dk}$ is the Psi (or digamma) function, which has the definite-integral representation that leads to

$$\psi(k) = \int_0^\infty \left( e^{-t} - \frac{1}{(1+t)^k} \right) \frac{dt}{t} = \ln k - \frac{1}{2k} - 2 \int_0^\infty \frac{tdt}{(t^2 + k^2)(e^{2\pi t} - 1)}$$

(see, for example, Abramowitz and Stegun (1964) p.259). The last expression leads to

$$\frac{d}{dk} \ln \left( \frac{k^k e^{-k}}{\Gamma(k)} \right) = \frac{1}{2k} + 2 \int_0^\infty \frac{tdt}{(t^2 + k^2)(e^{2\pi t} - 1)} > 0,$$

for all $k \in [1, \infty)$. Since $\frac{k^k e^{-k}}{\Gamma(k)} = e^{-1} \approx 0.37 > \frac{1}{4}$ when $k = 1$, this implies that the term $\frac{k^k e^{-k}}{\Gamma(k)}$ is greater than $\frac{1}{4}$. This further implies $\Phi_g(k, k) > 0$ for all $k \in [1, \infty)$ and $\Phi_g(x, k) > 0$ for all $x \in [k, \infty)$. Therefore, it has been shown that $\frac{\partial \Phi(x, k)}{\partial x} > 0$, for all $x_i \in (0, \infty)$, $i = s, m$, and the second term in (11) is negative. The third term in (11) is also negative, because

$$\frac{dN}{dx_s} = \frac{d}{dx_a} \left( \frac{1 - \beta}{1 - \beta x_a e^{-x_s}} \right) = \frac{\beta(1 - \beta) e^{-x_s} (1 - x_s)}{(1 - \beta x_a e^{-x_s})} > 0.$$

Combining all these terms, it follows that $\frac{dN}{dx_s} < 0$, $i = s, m$. The same procedure applies to prove $\frac{dx_s}{ds} < 0$, $i = s, m$. ■

\(\circ\) Bid-ask spread $p_m - \beta V^s$: Given the above results, it is sufficient to observe that

$$\frac{d(p_m - \beta V^s)}{dS} = \frac{dp_m}{dS} - \frac{d\beta V^s}{dS} \frac{dx_s}{dS} = \frac{\partial \phi^m(x_m, \cdot)}{\partial x_m} \frac{dx_m}{dS} N + \phi^m(\cdot) \frac{dN}{dx_s} \frac{dx_s}{dS} < 0.$$

The results on parameter $M$ are immediate. ■
Proof of Proposition 2

\( \circ \) Retail price of middlemen \( p_m \): For the expositional ease, let

\[
\nabla_1 \equiv \frac{x_m}{k_m} \Gamma(k_m, x_m); \quad \nabla_2 \equiv 1 - \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)}.
\]

Differentiating \( p_m = \varphi^m(x_m, k_m) \) with respect to \( k_m \in [1, \infty) \subset \mathbb{R}_+ \),

\[
(\nabla_1 + \nabla_2)^2 \frac{d\varphi^m(x_m, k_m)}{dk_m} = (\nabla_1 + \nabla_2)^2 \frac{d}{dk_m} \left( \frac{\nabla_2}{\nabla_1 + \nabla_2} \right)
= -\nabla_1 \frac{\partial \Gamma(k_m + 1, x_m)}{\partial k_m} \left( \frac{1}{k_m} \right) + \nabla_2 \left( \frac{x_m \partial \Gamma(k_m, x_m)}{\partial k_m} \right) + \frac{dx_m}{dk_m} \left( (\nabla_1 + \nabla_2) \frac{x_m e^{-x_m}}{\Gamma(k_m + 1)} - \frac{\nabla_1 \nabla_2}{x_m} \right).
\]

In Step 1 in the proof of Theorem 1, it has been shown that

\[
\frac{dx_m}{dk_m} = \frac{\partial \Gamma(k_m, x_m) / \Gamma(k_m)}{\partial k_m} \Bigg|_{k_m=1} \frac{\Gamma(k_m + 1)}{\Gamma(k_m)} - \Gamma(k_m, x) \frac{\partial \Gamma(k_m)}{\partial k_m} \Bigg|_{k_m=1} = e^{-x} \ln x + E_1(x) + e^{-x} \gamma,
\]

where in the second equality I have used:

\[
\frac{\partial \Gamma(k_m, x)}{\partial k_m} \Bigg|_{k_m=1} = \frac{\partial \Gamma(k_m, x)}{\partial k_m} \bigg|_{k_m=1} = e^{-x} \ln x + E_1(x); \quad \frac{\partial \Gamma(k_m)}{\partial k_m} \bigg|_{k_m=1} = -\gamma.
\]

(see Geddes, Glasser, Moore, and Scott (1990) for the former, and Abramowitz and Stegun (1964) p.228 for the latter, for example), where

\[
E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt
\]

is the exponential integral and \( \gamma (= 0.5772..) \) is the Euler-Mascheroni constant. Similarly, observe that

\[
\frac{\partial \Gamma(k_m + 1, x)}{\partial k_m} \bigg|_{k_m=1} = \frac{\partial \Gamma(k_m, x)}{\partial k_m} + \frac{x_k e^{-x}}{\Gamma(k_m + 1)} \bigg|_{k_m=1} = e^{-x} (1 + x)(\ln x + \gamma) - xe^{-x} + E_1(x).
\]

Applying these derivative expressions, and noting \( \nabla_1 = e^{-x} \) and \( \nabla_2 = 1 - e^{-x} - xe^{-x} \) when \( k_m = 1 \), one obtains

\[
(\nabla_1 + \nabla_2)^2 \frac{d\varphi^m(x, k_m)}{dk_m} \bigg|_{k_m=1} = -xe^{-x} \left( E_1(x)(e^x - x) - 1 + e^{-x} + \ln x + \gamma \right) + \frac{x - 1 + e^{-x}}{1 + M} \left( E_1(x) + e^{-x} (\ln x + \gamma) \right).
\]

32
In the above expression, the terms in the first bracket, denoted by $\Theta_1(x) \equiv E_1(x)(e^x - x) - 1 + e^{-x} + \ln x + \gamma$, satisfy:

$$\lim_{x \to 0} \Theta_1(x) = \lim_{x \to 0} (E_1(x) + \ln x + \gamma) = \lim_{x \to 0} E_{in}(x) = 0,$$

where I used $\lim_{x \to 0} E_1(x) = \lim_{x \to 0} xe^{-x} = 0$ (by the l'Hospital rule) in the first equality, and $E_1(x) = -\gamma - \ln x + E_{in}(x)$ in the second equality, where

$$E_{in}(x) = \int_0^x (1 - e^{-t}) \frac{dt}{t}$$

is the entire function (see footnote 3, p.228 in Abramowitz and Stegun (1964)); $E_{in}(x) = \left\{ \begin{array}{ll} 0 & \text{for } x \in (0,1] \setminus \gamma \equiv 1 \equiv (1) (e^x - x) - 1 + e^{-x} \Theta_1(x) = \int_0^\gamma (1 - e^{-t}) dt \end{array} \right.$

Hence, $\Theta_1 > 0$ for all $x \in (0,1]$. The terms in the second bracket, denoted by $\Theta_2(x) \equiv E_1(x) + e^{-x}(\ln x + \gamma)$, satisfy:

$$\lim_{x \to 0} \Theta_2(x) = \lim_{x \to 0} (E_1(x) + \ln x + \gamma) = \lim_{x \to 0} E_{in}(x) = 0; \quad \Theta_2(1) = E_1(1) + e^{-1}\gamma > 1;$$

$$d\Theta_2(x) \over dx = -e^{-x}(\ln x + \gamma) \gg 0 \iff x \ll e^{-\gamma}; \quad \Theta_2(e^{-\gamma}) = E_1(e^{-\gamma}) > 0.$$

Hence, $\Theta_2(x)$ achieves the unique minimum at $x = 0$ within $x \in [0,1]$, which equals to zero, thereby $\Theta_2(x) > 0$ for all $x \in (0,1]$. Now, since $\Theta_1(x) > 0, \Theta_2(x) > 0$ for all $x \in (0,1]$, the condition of price increase is given by

$$d^2\Theta(x) \over dx^2 \mid_{x=1} > 0 \iff M < \frac{(x - 1 + e^{-x})\Theta_2(x) - x e^{-x}\Theta_1(x)}{x e^{-x}\Theta_1(x)}.$$

In what follows, I identify the values of $x = 1/(M + 1) \in (0,1]$ (and hence $M \in (0,\infty)$) that satisfy the condition of price increase (13). For this purpose, define

$$\Omega(x) \equiv (x - 1 + e^{-x})\Theta_2(x) - x e^{-x}\Theta_1(x).$$

Note the inequality (13) holds true if and only if $\Omega(x) > 0$. $\Omega(\cdot)$ satisfies: $\lim_{x \to 0} \Omega(x) = 0$;

$$\Omega(1) = e^{-1}(\Theta_2(1)) - \Theta_1(1)) = e^{-1}[-E_1(1)(e^1 - 2) + (1 - e^{-1})(1 - \gamma)] > 0$$

since $E_1(1)(e^1 - 2) \approx 0.22 \cdot 0.72 \approx 0.16 < 0.27 \approx (1 - e^{-1})(1 - \gamma)$;

$$\frac{d\Omega(x)}{dx} = e^{-x}\Theta_1(x) + (2(1 - e^{-x}) - x) e^{-x} (\ln x + \gamma).$$

From the last expression, it follows that $\lim_{x \to 0} \frac{d\Omega(x)}{dx} = \lim_{x \to 0} (2(1 - e^{-x}) - x) \ln x = 0$ (by using the l'Hospital’s rule twice) and $\frac{d\Omega(x)}{dx} > 0$ for $x > e^{-\gamma}$. To identify the sign of the derivative for $x \leq e^{-\gamma}$, suppose that $\Omega(x) \geq 0$ for $x \in (0,e^{-\gamma}]$. Then, it has to hold that $e^{-x}\Theta_1(x) \leq (x - 1 + e^{-x})\Theta_2(x)$, which further implies

$$\frac{d\Omega(x)}{dx} \leq (x - 1 + e^{-x})\Theta_2(x) + (2(1 - e^{-x}) - x) e^{-x} (\ln x + \gamma) = (x - 1 + e^{-x})E_1(x) + (1 - e^{-x})e^{-x} (\ln x + \gamma) \equiv \Psi(x)$$

for $x \in (0,e^{-\gamma}]$. Observe that $\lim_{x \to 0} \Psi(x) = 0$ (by using the l'Hospital’s rule thrice on the first term and twice on the second term) and $\Psi(e^{-\gamma}) > 0$. Further,

$$\frac{d\Psi(x)}{dx} = (1 - e^{-x})\Theta_2(x) - (2 - 3e^{-x}) e^{-x} (\ln x + \gamma) + \frac{e^{-x}(2(1 - e^{-x}) - x)}{x} \to -\infty < 0$$
as \( x \to 0 \). This implies there exists some \( x' \in (0, e^{-\gamma}) \) such that \( \Upsilon(x') = 0 \) and \( \Upsilon(x) < 0 \) for \( x < x' \).

The latter further implies \( \frac{d\Omega(x)}{dx} < 0 \) for \( x < x' \), a contradiction to \( \frac{d\Omega(x)}{dx} \geq 0 \), which must be the case if \( \Omega(x) \geq 0 \) and \( \lim_{x \to 0} \Omega(x) = 0 \) for an interval of \( x \) close to 0. Hence, \( \Omega(x) < 0 \) for an interval \( x \) close to zero. As \( \Omega(x) \) is continuous in \( x \in (0, 1) \) and \( \Omega(1) > 0 \), this implies that there exists some \( x^* \in (0, 1) \) such that \( \Omega(x^*) = 0 \) and \( \Omega(x) < 0 \) for \( x \in (0, x^*) \).

Observe now that \( \Omega(e^{-\gamma}) = -(2 - x - e^{-x} - xe^{-x})E_1(x) + e^{-x}(1 - e^{-x}) |_{x = e^{-\gamma}} \approx -0.55 \cdot 0.49 + 0.25 < 0 \). This implies, since \( \Omega(x) \) is increasing in \( x \in (e^{-\gamma}, 1) \), it has to be that \( x^* \in (e^{-\gamma}, 1) \). This further implies that \( \Omega(x) \) must cross the horizontal axis (of \( \Omega(\cdot) = 0 \)) from below and only once at \( x^* \in (e^{-\gamma}, 1) \). As \( \lim_{x \to 0} \Omega(x) < 0 < \Omega(1) \), it should hold that

\[
\Omega(x) \leq 0 \quad \text{for} \quad x < x^* \in (0, 1) \quad \text{and} \quad \Omega(x) > 0 \quad \text{for} \quad x > x^*.
\]

Therefore, the condition of price increase (13) holds true if and only if \( x \in (x^*, 1) \), and since \( x = X \) when \( k_m = 1 \), this proves the first claim in the proposition with \( x^* = X^* \in (0, 1) \).

To prove the second claim, it is sufficient to observe that since \( x_m \to 1/M \), \( x_m \eta(x_m, k_m) \to 1/M \), \( k_m \nabla_2 \to 0 \) as \( k_m \to \infty \), it holds that \( \varphi^m(x_m, k_m) \to 0 \) as \( k_m \to \infty \). ■

\( \circ \) **Retail price of sellers** \( p_s \): It is sufficient to observe that \( x_s(\cdot) \) is strictly decreasing in all \( k_m \geq 1 \) (as shown in Step 1 in the proof of Theorem 1) and \( p_s(= \varphi^m(\cdot)) \) is strictly increasing in all \( x_s \in (0, 1) \) (as shown in the proof of Proposition 1).

\( \circ \) **Retail market premium** \( p_m - p_s \): Remember that with \( \beta = 0, (3) \) for \( i = s, m \) imply

\[
p_m - p_s = \varphi^m(\cdot) - \varphi^s(\cdot) = \frac{k_m}{x_m} \left( \frac{1 - \Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)} \right) - \frac{x_m(1 - e^{-x} - xe^{-x})}{\eta(x_s, 1)}.
\]

From this, it follows that \( \eta(x_m, k_m) \eta(x_s, 1)(p_m - p_s) \)

\[
= \left[ \frac{k_m}{x_m} \left( 1 - \frac{1}{\Gamma(k_m + 1)} \right) - \frac{x_m(1 - e^{-x} - xe^{-x})}{\eta(x_s, 1)} \right] e^{-x_s}.
\]

\[
\left[ \frac{1 - e^{-x_s}(k_m x_s - x_m)}{x_s} + \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} - \frac{x_m e^{-x_m}}{\Gamma(k_m)} \right] e^{-x}.
\]

\[
= \left[ \frac{(1 - e^{-x_s})(\Gamma(k_m + 1, x_s) - 1) + \frac{1}{\Gamma(k_m - 1)} x_s(1 - x_s)}{\Gamma(k_m - 1)} \right] e^{-x_s}.
\]

where I have used (8) for the first equality, (8) and \( \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)} = \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} + \frac{x_m e^{-x_m}}{\Gamma(k_m + 1)} \) for the second equality, and (1) and \( \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} = \frac{\Gamma(k_m - 1, x_m)}{\Gamma(k_m - 1)} + \frac{x_m e^{-x_m}}{\Gamma(k_m + 1)} \) for the third equality. Define \( \Lambda_x(x_s) \) as the parenthesis terms in the last expression above for \( x_s \in (0, 1) \) and \( k_m > 1 \). Then, it satisfies \( \Lambda_x(x) \to 0 \) as \( x_s \to 0 \), \( \Lambda_x(x) \to (1 - e^{-1})MK_m > 0 \) as \( x_s \to 1 \), and

\[
\frac{d\Lambda_x(x_s)}{dx_s} = \frac{MK_m(1 - e^{-x_s} - xe^{-x_s}) + 2x_s e^{-x_s} - \frac{\Gamma(k_m - 1, 1 - x_s)}{\Gamma(k_m - 1)} (3x_s - 1)}{\Gamma(k_m - 1)} > 0,
\]

where I have used \( \frac{\Gamma(k_m - 1, x_s)}{\Gamma(k_m - 1)} < \frac{\Gamma(k_m, x_s)}{\Gamma(k_m)} (= e^{-x_s} \text{ by (8)}) \) for the first inequality. Hence, \( p_m - p_s > 0 \) for all \( k_m > 1 \). Since \( p_m - p_s = 0 \) when \( k_m = 1 \) and \( p_m - p_s \to 0 \) as \( k_m \to \infty \), this implies that the retail premium is increasing in low \( k_m \) and is decreasing in high \( k_m \). ■
Proof of Proposition 3

With the fixed total supply of middlemen $G = Mk_m$, the only modification appears in the adding-up restriction (1), which now becomes (with normalization $S = 1$)

$$\frac{G}{k_m} x_m + x_s = 1.$$ 

This affects the analysis in Step 1 in the proof of Theorem 1, so that now I have

$$\frac{dx_m}{dk_m} = \frac{\partial \left(\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)}\right)}{k_m x_m} + \frac{G x_m e^{-x_s}}{k_m}.$$

Observe that there is an additional, positive term in the numerator of this expression. This modification further affects the following parts of the analysis: the derivative in question becomes

$$(\nabla_1 + \nabla_2)^2 \frac{d^2 \phi(x, k_m)}{dk_m^2} \bigg|_{k_m = 1 \& \ G = M k_m} = -xe^{-x} \left( E_1(x)(e^x - x) - 1 + e^{-x} \ln x + \gamma \right) + \frac{x - 1 + e^{-x}}{1 + G} \left( E_1(x) + e^{-x} \ln x + \gamma \right) + Gxe^{-x},$$

where a positive term is added inside the second bracket; the condition for price increase (13) is then modified to

$$G \left( 1 - \frac{x - (1 - e^{-x})}{\Theta_1(x)} \right) < \frac{(x - 1 + e^{-x}) \Theta_2(x) - xe^{-x} \Theta_1(x)}{xe^{-x} \Theta_1(x)}.$$

Observe here that the R.H.S. remains the same as before, while the L.H.S. is now multiplied by a new term which is less than one. As $G = M$ when $k_m = 1$, this implies that the above inequality holds for all $x \in (x^*, 1)$ (see the proof of Proposition 2) and so $\frac{d^2 \phi(x, k_m)}{dk_m^2} |_{k_m = 1 \& \ G = M k_m} > 0$ for all $x \in (x^*, 1)$. This proves the first claim in the proposition.

The second claim can be shown by using the following property (see Temme (1996) p.285):

$$\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \to D \quad \text{as} \quad k_m \to \infty$$

where $D \in [0, 1]$ satisfies: $D = 1$ if and only if $x_m < k_m$; $D = 0$ if and only if $x_m > k_m$.

Throughout the proof given below, keep in mind that with the fixed total supply $G = Mk_m \in (0, \infty)$, it has to be that $M/G/k_m \to 0$ as $k_m \to \infty$, thus $x_m \to \infty$ as $k_m \to \infty$. There are three cases. Consider first the case $G < 1$. Suppose $x_m > k_m$ as $k_m \to \infty$. This leads to $\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \to 0$ as $k_m \to \infty$ by (14) and so $x_s \to \infty$ as $k_m \to \infty$ by (8). However, this contradicts to (1) which requires $x_s \in [0, 1]$. Suppose $x_m < k_m$ as $k_m \to \infty$. Then, $\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \to 1$ as $k_m \to \infty$ by (14) and so $x_s \to 0$ as $k_m \to \infty$ by (8). However, this contradicts to (1) and $G < 1$, or

$$M(x_m - k_m) + x_s = 1 - G > 0$$

which requires $x_s > 0$, if $x_m < k_m$. Therefore, the only possible solution when $G < 1$ is $x_m = k_m$ as $k_m \to \infty$, which in turn leads to $x_s = 1 - G$ by (1), as is consistent with (14), requiring $\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} = e^{-x_s} \in (0, 1)$ as $k_m \to \infty$ and $x_m = k_m$. In this solution, it holds that:

$$\eta(x_m, k_m) = \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} = \frac{k_m + x_m \left( 1 - \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \right)}{x_m} = \frac{x_m k_m^{-1} e^{-x}}{\Gamma(k_m)} \to 1 \quad \text{as} \quad k_m \to \infty$$

because $\frac{x_m k_m^{-1} e^{-x}}{\Gamma(k_m)} \to 0$ as $k_m \to \infty$ for any $\frac{x_m}{k_m} \in (0, \infty)$; for all $k_m \geq 1$,

$$\frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} < \lim_{k_m \to \infty} \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)},$$

which completes the proof of the second part of Proposition 3.

Therefore, the conclusion holds for all middlemen $M$.
which follows from \( x^m > \lim_{k_m \to \infty} x^s = 1 - G \) for all \( k_m \geq 1 \), or \( e^{-x^m} < \lim_{k_m \to \infty} e^{-x^s} = e^{-(1-G)} \) for all \( k_m \geq 1 \) in (8). It then follows that for all \( k_m \geq 1 \),

\[
\varphi^m(x_m, k_m) = 1 - \frac{\Gamma(k_m-1, x_m)}{\Gamma(k_m)} > 1 - \lim_{k_m \to \infty} \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} = \lim_{k_m \to \infty} \varphi^m(x_m, k_m).
\]

Consider next the case \( G = 1 \). Suppose \( x_m < k_m \) as \( k_m \to \infty \). Then, \( \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \to 1 \) as \( k_m \to \infty \) by (14) and so \( x_s \to 0 \) as \( k_m \to \infty \) by (8). However, this contradicts to (1) and \( G = 1 \), or

\[
M(x_m - k_m) + x_s = 1 - G = 0
\]

which requires \( x_s > 0 \), if \( x_m < k_m \). Similarly, \( x_m \geq k_m \) as \( k_m \to \infty \) cannot be the solution. Therefore, there is no limiting solution as \( k_m \to \infty \) with the fixed total supply \( G = M k_m \) when \( G = 1 \).

Consider finally the case \( G > 1 \). Then, by (1),

\[
M(x_m - k_m) + x_s = 1 - G < 0,
\]

implying that \( x_m < k_m \) as \( k_m \to \infty \), leading to \( \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \to 0 \) as \( k_m \to \infty \) by (8) and (14), is the only solution. Therefore, \( 1 - \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} \to 0 \) as \( k_m \to \infty \), which implies \( \varphi^m(\cdot) \to 0 \) as \( k_m \to \infty \) and thus \( \varphi^m(\cdot) \geq \lim_{k_m \to \infty} \varphi^m(\cdot) \) for all \( k_m \geq 1 \).

---

**Proof of Proposition 4**

○ Retail market premium \( p_m - p_s \): It is sufficient to remember that (3) for \( i = s, m \) imply

\[
p_m - p_s = (\varphi^m(\cdot) - \varphi^s(\cdot)) N,
\]

where \( N > 0 \) for all \( \beta \in [0, 1] \) and \( k_m \geq 1 \), and that the proof of Proposition 2 shows that \( \varphi^m(\cdot) - \varphi^s(\cdot) = 0 \) for \( k_m = 1 \), \( \varphi^m(\cdot) - \varphi^s(\cdot) > 0 \) for \( k_m > 1 \), and \( \varphi^m(\cdot) - \varphi^s(\cdot) \to 0 \) as \( k_m \to \infty \).

○ Part 1 (Without fixed total supply): Differentiation yields

\[
(\varphi^m(x_m, k_m)N)^{-1} \frac{d\varphi^m(x_m, k_m)N}{dk_m} = \frac{d\varphi^m(x_m, k_m)/dk_m}{\varphi^m(x_m, k_m)} + \frac{dN/dk_m}{N}.
\]

Using the result obtained in the proof of Proposition 2, the first term in the R.H.S. above is computed as

\[
\frac{d\varphi^m(\cdot)/dk_m}{\varphi^m(\cdot)} \bigg|_{k_m=1} = \frac{-xe^{-x} \Theta_1(x) + e^{-x}(x-1 + e^{-x}) \frac{dx_m}{dk_m}}{(1 - e^{-x} - xe^{-x})(1 - e^{-x})},
\]

where \( \frac{dx_m}{dk_m} \bigg|_{k_m=1} = \frac{e^{-\Theta_2(x)}}{e^{\Theta_2(x)}(1+M)} \), and the second term as

\[
\frac{dN/dk_m}{N} \bigg|_{k_m=1} = \frac{\beta e^{-x}(1-x)M \frac{dx_m}{dk_m}}{1 - \beta e^{-x}} \frac{dx_m}{dk_m} \bigg|_{k_m=1}.
\]

Here, \( \Theta_1(x) \equiv E_1(x)(e^x - x) - 1 + e^{-x} + \ln x + \gamma > 0 \) and \( \Theta_2(x) \equiv E_1(x) + e^{-x}(\ln x + \gamma) > 0 \) are introduced in the proof of Proposition 2. Combining these two terms,

\[
\frac{(1 - e^{-x})^2}{N} \frac{d\varphi^m(\cdot)N}{dk_m} \bigg|_{k_m=1} = -xe^{-x} \Theta_1(x) + \frac{x - (1 - e^x)}{1 + M} \Theta_2(x) - \frac{M}{1 + M} \Theta_2(x) \Theta_3(x).
\]
where
\[ \Theta_3(x, \beta) \equiv \beta (1 - x) (1 - e^{-x}) (1 - e^{-x} - xe^{-x}) / (1 - \beta xe^{-x}) > 0 \]

for all \( x \in (0, 1) \), \( \beta \in [0, 1) \).

The condition for price increase is now given by
\[ \frac{d\varphi^m(x, k_m) N}{dk_m} \bigg|_{k_m = 1} > 0 \iff M < \frac{(x - 1 + e^{-x}) \Theta_2(x) - xe^{-x} \Theta_1(x)}{xe^{-x} \Theta_1(x)} \]

(15).

Observe that the denominator in the R.H.S. of (15) is positive and the numerator is identical to \( \Omega(\cdot) \) given in the proof of Proposition 2, which satisfies \( \Omega(1) > 0 \). Hence, the above inequality holds with sufficiently high (low) values of \( x = 1/(M+1) < 1 \) (\( M > 0 \)). Further, the condition of price increase (15) can be written as
\[ M \left( 1 + \frac{\Theta_2(x) \Theta_3(x, \beta)}{xe^{-x} \Theta_1(x)} \right) < \frac{(x - 1 + e^{-x}) \Theta_2(x) - xe^{-x} \Theta_1(x)}{xe^{-x} \Theta_1(x)} \]

(16).

Comparing it with the previous condition (13), one can see that the R.H.S. of this expression (15) is the same as before, but now the L.H.S. is multiplied by a positive term, which is no less than one. This implies, the critical value, denoted by \( x^*(\beta) < 1 \), above which (16) holds true, should be such that \( x^*(\beta) \geq x^* \), with equality if and only if \( \beta = 0 \) (since \( \Theta_3(x, 0) = 0 \)). This proves the first claim in the proposition (with \( x^*(\beta) = X^*(\beta) \in (0, 1) \)).

The second claim follows from \( N \rightarrow 1 - \beta \) and \( \varphi^m(x_m, k_m) \rightarrow 0 \) as \( k_m \rightarrow \infty \). □

○ Part 2 (With fixed total supply \( G = Mk_m \)): As in the proof of Proposition 3, the fixed total supply of middlemen \( G = Mk_m \) leads to the modification,
\[ \frac{dx_m}{dk_m} \bigg|_{k_m = 1 \ and \ G = Mk_m} = \frac{E_1(x) + e^{-x} (\ln x + \gamma) + Gxe^{-x}}{e^{-x}(1 + G)} > 0, \]

which further modifies the derivative in question (derived in Part 1) to
\[ \frac{(1 - e^{-x})^2}{N} \frac{d\varphi^m(\cdot) N}{dk_m} \bigg|_{k_m = 1 \ and \ G = Mk_m} = -xe^{-x} \Theta_1(x) + \frac{x - (1 - e^{-x})}{1 + G} (\Theta_2(x) + Gxe^{-x}) - \frac{G}{1 + G} (\Theta_2(x) - xe^{-x}) \Theta_3(x), \]

and the condition (15) to
\[ G \left( 1 + \frac{\Theta_2(x) \Theta_3(x, \cdot)}{xe^{-x} \Theta_1(x)} \right) \frac{x - (1 - e^{-x}) + \Theta_3(x, \cdot)}{\Theta_1(x)} < \frac{(x - 1 + e^{-x}) \Theta_2(x) - xe^{-x} \Theta_1(x)}{xe^{-x} \Theta_1(x)} \]

(15).

Comparing it with the previous condition (16), one can see that the R.H.S. of this expression is the same as before, but now the L.H.S. is smaller than that of (16). This implies, the above inequality holds for all \( x \in (x^*(\beta), 1) \), proving the first claim in the proposition (with \( x^*(\beta) = X^*(\beta) \in (0, 1) \)).

The second claim follows from \( N \rightarrow \lim_{k_m \rightarrow \infty} N \) for all \( k_m \geq 1 \) (as \( N \) is decreasing in \( k_m \)) and the results obtained in the proof of Proposition 3, showing that \( \varphi^m(\cdot) \geq \lim_{k_m \rightarrow \infty} \varphi^m(\cdot) \) for all \( k_m \geq 1 \). □
Proof of Proposition 5

From the free entry condition (9), the fixed point condition for the equilibrium number of middlemen $M \in (0, \infty)$ is given by

$$
\Phi_m(M, \cdot) \equiv k_m^{1-\alpha} \left( 1 - \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)} \right) \frac{1 - \beta}{1 - \beta x \alpha e^{-x}} = c
$$

(17)

where $x_i = x_i(M), i = s, m$, is strictly decreasing in $M$ and satisfies $x_i \to 0$ as $M \to \infty$, as shown in Step 1 in the proof of Proposition 1. It then follows that $\Phi_m = \Phi_m(M, \cdot)$ is continuous and strictly decreasing in $M \in (0, \infty)$ and satisfies $\Phi_m \to -c k_m < 0$ as $M \to \infty$. Therefore, with $c \equiv \lim_{M \to 0} \Phi_m \in (0, \infty)$, there exists a unique $M \in (0, \infty)$ that satisfies (17) given $c \in (0, c)$. The comparative statistics of $c$, which is negative on $M$ satisfying $M \to \infty$ as $c \to 0$ and $M \to 0$ as $c \to c$, and of $\alpha$, which is negative satisfying $M > 0$ at $\alpha = 0$ and $M \to 0$ as $\alpha \to \infty$, is immediate.

For the comparative statics of $k_m$, observe that

$$
N^{-1} \frac{d\Phi_m}{dk_m} = (1 - \alpha) k_m^{-\alpha} \nabla_2 - k_m^{-\alpha} \frac{\partial}{\partial k_m} \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)}
$$

+ $\left[ \frac{x_m}{\Gamma(k_m + 1)} \right] \frac{\partial}{\partial k_m} \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)}

where I used (1) and $N = \frac{1 - \beta}{1 - \beta e^{-x}}$, and $\nabla_2 \equiv 1 - \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)}$ is introduced in the proof of Proposition 2. Evaluating this derivative at $k_m = 1$,

$$
N^{-1} \frac{d\Phi_m}{dk_m} \big|_{k_m = 1} = -\Theta_4(x) + x \left( \frac{\Theta_3(x)}{1 - e^{-x}} \right) \Theta_2(x)
$$

where $\Theta_2(x) \equiv E_1(x) + e^{-x}(\ln x + 1)$ is introduced in the proof of Proposition 2, $\Theta_3(x, \beta) \equiv \beta(1-x)(1-e^{-x})(1-e^{-x-\beta}) > 0$ in the proof of Proposition 3, and now

$$
\Theta_3(x) \equiv E_1(x) + e^{-x}(\ln x + 1) + \alpha(1 - e^{-x} - x e^{-x}) - (1 - e^{-x}) > 0
$$

for all $x > 0$, since $\lim_{x \to 0} \Theta_3(x) = 0$, $\Theta_3(1) = E_1(1) + 2e^{-1} + \alpha(1 - 2e^{-1}) - (1 - e^{-1}) \geq E_1(1) + 2e^{-1} - (1 - e^{-1}) \simeq 0.22 + 0.42 - 0.63 > 0$, and $\frac{d\Theta_3(x)}{dx} = -xe^{-x}(\ln x + 1 - \alpha) > 0$ if and only if $x < e^{\alpha - 1}$. Applying $M = \frac{1}{x} \frac{\Phi_m}{\Phi_m} \big|_{k_m = 1} > 0 \iff \Omega_m(x) \equiv \frac{-\Theta_4(x)}{x \Theta_2(x)} + x \left( \frac{1}{x} \Theta_3(x) \right) > 0$.

Observe that $\Theta_2(1) - \Theta_3(1) = -\alpha(1 - 2e^{-1}) + (1 - e^{-1}) \geq 0$, implying $\Omega_m(1) > 0$ if and only if $\alpha < \alpha^* \equiv \frac{1 - e^{-1} + (1 + \alpha)}{1 - 2e^{-1}} \simeq 1.61$, whereas $\lim_{x \to 0} \Theta_4(x) \Theta_2(x) \equiv \frac{1}{2}$ (by using the l'Hospital’s rule twice) and $\lim_{x \to 0} \Theta_4(x) \Theta_2(x) = 0$ (by using the l'Hospital’s rule once), implying $\lim_{x \to 0} \Omega_m(x) = -\frac{1}{2} < 0$. Hence, the above inequality holds for relatively high $x$ (or low $M$) and $\alpha < \alpha^*$, but not for relatively low $x$ (or high $M$). As there is one-to-one negative relationship between $M$ and $c$, and $\Phi_m$ is strictly decreasing in $M$, it then follows that $M$ is increasing in $k_m$ for relatively large $c$ and only if $\alpha < \alpha^*$, and $M$ is decreasing in $k_m$ for relatively small $c$, at $k_m = 1$.

To demonstrate the result with large $k_m$, rewrite the fixed point condition (17) as

$$
1 - \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)} \frac{1 - \beta}{1 - \beta x \alpha e^{-x}} = c k_m^{\alpha - 1}.
$$

(18)

There are three cases depending on the value of $\alpha \in [0, \infty)$. Consider first the case $\alpha = 1$. In this case, the R.H.S. of (18) is constant $c \in (0, 1)$, thus if $\alpha = 1$ then it has to hold that

$$
\frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)} \in (0, 1)
$$

as $k_m \to \infty$. Note $\frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)} \to 0$ is impossible since it requires $x_m \to \infty$ by (8), violating (1) (see the
property (14) in the proof of Proposition 3). From the property (14), it follows that \( x_m = k_m \) as \( k_m \to \infty \). Since (1) implies, \( x_m \to \infty \) is possible only when \( M \to 0 \), the result follows: \( M > \lim_{k_m \to \infty} M \) for all \( k_m \geq 1 \). Consider next the case \( \alpha < 1 \). Observe that the R.H.S. of (18) approaches to 0 as \( k_m \to \infty \) if \( \alpha < 1 \). Thus, for \( \alpha < 1 \), it has to hold that \( \frac{\Gamma(k_m+1,x_m)}{\Gamma(k_m+1)} \to 1 \) as \( k_m \to \infty \), which holds true if and only if \( x_m < k_m \) as \( k_m \to \infty \). In this case, (8) requires that \( x_m \to 0 \) as \( k_m \to \infty \), implying \( x_m \to 1/M \) and \( M \to 0 \) as \( k_m \to \infty \) by (1). Therefore, \( M > \lim_{k_m \to \infty} M \) for all \( k_m \geq 1 \).

Consider finally the case \( \alpha > 1 \). In this case, an extra care is needed on the upper bound of the per-unit cost, since it gets arbitrary small for large \( k_m \); \( c \equiv \lim_{M \to 0}(1 - \beta)\frac{V^m}{k_m} \to 0 \) as \( k_m \to \infty \). Hence, the value of \( c > 0 \) needs to be kept small to consider large \( k_m \). For this purpose, define for given values of \( c \in (0, \bar{c}) \) the upper bound \( k_m \geq 1 \) to be such that

\[
ck_m = \lim_{M \to 0} \left( 1 - \frac{\Gamma(k_m+1,x_m)}{\Gamma(k_m+1)} \right) \frac{1 - \beta}{\beta xe^{-x}} \in (0,1).
\]

Then, consider the case as \( k_m \to k_m \) and \( c \to 0 \), which implies \( k_m \to \infty \) and \( c k_m^{-1} \in (0,1) \). Then, since the L.H.S. of (18) has to lie within \((0,1)\), the limit as \( k_m \to k_m \) has to accompany \( \frac{\Gamma(k_m+1,x_m)}{\Gamma(k_m+1)} \in (0,1) \), as in the case \( \alpha = 1 \), thus the result obtained there applies: \( M \to 0 \) as \( k_m \to \infty \). Thus, there exists some \( c \in (0, \bar{c}) \) such that \( M > 0 \) is decreasing in sufficiently large \( k_m < \infty \) for \( c \in (0, \bar{c}) \).

**Proof of Proposition 6**

Using the analysis in the proof of Proposition 5, one can compute the derivative,

\[
\frac{dM}{dk} \bigg|_{k_m=1} = -\frac{d\Phi_m}{dM} \bigg|_{k_m=1} = -\frac{\Theta_1(x) + \left( x^2 - \frac{(1-x)\Theta_2(x)}{e^{-x}} \right) \Theta_2(x)}{x^2 e^{-x} \left( x + \frac{\Theta_1(x)}{1-e^{-x}} \right)}.
\]

This leads to the followings:

\[
\frac{dx_m}{dk} \bigg|_{k_m=1} = \frac{x \Theta_2(x) + \frac{dM}{dk} \bigg|_{k_m=1} x^2 e^{-x}}{e^{-x}} = \frac{\Theta_1(x) + \frac{\Theta_2(x) \Theta_3(x)}{1-e^{-x}}}{e^{-x} \left( x + \frac{\Theta_1(x)}{1-e^{-x}} \right)} > 0;
\]

\[
\frac{dx_s}{dk} \bigg|_{k_m=1} = -\frac{dM}{dk} \bigg|_{k_m=1} - M \frac{dx_m}{dk} \bigg|_{k_m=1} = -\frac{x \Theta_2(x) - \Theta_1(x)}{e^{-x} \left( x + \frac{\Theta_1(x)}{1-e^{-x}} \right)}.
\]

To repeat, \( \Theta_2(x) \equiv E_1(x) + e^{-x}(\ln x + \gamma) > 0 \), \( \Theta_3(x, \beta) \equiv \frac{\beta(1-x)(1-e^{-x})(1-e^{-x}-xe^{-x})}{1-\beta e^{-x}} > 0 \), and \( \Theta_4(x) \equiv E_1(x) + e^{-x}(1 + x)(\ln x + \gamma) + \alpha(1 - e^{-x} - xe^{-x}) - (1 - e^{-x}) > 0 \).

Using the above expressions, the derivative in question (see the proof of Proposition 4) can be computed as

\[
\frac{(1 - e^{-x})^2}{N} \frac{d(e^m(x,k_m)N)}{dk_m} \bigg|_{k_m=1} = -xe^{-x} \Theta_1(x) + (x - (1 - e^{-x})) \left( \Theta_1(x) + \frac{\Theta_2(x) \Theta_3(x)}{1-e^{-x}} \right) - \Theta_3(x) \left( x \Theta_2(x) - \Theta_1(x) \right) \frac{x \Theta_2(x) - \Theta_1(x)}{x + \frac{\Theta_1(x)}{1-e^{-x}}},
\]

where \( \Theta_1(x) \equiv E_1(x)(e^x - x) - 1 + e^{-x} + \ln x + \gamma \). The condition for the spread increase is now given by

\[
\Omega_f(x) \equiv (x - (1 - e^{-x})) \Theta_4(x) - x^2 e^{-x} \Theta_1(x) + \frac{\Theta_3(x)}{1-e^{-x}} \left[ (1 - e^{-x}) \Theta_4(x) - (1 - e^{-x} - xe^{-x}) \Theta_2(x) - xe^{-x} \Theta_1(x) \right] > 0.
\]

39
Notice that term \( \Theta_2(x) \) can be made arbitrary large with \( \alpha \), while the other terms are bounded above with large \( \alpha \). Hence, it holds that \( \Omega_f(x) > 0 \) for all \( c \in (0, \hat{c}) \), if \( \alpha \) is sufficiently large. In the case \( x \to 1 \) (so \( \Theta_3(x) \to 0 \)) as \( c \to \hat{c} \), we have \( \Omega_f(x) \to \Omega_f(1) = (1 - 2e^{-1}) (\alpha - E_1(1)e^{-1} - \gamma) > 0 \), as \( c \to \hat{c} \), which holds true if and only if \( \alpha > E_1(1)e^{-1} + \gamma \) (\( \simeq 1.18 \)). Therefore, the bid-ask spread \( p_m - \beta V^* \) is increasing in \( k_m \) at \( k_m = 1 \) for all \( c \in (0, \hat{c}) \), if \( \alpha \) is sufficiently large.

To examine the effect of large \( k_m \)'s, it is convenient to write

\[
p_m - \beta V^* = \frac{k_m (1 - \frac{\Gamma(k_m+1,x_m)}{\Gamma(k_m+1)})}{x_m \eta(x_m,k_m) / k_m} 1 - \beta x_s e^{-x_s} = \frac{c k_m^{\alpha-1}}{x_m \eta(x_m,k_m) / k_m},
\]

using (18). As shown in the proof of Proposition 5, when \( \alpha = 1 \), it holds that \( x_m = k_m \) and \( \eta(\cdot) \to 1 \) as \( k_m \to \infty \). Hence, the denominator of the last expression above approaches 1, the highest possible value, while the numerator is constant at \( e \) when \( \alpha = 1 \). Therefore, \( p_m - \beta V^* > \lim_{k \to \infty} p_m - \beta V^* \) for all \( k_m \geq 1 \). When \( \alpha < 1 \), it holds that \( x_m/k_m > 0 \) and \( \eta(\cdot) \to 1 \) as \( k_m \to \infty \), while \( c k_m^{\alpha-1} \to 0 \) as \( k_m \to \infty \), thereby the result follows.

When \( \alpha > 1 \), define \( \hat{k}_m \) to be such that

\[
\hat{k}_m^{\alpha-1} = \lim_{M \to 0} \left( 1 - \frac{\Gamma(k_m+1,x_m)}{\Gamma(k_m+1)} \right) 1 - \beta x_s e^{-x_s} = \frac{c k_m^{\alpha-1}}{x_m \eta(x_m,k_m) / k_m},
\]

and \( \epsilon > 0 \) to be such that \( \hat{k}_m^{\epsilon} = \frac{k_m}{x_m \eta(x_m,k_m) / k_m} \). Then, as \( k_m \to \hat{k}_m \) and \( c \to 0 \), it has to be that \( \hat{k}_m \to \infty \)

\[
\frac{c k_m^{\alpha-1}}{x_m \eta(x_m,k_m) / k_m} \to \frac{c k_m^{\alpha-1}}{x_m \eta(x_m,k_m) / k_m} = \hat{k}_m^{\epsilon} > \frac{c k_m^{\alpha-1}}{x_m \eta(x_m,k_m) / k_m}
\]

for any \( k_m < \hat{k}_m \), since \( \frac{k_m}{x_m \eta(x_m,k_m)} = \hat{k}_m^{\epsilon} > \frac{k_m}{x_m \eta(x_m,k_m)} \) for any \( k_m \geq 1 \) as \( c \to 0 \). Therefore, there exists some \( \epsilon \in (0, \hat{c}) \) such that \( p_m - \beta V^* \) increases in sufficiently large \( k_m \) for all \( c \in (0, \hat{c}) \). ■

**Proof of Proposition 7**

Observe that

\[
T - \bar{T} = M x_m \eta(x_m,k_m) + x_s \eta(x_s,1) - S \bar{Z}_s \eta(Z_s,1) = M x_m (\eta(x_m,k_m) - \eta(x_s,1)) + (\eta(x_s,1) - \eta(Z_s,1)),
\]

where the second equality is by (1) and \( S \bar{Z}_s = \frac{1}{M k_m} (M k_m + 1) \). For \( k_m = 1 \), it holds that \( x_m = x_s = \bar{Z}_s = \frac{1}{M+1} \), implying \( \eta(x_m,k_m) = \eta(x_s,1) \) and \( \eta(x_s,1) = \eta(Z_s,1) \), thereby \( T = \bar{T} \). For \( k_m > 1 \), \( \eta(x_m,k_m) > \eta(x_s,1) \) and \( \eta(x_s,1) > \eta(Z_s,1) \), since \( \eta(x_s,1) \) is decreasing in \( x_s \) and \( x_s < Z_s \) for \( k_m > 1 \), thereby \( T > \bar{T} \). ■

**Proof of Proposition 8**

Differentiating \( T = M x_m \eta(x_m,k_m) + x_s \eta(x_s,1) \) with respect to \( k_m \in \mathbb{R}_+ \),

\[
\frac{dT}{dk_m} = M x_m \frac{\partial \eta(\cdot)}{\partial k_m} + M \frac{dx_m}{dk_m} \frac{\Gamma(k_m,x_m)}{\Gamma(k_m)} + \frac{dx_s}{dk_m} e^{-x_s} = M x_m \frac{\partial \eta(\cdot)}{\partial k_m} > 0,
\]

40
where the second equality follows from (1) and (8), proving the first claim. Similarly,

\[
\frac{dT}{dM} = x_m \eta(\cdot) + M \frac{dx_m}{dM} \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} + \frac{dx_s}{dM} e^{-x_s} = x_m \eta(\cdot) > 0,
\]

proving the second claim.

To examine the effect of small \( k_m \)'s on \( T \) given fixed supply \( G = M k_m \), observe that

\[
\frac{dT}{dk_m} \bigg|_{G=M k_m} = M x_m \left[ \frac{\partial}{\partial k_m} \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} - \frac{k_m}{x_m} \frac{\partial}{\partial k_m} \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m + 1)} \right].
\]

Evaluating this derivative at \( k_m = 1 \),

\[
\frac{dT}{dk_m} \bigg|_{G=M k_m} = M \left[ e^{-x}(x - \ln x - \gamma) - (1 - x)E_1(x) \right],
\]

where \( x \equiv x_m = \frac{1}{r+1} \in (0, 1) \) at \( k_m = 1 \). Define the parenthesis terms above as \( \Psi_T(x) \) for \( x \in (0, 1) \).

It satisfies: \( \lim_{x \to 0} \Psi_T(x) = \lim_{x \to 0} - (E_1(x) + \ln x + \gamma) = \lim_{x \to 0} - E_1(x) = 0 \) (see the proof of Proposition 2); \( \lim_{x \to 1} \Psi_T(x) = e^{-1}(1 - \gamma) > 0 ; \)

\[
\frac{d\Psi_T(x)}{dx} = - e^{-x}(x - \ln x - \gamma) + E_1(x).
\]

In the last expression, observe that \( \frac{d^2\Psi_T}{dx^2} \to 0 \) as \( x \to 0 \) and \( \frac{d^2\Psi_T}{dx^2} \to e^{-1}(\gamma - 1) + E_1(1) \approx 0.37 \ast (0.58 - 1) + 0.22 \approx 0.064 > 0 \), and that \( \frac{d^2\Psi_T}{dx^2} = - e^{-x}(1 - x + \ln x + \gamma) \), satisfying \( \frac{d^2\Psi_T}{dx^2} \to +\infty > 0 \) as \( x \to 0 \), \( \frac{d^2\Psi_T}{dx^2} \to - e^{-1} \gamma < 0 \) as \( x \to 1 \) and \( \frac{d^2\Psi_T}{dx^2} = 0 \) at some \( \hat{x} \in (0, 1) \) such that \( \hat{x} = 1 + \ln \hat{x} + \gamma \). Then, \( \frac{d^2\Psi_T}{dx^2} = - e^{-x} + E_1(\hat{x}) > 0 \) implies \( \frac{d^2\Psi_T}{dx^2} > 0 \) for all \( x \in (0, 1) \), which further implies \( \Psi_T > 0 \) for all \( x \in (0, 1) \) and so \( \frac{dT}{dk_m} \bigg|_{G=M k_m} > 0 \) for all \( G = M k_m \in (0, \infty) \).

Finally, to examine the effect of large values of \( k_m \) for given fixed supply, it is convenient to write

\[
T = M x_m \eta(x_m, k_m) + x_s \eta(x_s, 1) = (1 - x_s) e^{-x_s} + (G + 1)(1 - e^{-x_s}) - \frac{x_m e^{-x_m}}{\Gamma(k_m)} G,
\]

using (1), (8), and \( \frac{\Gamma(k_m + 1, x_m)}{\Gamma(k_m)} = \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)} + \frac{x_m e^{-x_m}}{\Gamma(k_m + 1)} \). Note that the last term above approaches to zero as \( k_m \to \infty \). There are two cases. Suppose \( G > 1 \). Then, \( x_s \to 0 \) as \( k_m \to \infty \) (see the proof of Proposition 3), and so \( T \) approaches to one, the highest possible values, as \( k_m \to \infty \). Suppose next \( G < 1 \). Then, \( x_s \to 1 - G > 0 \) and \( T \to 2 - x_s - e^{-x_s} \) as \( k_m \to \infty \). Since the term \( 2 - x - e^{-x} \) is monotone decreasing in \( x \in (0, 1) \) and \( 1 - G > 0 \) is the lowest possible value \( x_s \) can approaches to given \( G < 1 \), this implies that \( T \) approaches to the highest possible value as \( k_m \to \infty \). ■

**Proof of Proposition 9**

Since the welfare \( W \) is monotone decreasing in all \( x_s \in (0, 1) \), the first claim is immediate from the fact that \( x_s \) is strictly decreasing in all \( k_m \geq 1 \) and \( M \in (0, \infty) \) (as shown in the proof of Theorem 1). To prove the second claim on \( W \) for fixed supply \( G = M k_m \), observe that

\[
\frac{dx_s}{dk_m} \bigg|_{G=M k_m} = \frac{\frac{\partial}{\partial k_m} \frac{\Gamma(k_m, x_m)}{\Gamma(k_m)}}{\frac{x_m e^{-x_m}}{\Gamma(k_m)} + e^{-x_s}}.
\]

(see the proof of Proposition 3). The denominator of the above expression is clearly positive. To examine the sign of the denominator terms evaluated at \( k_m = 1 \), define \( \Theta_W(x) \equiv e^{-x}(-x + \ln x + \gamma) + E_1(x) \) for \( x \in (0, 1) \). Note that \( \frac{dx_s}{dk_m} < 0 \) at \( k_m = 1 \) if and only if \( \Theta_W(x) > 0 \). Observe that:
\[
\lim_{x \to 0} \Theta_W(x) = \lim_{x \to 0} E_1(x) + \ln x + \gamma = \lim_{x \to 0} E_{1i}(x) = 0 \text{ (see the proof of Proposition 2)};
\]
\[
\lim_{x \to 1} \Theta_W(x) = e^{-(\gamma - 1)} + E_1(1) \simeq 0.37 * (0.58 - 1) + 0.22 \simeq 0.064 > 0;
\]
\[
\frac{d\Theta_W}{dx} = -e^{-x}(\ln x + \gamma - x + 1).
\]

The parenthesis terms in the last derivative are monotone increasing in \(x \in (0,1)\), being negative as \(x \to 0\) and positive as \(x \to 1\). Therefore, \(\Theta_W(x) > 0\) for all \(x \in (0,1)\), implying \(\frac{d\Theta_W}{dx} < 0\) at \(k_m = 1\). As for large values of \(k_m\), it is sufficient to note that \(x_s\) approaches to a lowest possible value as \(k_m \to \infty\) for any given value of \(G \in (0,\infty)\) (see the proof of Proposition 3).

I now prove the last claim in the proposition on the welfare comparison. Note first that \(x_s = \frac{1}{W^{1/2}} = x_s\) and \(\Delta_W = W - \bar{W} = 0\) when \(k_m = 1\). Then, given any fixed values of \(G = Mk_m \in (0,\infty)\), differentiation yields
\[
\frac{dW}{dk_m} \bigg|_{k_m = 1} = \frac{(1 - x_s)^2(1 - \beta)}{(1 - \beta x_se^{-x_s})} \Theta_W(x_s) > 0
\]
where \(\Theta_W(x_s)\) is, as defined above, positive valued for all \(x_s \in (0,1)\). As for large value of \(k_m\), there are two cases. Consider first the case \(G > 1\), where \(x_s \to 0\) as \(k_m \to \infty\). Then, \(\Delta_W\) approaches to the highest possible value \(1 - \bar{W} > 0\) as \(k_m \to \infty\) (since \(\bar{W}\) is constant across all \(k_m\) when \(G = Mk_m\) is fixed). Consider next the case \(G < 1\) where \(x_s\) approaches to the lowest possible value \(1 - G > 0\) as \(k_m \to \infty\). Since \(1 - G < \frac{1}{1+G} = x_s\) for all \(G \in (0,1)\), this implies that the welfare difference \(\Delta_W\) must approach to a positive, highest possible value as \(k_m \to \infty\). ■
References


[5] ———— (2003), Strategic Stockouts in Supermarkets, Boston University manuscript.


43


[33] Hann, I.H., and C. Terwiesch (2003), Measuring the frictional costs of online transactions: The case of a reverse buying site, Management Science, 49(11), 1563-1579.


