Inventory Management in Decentralized Markets*

Fei Li Charles Murry Can Tian Yiyi Zhou†

March 18, 2023

Abstract

We present a directed search model of intermediaries with dynamic inventory and revenue management. Buyers purchase goods produced by sellers through intermediaries. Search frictions create demand uncertainty faced by each intermediary and make instantaneous replenishment impossible. To profit from the stochastic demand-supply misalignment, intermediaries hold inventory and employ inventory-based pricing and ordering policies. In equilibrium, when inventory is high, an intermediary posts a lower retail price to speed up sales and depresses wholesale price to slow down purchases. We characterize the evolution dynamics of inventory holdings and their steady-state distribution across intermediaries. We consider several extensions including multi-unit wholesale order, product differentiation, and heterogeneous intermediaries. Finally, we apply our framework to used-car dealer data from Ohio. We measure important unobservable characteristics of the decentralized markets, quantitatively evaluate the welfare consequence of used-car dealers' inventory management practice, and examine several policy-relevant transition dynamics due to the changes in market characteristics.

Keywords: Directed Search, Inventory Management, Revenue Management, Intermediary, Size Distribution, Used-Car Dealers.

JEL Classification Codes: D82, D83, L15, L62

---

*We are grateful to Victor Aguirregabiria, Simon Anderson, Gary Biglaiser, Michael Choi, Luca Flabbi, Chao Fu, Qing Gong, Michael Grubb, Naijia Guo, Veronica Guerrieri, Guido Menzio, Giuseppe Moscarini, Peter Norman, Stan Rabinovich, Paulo Somaini, Andrei Shevchenko, David Wiczer, Makoto Watanabe, Randall Wright, Shouyong Shi, Valentin Verdier, Jidong Zhou, Qiankun Zhou, conference and seminar participants of 2019 Mini conference on search and money at Madison, 2019 Midwest Macroeconomic conference at East Lansing, 2020 SaMMF Workshop: Advances in Search Theory, 2021 IIOC, Stanford University, University of North Carolina, University of Virginia, and Zhejiang University for helpful comments.

†Fei Li: Economics Department, University of North Carolina at Chapel Hill, lifei@email.unc.edu. Charles Murry: Economics Department, Boston College, charles.murry@bc.edu. Can Tian: Economics Department, University of North Carolina at Chapel Hill, tiancan@email.unc.edu. Yiyi Zhou: Department of Economics, Stony Brook University, yiyi.zhou@stonybrook.edu.

1 Introduction

Intermediaries play a prominent role in well-functioning markets. According to Spulber (1996), about a quarter of U.S. economic activity has been contributed by intermediaries. A common rationale of intermediation is to mitigate search frictions (see, e.g. Rubinstein and Wolinsky, 1987; Gavazza, 2016), but intermediaries still face uncertain demand and uncertain supply when search frictions cannot be fully eliminated. Such uncertainty leads to stochastic misalignment between idiosyncratic demand and supply, necessitating intermediaries to hold inventory and manage revenue through dynamic pricing policies. This observation presents a challenge to canonical decentralized-market intermediation models, which do not account for inventory management or its interaction with market frictions.

This paper develops a tractable equilibrium model where intermediaries face search frictions in both retail and wholesale markets and manage inventory. It highlights the interplay between multiple important economic forces related to inventory management that have welfare implications. The model generates inventory and price dynamics and distributions consistent with real markets. As a demonstration, we calibrate the model to the used-car market to measure important unobservable market characteristics, such as inventory costs and the quantitative importance of search frictions. We also quantitatively evaluate the welfare contribution of used-car dealers. Our empirical exercise generates two main findings. First, the welfare contribution of dealers is significant, large dealers almost double the surplus in the trade for used cars. Second, we find that actual inventory holding costs are small, and incentives for inventory management come from search frictions and resulting uncertainty dealers have about future tradings.

We borrow elements from the directed on-the-job search literature *à la* Menzio and Shi (2011) to model intermediaries' dynamic inventory management and pricing in the presence of market frictions in both demand and supply. Following the literature on frictional intermediation, we assume that it takes time for buyers and sellers to meet intermediaries in both retail and wholesale markets. At any time, each intermediary decides on a retail price and a wholesale price given its inventory level. Search is directed by prices in the following sense. In retail markets, buyers observe all retail prices and decide the set of intermediaries to seek. Analogously, in wholesale markets, sellers observe all wholesale prices and decide the set of intermediaries to search for. The model shares some flavor of Chamberlin's monopolistic competition insight: each intermediary faces a downward-sloping demand curve in retail markets and an upward-sloping supply curve in wholesale markets, but each intermediary is negligible in the sense that it can ignore its impact on, and hence reactions from, other intermediaries, making each intermediary's dynamic pricing and inventory management decision a monopolistic control problem. Intermediaries face idiosyncratic uncertainty originating from search frictions in both retail and wholesale markets; their individual inventory levels evolve stochastically, resulting in cross-sectional heterogeneity. Despite this complexity, the model remains tractable, enabling us to solve each agent's equilibrium policy function separately and characterize the inventory distribution's law of motion and its stationary limit. This simple structure makes the model an ideal platform to perform counterfactual experiments to understand the transitory and permanent effects of various important shocks.

In equilibrium, an intermediary's optimal pricing rules in retail and wholesale are deterministic functions of the inventory level. Pricing affects trading speed in respective markets and thus has stochastic influences on future inventory level. When inventory level is high, the intermediary charges a low retail price to speed up sales, and it offers a low wholesale price and waits for a willing seller. When inventory level is low, the intermediary charges a high retail price and slows down sales, and it also offers a high wholesale price to replenish faster. That is, the optimal retail and wholesale prices comove, and intermediaries profit from active price adjustments to reduce idiosyncratic misalignment in demand and supply. Consequently, high inventory is more likely to fall, and low inventory is more likely to rise. As such, each intermediary's inventory follows a controlled stochastic process, giving rise to a stationary cross-sectional distribution in the steady state.

We consider a few extensions of the baseline model and show that the general insight applies in richer settings. In the extension that allows for multiple wholesale units, the familiar (s, S)-rule for inventory acquisition policy naturally emerges in equilibrium. Other extensions consider product or intermediary heterogeneity; the model's tractability allows us to incorporate these features without altering the equilibrium structure.

As a demonstration, we apply our theory to used-car dealer inventory management and pricing to illustrate how to use our model to perform quantitative analysis. We have the access to data from an online used-car platform that contains weekly information of used-car dealers' inventories and list prices. To the best of our knowledge, we are the first to quantitatively examine intermediaries' inventory and pricing in a large decentralized market. The used-car industry is a natural laboratory to study the relationship between inventory and pricing in two-sided decentralized settings. Car dealers face inventory costs, dealers can adjust prices quickly, the wholesale market is relatively liquid, and a majority of used-car sales happen through dealers with a magnitude of tens of millions annually.[1] Moreover, the used-car industry is highly decentralized. Used-car dealers face

[1] We do not consider adverse selection in this paper. Although used cars are the canonical example of a lemons market (Akerlof, 1970), there is more recent research that suggests asymmetric information prob-

substantial uncertainty and frictions in both selling and buying cars.[2]

We detail how to identify the model primitives using panel data on dealer inventory and retail prices. We calibrate the model to the most popular and relatively homogeneous car category in our sample: 4-6 year-old non-luxury sedans. We allow for dealer heterogeneity and attribute the size difference among used-car dealers to differential characteristics such as matching efficiencies and inventory costs, which are not directly observed. We also quantify the welfare contribution of car dealers. We find that large dealers are able to create more surplus than small dealers by facilitating more tradings, and the difference is mainly due to their differential search and matching efficiency, not an inventory cost advantage. Finally, we conduct experiments with the calibrated model by making a 10% permanent change to each of the model parameters, one at a time, and analyze the transition dynamics of inventory and price. First, transitions are sluggish due to the stickiness in inventory adjustments resulting from quantitatively important matching frictions. Second, most changes to the parameters cannot rationalize the large changes to the used car industry in the wake of the COVID-19 pandemic. Our conclusion is that the decreased inventory and increased prices in 2021 and 2022 were caused by decreased supply, as opposed to changes in matching efficiency, demand, or inventory costs.

Related Literature and Contribution. Our paper contributes to the literature on intermediaries' role of mitigating search frictions. There are numerous theoretical studies on this topic in various settings. See, e.g., Rubinstein and Wolinsky (1987), Rust and Hall (2003), Duffie et al. (2005), Watanabe (2010), Wright and Wong (2014), Nosal et al. (2019), and Hugonnier et al. (2020, 2022). This theoretical hypothesis has also been supported by recent empirical research. See, e.g., Gavazza (2016) and Salz (2017). We recommend Gavazza and Lizzeri (2021) for a comprehensive survey of this literature. Our novelty is to introduce intermediaries' inventory and revenue management and study (i) their implications on individual-level inventory and price dynamics and (ii) their roles in shaping the cross-sectional distributions and evolution dynamics in the aggregate. Our model offers new empirical implications such as the relationship between inventory and retail and wholesale prices, and the co-movement of the retail and wholesale price time series.

We are certainly not the first to add inventory into search models. Many previous studies focus on the scale effect of holding multiple inventories and the corresponding

lems are not severe, particularly for late model vehicles – see Adams et al. (2011) and Biglaiser et al. (2020). Also, a literature on informational intermediaries demonstrates that, both theoretically and empirically, one of the most important functions of car dealers is to mitigate (if not fully resolve) information asymmetry between buyers and sellers. See, e.g., Biglaiser (1993), Lizzeri (1999), and Biglaiser et al. (2020). The basic argument is that dealers have the expertise to effectively detect lemons, and they have the incentive not to sell lemons due to the standard reputation concerns.

[2] This is in contrast to new car dealers, who have long-term relationships with manufacturers and with much less uncertainty regarding inventory.

benefit to the intermediaries. For instance, Johri and Leach (2002), Shevchenko (2004) and Smith (2004) introduce consumer preference heterogeneity and highlight the benefit of holding multiple units of inventory to satisfy diverse preferences. Watanabe (2020) attributes the emergence of intermediaries to their low inventory costs. In a recent paper, Rhodes et al. (2021) introduce multiple products and study the optimal portfolio choice of intermediaries.

While most previous studies adopt random search models and restrict their attention to the steady state of the economy, our paper explores a more tractable directed search approach *à la* Menzio and Shi (2011) that exploits directed search and block recursivity.[3] This tractability enables a straightforward characterization of the equilibrium inventory and price dynamics and a transparent discussion of the main trade-off of inventory management in the presence of search frictions both at and off the steady state. In addition, our model can easily accommodate rich heterogeneities and even aggregate uncertainty and transition dynamics, making it a tractable tool for applications.

Our paper also contributes to the literature on inventory management and pricing. While the idea to combine pricing and inventory management of consumption goods was first proposed by Whitin (1955), few studies have been done to understand the impact of these practices on equilibrium price dynamics and dispersion in a competitive environment.[4] See a recent survey by Chen and Simchi-Levi (2012). Also, there is a finance literature utilizing the similar continuous-time Markov chain technique to discuss dealers' optimal inventory-contingent pricing in equity security markets. The focus is to understand the existence of ask-bid spread and why securities' transaction prices deviate from their fundamentals through the lens of the inventory channel. Most papers focus on a monopolist dealer's dynamic decision problem, and the demand and supply of securities are often assumed to be exogenous. See, e.g., Amihud and Mendelson (1980), Stoll (1978), and Ho and Stoll (1981). One exception is Ho and Stoll (1983), which considers a model with two dealers trading two stocks. The current paper differs from these models by providing a tractable equilibrium search-and-matching framework to study inventory management with multiple dealers and endogenous arrivals of buyers and sellers. Our model can easily incorporate dealers' heterogeneity and conduct quantitative analysis, study the impact of search frictions in retail and wholesale markets, and analyze endogenous spill-over effects between the demand and supply sides. To the best of our knowledge, our paper is the first one that fills the gap between the literature on dynamic

[3] This tractable framework has been successfully applied in various contexts. We refer readers to Wright et al. (2017) for a comprehensive survey of the literature.

[4] One exception is the literature that combines demand uncertainty and costly capacity *à la* Prescott (1975) to generate price dispersion and inventory holding. See, e.g., Bental and Eden (1993) and Deneckere et al. (1996).

inventory management and equilibrium search theory.[5]

Finally, our model generates price dispersion in a search-theoretic model of intermediates' inventory management without any ex-ante heterogeneity. The price dispersion emerges as a pure-strategy equilibrium. It is an addition to the search-theoretic literature aiming to rationalize the well-documented empirical fact that observationally equivalent products are sold at different prices in many industries. The literature typically (i) requires buyers with (essentially) heterogeneous information (Burdett and Judd 1983, and Stahl 1989) or sellers with heterogeneous cost or visibility (Reinganum 1979) to generate price dispersion in mixed-strategy equilibria,[6] or (ii) relies on non-stationarity of search (Coey et al. 2020). See Kaplan and Menzio (2015) for a recent study and Baye et al. (2006) for a survey of the literature.

Organization. The rest of the paper is organized as follows. Section 2 introduces the theoretical model. Section 3 characterizes the equilibrium and derives empirical implications. Section 4 study some extensions of the benchmark model. Section 5 calibrates the model using the data from used-car markets. Extensions of the baseline model are in Section 4. Section 6 concludes. Omitted proofs are relegated to Appendix A. Supplementary empirical evidences are provided in Appendix B.

2 Model

2.1 Environment

We consider a continuous-time model with infinite horizon; i.e., the calendar time $t \in \mathbb{R}_+$. The economy is populated by buyers, sellers, and intermediaries.

Agents. There is a large pool of atomistic buyers and sellers. Each seller has a unit supply of the indivisible (consumption) good, and he receives zero utility by consuming the good by himself. Each buyer has a unit demand of the good, and by consuming the good, his utility is $u > 0$. A buyer leaves the market after his demand is satisfied and a seller leaves the market whenever his good is sold. To maintain the size of the potential buyers and sellers pool, we assume a new buyer (seller) arrives whenever an existing buyer (seller)

[5]Contemporaneous with our paper, Yang and Zeng (2021) consider trade between dealers and dealers' inventory choices into the framework of Duffie et al. (2005) and explore the equilibrium multiplicity. Colliard et al. (2021) propose a stylized three-period dealer network model and examine the joint effect of dealers' network connections and inventory management on prices and allocations in over-the-counter markets.

[6]The qualification "essentially" is added because the heterogeneous information structure can be endogenized by adding a stage of costly information acquisition of homogeneous consumers as in Burdett and Judd (1983).

leaves. Despite gains from trade, buyers and sellers face some obstacles to trade, creating a role of intermediaries. Our focus is to model the equilibrium intermediated transaction mechanism where the consumption goods are delivered from sellers to buyers through intermediaries.

There is a unit measure of ex-ante identical intermediaries (dealers), each of whom purchases consumption goods from sellers in the wholesale market and sells goods to buyers in the retail market. An intermediary can sell only if his current inventory is positive. The flow cost of holding x units of inventory is $c(x)$ for $x = 0, 1, 2, \ldots$. The cost function $c : \mathbb{N} \to \mathbb{R}$ is increasing with $c(0) = 0$, and the marginal cost $c(x+1) - c(x)$ weakly increases in x. All agents are risk-neutral and share a common discount rate $\rho > 0$.

Let $g_t : \mathbb{N} \to [0, 1]$ be the probability mass function of the distribution of inventory holding across intermediaries at time t. Specifically, $g_t(x)$ represents the measure of intermediaries who holds x units of inventory at time t. Therefore, $g_t(x) \geq 0, \forall t, x$ and $\sum_{x \in \mathbb{N}} g_t(x) = 1, \forall t$. For notation convenience, we use g_t to denote the vector $\{g_t(x)\}_{x \in \mathbb{N}}$ for each t.

Markets. The retail market is organized in multiple submarkets indexed by the retail price $p \in \mathbb{R}$. In each retail submarket p, the ratio of buyers to intermediaries is denoted by $\theta(p)$. Retail submarket p can therefore be viewed as a group of agents who wish to trade at price p. Similarly, the wholesale market is organized in multiple submarkets indexed by the wholesale price $w \in \mathbb{R}$. In each wholesale submarket w, the ratio of sellers to intermediaries is denoted by $\lambda(w)$. Following Pissarides (1985), we refer to $\theta(p)$ and $\lambda(w)$ as the *tightness* of the corresponding retail and wholesale submarkets.

Search and Matching. Search is directed in the sense of Moen (1997) and Acemoglu and Shimer (1999). At each moment, an intermediary can choose to enter at most one retail submarket and one wholesale submarket simultaneously. In this way, we capture the intermediary's retail/wholesale pricing problem as a choice of the corresponding submarkets.

At each instant, a buyer sees all the retail submarkets (prices) and chooses to enter at most one retail submarket to search for intermediaries. If he does not enter any retail submarket, he receives a *flow* outside option $\kappa_b > 0$. That is, a buyer's opportunity cost of searching in any retail submarket for time length $dt > 0$ is $\kappa_b dt$. The outside option can be interpreted in many ways. For example, one possibility is that the buyer searches in a decentralized market without intermediation (via, e.g., Craigslist) to look for a seller, where he meets a seller at a rate which is normalized to be 1 and receives an expected surplus κ_b from a meeting with a seller. Similarly, a seller sees all wholesale submarkets (prices) and chooses to enter at most one wholesale submarket at each moment. If the seller does not enter any wholesale submarket, he receives a flow outside option $\kappa_s > 0$.[7] For example, in a used-car setting, the seller's outside options also include his flow utility from keeping the used car. See section 2.2 for more discussion.

There are frictions in submarkets. In each submarket, the matching process is determined by a matching function. Following the literature (see, e.g., Pissarides (1985) and Moen (1997)), we assume that the matching function is homogeneous of degree one so that the matching process in each submarket is fully determined by its tightness. Specifically, at each instant, an intermediary meets a buyer at Poisson rate $\phi_r(\theta(p))$ in retail submarket p. We further assume that the function $\phi_r : \mathbb{R}_+ \to \mathbb{R}_+$ is bounded, twice-differentiable, strictly increasing, and strictly concave, such that $\phi_r(0) = 0$. On the other side of a retail submarket, a buyer makes a contact with an intermediary at Poisson rate $\psi_r(\theta(p))$ where the function $\psi_r : \mathbb{R}_+ \to \mathbb{R}_+$ is twice-differentiable and strictly decreasing. Due to the homogeneity of the matching function, we have $\psi_r(\theta) = \phi_r(\theta)/\theta, \forall\theta > 0$, which reflects the fact that the number of intermediaries who meet buyers must equal the number of buyers who meet intermediaries. Finally, the matching function satisfies $\lim_{\theta\to\infty} \psi_r(\theta) = 0$, and $\lim_{\theta\to 0} \psi_r(\theta) = +\infty$. Similarly, in a wholesale submarket w, an intermediary meets a seller at Poisson rate $\phi_w(\lambda(w))$ where $\phi_w : \mathbb{R}_+ \to \mathbb{R}_+$ is bounded, twice-differentiable, strictly increasing, and strictly concave, such that $\phi_w(0) = 0$. On the other side, a seller meets an intermediary at Poisson rate $\psi_w(\lambda(w))$ where $\psi_w : \mathbb{R}_+ \to \mathbb{R}_+$ is twice-differentiable and strictly decreasing such that $\psi_w(\lambda) = \phi_w(\lambda)/\lambda, \forall\lambda > 0$, $\lim_{\lambda\to\infty} \psi_w(\lambda) = 0$, and $\lim_{\lambda\to 0} \psi_w(\lambda) = +\infty$.

When an intermediary and a buyer meet in retail submarket p, the buyer buys one unit of the good from the intermediary at price p. When an intermediary and a seller meet in wholesale submarket w, the seller sells one unit of the good to the intermediary at price w.

2.2 Discussion of Assumptions

Before moving forward, we discuss some assumptions. First, we assume that search is directed, such that an agent is fully aware of the price and the matching probability of each submarket. The model both captures search friction and also preserves the familiar trade-off between transaction speed and price in a competitive environment. Specifically, we model an intermediary' pricing decision as choosing which submarket to enter. To sell faster, the intermediary has to enter a retail submarket with higher tightness, implying

[7] The assumption that an agent can visit at most one submarket at each time can be relaxed. Alternatively, one can allow a buyer or seller to visit n submarkets by foregoing a flow outside option $n\kappa_i$ where $i = b, s$, and $n > 0$ is an integer that can be either exogenously specified or endogenously chosen *à la* Stigler (1961). In this case, the rate at which the buyer or seller meets an intermediary is proportional to n as well.

a lower equilibrium retail price. Similarly, to buy faster, the intermediary has to enter a wholesale submarket with higher tightness, implying a higher equilibrium wholesale price. The assumption of observable prices is consistent with the idea that buyers shop online to discover prices. A buyer's choice of submarket reflects the idea that a buyer understands that transaction speeds vary with the listed price.[8]

As noted by Acemoglu and Shimer (1999) and Faig and Jerez (2005), the directed search paradigm encompasses many reasonable possibilities. One possibility is that submarkets are located in different places (malls, streets, or online platforms) and that in each submarket the good is required to be traded at an identical term. Agents are aware of the trading term in each submarket and understand that better terms of trade are associated with greater degree of congestion within a submarket. Another one is to think of a submarket as a set of trading opportunities with identical trading term and agents randomly select one of them as in the frictional assignment literature (Peters, 1991, 2000; Burdett et al., 2001).

Second, we assume sufficiently large pools of potential buyers and sellers, so the entry of buyers and sellers has infinite elasticity. This assumption is for simplicity. A straightforward implication of this assumption is that buyers and sellers will break even and act myopically in equilibrium, and so the outside option parameters κ_b and κ_s capture the expected flow payoff of buyers and sellers respectively. We view this modeling choice as a simple way to close our partial equilibrium model which focuses on the dynamics of intermediaries instead of buyers and sellers.[9] In the quantitative exercise, these outside option parameters are shown to be identified with limited data, and they serve as a natural baseline to gauge intermediaries' welfare contribution.

Third, we assume that an intermediary can order at most one unit at each moment. This assumption can be relaxed. In subsection 4.1, we extend our baseline model and allow a seller to carry multiple units, and the outcome of intermediaries' equilibrium policies resembles the familiar (s, S)-rule and non-linear pricing. Also, we assume a homogeneous product and ex-ante homogeneous intermediaries. The first assumption is made to emphasize the search frictions resulting from the uncertainty about how quickly agents are matched. This is a deliberate simplification to highlight our main mechanism.

[8] That being said, observing perfect price information is not essential for the negative relationship between price and trading speed. One can obtain a similar trade-off in a random search model by introducing random utility (demand curve) and production cost (supply curve) and analyze steady-state equilibria where the equilibrium inventory distribution is constant over time. However, the directed search formalization, as discussed in Menzio and Shi (2011), is necessary to make agents' decisions independent of the distribution of inventory holdings. It makes our model suitable for considering the transaction dynamics of a permanent shock or the implication of aggregate uncertainty.

[9] A similar modeling choice is used in canonical labor search model where firms' intertemporal job posting trade-off is trivialized and the focus is on the workers' job search dynamics. See Pissarides (2000) for a textbook treatment.

In the search theory literature, another important source of search frictions comes from the uncertainty regarding the match quality between agents and products, which can also be accommodated by a stylized extension of our model (see subsection 4.2). The extension on vertical product differentiation is also discussed (see subsection 4.3). We assume that intermediaries are ex-ante homogeneous to highlight the contribution of the ex-post inventory dynamics to price dynamics and endogenous cross-sectional heterogeneity. This assumption leads to a common optimal inventory-based pricing policy among intermediaries. It is straightforward to extend our model to allow for ex-ante heterogeneous inventory-price relationship (see subsection 4.4).

Finally, we fix the measure of intermediaries but endogenize buyers' and seller's participation decisions. This assumption reflects our belief that the participation decisions of intermediaries are significantly less flexible than those of buyers and sellers in many intermediated decentralized markets (e.g., real estate, used car, and financial asset markets) due to non-trivial entry/exit cost. It certainly limits the applicability of our model to understanding long-run industry dynamics and firms' turnover in these markets.

2.3 Individual Problem and Equilibrium

This subsection formulates the competitive search equilibrium. In our model, intermediaries are ex post heterogeneous in their inventory holdings. Therefore, as in other continuous-time heterogeneous-agent models with a continuum of atomistic agents (see, e.g., Shi (2009); Nuño and Moll (2018); Achdou et al. (2022)), a competitive equilibrium can be characterized by two coupled differential equations: a Hamilton-Jacobi-Bellman (HJB) equation for the optimal choices of each atomistic individual who takes the evolution of the inventory distribution as given, and a Kolmogorov Forward (KF) equation describing the law of motion of the inventory distribution induced by agents' optimal choices.

The Seller's Problem. Let S_t denote a seller's life-time expected surplus gaining from the intermediary sector. At each instant t, he decides whether and where to search. By the standard argument, S_t obeys the following HJB equation

$$\rho S_t = \dot{S}_t + \max\Big\{ -\kappa_s + \max_{w>0}\{\psi_w(\lambda_t(w))(w - S_t)\},\ 0\Big\}. \tag{1}$$

There are two terms on the right-hand side of the HJB. The first term is the value function's partial derivative with respect to the calendar time, absorbing the effect of aggregate state g_t. The second term captures the payoff corresponding to the seller's choice. If he chooses to enter wholesale submarket w, he foregoes a flow outside option κ_s and meets an intermediary at a rate $\psi_w(\lambda_t(w))$ and sells his product, receiving a payoff change from S_t to w. The time index of $\lambda_t(w)$ allows for potentially time-dependent mapping between wholesale price w and submarket tightness. If he decides not to enter any wholesale submarket, he neither meets any intermediary nor foregoes the flow outside option, receiving a flow surplus 0. Obviously, it is strictly suboptimal to give up the outside option $\kappa_s > 0$ to search in a wholesale market with $w \leq 0$, so it is without loss to focus on wholesale submarkets such that $w > 0$.

The standard free-entry argument implies that in equilibrium, at any time t, there is no room for extra gains, so a seller never derives positive surplus from the intermediary sector, i.e.,

$$S_t = \dot{S}_t = 0, \ \forall t.$$

As a result, the tightness $\lambda_t(w)$ in any wholesale submarket market w must satisfy the following free-entry (FE) condition

$$\kappa_s \geq \psi_w(\lambda_t(w))w, \tag{2}$$

and $\lambda_t(w) \geq 0$ with complementary slackness at each instant t. The left-hand side of condition (2) is the seller's outside option κ_s; the right-hand side corresponds to the expected revenue of entry, which is given by the product between the rate at which the seller meets an intermediary $\psi_w(\lambda_t(w))$ and the selling price of the good w. If the tightness is positive in any submarket w, then the equality must hold in (2); otherwise either more sellers have the incentive to enter the wholesale submarket or some sellers who are supposed to enter the wholesale submarket have the incentive not to do so. Therefore, for each wholesale price $w > 0$, there exists a unique submarket tightness that is strictly positive and satisfies the FE condition, given as

$$\lambda(w) = \psi_w^{-1}\Big(\frac{\kappa_s}{w}\Big), \tag{3}$$

which describes a one-to-one mapping between wholesale price and submarket tightness that is independent of time t and the aggregate g_t.

The Buyer's Problem. Similarly, let B_t denote a buyer's surplus gaining from the intermediary section. At each instant, the buyer decides whether to enter a retail submarket and which one to enter. The buyer's B_t satisfies an HJB equation, given as

$$\rho B_t = \dot{B}_t + \max\Big\{ -\kappa_b + \max_{p\in[0,u)} \{\psi_r(\theta_t(p))(u - p - B_t)\},\ 0\Big\}. \tag{4}$$

The first term on the right-hand side of the HJB is the time derivative of the buyer's value function, reflecting the effect of change in the aggregate state g_t. The second term captures the impact of the buyer's choice. If he chooses to enter retail submarket p, he forgoes a flow outside option κ_b and meets an intermediary at a rate $\psi_r(\theta_t(p))$ and buys at price p, receiving a payoff gain $u - p - B_t$. The time index in $\theta_t(p)$ allows for generic time-dependent mapping between retail price p and submarket tightness. If he decides not to enter any retail submarket, he neither foregoes the outside option nor meet any intermediary. Obviously, it is suboptimal to search in a retail market with $p \geq u$, so it is without loss to focus on retail submarkets such that $p \in [0, u)$.

As in the seller's problem, the usual free-entry argument implies that $B_t = \dot{B}_t = 0$ in equilibrium, $\forall t$, and the tightness in each retail submarket must satisfy the following FE condition

$$\kappa_b \geq \psi_r(\theta_t(p))(u - p), \tag{5}$$

and $\theta_t(p) \geq 0$ with complementary slackness. Condition (5) guarantees that the tightness $\theta_t(p)$ is consistent with the buyer's incentive to search. The opportunity cost of search is given by κ_b, and the benefit of search is given by the product between the rate at which the buyer meets an intermediary $\psi_r(\theta_t(p))$ and the surplus from buying the good at price p. The stationary one-to-one mapping between the retail market price and the market tightness is described by

$$\theta(p) = \psi_r^{-1}\Big(\frac{\kappa_b}{u - p}\Big), \tag{6}$$

such that for each retail price $p \in [0, u)$, there is a unique submarket tightness that is strictly positive and satisfies the FE condition, and it is independent of time t and the aggregate g_t.

The Intermediary's (Pricing) Problem. Consider an intermediary with inventory x at time t. It decides whether and where (at what wholesale price) to make new orders and whether and where (at what retail price) to sell its inventory. If the intermediary has at least one unit of inventory $x \geq 1$ at t, its expected value $V_t(x)$ obeys the following HJB equation

$$\begin{aligned} \rho V_t(x) = \dot{V}_t(x) - c(x) \quad &+ \quad \underbrace{\max\Big\{0, \max_{w \in \mathbb{R}_+} \phi_w(\lambda(w))[-w + V_t(x+1) - V_t(x)]\Big\}}_{\text{wholesale problem}} \\ &+ \quad \underbrace{\max\Big\{0, \max_{p \in [0,u)} \phi_r(\theta(p))[p + V_t(x-1) - V_t(x)]\Big\}}_{\text{retail problem}}. \end{aligned} \tag{7}$$

At each moment, an atomistic intermediary with positive inventory chooses a retail submarket and a wholesale submarket to enter, associated with a pair of retail and wholesale prices, treating the functional forms of the corresponding market tightnesses $\lambda(\cdot), \theta(\cdot)$ defined in equations (3) and (6) as given. There are four terms on the right-hand side of equation (7). The first term is the value function's partial derivative with respect to the calendar time, absorbing the effect of aggregate state g_t on the intermediary's problem. The second term is the flow cost of holding x units of inventories. The third and fourth items concern whether and which wholesale and retail submarkets to enter, respectively. Choosing not to enter any submarket yields zero value flow. In the third term, the expected value flow of entering wholesale submarket w is the rate at which the intermediary meets a seller $\phi_w(\lambda(w))$ in the submarket times the change in continuation value when the intermediary buys one unit of good from the seller, $-w + V_t(x+1) - V_t(x)$. Analogously, in the last term, the expected value flow of entering retail submarket p is the rate at which the intermediary meets a buyer $\phi_r(\theta(p))$ in the submarket times the change in continuation value when the buyer purchases one unit of good from the intermediary, $p + V_t(x-1) - V_t(x)$.

In words, at each moment t and the inventory level $x \geq 1$, an intermediary takes as given the price-tightness mapping in each submarket and the evolution of the aggregate state g_t and controls the Poisson arrival rates of buyers $(\phi_r(\theta(p)))$ and sellers $(\phi_w(\lambda(w)))$ by choosing retail and wholesale prices.

An intermediary cannot sell when stocking out, so the HJB equation at $x = 0$ does not include the choice of a retail submarket,

$$\rho V_t(0) = \max_{w \in \mathbb{R}_+} \dot{V}_t(0) + \phi_w(\lambda(w))[-w + V_t(1) - V_t(0)]. \tag{8}$$

Denote the optimal control as $p_t(x), w_t(x)$ for each t.

Evolution of the Aggregate State. Given $\{\theta(\cdot), \lambda(\cdot), p_t(\cdot), w_t(\cdot)\}_{t \in \mathbb{R}_+}$, the distribution of inventory across intermediaries g_t evolves according to the following KF equation:

$$\begin{aligned} \dot{g}_t(x) \quad &= \quad \underbrace{g_t(x-1)\phi_w(\lambda(w_t(x-1))) + g_t(x+1)\phi_r(\theta(p_t(x+1)))}_{\text{inflows}} \\ &- \quad \underbrace{g_t(x)[\phi_r(\theta(p_t(x))) + \phi_w(\lambda(w_t(x)))]}_{\text{outflows}}, \end{aligned} \tag{9}$$

for every $x \in \mathbb{N}$, and

$$\sum_{x \in \mathbb{N}} g_t(x) = 1, \ \forall t. \tag{10}$$

The left-hand side of equation (9) is the time derivative of the measure of intermediaries who hold x units of inventory at time t. The right-hand side of equation (9) has three parts. The first two terms are positive, but the last term is negative. First, $g_t(x-1)$ of intermediaries hold $x-1$ units of inventory each and search in a wholesale submarket with tightness $\lambda(w_t(x-1))$ at time t, and $\phi_w(\lambda(w_t(x-1)))$ of them find sellers, trade, and increase their stock to x. Second, $g_t(x+1)$ of intermediaries hold $x+1$ units of inventory each and search in a retail submarket with tightness $\theta(p_t(x+1))$ at time t, and $\phi_r(\theta(p_t(x+1)))$ of them find buyers, trade, and decrease their stock to x. Finally, $g_t(x)$ of intermediaries hold x units of inventory at time t and $\phi_r(\theta(p_t(x)))$ of them meet buyers and $\phi_w(\lambda(w_t(x)))$ of them meet sellers, changing their inventory from x to $x-1$ and $x+1$, respectively.

Equilibrium. A competitive search equilibrium is a value function $V_t(x)$, a pair of controls $(p_t(x), w_t(x))$, a pair of market tightness function $(\theta(p), \lambda(w))$ and a probability mass function $g_t(x)$ for each calendar time t such that

1. $\theta(p)$ and $\lambda(w)$ satisfy the free-entry (FE) conditions (3) and (6) for any $p \in [0, u)$, $w \in \mathbb{R}_{++}$,

2. given $\{\theta(p), \lambda(w)\}$, $V_t(x)$ is the solution of the intermediary's Hamilton-Jacobi-Bellman (HJB) equations (7) and (8), and the associated optimal controls are $p_t(x), w_t(x)$ for any t and $x \in \mathbb{N}$, and

3. given $\{\theta(p), \lambda(w), p_t(x), w_t(x)\}$, $g_t(x)$ is a solution of the Kolmogorov Forward (KF) equations (9) and (10).

The system of equations (3), (6), (7), (8), (9), and (10) fully characterize the evolution dynamics of our economy given an initial inventory distribution g_0, which is degenerate when all intermediaries are ex ante identical. In general, two systems need to be pinned down simultaneously: the KF equations are determined by individual optimal policy, and the evolution of the inventory distribution affects individual's optimal choice through the calendar time. We are particularly interested in the competitive search equilibrium with the so called *block recursive structure* where each intermediary's problem is distribution-free (see, e.g., Shi, 2009; Menzio and Shi, 2010, 2011), i.e.,

$$p_t(x) = p(x), \quad w_t(x) = w(x), \quad V_t(x) = V(x), \quad \forall t, x. \tag{11}$$

Solving a block recursive equilibrium is both analytically and computationally convenient. As is standard in competitive search models, it is without loss of generality to focus on block recursive equilibria in our setting. We will argue that (i) all competitive search equilibria are block recursive, and (ii) there is a unique block recursive equilibrium.

3 Analysis

In this section, we provide equilibrium analysis of the model. First, we characterize the equilibrium and introduce our main proposition which establishes that prices decrease with inventory. Next, we analyze the steady state and transition dynamics and show the existence of a unique stationary distribution of inventory holdings.

3.1 Equilibrium Characterization

An intermediary faces a trade-off between the expected speed of trade and the transaction price. Specifically, equation (6) implies that each buyer's expected benefit of search must be constant in every retail submarket in an equilibrium. Hence, $\theta(p)$ must decrease in p. That is, if a retail submarket features a higher price, its equilibrium buyer-to-intermediary ratio must be lower, making it more likely for each consumer to meet an intermediary. If an intermediary wants to sell faster (larger $\phi_r(\theta(p))$), he must enter a retail submarket featuring a lower price p. The same reasoning applies to the trade-off between transaction speed and price in wholesale submarkets. The one-to-one equilibrium relationship between price and market tightness implies that one can reformulate each atomistic intermediary's (pricing) problem as choosing the tightness of the submarket he plans to enter.

Specifically, the intermediary's equilibrium choice of retail policy $p_t(x)$ necessarily solves

$$\max_{p\in[0,u)} \phi_r(\theta(p))[p + V_t(x-1) - V_t(x)], \tag{12}$$

where $\theta(p)$ is given by (6) and $V_t(x)$ solves the HJB equation (7). It is mathematically equivalent to a text-book monopoly pricing problem where the cost is $V_t(x) - V_t(x-1)$ and the demand function is $\phi_r \circ \theta(\cdot)$. One can equivalently write the problem in (12) as

$$\max_{\theta\in\Theta} \phi_r(\theta)[p(\theta) + V_t(x-1) - V_t(x)],$$

where the tightness feasible set is

$$\Theta \equiv [0, \psi_r^{-1}(\kappa_b/u)] = \{0\} \cup \theta([0,u)]),$$

which is the union between 0 and the image of function $\theta(\cdot)$, and for any $\theta > 0$, $p(\theta)$ is the inverse of $\theta(\cdot)$ in (6), i.e.,

$$p(\theta) = u - \frac{\kappa_b}{\psi_r(\theta)} = u - \frac{\kappa_b\theta}{\phi_r(\theta)}, \tag{13}$$

where $\psi_r(\cdot)$ is decreasing. Condition (13) thus immediately implies that $p(\cdot)$ is decreasing. It can be viewed as the inverse "demand curve" an intermediary faces. We extend the feasible set of Θ to include $\theta = 0$ to capture the idea that the intermediary is free not to search in any retail submarket. We assume $p(0)$ is an arbitrary constant so that choosing $\theta = 0$ leads to a zero value of the retail problem.

Similarly, the intermediary's wholesale problem in equation (7) can be equivalently written as

$$\max_{\lambda \in \Lambda} \phi_w(\lambda)[-w(\lambda) + V_t(x+1) - V_t(x)],$$

where $\Lambda = \mathbb{R}_+$ is the feasible set for wholesale submarket tightness, and for any $\lambda > 0$, $w(\cdot)$ is the inverse of $\lambda(\cdot)$ in equation (3), given as

$$w(\lambda) = \frac{\kappa_s}{\psi_w(\lambda)} = \frac{\kappa_s \lambda}{\phi_w(\lambda)}, \tag{14}$$

which is increasing in λ and has the flavor of the inverse "supply curve" faced by an intermediary. The feasible set includes $\lambda = 0$ to allow the intermediary not to search in any wholesale submarket. Assume $w(0)$ to be an arbitrary constant so that choosing $\lambda = 0$ leads to a zero value of the wholesale problem.

Plugging equations (13) and (14) into the intermediary's retail and wholesale problems in equation (7) implies that the intermediary's equilibrium value function $V_t(x)$ solves the following HJB equation,

$$\begin{aligned} \rho V_t(x) = \dot{V}_t(x) - c(x) \quad &+ \quad \max_{\theta \in \Theta} \underbrace{\phi_r(\theta)[u + V_t(x-1) - V_t(x)] - \kappa_b \theta}_{\text{retail market surplus}} \\ &+ \quad \max_{\lambda \in \Lambda} \underbrace{\phi_w(\lambda)[V_t(x+1) - V_t(x)] - \kappa_s \lambda}_{\text{wholesale market surplus}}. \end{aligned} \tag{15}$$

In equilibrium, it is *as if* the intermediary solves a decision problem based on the inventory level at each moment, such that the intermediary chooses the tightness of the retail submarket where he looks for buyers and the tightness of the wholesale submarket where he looks for sellers.

The optimal retail market tightness, denoted by $\theta_t^*(x)$, maximizes the expected flow surplus generated by the intermediary and a mass of buyers with measure $\theta_t^*(x)$. To be specific, maintaining the market tightness to be θ incurs a social opportunity cost $\kappa_b\theta$, but the intermediary and a buyer will meet at a rate $\phi_r(\theta)$ and generate gains from trade $u + V_t(x-1) - V_t(x)$. Similarly, the optimal wholesale market tightness, denoted by $\lambda_t^*(x)$, maximizes the expected surplus generated by the intermediary and a mass of sellers with measure $\lambda_t^*(x)$. Given that all buyers and sellers break even in equilibrium, the value function $V_t(x)$ thus corresponds to the discounted expected social surplus that an intermediary with inventory x at t generates from time t on. The optimal controls $\theta_t^*(x)$ and $\lambda_t^*(x)$ for the HJB equation (15) correspond to the equilibrium market tightnesses in retail submarket $p(\theta_t^*(x))$ and wholesale submarket $w(\lambda_t^*(x))$, respectively. Therefore, solving a competitive search equilibrium is equivalent to solving the decision problem (15) and to plug the optimal controls into the conditions (13) and (14).

Now we argue that the solution $V_t(\cdot)$ to HJB equation (15) is stationary. In principle, the solution to problem in equation (15) is allowed to be non-stationary ($\dot{V}_t \neq 0$) to capture the impact of the evolution of the inventory distribution g_t, but a closer look at the problem in equation (15) reveals that the calendar time plays a role in the dynamic control problem only through the control variables $\{\theta_t(x), \lambda_t(x)\}$. Specifically, in (15), an intermediary maximizes the expected life-time total utility that he delivers to buyers, net of the expected life-time inventory cost, and $\{\theta_t(x), \lambda_t(x)\}$ solely pins down the stochastic process of the intermediary's life-time inventory $\{x_t\}$ and hence the process of flow payoff. Towards a contradiction, suppose that $V_t(x) < V_{t'}(x)$ for some x and $t \neq t'$, the intermediary at time t has a profitable deviation by mimicking its time-t' self's continuation play. As a consequence, the calendar time t (and thus the distribution g_t) cannot affect the intermediary's optimal continuation payoff, i.e., $V_t(x) = V(x), \dot{V}_t(x) = 0, \forall t, x$, and equation (15) can be rewritten as the following stationary HJB,

$$\begin{aligned} \rho V(x) = -c(x) \quad &+ \quad \max_{\theta \in \Theta} \phi_r(\theta)[u + V(x-1) - V(x)] - \kappa_b \theta \\ &+ \quad \max_{\lambda \in \Lambda} \phi_w(\lambda)[V(x+1) - V(x)] - \kappa_s \lambda. \end{aligned} \tag{16}$$

Naturally, the optimal controls $\theta^*(x), \lambda^*(x)$ must be stationary. By conditions (13) and (14), the equilibrium retail and wholesale prices must be stationary as well, i.e., $p_t^*(x) = p(\theta^*(x)), w^*(x) = w(\lambda^*(x))$. In sum, all agents' problems in equilibrium are independent of the aggregate state g_t, and so all competitive equilbiria are BRE.

The existence and uniqueness of BRE boils down to the existence and uniqueness of the solution to the stationary HJB function (16), which can be verified by the standard argument (see, e.g., Chapter 4 of Guo and Hernández-Lerma (2009)). The optimal value function $V(x)$ must be unique. Therefore, given $V(x)$, it is straightforward to see from (15) that the intermediary's optimal retail policy $\theta(x)$ and optimal wholesale policy $\lambda(x)$ can be characterized separately as two optimization problems and are unique due to the strict concavity of the matching function.[10]

[10] The lack of multiplicity is due to the free-entry specification in both retail and wholesale markets, making the equilibrium allocation socially efficient, as in many competitive search models, e.g. Moen (1997) and Menzio and Shi (2011).

From the problem in (16), it follows that the optimal $\theta^*(x)$ must satisfy the first-order condition (FOC) given as

$$\kappa_b \geq \phi_r'(\theta^*(x))[u + V(x-1) - V(x)], \tag{17}$$

and $\theta^*(x) \geq 0$ with complementary slackness. The FOC in (17) says that the marginal opportunity cost κ_b to the economy to maintain the tightness to be $\theta^*(x) > 0$ must equal the social marginal benefit of doing so in any retail submarket with positive tightness. Here, the social opportunity cost is incurred by buyers and the social benefit is the expected surplus of a transaction. Similarly, the optimal $\lambda^*(x)$ must satisfy

$$\kappa_s \geq \phi_w'(\lambda^*(x))[V(x+1) - V(x)], \tag{18}$$

and $\lambda^*(x) \geq 0$ with complementary slackness. It says that the marginal social opportunity cost κ_s incurred by sellers equals the social marginal benefit of maintaining the tightness to be $\lambda^*(x) > 0$.

Notice that conditions (17) and (18) imply that $\theta(x)$ and $\lambda(x)$ depend on gains from trade, $u + V(x-1) - V(x)$ and $V(x+1) - V(x)$, respectively, which depends on the intermediary's current inventory size x. The following lemma characterizes how inventory size affects gains from trade for an intermediary in both retail and wholesale markets.

Lemma 1. *In the equilibrium, whenever $V(x)$ increases in x, the difference $V(x) - V(x-1)$ decreases in x; if $V(x)$ starts to decrease at some $x = S \in \mathbb{N}$, $V(x)$ decreases over all $x > S$.*

Lemma 1 says that any positive marginal benefit of accumulating inventory decreases in the level of inventory whenever it exists.[11] Intuitively, this property is due to the combination of two factors. The first one is the diminishing risk of stocking out. With two-sided search frictions, an intermediary faces uncertainty about both the demand in retail markets and the supply in wholesale markets. If his inventory is reduced to zero, he can neither immediately order goods from sellers nor trade with buyers. As the intermediary's inventory size increases, the stockout risk in the near future falls, lowering the marginal benefit of increasing inventory. The second one is the increasing inventory cost function. As x increases, the marginal inventory cost erodes the benefit of reduced stockout risk, also contributing to the diminishing return of adding inventory. Moreover, with a weakly convex inventory cost function, the benefit of adding more inventory $V(x) - V(x-1)$ becomes negative at sufficiently large x. Therefore, the maximum of $V(x)$ exists, and we

[11]This property of diminishing returns to inventory is quite robust. It was first found in the multi-unit search paper of Carrasco and Smith (2017). Their paper is different in that it is a single-agent search model. The model is extended by Carrasco and Harrison (2022) by introducing operational cost. A similar property has been obtained in Chen et al. (2020) where search frictions are absent.

define

$$S = \min\{\arg\max_{x \in \mathbb{N}} V(x)\} \tag{19}$$

as the minimal inventory level at which an intermediary's value achieves the maximum.[12] Then, by Lemma 1, the benefit of adding more inventory, $V(x) - V(x-1)$ will be negative for any $x > S$ as well, resulting in an equilibrium upper bound for the intermediary's inventory level.

Now we are ready to derive the relationship between inventory and prices. Let

$$p^*(x) \equiv p(\theta^*(x)) \text{ and } w^*(x) \equiv w(\lambda^*(x))$$

denote the equilibrium retail and wholesale pricing policy where $p(\cdot)$ and $w(\cdot)$ are specified in conditions (13) and (14).

Proposition 1. *In the equilibrium, the intermediary's choice of submarkets is such that*

1. *$\theta^*(x)$ increases in x, and retail price $p^*(x)$ decreases in x, and*
2. *$\lambda^*(x)$ and the wholesale price $w^*(x)$ decrease in x.*

Proposition 1 says that at a higher inventory level, the intermediary will enter a retail submarket with lower price and higher matching probability (easier to sell), and enter a wholesale submarket with lower wholesale price and lower matching probability (harder to buy). This is intuitive. An intermediary trades off between the risk of stockout and the cost of inventory and new orders. When the inventory stock becomes higher, the risk of stockout decreases, but the inventory cost becomes higher, so it is optimal to lower future inventory by selling more and buying less. To do so, the intermediary needs to lower both the retail and wholesale prices. Similarly, when his stock becomes too low, the concern of stockout grows, and the intermediary raises both the retail price and the wholesale price to slow down the sales and speed up new orders, increasing his future inventory holding in expectation.

The empirical implication of Proposition 1 is that when the intermediary's inventory increases, (i) the retail price decreases and the sales increase on average, and (ii) the wholesale price decreases and new orders decrease on average.

Corollary 1. *An intermediary's equilibrium retail and wholesale prices co-move over time.*

Corollary 1 is an immediate implication of Proposition 1. Driven by the change in inventory, an intermediary's retail price and wholesale price should move in the same

[12] The set $\arg\max_{z \in \mathbb{N}} V(z)$ is a singleton at a generic point in the parameter space and may contain up to two adjacent elements, in which case we select the smallest element. This selection is the only robust choice to the perturbation of an arbitrarily small marginal cost of production and delivery.

direction. Depending on the elasticity of the matching functions in retail and wholesale markets and the search and entry cost, the retail price and the wholesale price may respond to the inventory change differently. When the wholesale price is more sensitive to the change of inventory, the equilibrium exhibits *incomplete pass-through* (Nakamura and Zerom 2010). Also, because of the co-movement, the *markup*, which is the difference between the retail price and the wholesale price, can be either positively or negatively correlated with the inventory, depending on the matching function elasticity in the retail and wholesale markets.

Recall that Lemma 1 implies that $V(x)$ increases if and only if $x < S$, where S is defined in equation (19), so $\lambda^*(x) = 0$ for any $x \geq S$. However, even if $x < S$, the marginal benefit of increasing inventory may be sufficiently small so that

$$\kappa_s > \phi_w'(\lambda)[V(x+1) - V(x)],$$

for any λ, making it impossible to generate gains from trade in the wholesale market. In this case, it is still optimal to set $\lambda = 0$. We denote by

$$s = \max\{x \in \mathbb{N} : \lambda^*(x) > 0\}, \tag{20}$$

in the equilibrium, which is referred as the *base level of stock* in the literature. Notice that $s < S$, and $\lambda^*(x) > 0$ for any $x \leq s$. Therefore, the equilibrium resembles the classic *base stock policy* in the inventory management literature (see, e.g., Porteus 2002).

Corollary 2. *In the equilibrium, the intermediary employs a base stock policy, i.e., $\lambda^*(x) > 0$ if and only if $x \leq s$.*

3.2 Steady-State Distribution and Transition Dynamics

Now we study the steady-state distribution of inventory holding and retail price. In equilibrium, the optimal controls $\theta^*(x), \lambda^*(x)$ govern the law of motion of the inventory distribution across intermediaries. With any given initial distribution g_0 at $t = 0$, the following KF equation fully describes the equilibrium dynamics as

$$\begin{aligned} \dot{g}_t(x) &= \underbrace{g_t(x-1)\phi_w(\lambda^*(x-1)) + g_t(x+1)\phi_r(\theta^*(x+1))}_{\text{inflows}} \\ &- \underbrace{g_t(x)[\phi_r(\theta^*(x)) + \phi_w(\lambda^*(x))]}_{\text{outflows}}, \end{aligned} \tag{21}$$

for every $x \in \mathbb{N}$ and any time t, with $\sum_{x=0}^{\infty} g_t(x) = 1$.

Steady-State Distribution. At steady state, $\dot{g}_t(x) = 0$ for every x, and so the distribution of inventory holding across intermediaries is constant over time.

Proposition 2. *There exists a unique steady-state distribution g_{ss} of inventory holdings across intermediaries, such that $\lim_{t\to\infty} g_t = g_{ss}$, with $g_{ss}(x) > 0$ if $0 \leq x \leq s+1$ and $g_{ss}(x) = 0$ otherwise. The distribution is unimodal and satisfies*

$$g_{ss}(x) = g_{ss}(0) \prod_{i=1}^{x} \frac{\phi_w(\lambda^*(i-1))}{\phi_r(\theta^*(i))}, \ \forall x \geq 1,$$

where

$$g_{ss}(0) = \left(1 + \sum_{x=1}^{s+1} \prod_{i=1}^{x} \frac{\phi_w(\lambda^*(i-1))}{\phi_r(\theta^*(i))}\right)^{-1}.$$

Proposition 2 says that the unique stationary inventory distribution has positive probability masses over finite $(s+2)$ inventory levels, and the probability mass function g_{ss} has a single peak, because the retail rate $\phi_r(\theta^*(x))$ increases in inventory level x, whereas the wholesale rate $\phi_w(\lambda^*(x))$ decreases.[13] The intuition behind is very simple. By Proposition 1, in equilibrium, an intermediary's expected increment of inventory is decreasing in his current inventory. Therefore, there exists a *cutoff inventory level* denoted by x^* such that the intermediary's expected increment is negative if and only if his current inventory stock is above x^*. As a result, the equilibrium inventory dynamics behaves as if a "mean" regression process: Whenever an intermediary's inventory deviates from x^*, he adjusts the retail or wholesale policy θ and λ to push the future stock back to x^*. The more the stock deviates from the mean level, the faster the speed of the regression is. In the steady state, the mass of intermediaries at the cutoff level of inventory x^* is the highest, and the mass monotonically decreases as the stock becomes farther and farther away from x^*. As a consequence, x^* is the unique mode of the steady-state distribution.

Because the intermediary retail price is monotone in his inventory size (Proposition 1), it is immediate that the equilibrium inventory dynamics shapes the steady-state distribution of retail price.

Corollary 3. *There exists a unique steady-state distribution of retail prices across intermediaries, and it is unimodal.*

That is, our model predicts that the distribution of retail price in the steady state is single-peaked. Because the inventory is most likely to be around x^*, one should expect

[13] Following Hartigan and Hartigan (1985), we say a probability distribution is unimodal (or single-peaked) if there is a mode x^* such that the cumulative density or mass function of the probability distribution is convex for $x \leq x^*$ and concave for $x \geq x^*$.

that the intermediary's retail price is equal to or close to $p^*(x^*)$ most of the time. Extremely high or low prices will be observed rarely. Notice that our model has no ex-ante heterogeneity among buyers, among sellers, or among intermediaries. The retail price dispersion is generated even if no agent randomizes, which distinguishes our model from most search models that rely on agents' heterogeneity and mixed-strategy to generate price dispersion.

We want to point out that at the steady state, an individual intermediary's price still changes over time due to inventory changes. Therefore, the equilibrium price exhibits *intra-distribution dynamics*. That is, the rank of an intermediary's price varies over time within the price distribution. This is because we assume that intermediaries are identical, so the model only generates a temporal price dispersion rather than a "spatial" or persistent price dispersion across intermediaries. This is consistent with a number of empirical studies such as Lach (2002) and Chandra and Tappata (2011). In the literature, such a phenomenon is often used to support the mixed-strategy pricing equilibrium suggested by consumer search models. Our result suggests that, to test whether firms play mixed strategies (at least in industries where inventory costs and stockout risks are non-trivial), one may also need to take into account their inventory dynamics.

Transition Dynamics. Many economic and policy relevant questions involve the transition dynamics, that is, the endogenous evolution of the economy from some initial inventory distribution. We close this section by briefly discussing the transition dynamics.

In our model, thanks to the block recursive structure, the individual equilibrium policy is independent of the inventory distribution. It is therefore sufficient to keep tracking the solution to the differential equations (9) given $\theta^*(x), \lambda^*(x)$ and some initial distribution of inventory $g_0(x)$. This tractability makes it easy to use our model to study many interesting questions such as the role of frictional supply chain in the transmission of unexpected demand and supply shock. Specifically, suppose that the seller's entry cost κ_s permanently increases at time 0 when the economy's old steady state distribution is $g_0(x)$. After the supply shock, all individuals immediately adjust their policies and the equilibrium market tightnesses change accordingly to $\theta^*(x), \lambda^*(x)$. However, it will take time for the economy to converge to the new steady state due to frictions. We illustrate it further in Section 5 after calibrating the model.

4 Extensions

In this section, we enrich our baseline model by introducing multi-unit wholesale package, product differentiation, and intermediary heterogeneity. These extensions make our model applicable to many markets and demonstrate that the model can address some questions that are usually studied in static or decision frameworks.

4.1 Multi-Unit Wholesales and the Optimality of (s, S)-Rule

In many industries, it is reasonable to assume that an intermediary can purchase multiple units when he meets a seller. Our framework can easily incorporate this feature, and some classic inventory management properties such as (s, S)-rule and non-linear wholesale pricing naturally emerge in equilibrium.

Suppose that a wholesale submarket is indexed by a bundle $(w, y) \in \mathbb{R}_+ \times \mathbb{N}$ where y is the supply quantity of the bundle and w is the total price of the bundle. To focus on the impact of search frictions, we ignore the production cost by assuming the seller's fixed and marginal production cost to be zero, so the FE condition (2) still holds. An intermediary therefore decides not only the wholesale purchase price but also the wholesale purchase quantities y by choosing a wholesale submarket. Using a similar procedure, we conclude that the intermediary acts as if to solve the following problem at any $x \geq 1$,

$$\begin{aligned} \rho V(x) = -c(x) \quad &+ \quad \max_{\theta \in \Theta} \phi_r(\theta)[u + V(x-1) - V(x)] - \kappa_b \theta \\ &+ \max_{\lambda \in \Lambda,\, y \in \mathbb{N}} \phi_w(\lambda)[V(x+y) - V(x)] - \kappa_s \lambda, \end{aligned} \tag{22}$$

and (8) at $x = 0$. The optimal $\theta^*(x)$ still satisfies condition (17), but the optimal wholesale policy $y^*(x), \lambda^*(x)$ satisfy the following necessary conditions

$$\kappa_s \geq \phi_w'(\lambda^*(x))[V(x + y^*(x)) - V(x)], \tag{23}$$

$$y^*(x) \in \arg\max_{y \in \mathbb{N}} \{\phi_w(\lambda^*(x))[V(x+y) - V(x)] - \kappa_s \lambda^*(x)\}. \tag{24}$$

The rest of the equilibrium analysis is straightforward. Using the same argument, one can show the analogy of Lemma 1 and that $\theta^*(x)$ increases and $\lambda^*(x)$ decreases.

Given the optimal controls $\theta^*(x), \lambda^*(x), y^*(x)$, the corresponding optimal retail price for the intermediary with x units of inventory can be still computed by plugging $\theta^*(x)$ into equation (13), and if he search for ordering new inventory, i.e., $\lambda^*(x) > 0, y^*(x) > 0$, his optimal wholesale price of the $y^*(x)$-unit bundle is given by plugging $\lambda^*(x)$ into equation (14).

The Optimality of (s, S)-Rule. One interesting implication of this extension is that the classic (s, S)-rule (Scarf 1960) naturally emerges in our equilibrium search model. Under this policy, an intermediary makes wholesale orders whenever the inventory level falls to or below some $s > 0$ and replenishes to a target level $S > s$. An exogenous assumption of a concave cost of ordering is often necessary to ensure the optimality of an (s, S)-rule.

In our setting, however, such a concave structure naturally emerges as an equilibrium outcome due to search frictions. The formal statement is as follows.

Proposition 3. *The optimal wholesale policy $\lambda^*(x), y^*(x)$ is an (s, S)-rule. Specifically, define $S \in \mathbb{N}$ as the minimal inventory level at which an intermediary's value achieves the maximum as in (19). Then, $\exists s \in \mathbb{N}$, $s < S$, such that*

1. *an intermediary with inventory x searches in the wholesale market if and only if its inventory is at or lower than the replenishment point, i.e., $\lambda^*(x), y^*(x) > 0$ if and only if $x \leq s$, and*
2. *whenever $x \leq s$, the intermediary seeks to raise its inventory level to the order-up-to level S, i.e., $y^*(x) = S - x, \forall x \leq s$.*

The logic is as follows. Suppose that $\lambda^*(x), y^*(x) > 0$ for some $x \leq s$, then by (24) the intermediary's optimal order quantity $y^*(x)$ must satisfy

$$y^*(x) + x = \arg\max_{z \geq x} \phi_w(\lambda^*(x))[V(z) - V(x)] - \kappa_s \lambda^*(x).$$

With $\lambda^*(x)$ being independent of the order quantity, it must be optimal to set $y^*(x) = S - x$. Therefore, if the intermediary decides to search in the wholesale market for replenishment, it seeks to order to the level that maximizes the equilibrium continuation value.

When $x \geq S$, the value function $V(x)$ decreases, and there is no benefit to order more inventory, so $\lambda^*(x) = 0$. When $x < S$, the gain from ordering up $V(S) - V(x)$ is positive and increases as the inventory level x goes down due to the concavity of the value function. When x is sufficiently low, it is optimal to set $\lambda^*(x) > 0$. We define s as in equation (20). Notice that $s = S - 1$ if κ_s is sufficiently small. The idea is visualized in Figure 1.

Remark 1. *When the marginal production cost is $\delta > 0$, the optimal ordering policy will satisfy an adjusted (s, S)-rule. First, it is still optimal to set $\lambda^*(x) > 0$ only if $x \leq s$, but when $x \leq s$, the optimal quantity of order will not be constant but satisfies $y^*(x) + x = S(x) = \arg\max_{z \in \mathbb{N}} V(z) - \delta z$.*

Equilibrium Non-Linear Pricing. Another interesting feature of the equilibrium is that the equilibrium price-quantity relationship in the wholesale market is in general non-linear. When $x \leq s$, the equilibrium price $w(\lambda^*(x))$ decreases in x, and the quantity-price relationship is non-linear across bundles being traded in the wholesale market, that is, the "unit wholesale price" $w(\lambda^*(x))/y^*(x)$ is not constant across $x \in \{0, 1, ..., s\}$. Specifically, for any $x \leq s$, the intermediary aims to place a wholesale order of $y^*(x) = S - x$ units, and the equilibrium cost of the bundle is $w(\lambda^*(x)) = \kappa_s / \psi_w(\lambda^*(x))$ according to

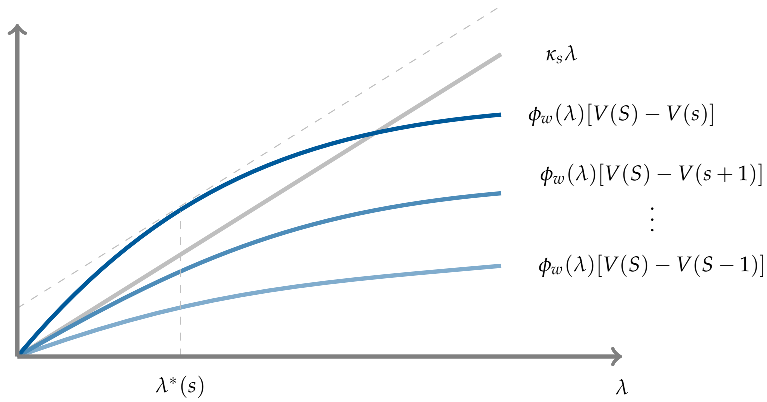

Figure 1: Illustration of the optimality of the (s, S)-rule.
The horizontal axis represents the choice of λ, and the vertical axis represents the value of corresponding benefit and cost of each λ.

(14). Consequently, the unit price becomes $w(\lambda^*(x))/y^*(x) = \kappa_s/[(S-x) \times \psi_w(\lambda^*(x))]$, a nontrivial function of x rather than a constant.

Remark 2. *The non-linear pricing result should be expected given the optimality of the (s,S)-rule. In the literature, to sustain the (s,S)-rule as an equilibrium choice, it is often assumed that the intermediary faces a non-linear price-quantity relationship. In contrast, the price non-linearity endogenously emerges in our equilibrium search model.*

Steady-State Distribution. Now we study the cross-sectional distributions of inventory holdings and retail prices. Using the equilibrium policies $\theta^*(x), \lambda^*(x), y^*(x)$ and starting from any initial distribution g_0, the law of motion of the inventory distribution across intermediaries can be written as the following KF equations:

$$\dot{g}_t(x) = \begin{cases} \sum_{x' \leq s} g_t(x')\phi_w(\lambda^*(x')) + g_t(S+1)\phi_r(\theta^*(S+1)) - g_t(S)\phi_r(\theta^*(S)) & \text{if } x = S, \\ g_t(x+1)\phi_r(\theta^*(x+1)) - g_t(x)[\phi_r(\theta^*(x)) + \phi_w(\lambda^*(x))] & \text{if } x \neq S, \end{cases} \tag{25}$$

and $\sum_{x=0}^{\infty} g_t(x) = 1, \forall t$. The left-hand side of equation (25) is the time derivative of the measure of intermediaries who hold x units of inventory at time t. The right-hand side of equation (25) depends on the value of x. When $x = S$, the inflow $g_t(x')\phi_w(\lambda^*(x'))$ is the mass of intermediaries who are holding $x' \leq s$ units of inventory and successfully find sellers, trade, and increase their inventory up to S. Such inflows occur at $x = S$ only due to the optimal (s, S)-rule. Another inflow $g_t(S+1)\phi_r(\theta^*(S+1))$ accounts for any mass of intermediaries holding $S+1$ units of inventory and successfully find buyers, trade, and lower their inventory to S. The outflow $g_t(S)\phi_r(\theta^*(S))$ is similarly the mass of intermediaries who are holding S units of inventory and successfully find buyers, trade, and lower their inventory to $S-1$. When $x \neq S$, there is no inflow due to inventory replenishment in the wholesale market, but there may be an additional outflow of intermediaries $g_t(x)\phi_w(\lambda^*(x))$. Specifically, $\lambda^*(x) > 0$ if $x \leq s$, and this additional outflow is positive because these low inventory intermediaries may successfully order up to S; otherwise, intermediaries with sufficient inventory do not place wholesale order, and $\lambda^*(x) = 0$.

At steady state, $\dot{g}_t(x) = 0$ for every x and t, and the distribution of inventory holdings across intermediaries is stationary over time, denoted as g^m_{ss}. The following proposition characterizes it.

Proposition 4. *Suppose that intermediaries can make multi-unit wholesale orders. There exists a unique stationary distribution g^m_{ss} of inventory holdings across intermediaries, such that $\lim_{t\to\infty} g_t = g^m_{ss}$, with $g^m_{ss}(x) > 0$ if $0 \leq x \leq S$, and $g^m_{ss}(x) = 0$ otherwise. The distribution satisfies*

$$g^m_{ss}(x) = g^m_{ss}(S) \prod_{i=x}^{S-1} \frac{\phi_r(\theta^*(i+1))}{\phi_r(\theta^*(i)) + \phi_w(\lambda^*(i))}, \ \forall 0 \leq x < S,$$

where $\phi_w(\lambda^*(i)) = 0$ if $i \geq s+1$, $\phi_r(\theta^*(0)) = 0$, and

$$g^m_{ss}(S) = \left(1 + \sum_{x=0}^{S-1} \prod_{i=x}^{S-1} \frac{\phi_r(\theta^*(i+1))}{\phi_r(\theta^*(i)) + \phi_w(\lambda^*(i))}\right)^{-1}.$$

The distribution's mode is between $x = 0$ and $x = s+1$.

Proposition 4 says that the unique stationary inventory distribution has positive probability mass over finite $(S+1)$ inventory levels. One may conjecture that there is a mode at $x = S$ because all intermediaries with $x \leq s$ aim to bring their inventory up to level S. However, this is only half of the story. An intermediary with S units of inventory will sell at a sufficiently low retail price, and so the transition probability from $x = S$ to $x = S-1$ will be sufficiently large, bringing up the mass of intermediaries holding $S-1$ units of inventory in the steady state. A similar argument applies to intermediaries with $S-2, S-3, ...$ units of inventory. Moreover, for any $s < x < S$, the steady-state condition $\dot{g}_t = 0$ implies that

$$g^m_{ss}(x+1)\phi_r(\theta^*(x+1)) = g^m_{ss}(x)\phi_r(\theta^*(x)).$$

Since both $\theta^*(\cdot)$ and $\phi_r(\cdot)$ are increasing, we must have $g^m_{ss}(x+1) \leq g^m_{ss}(x)$ for $x > s$, and so the mode is between $x = 0$ and $x = s+1$.

As in the benchmark model, the intermediary retail price is monotone in his inventory size, so the steady-state distribution of retail price can be easily characterized by the retail policy $p(\theta^*(s))$ and the inventory steady-state distribution, which is omitted.

4.2 Horizontal Product Differentiation

In the literature of industrial organization, a popular way to capture (horizontal) product differentiation and consumer taste heterogeneity is to introduce idiosyncratic utility into the model: a buyer's payoff by consuming a product is a random variable $\tilde{u}$ (see, e.g., Anderson et al. (1992) for a textbook treatment). In this section, we introduce random utility into our framework and demonstrate how the presence of horizontal product differentiation may alter the equilibrium characterization of the benchmark model.

Suppose the buyer-product match-specific utility is independently and identically distributed across buyers and products. When a buyer and an intermediary meet, the buyer will pick his favorite product that delivers positive payoff. For simplicity, assume that the match between a buyer and a product is randomly good or bad. A good match occurs with probability $\alpha \in (0, 1]$, such that the buyer receives utility u by consuming the product; the match is bad with complementary probability, and consumption delivers zero utility to the buyer. The match between a buyer and an intermediary with inventory x is good if the buyer finds at least one good match with the x products, with probability

$$\Phi(x) = 1 - (1 - \alpha)^x,$$

which is strictly increasing and concave in x. Therefore, holding a large number of inventory endows the intermediary another advantage: reducing the possibility of *mismatch*. In this case, a retail submarket is indexed by (p, x), the price and the inventory size of the intermediaries who trade in this market. In equilibrium, a match between a buyer and an intermediary will lead to a transaction if and only if the match is good and generates strictly positive gain from trade. Therefore, the intermediary's problem (16) becomes

$$\begin{aligned} \rho V(x) = -c(x) \quad &+ \quad \max_{\theta \geq 0} \phi_r(\theta)\Phi(x)[u + V(x-1) - V(x)] - \kappa_b \theta \\ &+ \quad \max_{\lambda \geq 0} \phi_w(\lambda)[V(x+1) - V(x)] - \kappa_s \lambda. \end{aligned} \tag{26}$$

The equilibrium FOC that the optimal $\lambda^*(x)$ must satisfy is unchanged, whereas the optimal $\theta^*(x)$ must satisfy

$$\kappa_b \geq \phi_r'(\theta^*(x))\Phi(x)[u + V(x-1) - V(x)]. \tag{27}$$

Because $\Phi(x)$ is strictly increasing and concave in x, one can verify that the optimal $\theta^*(x)$ is still increasing in x. This is because the expected gain from trade between a matched intermediary-buyer pair, $\Phi(x)[u + V(x-1) - V(x)]$, increases in x. In words, when x is higher, each match between a buyer and the intermediary will more likely lead to a transaction, so it is socially optimal to let more buyers search. In fact, the extra term $\Phi(x)$ gives the intermediary a stronger incentive to hold a large amount of inventory, which is to increase the probability of a good match. Decreasing α will naturally intensify this scale effect and the intermediaries' incentive to become big, which shifts the steady-state distribution of inventory toward the right.

In the benchmark model, the equilibrium relationship between an intermediary's inventory and the optimal choice of retail price is monotone. With product differentiation, this monotonicity is no longer guaranteed.

The tightness of each retail submarket must satisfy

$$\kappa_b \geq \psi_r(\theta(p))\Phi(x)(u - p), \tag{28}$$

and $\theta(p) \geq 0$ with complementary slackness, so the equilibrium price in each retail submarket with a positive tightness is given by

$$u - \frac{\kappa_b}{\psi_r(\theta^*(x))\Phi(x)}. \tag{29}$$

In the equilibrium, $\psi_r(\theta^*(x))$ is decreasing in x while $\Phi(x)$ is increasing in x, so the retail price may no longer be monotone in the inventory x. The intuition is as follows. When his inventory increases, the intermediary wants to sell faster, so he enters a retail submarket with higher θ. From the perspective of buyers, it is less likely to meet an intermediary in a submarket with higher θ, but conditional on meeting an intermediary, it is more likely to find a desired product due to the intermediary's larger inventory size. Therefore, the effective matching probability $\psi_r(\theta)\Phi(x)$ and the buyer's willingness to pay may not be monotone in x in the equilibrium. We summarize the above discussion as follows.

Proposition 5. *In equilibrium, an intermediary's optimal retail price is given by expression* (29), *which may be non-monotone in x.*

The non-monotone relationship between retail price and inventory implies that even though the steady-state distribution of inventory g_{ss} is still unimodal, the distribution of retail prices may not be.[14]

4.3 Vertical Product Differentiation

This section discusses how to introduce vertical product differentiation into our framework. Suppose that there are J quality types, each of which delivers utility u_j to the buyer, such that $u_{j+1} > u_j, \forall 1 \leq j < J$. We assume symmetric information, so different types of

[14]Numerical examples of non-single peaked steady-state price distribution are available upon request.

products are traded in different submarkets. Then the retail and wholesale submarkets can be indexed by the price and the type of product being traded. In equilibrium, the free-entry condition still holds for any submarket with positive market tightness.

It is natural to allow intermediaries to hold multiple types of product, i.e., an intermediary's inventory is a vector $x = (x_1, \ldots, x_J) \in \mathbb{N}^J$ where x_j denote the inventory of type-j product. Following exact the same argument presented in the baseline model, the intermediary acts as if to decide its retail and wholesale policy for each type of product $\theta^j(x), \lambda^j(x)$. The corresponding HJB becomes

$$\begin{aligned} \rho V(x) = -c(x) \quad &+ \quad \sum_{j=1}^{J} \Big\{ \max_{\theta^j \in \Theta^j} \phi_r(\theta^j)[u_j + V(x_j - 1, x_{-j}) - V(x)] - \kappa_b \theta^j \Big\} \\ &+ \quad \sum_{j=1}^{J} \Big\{ \max_{\lambda^j \in \Lambda} \phi_w(\lambda^j)[V(x_j + 1, x_{-j}) - V(x)] - \kappa_s \lambda^j \Big\}, \end{aligned} \tag{30}$$

where the cost function is $c : \mathbb{N}^J \to \mathbb{R}_+$, increasing in each argument, and x_{-j} represents the vector $(x_1, .., x_{j-1}, x_{j+1}, ...x_J)$. By similar argument, we can show that the corresponding equilibrium retail and wholesale prices are given by $u_j - \kappa_b/\psi_r(\theta^j(x))$ and $\kappa_s/\psi_w(\lambda^j(x))$ respectively. Unfortunately, this multi-dimensional dynamic optimization problem is analytically intractable in general. The following proposition characterizes the equilibrium policy when the cost function is additive.

Proposition 6. *Suppose that the cost function is additive, i.e.,*

$$c(x) = \sum_{j=1}^{J} c^j(x_j), \tag{31}$$

where $c^j : \mathbb{N} \to \mathbb{R}_+$ is the inventory cost for product type j. Then the optimal policy is such that $\theta^j(x)$ and $\lambda^j(x)$ depend on x through x_j only, and the value function satisfies

$$V(x) = \sum_{j=1}^{J} V^j(x_j), \tag{32}$$

where $V^j(x_j)$ corresponds to the type-j product problem such that

$$\begin{aligned} \rho V^j(x_j) = -c^j(x_j) \quad &+ \quad \max_{\theta^j \in \Theta^j} \phi_r(\theta^j)[u_j + V^j(x_j - 1) - V^j(x_j)] - \kappa_b \theta^j \\ &+ \quad \max_{\lambda^j \in \Lambda} \phi_w(\lambda^j)[V^j(x_j + 1) - V^j(x_j)] - \kappa_s \lambda^j. \end{aligned} \tag{33}$$

Proposition 6 says that when the cost function is additive, a multi-product intermediary acts as multiple single-product intermediaries. The proof is to use the standard verification argument. By plugging equations (31) and (32) into equation (30), one can verify that the HJB equation is balanced. A simple parametric example of the inventory cost being additive is to specify the inventory cost as a linear function of x; i.e., $c(x) = \sum_j c^j x_j$ where the marginal inventory cost $c^j \geq 0, \forall j$. If c^j is constant across j, the inventory cost depends only on the total inventory $\sum_{j=1}^{J} x_j$. In this case, the KF equation and the steady-state distribution of inventory can be characterized separately for each product as in the baseline model, which is omitted.

4.4 Heterogeneous Intermediaries

Intermediaries may be heterogeneous in their inventory costs and matching technologies. For example, some intermediaries have outstanding marketing and sales managers, bringing them high visibility to buyers; some have effective purchasing departments and maintain good relationship with manufacturers, allowing them to be part of an efficient supply chain; some have superior transportation or handling teams or low opportunity cost of the money, admitting low marginal inventory cost. These heterogeneities can lead to variations in expected inventory sizes, sales and profitability, and, therefore, different inventory-price relationships among intermediaries.

It is straightforward to extend our model to accommodate intermediary heterogeneity. Specifically, there are J types of intermediaries, and the proportion of each type j is denoted by f_j. Denote $c^j(\cdot)$, $\phi_r^j(\cdot)$, and $\phi_w^j(\cdot)$ as respective functions for type-j intermediaries' inventory cost, retail matching rate, and wholesale matching rate. A retail submarket is indexed by (p, j); a wholesale submarket is indexed by (w, j). Accordingly, the market tightnesses are $\theta^j(p)$ and $\lambda^j(w)$. Using an almost identical argument, we can show that, in the unique block recursive equilibrium, each type-j intermediary's value solves the type-specific dynamic optimization problem

$$\begin{aligned} \rho V^j(x) = -c^j(x) \quad &+ \quad \max_{\theta \in \Theta^j} \phi_r^j(\theta)[u + V^j(x-1) - V^j(x)] - \kappa_b \theta \\ &+ \quad \max_{\lambda \geq \Lambda^j} \phi_w^j(\lambda)[V^j(x+1) - V^j(x)] - \kappa_s \lambda. \end{aligned} \tag{34}$$

Given the optimal control $\theta^j(x), \lambda^j(x)$, the corresponding retail and wholesale prices for type-j intermediary can be computed using conditions (13) and (14) as in the benchmark model. That is, a type-j intermediary's optimal retail and wholesale prices are given by $p(\theta^j(x)) = u - \frac{\kappa_b \theta^j(x)}{\phi_r(\theta^j(x))}$, and $w(\lambda^j(x)) = \frac{\kappa_s \lambda^j(x)}{\phi_w(\lambda^j(x))}$, respectively. The within-type inventory distribution evolves according to a type-specific KF equation, given as

$$\begin{aligned} \dot{g}_t^j(x) &= g_t^j(x-1)\phi_w^j(\lambda^j(x-1)) + g_t^j(x+1)\phi_r^j(\theta^j(x+1)) \\ &- g_t^j(x)[\phi_r^j(\theta^j(x)) + \phi_w^j(\lambda^j(x))], \end{aligned} \tag{35}$$

for every $x \in \mathbb{N}$ and at each moment t, with $\sum_{x=0}^{\infty} g_t^j(x) = 1, \forall t, j$. The steady-state distribution for type-j intermediary's inventory, denoted by $g_{ss}^j(x)$ can be computed accordingly, and it has a single peak by the same argument. The overall steady-state distribution of inventory is $g_{ss}(x) = \sum_{j=1}^{J} f_j g_{ss}^j(x), \forall x$. The cross-sectional retail price distribution follows, which necessarily depends on each g_{ss}^j and the type distribution f_j.

5 Application to Used-Car Markets

In this section, we apply the model to study used-car dealers' inventory management and dynamic pricing by using detailed information on used-car listings (inventories) by a large number of car dealers. The empirical exercise serves multiple purposes. First, the empirical exercise serves as a test case for our model. We show (1) how to identify key parameters with limited information, and (2) our model of search frictions and inventory predicts transition dynamics. Second, our focus on the used-car market is policy relevant. We quantify some important unobservable characteristics of market participants and the contribution of frictional intermediaries managing inventories to welfare and analyze how changing primitives in this market leads to different outcomes – which is particularly important given recent market disruptions in this industry. There is a growing literature studying the economics of car dealers, and we are the first to focus on the role of inventories (see, e.g., Gavazza et al. (2014), Biglaiser et al. (2020), Larsen (2021), and Gillingham et al. (2022)).

5.1 On Used-Car Markets

While the practice of inventory management plays out in many real-world settings, several factors make the used-car market suitable for our study. First, the market is highly decentralized, making the search and matching frictions non-trivial. As a result, many transactions are intermediated. Nationally, about two-thirds of used-car sales are made by dealers. Second, inventory management is important for used-car dealers. In general, dealers must manage both value erosion as assets age and holding costs, which include floor-plan inventory investment and cost of capital.[15] Third, cars are durable goods. Most buyers and sellers do not make frequent transactions, so it is uncommon for dealers to manage inventory acquisition with long-term contracts.[16] Fourth, stocking decisions can be made frequently. Dealers face substantial uncertainty and typically acquire used cars from individuals or at wholesale auctions. Dealers may have access to multiple auctions a week at multiple auction locations. Fifth, dealers frequently adjust prices. These features suggest that the interaction between inventory control and search friction is important in the used-car market, making our theory applicable.[17]

Before moving forward, we would like to further elaborate the applicability (and limitation) of our model to the used-car market. First, our model does not consider intermediaries' entry/exit. Our data includes dealers' listing and pricing in one year. In this relatively short-term period, we do not anticipate significant structural change in the market. Second, our directed search model assumes all prices are observable. In the used-car application, this information assumption is materialized by the aggregator such as `cars.com`: we implicitly assume that all agents check and compare prices online before visiting dealers. Third, as we discussed in the model section, a submarket corresponds a set of agents whose target purchase/selling price is consistent with the submarket price. See more discussion in section 2.2. In Appendix B, we provide some preliminary evidence supporting the hypothesis of directed search models.

5.2 Data

We obtain information on used-car listings from a large car listings platform, `cars.com`. We observe the daily listings for dealers who list inventory on the platform in the state of Ohio in 2017. For each car, we know the Vehicle Information Number (VIN), which is a unique number assigned to a vehicle that contains information to describe and identify the vehicle, make, model, model year, and trim with a particular set of options, exterior color, odometer mileage, whether it is certified by the OEM, and the daily listing price from the date when it is initially listed to the date when it is removed from the website.

[15] An inventory management expert Jasen Rice of LotPop said "For a dealer having 50 units or fewer on the lot, one or two inventory management mistakes can crush their month." See https://www.cbtnews.com/dealers-experts-discuss-inventory-holding-cost-erosion/ for details.

[16] On the contrary, new car dealers sign long-term contracts, e.g., dealership agreement, with manufactures and essentially act as their representatives.

[17] Our general understanding of the industry is from various industry reports, including Edmunds' "Used Vehicle Market Report," Manheim's "Used-Car Market Report," and Murry and Schneider (2015). For industry reports, see `https://dealers.edmunds.com/static/assets/articles/2017_Feb_Used_Market_Report.pdf` and `https://publish.manheim.com/content/dam/consulting/2017-Manheim-Used-Car-Market-Report.pdf`

Notably, the platform's pricing is not marginal to the number of cars listed, and the platform reports that dealers typically list their entire inventory on the platform. Moreover, according to our conversations with `cars.com`, most car dealers update their listings on the platform immediately. Therefore, we are confident that a dealer's new listings, listing removals, and active listings at a point of time are the actual new orders, car removals, and inventory in the dealership at that time, respectively. Although our data do not allow us to identify where a newly added car is obtained from and where a removed car goes to, a dealer's new orders and car removals at a point of time are good measures of the inflows and outflows of that dealer's inventory, which is our primary focus.[18]

We focus on four to six years gasoline sedans of non-luxury brands and treat them as the same product.[19] This group of cars accounts for 12.4% of all listings on cars.com during the sample period. We choose this group of cars for our analysis for the following two reasons. First, the majority of all transactions of this group of cars are sold by dealers (see Figure 1 of Biglaiser et al. (2020)). Second, this group of cars are relatively homogeneous compared to older cars and luxury cars. We count each dealer's inventory as the number of these cars. We implicitly assume that dealers make stocking and pricing decisions for this product segment independently of decisions for other segments. This assumption is reasonable if the dealer's inventory cost is additive according to Proposition 6. We also acknowledge dealers' heterogeneous retail and wholesale behavior pattern due to their size difference. We consider two groups of dealers according to their average inventories. Small dealers are those whose average inventory during the sample year is fewer than 10 cars, while large dealers are those whose average inventory is between 10 and 20 cars.[20]

After selecting car types, we end up with 16,239 used cars listed by 259 small dealers and 15,551 listed by 133 large dealers over the course of a year. Table 1 reports the sample statistics of dealer-week-level inventory and inventory change and car-level prices separately for small dealers (Panel A) and large dealers (Panel B). On average, small dealers hold 7 units of non-luxury 4-6-year-old sedans and large dealers hold 13 vehicles in this product segment. Moreover, the list price of smaller dealers is higher than that of large dealers.

[18]A newly added car can come from a wholesale trade-in or dealer-to-dealer auction market, or just be allocated from another site if the dealer is a chain store. Similarly, a removed car can be sold to an individual or another dealer or reallocated to another site if the dealer is a chain store.

[19]The brands we consider include Chevrolet, Chrysler, Dodge, Ford, GMC, Honda, Hyundai, Jeep, Kia, Mazda, Mercury, Mitsubishi, Nissan, Pontiac, Saturn, Subaru, Suzuki, Toyota, and Volkswagen.

[20]We drop 29 very large dealers whose average inventory of this particular type of car ranges from 20 to 54, including the five CarMax stores in Ohio. It is tempting to include them in our quantitative analysis, but unfortunately, there are too few observations and these dealers substantially differ from each other in sizes and inventory patterns. For example, their inventory of the 4-6 year old sedan segment ranges from zero to 88. So we drop them from our analysis.

Table 1: Descriptive Statistics

Panel A. Small Dealers[†]	Mean	SD	Min	P25	P50	P75	Max
Inventory (dealer-week)	7.158	3.377	0	5	7	9	26
Inventory change (dealer-week)	-0.038	1.482	-18	-1	0	1	11
List price ($, car listing)	11,513	5,399	5,900	8,835	10,288	12,988	34,898
Weeks on market (car listing)	7.441	7.082	1	2	5	10	35
Panel B. Large Dealers[†]	Mean	SD	Min	P25	P50	P75	Max
Inventory (dealer-week)	13.218	5.611	0	10	12	16	52
Inventory change (dealer-week)	-0.127	2.339	-16	-1	0	1	12
List price ($, car listing)	11,339	4,229	5,494	8,920	10,500	12,990	27,990
Weeks on market (car listing)	6.627	6.284	1	2	5	9	30

Notes. Data source: Cars.com. Sample selection is described in text.
† The sample of small dealers includes 13,209 dealer-week observations and 16,239 car listings.
‡ The sample of large dealers includes 6,783 dealer-week observations and 15,551 car listings.

5.3 Parametric Specification

As demonstrated in Section 4, our benchmark model can be enriched in many ways, so we tailor the model to our empirical application. The model that we take to the data is the model described in Section 4.4, which extends the baseline model to heterogeneous intermediaries. We allow for two types of dealers that differ in their inventory costs and matching functions. In our data, there is a lot of variation in dealer size, and given this there is good reason to believe that dealers have different primitives in their objetive functions.

There are other potential features of used car markets that we do not capture in the model that we take to the data. We focus on a single vehicle segment, 4-6 year-old non-luxury sedans. For this reason, we don't further model vertical differentiation. Because of our choice of vehicle segment, we also ignore issues related to asymmetric information. Although used cars are the canonical example of a lemons market, there is more recent research that suggests asymmetric information problems are not severe, particularly for late model vehicles – see Adams et al. (2011) and Biglaiser et al. (2020). Used-car dealers have a variety of channel to acquire inventories including trade-in of new buyers, participating auctions, etc. A dealer with low inventory may be able to order multiple units at once. Unfortunately, we do not observe the source of dealers' inventory addition. If a dealer adds multiple inventory within a week, the data does not allow us to distinguish whether they are purchase in one order or multiple ones. For simplicity, we keep the single-unit order assumption as in the benchmark model. This brings the risk of the seller side outside option and surplus being misspecified. Lastly, although there is likely unobserved horizontal product taste across consumers, but we can not separately identify cars that are rejected by buyers due to a bad match from the matching function itself because we don't observe failed dealer visits. We ignore the role of unobserved product tastes, out extension in Section 4.2.

We assume that search frictions on both the retail and the wholesale markets are summarized by type-specific scaled urn-ball matching functions given as

$$\phi_r^j(\theta) = \mu_r^j(1 - e^{-\theta}),$$
$$\phi_w^j(\lambda) = \mu_w^j(1 - e^{-\lambda}),$$

where $\mu_r^j, \mu_w^j > 0$ are scaling parameters to capture search frictions for each type $j = 1, 2$, and they ensure that the matching rates are bounded. Bounded matching rates further ensures that we can transform the continuous-time Markov decision process described by the HJB equation (16) into an equivalent discrete-time problem using an uniformization technique (see, e.g., Guo and Hernández-Lerma, 2009, Chapter 6). As Poisson matching rates rather than probabilities, ϕ_r^j and ϕ_w^j can be greater than 1 if μ_r^j or μ_w^j is above 1.[21] Our matching-function choice is motivated by Peters (2000) and Burdett et al. (2001), who provide the micro foundation of an urn-ball matching function as a limit result of a finite directed search game as the number of traders goes to infinity.

We assume linear inventory costs with type-specific *marginal* cost parameters $c^j \geq 0$, for $j = 1, 2$, such that

$$c^j(x) = c^j x.$$

Notice that the linear cost specification satisfies the additive condition (31). By Proposition 6, it is without loss to treat a multi-product intermediary as multiple single-product intermediaries.

5.4 Parameter Identification

Our model is in continuous time. Accordingly, we treat the data as a finite sample of periodic observations of a continuous-time data-generating process. More specifically, a continuous-time process has a realized path $x(t)$ with $t \in [0, T]$, and our sample consists of observations $\{x(0), x(\Delta), \ldots, x(n\Delta)\} \subset x([0, T])$, where $\Delta > 0$ is the time interval between two observations. We normalize $\Delta = 1$ to be a week to match the data frequency.

The discount rate ρ is predetermined and matches a 5% annual rate, such that

$$1 - e^{-52\rho} = 5\%.$$

[21] Recall that the probability of a type-j intermediary meeting a buyer (or seller) within a small time period of length $dt > 0$ is roughly $\phi_r^j dt$ (or $\phi_w^j dt$).

The remaining parameters to be calibrated are a buyer's utility u, matching function parameters (μ_r, μ_m), the buyer and seller's respective outside options κ_b and κ_s, and the marginal inventory cost c^j for each type $j = 1, 2$. In the used-car setting, a buyer's outside options include keeping their current car, buying a new car, buying an old car from other sources such as friends or relatives, etc, or using other transportation options. The seller's outside options include selling the car by himself or keeping it.

To understand the challenge of model identification, notice that although the theoretical model considers a two-sided market with three types of agents: buyers, sellers, and intermediaries (dealers), who interact in retail and wholesale markets, the data we use only contains information of dealers' inventory and list retail prices. Specifically, we do not observe wholesale prices. We show in the following discussion how to utilize analytical implications of the model to identify the relevant parameters step by step. Identification is sequential, such that in each step, we show how to express a subset of unknown parameters as a closed-form function of moments of the data and parameters that are already shown to be identified from a previous step. Identification does not rely on, and is not complicated by, the two types of dealers, so we drop the dealer type indexing in the discussion.

Step 1: Matching rates. First, we show that the transition probability matrix of inventory levels identifies the unobserved Poisson rates of matching in both markets at each inventory level $\phi_r^*(x) \equiv \phi_r(\theta^*(x)), \phi_w^*(x) \equiv \phi_w(\lambda^*(x))$, fully determining the pattern of dealers' inventory transition. Our approach resembles the standard approach in the labor search literature (see, e.g., Menzio and Shi (2011) and Guo (2018) where the state variable is workers' employment status). However, unlike the employment status that switches infrequently, an intermediary's state is the inventory level, which may increase or decrease quickly.

In equilibrium, each intermediary's inventory level follows a continuous-time Markov chain over $s+2$ states $\{0, 1, \ldots, s+1\}$, where s is given in equation (20). The process is induced by intermediaries optimal controls $\theta^*(x)$ and $\lambda^*(x)$ in Proposition 1, such that the transition rates are captured by the matching rates $\phi_r^*(x)$ and $\phi_w^*(x)$. An $(s+2) \times (s+2)$ square matrix $Q = [Q_{xy}]$ summarizes these transition rates, where each entry Q_{xy} is given as

$$Q_{xy} = \begin{cases} \phi_r^*(x) & \text{if } y = x - 1 \geq 0, \\ -[\phi_r^*(x) + \phi_w^*(x)] & \text{if } y = x, \\ \phi_w^*(x) & \text{if } y = x + 1 \leq s + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Each column of Q contains at most three non-zero entries, reflecting the inflow and outflow rates in equation (21) at the corresponding inventory level.[22] Therefore, using Q, the transition dynamics for the cross-sectional inventory distribution in Kolmogorov Forward (KF) equation (21) can be written as $\dot{g}_t = g_t Q$, where $g_t = (g_t(x))_{x=0,...,s+1}$ is a $1 \times (s+2)$ vector for the inventory distribution at t, and $\dot{g}_t(x)$ is another $1 \times (s+2)$ vector for the law of motion. The stationary inventory distribution g_{ss} thus satisfies $0 = g_{ss}Q$.

If Q is observable, the matching rates are readily available. However, Q as a matrix for Poisson transition rates does not have an immediate data counterpart. What can be computed from the data is the weekly inventory transition *probability* matrix, denoted by another $(s+2) \times (s+2)$ square matrix $P = [P_{xy}]$. Each entry of the matrix is

$$P_{xy} = \Pr(X_{\tau+1} = y \mid X_\tau = x), \quad \forall x, y \in \{0, 1, ..., s+1\},$$

which is the probability that inventory level X_t changes from x to y in a week. See the two top panels of Figure 2 for a visualized illustration of each dealer type's inventory transition matrix.

We transform the transition rate matrix Q into a weekly inventory transition probability matrix P. For such a continuous-time Markov process, i.e., a general birth-death process over finite states, it is a known result that there is a one-to-one mapping between Q and an associated transition probability matrix $P(t)$ over time $t > 0$, such that each xy-entry represents $P_{xy}(t) = \Pr(X_{\tau+t} = y \mid X_\tau = x)$ for any $x, y \in \{0, 1, ..., s+1\}$ and any time $\tau \geq 0$. See, for example, Chapter 6 of Pinsky and Karlin (2010). The matrix $P(t)$ satisfies

$$P(t) = e^{tQ} \equiv \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k,$$

where $Q^0 = I$ is the identity matrix at $k = 0$, and Q^k is the k-th power of the square matrix Q. The weekly transition probability matrix is simply $P = P(1) = e^Q$ given our normalization. Notice that although a dealer is allowed to sell or buy at most one car at each instant, it may sell or buy *multiple* cars over a week.

The empirical counterpart of P thus disciplines the matching rates $\{\phi_r^*(x), \phi_w^*(x)\}_{x=0}^{s+1}$ with $\phi_r^*(0) = 0$ and $\phi_w^*(s+1) = 0$. Note that the empirical counterpart of the stationary distribution g_{ss} contains less information than that of P, as $g_{ss} = g_{ss}P$. The remaining identification argument treats these matching rates as observable. We rely on the assumed functional forms and the equilibrium conditions to recover the parameters.

Step 2: Retail-market parameters. Second, we show that the retail prices $p^*(x)$ and the intermediaries' retail-market matching rates $\phi_r^*(x)$, determined from the previous step, can identify the retail-side parameters u, κ_b, and μ_r via the buyers' free-entry condition

[22] See Chapter 6 of Pinsky and Karlin (2010) for an introduction of continuous-time Markov chain.

rewritten as equation (13).

Given $\phi_r^*(x)$ and the urn-ball matching function, the retail submarket tightness and the buyer-side matching rate become nonlinear functions of μ_r, respectively denoted as $\theta^*(x;\mu_r)$ and $\psi_r^*(x;\mu_r)$, satisfying

$$\theta^*(x;\mu_r) = \ln\frac{\mu_r}{\mu_r - \phi_r^*(x)}, \quad \psi_r^*(x;\mu_r) = \frac{\phi_r^*(x)}{\theta^*(x;\mu_r)}.$$

Then, the free-entry condition (13) can be written as

$$p(x) = u - \frac{\kappa_b}{\psi_r^*(x;\mu_r)}.$$

It follows that, when we observe retail prices $p(x)$ at more than three inventory levels with sufficient variations, parameters u, κ_b, and μ_r are jointly determined.

Step 3: Wholesale-market Parameters. Third, we turn to the wholesale side and show that the intermediaries' FOCs (17) and (18) identify the wholesale-side parameters κ_s and μ_w.

Given $\phi_r^*(x)$, parameter μ_r and the urn-ball matching function, the first-order derivative $\phi_r'(\theta(x))$ in equation (17) becomes $\mu_r - \phi_r^*(x)$. Then, given u, κ_b, and μ_r in the previous step, equation (17) pins down $V(x) - V(x-1)$ as

$$V(x) - V(x-1) = u - \frac{\kappa_b}{\mu_r - \phi_r^*(x)}.$$

Similarly, given the wholesale-side matching rate $\phi_w^*(x)$ and the urn-ball matching function, the first-order derivative $\phi_w'(\lambda(x))$ in equation (18) becomes $\mu_w - \phi_w^*(x)$, which is a function of μ_w. Then equation (18) becomes

$$V(x) - V(x-1) = \frac{\kappa_s}{\mu_w - \phi_w^*(x-1)},$$

which pins down κ_s and μ_w jointly.

Step 4: Marginal inventory cost. Lastly, the stationary HJB equation (16) identifies the marginal inventory cost c. Specifically, we take the first-order difference on both sides of equation (16), which results in an expression with c on the right-hand side being the only unknown element.

5.5 Parameter Values

We calibrate the parameter values using the simulated method of moments. Guided by the identification arguments, for each type $j = 1, 2$, we choose the weekly inventory transition matrix P^j, the cross-sectional inventory distribution g^j, and the logarithms of the average retail prices by inventory level $\ln p^j$ as the targets. We select the parameters $(u, \kappa_b, \kappa_s, (\mu_r^j, \mu_w^j, c^j)_{j=1,2})$ such that they solve

$$\min_{u,\kappa_b,\kappa_s,(\mu_r^j,\mu_w^j,c^j)_{j=1,2}} \sum_{j=1,2} \sum_{m=P,g,\ln(p)} \|m^j_{model} - m^j_{data}\|_2^2,$$

where $\|\cdot\|_2$ represents the L^2 norm.

Table 2 reports the parameter values. Figure 2 shows goodness of fit by comparing the transition probability matrices and the inventory distributions. The fit is reasonable, although the model produces slightly lower inventory levels. Interpretations of parameter values are as follows.

Table 2: Calibrated Parameter Values

Parameter	Value		Description
ρ	9.86×10^{-4}		Weekly interest rate to match a 5% annual rate
u	$17,614$		Unit utility (\$)
κ_b	$5,880$		A buyer's flow outside option to search
κ_s	$23,927$		A seller's flow outside option to search
	Small	Large	
μ_r^j	1.31	1.71	Retail-market matching function parameter
μ_w^j	3.73	8.55	Wholesale-market matching function parameter
c^j	14.78	4.55	A dealer's marginal cost of inventory (\$) per week

We begin with parameters common to both small and large dealers. The value of u captures the average monetary-measured utility of purchasing a 4-6-year-old non-luxury sedan. As the outside options to search, values of buyers' κ_b and sellers' κ_s are flow *rates* per unit of time. Examples of such outside options include buyers and sellers searching for direct trades, and sellers keeping the cars; see our discussion of assumptions in the model section. The values of these two parameters are indeed reasonable given our free-entry specification. To make sense of the numbers, recall that if there are buyers in a submarket p, we have the free-entry condition $\psi_r(\theta^*(p))(u-p) = \kappa_b$. Our data frequency is weekly, so $\psi_r(\theta^*(p))/7$ is a linear approximation of the daily probability of a buyer meeting a dealer, and $\mathbb{E}_p[\psi_r(\theta^*(p))(u-p)]/7 \approx \$840$ is the buyer's daily expected payoff. By the free-entry condition, it is also the daily opportunity cost he incurs. Observe in Table 3 that the average T_b is 1.09 (or 1.10) at small (or large) dealers, which means that it takes a buyer a little over a week to buy a car on average. Since our data frequency is also weekly, the value of κ_b is roughly equal to the buyer's expected total payoff of searching for buying a used car.

Similarly, for sellers in a wholesale submarket w, we have $\kappa_s = \psi_w(\lambda^*(w))w$, where w should be interpreted as the seller's surplus from a trade (price minus his value of owning the car which has been normalized to be zero). By the same logic, $\phi_w(\lambda^*(w))/7$ approximates the daily probability of a seller meeting a dealer, and $\phi_w(\lambda^*(w))w/7 \approx 3418$ is roughly a seller's daily expected payoff by searching for selling his car. From Table 3, we see that the average T_s is 0.30 (or 0.13) weeks at small (or large) dealers, i.e., it takes a seller about 1-2 days to make a sale, so the parameters implied average surplus of the seller is roughly between $\$3,000$ and $\$7,000$. We interpret the difference between large and small dealers as the difference in acquisition sources, speed and convenience. For example, cars are acquired through trade-ins and wholesale auctions, both of which vary across dealers.

We turn to the type-specific parameter values. Qualitatively, their relative scales are as expected. Specifically, small dealers face greater search frictions than large dealers in both retail and wholesale markets, reflected by the scaling parameters in matching functions $\mu_r^1 < \mu_r^2$ and $\mu_w^1 < \mu_w^2$; small dealers also have higher marginal cost of inventory such that $c^1 > c^2$. Furthermore, we have $\mu_r^j < \mu_w^j$, for $j = 1, 2$, so it is generally easier for both types of dealers to find sellers in wholesale markets than meeting retail buyers. These parameter differences jointly capture the heterogeneity in inventory distributions. The marginal inventory costs are roughly $15/week for a small dealers and $5/week for large dealers. Dealers typically debt-finance their inventory, and our calibrated costs correspond to weekly cost of funds on a $10,000 loan with 5% annual interest (which is about $10/week). However, these inventory marginal costs are relatively low, which implies that dynamic pricing and inventory management is more about search frictions and uncertainty than literal holding costs. The inventory cost parameter absorbs remaining dealer heterogeneity and other factors that affect dealers' flow revenue besides market frictions, however these factors appear to be small.

5.6 Small versus Large Dealers

This section explores the difference between small and large dealers. From Table 2, we learn small and large dealers differ in their inventory costs and search and matching technologies. This section takes a closer look at the difference between small and large dealers and the corresponding welfare implications. We report selected equilibrium statistics in

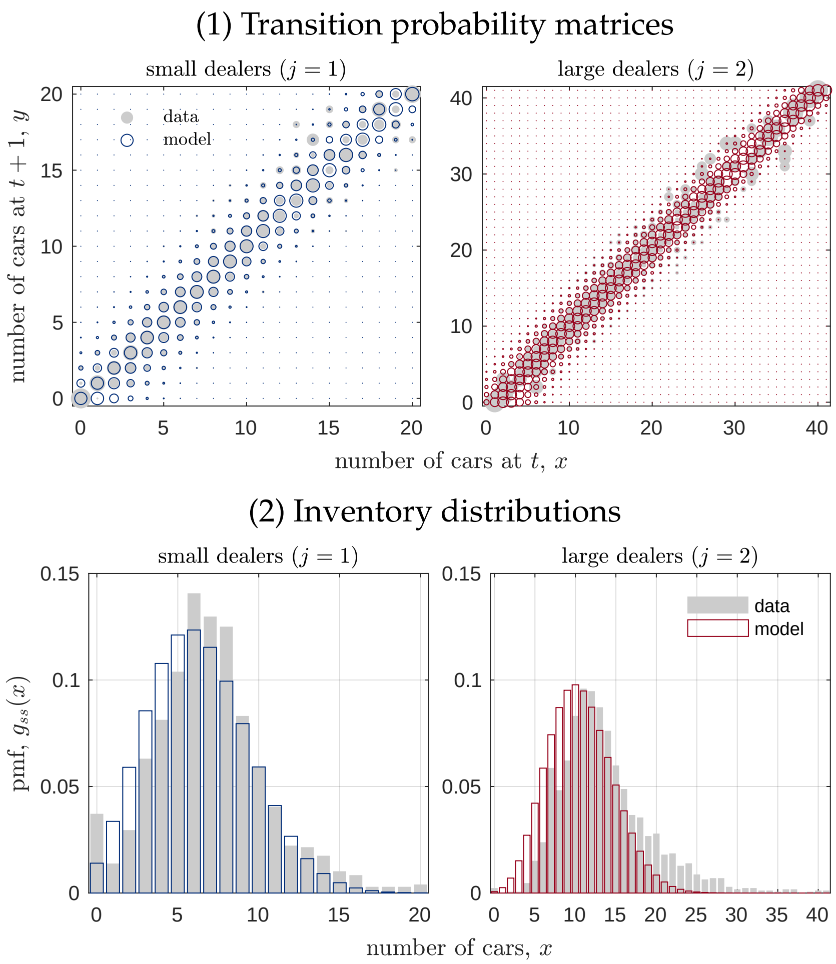

Figure 2: Goodness of fit: weekly transition probability matrix and inventory distribution by dealer type

Inventory levels are trimmed at the 99th percentile. Panel (1) plots the weekly type-specific transition matrices $P^j = [P^j_{xy}]$, where solid gray discs represent matrix entries in the data, colored circles represent those implied by the model, and marker sizes correspond to entry values. In panel (2), solid gray bars are empirical frequencies, colored empty bars represent model-implied steady state probability mass functions.

Table 3: Summary of Equilibrium Outcome by Dealer Type

	cross-sectional mean of...									relative
	x	$p(x)$	$\phi_r(x)$	$\theta^*(x)$	$T(x)$	$T_r(x)$	$T_w(x)$	$T_b(x)$	$T_s(x)$	surplus
Small	6.35	11,226 (\$)	0.69	0.76	0.74	1.44	1.55	1.09	0.30	129%
Large	10.65	11,170 (\$)	1.30	1.42	0.39	0.77	0.82	1.10	0.13	185%

x is a dealer's inventory level, $x = 0, 1, \dots, s$; $p(x)$ is the retail price posted by a dealer with inventory x, $\phi_r(x)$ is the corresponding retail matching rate, and $\theta^*(x)$ is the associated tightness, for $x \geq 1$. $T(x) = 1/[\phi_r(x) + \phi_w(x)]$ is a dealer's expected time spent at inventory level x before it increases or decreases. $T_r(x) = 1/\phi_r(x)$ is a dealer's expected time to sell a car when meeting buyers at rate $\phi_r(x)$, $x \geq 1$. $T_w(x) = 1/\phi_w(x)$ is a dealer's expected time to gain a car when meeting sellers at rate $\phi_w(x)$, $x \leq s-1$. $T_b(x) = 1/\psi_r(x)$ is a buyer's expected waiting time in retail submarket-x, $x \geq 1$. $T_s(x) = 1/\psi_w(x)$ is a seller's expected waiting time in wholesale submarket-x, $x \leq s-1$. Relative surplus is the ratio of the aggregate social surplus created by the intermediaries over the aggregate option values.

Table 3.

First, we compute the average time that a large (small) dealer spent at inventory level x. For each dealer type $j = 1, 2$, the time spent at inventory level x before it jumps is an exponential random variable with expectation $T^j(x) = 1/[\phi_r^j(x) + \phi_w^j(x)]$. Given the stationary distribution g_{ss}^j, the cross-sectional mean is simply $T^j = \sum_{x=0}^{s^j+1} g_{ss}^j(x) T^j(x)$. It is clear that large dealers spend less time at each x, and their inventory levels churn faster than small dealers'. On average, small dealers see changes in inventory levels every 5 days ($T^1 = 0.74$), and large ones need 3 days ($T^2 = 0.39$).

We take a closer look and separately examine the retail side and the wholesale side. In retail, suppose that a type-j dealer stays in submarket-x until meeting with a buyer, then the expected waiting time is $T_r^j(x) = 1/\phi_r^j(\theta^*(x))$, for $j = 1, 2$. Similarly, in wholesale, suppose that a type-j dealer stays in submarket-x until meeting with a seller, then the expected waiting time is $T_w^j(x) = 1/\phi_w^j(\lambda^*(x))$, for $j = 1, 2$. Not surprisingly, small dealers need to wait longer than large ones on both sides. However, despite the relatively small difference in μ_r^j, large dealers sell almost twice as fast as small ones, thanks to the more considerable difference in μ_w^j and large dealers' low inventory costs.

Next, we shift our focus to buyers in retail markets and sellers in wholesale markets, respectively. Recall that our data does not contain information about buyers and sellers. However, our model predicts buyers' and sellers' behavior patterns based on information about intermediaries. The expected waiting time for a buyer in type-j dealers' retail submarket x is $T_b^j(x) = 1/\psi_r^j(\theta^*(x))$, for $x = 1, \dots, s+1$; a seller's analog is $T_s^j(x) = 1/\psi_w^j(\lambda^*(x))$, for $x = 0, \dots, s$. Given the inventory distribution g_t at t, we can calculate the measure of buyers or sellers in each submarket, such that

$$g_t^b(x) = g_t(x)\theta^*(x), \quad g_t^s(x) = g_t(x)\lambda^*(x), \quad \forall x = 0, 1, \dots, s.$$

Normalizing yields the probability distributions conditional on submarkets being active, such that $g_t^b(x)/\sum_{y=1}^{s+1} g_t^b(y)$ for $x = 1, \ldots, s+1$ and $g_t^s(x)/\sum_{y=0}^{s} g_t^s(y)$ for $x = 0, \ldots, s$. We use these distributions at the steady state to calculate the cross-sectional means of $T_b^j(x)$ and $T_s^j(x)$ in Table 3. As mentioned in the previous section, on average, it takes buyers a similar amount of time to search for either type of dealers, and sellers need less time to search for a dealer, especially a large one.

5.7 Gains from Trade due to Dealers

We examine the fraction of gains from trade that is created by used-car dealers. We calculate the relative surplus as follows. At each moment t, given an inventory distribution g_t, the *gross* value flow created by the dealer sector is

$$G_t = \sum_{x=0}^{s+1} g_t(x)[\phi_r(\theta^*(x))u - c(x)]. \tag{36}$$

In words, this is the likelihood of a match multiplied by the buyer utility created minus the holding costs for a given inventory level x, integrated over the inventory distribution. The total outside-option value flow is

$$C_t = \sum_{x=0}^{s+1} g_t(x)[\theta^*(x)\kappa_b + \lambda^*(x)\kappa_s] = \sum_{x=0}^{s+1} [g_t^b(x)\kappa_b + g_t^s(x)\kappa_s], \tag{37}$$

which captures the values to buyers and sellers is no trade occurs for different inventory levels.

Consider the relative (gross) value at t as G_t/C_t. We compute them for both large and small dealers at the steady state where $\dot{g}_t = 0$. The results are reported in Table 3. Large dealers make significantly more welfare contribution than small dealers. From the Tables 2 and 3, we see that the large dealers advantage in improving social welfare mainly comes from two factors. First, they have lower inventory cost. Second, they have superior search and matching technology, making them much more efficient at allocating cars between sellers and buyers. Between these two factors, the crucial one is the difference in search and matching technologies. To see this, we simply remove the inventory costs by setting $c^1 = c^2 = 0$. Keeping other values unchanged, we re-simulate the model to obtain the respective stationary equilibrium without any inventory costs. At the zero-cost equilibrium, both small and large dealers' average inventory levels increase by more than 3 units. However, the change in the relative surplus created by either type of dealers is insignificant. Specifically, the increase in small dealers' surplus creation is only 1 percentage point, and the change is even less for large ones by about 0.4 percentage point. Therefore, we conclude that large dealers' greater contribution to welfare is due to their search technology.

5.8 Transition Paths

Lastly, we examines market transitions after changes to model primitives. This is particularly relevant for the used car market since the market underwent sudden changes in the wake of the COVID19 pandemic in 2021 and 2022. In 2021, used car inventories decreased substantially and prices rose.[23] Journalists and industry professionals attributed the changes in the used car market to many factors, including disruption that spilled over from the supply chain issues in the new-car market and changes in underlying demand and the behavior of used car shoppers. For the former, scarcity among new cars likely disrupted supply into the used car market (as consumers held their cars longer) and consumers who would typically be new-car customers were substituted to the used car market due to insufficient selection and high prices for new cars. For the latter, internal migration, generous fiscal and monetary policy, and work-from-home, prompted some consumers to value cars differently.

We examine the implications of inventory management to changes in a market by plotting the transition dynamics after a 10% increase in each of the models' primitives. Figure 3 and Figure 4 plot the resulting dynamics of three type-specific statistics: (a) average inventory level, (b) average retail price, and (c) relative surplus created. Observe that, in each experiment, (a) evolves continuously, whereas the paths of (b) and (c) see jumps at the time of shock $t = 0$. The reason is that, although the equilibrium policies adjust instantaneously upon the shock, which causes the jumps in (b) and (c), the cross-sectional inventory distribution evolves gradually according to the updated KF equation (21) induced by the new equilibrium policies, resulting in a continuous path of (a).

We begin with the responses to changes in market frictions in Figure 3. The first experiment considers the impact of increasing the buyer's outside option κ_b by 10%, which is essentially a negative demand shift. Buyers search less, so fewer inventories are necessary. As a result, dealers cut retail prices and wholesale order frequencies, gradually lowering their inventories. The relative surplus declines due to the direct effect of increasing κ_b and the indirect effect as fewer transactions are made through the dealer sector. In the second experiment, we increase the seller's outside option κ_s by 10% to capture a negative supply shift. As a response, dealers will order at a lower speed. Interestingly, the impact on the average inventory is starkly heterogeneous among dealers. To the large dealers, the

[23] See the CNBC report from the following link. https://www.cnbc.com/2020/10/15/used-car-boom-is-one-of-hottest-coronavirus-markets-for-consumers.html

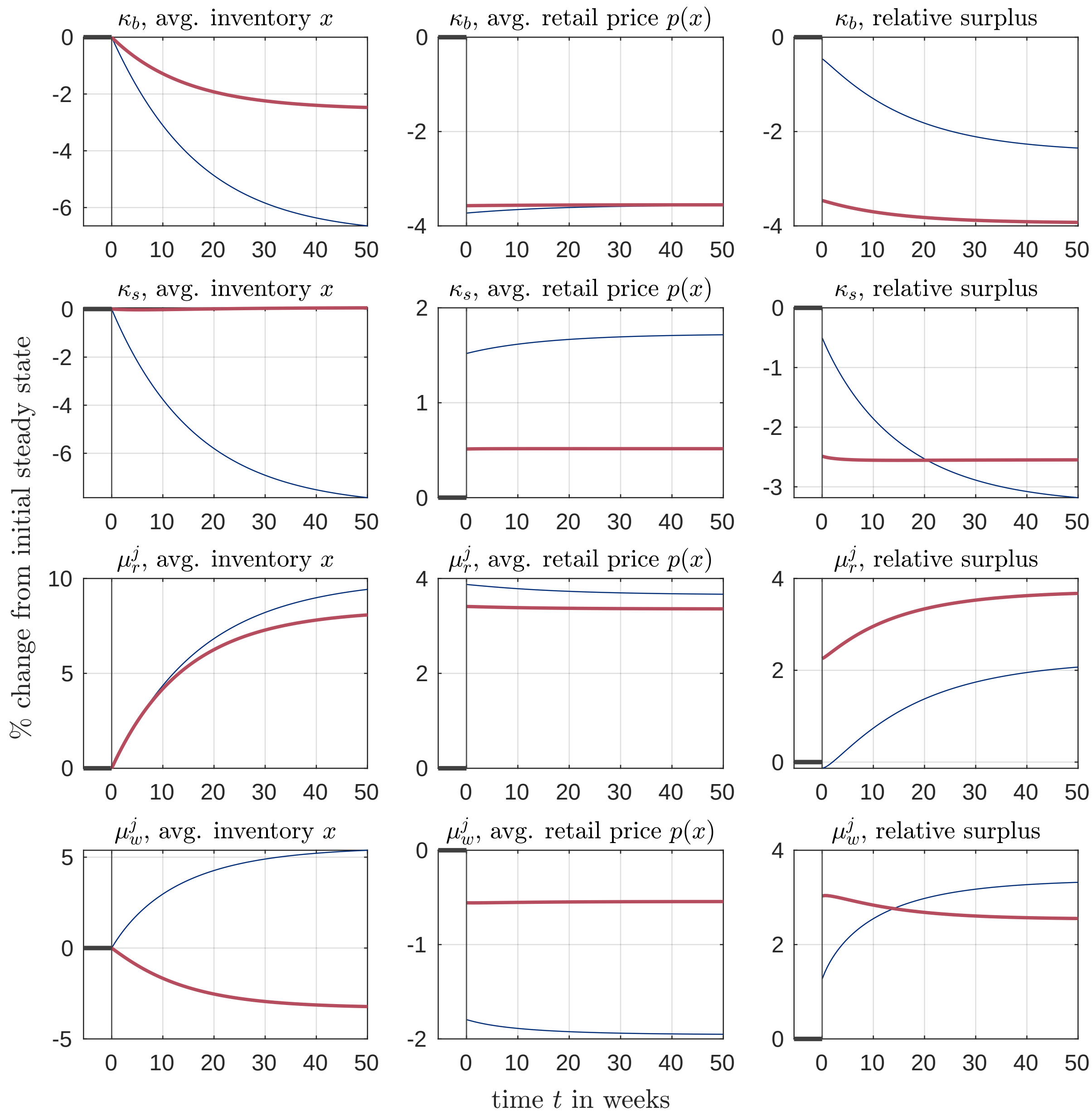

Figure 3: Responses to permanent changes in search frictions: type-specific transition paths of average inventory, average retail price, and relative surplus. At $t = 0$, matching parameter k_b, k_s, μ_r^j, or μ_w^j permanently increases by 10%.

In each panel, the *thin blue* line is for the small dealers ($j = 1$), and the *thick red* line the large ones ($j = 2$); time t is on the horizontal axis, and the vertical axis shows the percentage deviation from the baseline steady state before any parameter change. Each row corresponds to a shocked parameter, and each column contains the transition paths of an equilibrium statistic. Time of shock is $t = 0$. Before $t = 0$, the economy is at the baseline steady state with a stationary type-specific inventory distribution g_{ss}^j. At $t = 0$, a parameter (pair) permanently increases by 10%. After $t = 0$, the distributions evolve in continuous time.

impact can be ignored, and the average inventory eventually increases slightly, whereas to the small dealers, it is non-trivial, which means the large dealers must proportionally decrease both the buying and selling speed, keeping average inventory unchanged. This is also reflected in the difference between small and large dealers' average prices. For large dealers, the average retail price jumps less, meaning a smaller decline in the retail transactions rate; the transmission of the change from the wholesale side to the retail markets affects the large dealers less, mainly due to their less frictional search and matching technologies.

The next two experiments (third and fourth rows of Figure 3) examine the effect of a respective 10% increase in dealers' matching-function parameters. We simultaneously increase small dealers' retail matching-function parameter μ_r^1 and large dealers' μ_r^2. Matching efficiencies in retail markets improve for both dealers, and meeting buyers becomes easier. Dealers want to increase inventory holdings to avoid greater stockout risks and charge higher retail prices; small dealers respond more in inventory levels and prices. Retail transactions become more frequent despite the higher retail prices. Upon impact, smaller dealers' retail prices jump up more, partially offsetting the benefit of increased retail trading rates, which explains the initial small dip in the relative surplus created by small dealers. Dealers of both types are able to contribute more in welfare eventually.

A more striking result shows in the experiment of improving the wholesale matching as a 10% increase in small and large dealers' μ_w^1, μ_w^2, respectively. In response, large dealers reduce the average inventory holdings, and small ones do the opposite. The asymmetry occurs due to the interactions between frictions in retail and wholesale markets. For small dealers, the dominating effect of reduced frictions in wholesale is that they can better "insure" themselves against stockout risks by holding more inventories, which allows them to set lower retail prices. For large dealers, wholesale-side frictions are low even before the impact, and the dominating effect of smoother wholesale transactions is to reduce large dealers' need to hold inventories to cut inventory costs. In either dealer type's case, improved matching in wholesale spills over to retail markets, such that dealers post lower retail prices to increase retail rates. Consequently, both dealers create more surplus.

Now we move to the responses to changes in utility and inventory cost in Figure 4. The parameter u captures the buyer's life-time utility of owning a used car. In our model, when the buyers' utility of trading through the dealer sector increases, more buyers enter retail markets and this increases dealers' selling speed. To respond, dealers (i) increase the retail price, and (ii) hold more inventory to slow down the rise of the stockout risk.

Finally, we consider a shock to the dealers' inventory cost. If dealers' inventory cost permanently increases, they will immediately adjust their retail and wholesale policy to reduce their inventory levels. To boost the selling speed, the retail price immediately falls.

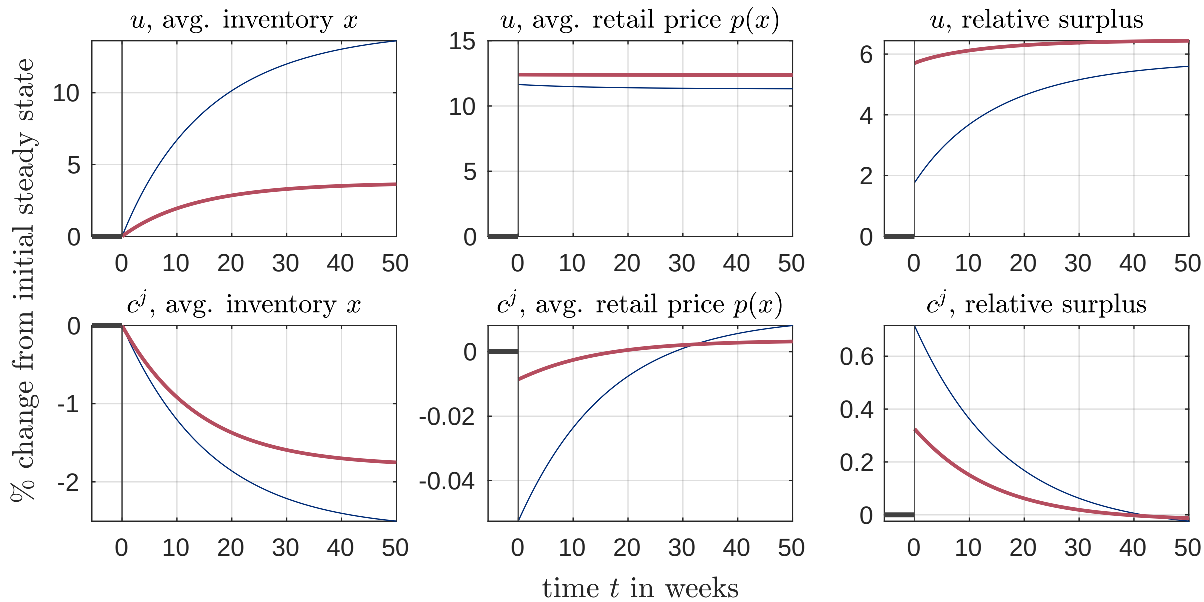

Figure 4: Responses to a permanent increase in unit utility u or inventory cost c^j: type-specific transition paths of average inventory, average retail price, and relative surplus.

See Figure 3 for figure notes.

This is the direct effect of increasing inventory cost on the average retail price. An indirect effect will take place in the long run: Dealers' inventories gradually decreases over time, and the retail prices increases due to the rise of the stockout risk accordingly. In the long-run, the average retail price may be either higher or lower than the level prior to the shock, depending on the competition between the two effects. In terms of welfare, the dealer sector's welfare contribution jumps up immediately following the inventory shock. Although increasing inventory cost has a direct negative effect on the dealer sector's welfare contribution, in the short-run, this effect is dominated by the positive effect due to the jump of the selling speed. However, in the long run, as the average inventory declines, the selling speed follows and the temporary welfare bump vanishes.

Which of these changes to model primitives best reflects the outcomes observed in the wake of the COVID-19 pandemic in 2021 and 2022? Two of the changes to model primitives generate an opposite co-movement in inventory levels and prices like the used car market experienced in 2021 and 2022. First, an increase in inventory cost (second row of Figure 4) leads to lower inventories and greater long-run retail prices. However, there don't seem to be any major changes to inventory costs during this time, so this mechanism seems like an unlikely source of market disruptions observed during this time. In fact, inventory costs may have decreased due to loosening monetary policy. Alternatively, a 10% increase in κ_s generates decreased inventory (particularly for our small dealers) and higher retail prices. This seems like a much more likely story, as supply problems with the new car market spilled into the supply of wholesale used cars. Our model and calibration find evidence that changes to used-car supply generated reduced inventory and increased prices in 2021 and 2022.

6 Concluding Remarks

This paper fills a gap between several active areas of literature: one on search theoretic models of intermediaries, and one on pricing and inventory control. We highlight the role of inventory dynamics in shaping retail price dynamics and dispersion in a search model. The natural combination of equilibrium search and inventory management has a significant logical consequence. Prices fluctuate in response to the inventories change, as intermediaries adjust prices to sell inventory or restock. The model is extended in various directions. We calibrate the model using used-car dealer data from Ohio and quantitatively highlight the important interaction between search frictions and inventory dynamics. The calibrated model is then used to study a number of important roles of used-car dealers' inventory management practice such as its welfare contributions to the economy and its effect in shaping transition dynamics caused by shocks of important market characteristics.

A Appendix: Proofs

Proof of Lemma 1. Suppose that the marginal value of holding one more unit of inventory is positive at x, or $V(x) - V(x-1) > 0$. We show that $V(x+1) - V(x) \leq V(x) - V(x-1)$. Consider, at time t, an intermediary with $x_t = x$ units of inventory that adopts the optimal policy of its $(x_\tau + 1)$-inventory self at $\tau \geq t$ until the first time its inventory drops to $x - 1$. Clearly, such a policy is suboptimal for this intermediary with $x_t = x$. Formally, suppose it employs the following policy $\Gamma \equiv \{\theta_\tau, \lambda_\tau\}_{\tau \geq t}$ that generates an inventory process $\{x_\tau\}_{\tau \geq t}$. The policy solves the problem (7) for the inventory being $x_\tau + 1$ until the first instant when its true inventory drops to $x - 1$. Afterwards, the intermediary employs the optimal policy. Denote

$$T \equiv \inf\{\tau \geq t : x_\tau \leq x - 1\}.$$

Note that $T - t$ is the first passage time from state x to state $x - 1$ of the Markov process induced by policy Γ. Crucially, $T - t$ is a continuous random variable with a non-negative support, and $\Pr(T - t > \epsilon) > 0$ for an arbitrary $\epsilon > 0$. The intuition is that, due to search frictions, it takes time for inventory level to change, regardless of whether the intermediary's policy is optimal or not.

Denote the associated life-time profit to be $V^{\Gamma}(x)$, then we must have

$$V^{\Gamma}(x) = V(x+1) + \mathbb{E}\left\{\int_t^T e^{-\rho(\tau-t)}[c(x_\tau+1)-c(x_\tau)]d\tau + e^{-\rho(T-t)}[V(x-1)-V(x)]\right\}$$

where the expectation is taken over the random time T. $V^{\Gamma}(x)$ differs from $V(x+1)$ in two aspects. First, in time interval $[t, T)$, the flow inventory cost is $c(x_\tau)$ rather than $c(x_\tau+1)$. Second, after the transaction at time T, the continuation value is $V(x-1)$ instead of $V(x)$. Because policy Γ is suboptimal, $V^{\Gamma}(x) \leq V(x)$; and therefore

$$\begin{aligned} V(x+1) - V(x) &\leq \mathbb{E}\left\{-\int_t^T e^{-\rho(\tau-t)}[c(x_\tau+1)-c(x_\tau)]d\tau + e^{-\rho(T-t)}[V(x)-V(x-1)]\right\} \\ &< V(x) - V(x-1). \end{aligned} \tag{38}$$

The second inequality holds because (i) $c(\cdot)$ is increasing and (ii) $\mathbb{E}[e^{-\rho(T-t)}] \in (0,1)$. As a consequence, $V(x) - V(x-1)$ decreases in x when it is positive. Note that the first inequality in (38) also implies that, whenever $V(x) - V(x-1)$ turns negative, then so does $V(x+1) - V(x)$. □

Proof of Proposition 1. Recall that both $\phi_r(\cdot)$ and $\phi_w(\cdot)$ are increasing. From Lemma 1, both $V(x) - V(x-1)$ and $V(x+1) - V(x)$ in FOCs (17) and (18) are decreasing in x, so the first part of the proposition immediately follows. The second part of the proposition is a direct consequence of the combination of part 1 and conditions (13) and (14). □

Proof of Proposition 2. First, we prove the existence and uniqueness by solving for the stationary probability distribution g_{ss}. In equilibrium, a dealer's inventory follows an general birth-death process $\{x_t\}$ over finite states, induced by the equilibrium policy $\theta^*(\cdot)$ and $\lambda^*(\cdot)$. For a general birth-death process, conditions for the existence and uniqueness of a stationary probability distribution are standard, and so is the probability mass function's form. See, e.g., Chapter 6 of Pinsky and Karlin (2010) for reference. Nonetheless, we show it here for completeness. The stationary distribution g_{ss} satisfies condition (21) and can be solved recursively. At $x = 0$, we have $0 = \phi_r(\theta^*(1))g_{ss}(1) - \phi_w(\lambda^*(0))g_{ss}(0)$, or

$$g_{ss}(1) = \frac{\phi_w(\lambda^*(0))}{\phi_r(\theta^*(1))} g_{ss}(0).$$

Consequently, at $x = 1$, $0 = \phi_w(\lambda^*(0))g_{ss}(0) + \phi_r(\theta^*(2))g_{ss}(2) - [\phi_r(\theta^*(1)) + \phi_w(\lambda^*(1))]g_{ss}(1)$, or

$$\begin{aligned}\phi_r(\theta^*(2))g_{ss}(2) =&[\phi_r(\theta^*(1)) + \phi_w(\lambda^*(1))]g_{ss}(1) - \phi_w(\lambda^*(0))g_{ss}(0)\\ =&\phi_w(\lambda^*(1))g_{ss}(1)\\ g_{ss}(2) =&\frac{\phi_w(\lambda^*(1))}{\phi_r(\theta^*(2))}g_{ss}(1) = \frac{\phi_w(\lambda^*(1))\phi_w(\lambda^*(0))}{\phi_r(\theta^*(2))\phi_r(\theta^*(1))}g_{ss}(0).\end{aligned}$$

The general formula follows as

$$g_{ss}(x) =\frac{\phi_w(\lambda^*(x-1))}{\phi_r(\theta^*(x))}g_{ss}(x-1) \tag{39}$$

$$=g_{ss}(0)\prod_{i=1}^{x}\frac{\phi_w(\lambda^*(i-1))}{\phi_r(\theta^*(i))}, \quad \forall x \geq 0. \tag{40}$$

Clearly, with s being the base level of the stock defined in equation (20), $g_{ss}(x) = 0$ for any $x > s+1$ because $\phi_w(\lambda^*(x-1)) = \phi_w(0) = 0$, and equation (39) implies $g_s s(x) = 0$. Plugging (40) into the constraint $\sum_{x=0}^{\infty} g_{ss}(x) = 1$ yields the expression for $g_{ss}(0)$. A unique distribution exists if and only if $g_{ss}(0)$ is well-defined, which requires

$$0 < \sum_{x=1}^{s+1}\prod_{i=1}^{x}\frac{\phi_w(\lambda^*(i-1))}{\phi_r(\theta^*(i))} < \infty,$$

and it naturally holds when $s < \infty$.

Second, we prove that the steady-state distribution is unimodal. The result holds trivially if $1 = s$, i.e., the distribution is over two points. If $1 < s$, there are at least three inventory levels with positive probability mass. Rearranging equation (21) yields

$$\phi_w(\lambda^*(x))g_{ss}(x) - \phi_w(\lambda^*(x-1))g_{ss}(x-1) = \phi_r(\theta^*(x+1))g_{ss}(x+1) - \phi_r(\theta^*(x))g_{ss}(x). \tag{41}$$

Because $\phi_w(\lambda^*(x))$ decreases in x, the left-hand side of equation (41) is less than $[g_{ss}(x) - g_{ss}(x-1)]\phi_w(\lambda^*(x-1))$. Because $\phi_r(\theta^*(x))$ increases in x, the right-hand side of equation (41) is greater than $[g_{ss}(x+1) - g_{ss}(x)]\phi_r(\theta^*(x))$. Therefore, we have $[g_{ss}(x) - g_{ss}(x-1)]\phi_w(\lambda^*(x-1)) \geq [g_{ss}(x+1) - g_{ss}(x)]\phi_r(\theta^*(x))$. That is, for any $x \geq 1$, whenever $g_{ss}(x+1) \geq g_{ss}(x)$, we have $g_{ss}(x) \geq g_{ss}(x-1)$, and whenever $g_{ss}(x) \leq g_{ss}(x-1)$, we have $g_{ss}(x+1) \leq g_{ss}(x)$. So the steady-state probability mass function $g_{ss}(\cdot)$ is single-peaked, or unimodal.

□

The omitted proof of Proposition 4. The argument of the existence and uniqueness is similar to the proof of Proposition 2. We solve for the stationary probability distribution recursively. To ease notation, we suppress the super- and sub-scripts of g^m_{ss} in this step and let $\phi_r(x), \phi_w(x)$ represent $\phi_r(\theta^*(x)), \phi_w(\lambda^*(x))$, respectively.

The KF equation at the steady state can be written as follows.

$$\begin{aligned}
g(0)\phi_w(0) =& g(1)\phi_r(1), \\
g(1)[\phi_r(1) + \phi_w(1)] =& g(2)\phi_r(2), \\
&\vdots \\
g(s)[\phi_r(s) + \phi_w(s)] =& g(s+1)\phi_r(s+1), \\
g(s+k)\phi_r(s+k) =& g(s+k+1)\phi_r(s+k+1), \ \forall k = 1, \dots, S-s+1, \\
g(S)\phi_r(S) =& \sum_{x=0}^{s} g(x)\phi_w(x) + g(S+1)\phi_r(S+1), \\
g(S+k)\phi_r(S+k) =& g(S+k+1)\phi_r(S+k+1), \ \forall k \geq 1.
\end{aligned}$$

Stationarity requires that $g(S+k) = 0, \forall k \geq 1$. We also observe that

$$\begin{aligned}
g(x+1)\phi_r(x+1) =& \sum_{i=0}^{x} g(i)\phi_w(i), \ \forall x = 1, \dots, s; \\
g(s+1)\phi_r(s+1) =& g(s+k)\phi_r(s+k) = g(S)\phi_r(S), \ \forall k = 1, \dots, S-s.
\end{aligned}$$

The system thus becomes functions of $g(S)$, such that

$$\begin{aligned}
g(s+k) =& g(S)\frac{\phi_r(S)}{\phi_r(s+k)}, \ \forall k = 1, \dots, S-s, \\
g(s) =& g(s+1)\frac{\phi_r(s+1)}{\phi_r(s) + \phi_w(s)} = g(S)\frac{\phi_r(S)}{\phi_r(s) + \phi_w(s)}, \\
g(s-k) =& g(s-k+1)\frac{\phi_r(s-k+1)}{\phi_r(s-k) + \phi_w(s-k)} \\
=& g(S)\frac{\phi_r(S)}{\phi_r(s) + \phi_w(s)}\frac{\phi_r(s)}{\phi_r(s-1) + \phi_w(s-1)} \cdots \frac{\phi_r(s-k+1)}{\phi_r(s-k) + \phi_w(s-k)}, \ \forall k = 1, \dots, s,
\end{aligned}$$

where $\phi_r(0) = 0$. The system can be written more concisely as

$$g(x) = g(S)\prod_{i=x}^{S-1} \frac{\phi_r(i+1)}{\phi_r(i) + \phi_w(i)}, \ \forall 0 \leq x < S,$$

with $\phi_w(i) = 0$ if $i \geq s+1$, and $\phi_r(0) = 0$. Applying the additional constraint $\sum g(x) = 1$

yields

$$g(S) = \left(1 + \sum_{x=0}^{S-1} \prod_{i=x}^{S-1} \frac{\phi_r(i+1)}{\phi_r(i) + \phi_w(i)}\right)^{-1}.$$

$\square$

B Appendix: Statistics Supporting Directed Search

This appendix examines the relationship between a car's time on market and list price. The directed search doctrine relies on a positive relationship between the list price and time to sell. Table B.1 reports the results of the regressions of the log of a car's weeks on market on the log of the car's list price and other controlling variables. Across all specifications with or without the weekly fixed effects and car model fixed effects, the relationship between the time on market and the price is significantly positive.

Table B.1: Time on Market v.s. List Price

	Small Dealers			**Large Dealers**		
	(1)	(2)	(3)	(1)	(2)	(3)
log (list price)	0.456	0.418	0.508	0.391	0.343	0.557
	(0.027)	(0.026)	(0.042)	(0.027)	(0.026)	(0.041)
log (mileage)	0.155	0.166	0.178	0.064	0.075	0.109
	(0.014)	(0.013)	(0.015)	(0.013)	(0.012)	(0.014)
Week FEs		✓	✓		✓	✓
Car model FEs			✓			✓
# of observations	16,239	16,239	16,239	15,551	15,551	15,551
R-square	0.025	0.083	0.100	0.026	0.089	0.106

Notes. Dependent variable is the log of weeks on market of a car. We use ✓ to differentiate additional controls.

References

Acemoglu, D. and R. Shimer (1999). Efficient unemployment insurance. *Journal of Political Economy 107*(5), 893–928.

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *The Review of Economic Studies 89*(1), 45–86.

Adams, C. P., L. Hosken, and P. W. Newberry (2011). Vettes and lemons on ebay. *Quantitative Marketing and Economics 9*, 109–127.

Akerlof, G. A. (1970). The market for" lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 488–500.

Amihud, Y. and H. Mendelson (1980). Dealership market: Market-making with inventory. *Journal of Financial Economics 8*(1), 31–53.

Anderson, S. P., A. De Palma, and J.-F. Thisse (1992). *Discrete choice theory of product differentiation*. MIT press.

Baye, M. R., J. Morgan, and P. Scholten (2006). Information, search, and price dispersion. *Handbook on Economics and Information Systems 1*, 323–375.

Bental, B. and B. Eden (1993). Inventories in a competitive environment. *Journal of Political Economy 101*(5), 863–886.

Biglaiser, G. (1993). Middlemen as experts. *The RAND Journal of Economics*, 212–223.

Biglaiser, G., F. Li, C. Murry, and Y. Zhou (2020). Intermediaries and product quality in used car markets. *The RAND Journal of Economics 51*(3), 905–933.

Burdett, K. and K. L. Judd (1983). Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society*, 955–969.

Burdett, K., S. Shi, and R. Wright (2001). Pricing and matching with frictions. *Journal of Political Economy 109*(5), 1060–1085.

Carrasco, J. and R. Harrison (2022). Costly multi-unit search. *Available at SSRN 3949785*.

Carrasco, J. A. and L. Smith (2017). Search at the margin. *American Economic Review 107*(10), 3146–81.

Chandra, A. and M. Tappata (2011). Consumer search and dynamic price dispersion: an application to gasoline markets. *The RAND Journal of Economics 42*(4), 681–704.

Chen, N., G. Gallego, P. Gao, and S. Kou (2020). Dealership or marketplace: A dynamic comparison.

Chen, X. and D. Simchi-Levi (2012, 06). Pricing and Inventory Management. In *The Oxford Handbook of Pricing Management*. Oxford University Press.

Coey, D., B. J. Larsen, and B. C. Platt (2020). Discounts and deadlines in consumer search. *American Economic Review 110*(12), 3748–85.

Colliard, J.-E., T. Foucault, and P. Hoffmann (2021). Inventory management, dealers' connections, and prices in over-the-counter markets. *The Journal of Finance 76*(5), 2199–2247.

Deneckere, R., H. P. Marvel, and J. Peck (1996). Demand uncertainty, inventories, and resale price maintenance. *The Quarterly Journal of Economics 111*(3), 885–913.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). Over-the-counter markets. *Econometrica 73*(6), 1815–1847.

Faig, M. and B. Jerez (2005). A theory of commerce. *Journal of Economic Theory 122*(1), 60–99.

Gavazza, A. (2016). An empirical equilibrium model of a decentralized asset market. *Econometrica 84*(5), 1755–1798.

Gavazza, A. and A. Lizzeri (2021). Frictions and intermediation in product markets. Technical report, London School of Economics.

Gavazza, A., A. Lizzeri, and N. Roketskiy (2014). A quantitative analysis of the used-car market. *American Economic Review 104*(11), 3668–3700.

Gillingham, K., F. Iskhakov, A. Munk-Nielsen, J. Rust, and B. Schjerning (2022). Equilibrium trade in automobiles. *Journal of Political Economy 130*(10), 2534–2593.

Guo, N. (2018). The effect of an early career recession on schooling and lifetime welfare. *International Economic Review 59*(3), 1511–1545.

Guo, X. and O. Hernández-Lerma (2009). *Continuous-Time Markov Decision Processes*. Springer.

Hartigan, J. A. and P. M. Hartigan (1985). The dip test of unimodality. *The annals of Statistics 13*(1), 70–84.

Ho, T. and H. R. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics 9*(1), 47–73.

Ho, T. S. and H. R. Stoll (1983). The dynamics of dealer markets under competition. *The Journal of finance 38*(4), 1053–1074.

Hugonnier, J., B. Lester, and P.-O. Weill (2020). Frictional intermediation in over-the-counter markets. *The Review of Economic Studies 87*(3), 1432–1469.

Hugonnier, J., B. Lester, and P.-O. Weill (2022). Heterogeneity in decentralized asset markets. *Theoretical Economics 17*(3), 1313–1356.

Johri, A. and J. Leach (2002). Middlemen and the allocation of heterogeneous goods. *International Economic Review 43*(2), 347–361.

Kaplan, G. and G. Menzio (2015). The morphology of price dispersion. *International Economic Review 56*(4), 1165–1206.

Lach, S. (2002). Existence and persistence of price dispersion: an empirical analysis. *Review of Economics and Statistics 84*(3), 433–444.

Larsen, B. J. (2021). The efficiency of real-world bargaining: Evidence from wholesale used-auto auctions. *The Review of Economic Studies 88*(2), 851–882.

Lizzeri, A. (1999). Information revelation and certification intermediaries. *The RAND Journal of Economics*, 214–231.

Menzio, G. and S. Shi (2010). Block recursive equilibria for stochastic models of search on the job. *Journal of Economic Theory 145*(4), 1453–1494.

Menzio, G. and S. Shi (2011). Efficient search on the job and the business cycle. *Journal of Political Economy 119*(3), 468–510.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy 105*(2), 385–411.

Murry, C. and H. Schneider (2015). The economics of retail markets for new and used cars. In E. Basker (Ed.), *Handbook on the Economics of Retail and Distribution*. Edward Elgar.

Nakamura, E. and D. Zerom (2010). Accounting for incomplete pass-through. *The Review of Economic Studies 77*(3), 1192–1230.

Nosal, E., Y.-Y. Wong, and R. Wright (2019). Intermediation in markets for goods and markets for assets. *Journal of Economic Theory 183*, 876 – 906.

Nuño, G. and B. Moll (2018). Social optima in economies with heterogeneous agents. *Review of Economic Dynamics 28*, 150–180.

Peters, M. (1991). Ex ante price offers in matching games non-steady states. *Econometrica (1986-1998) 59*(5), 1425.

Peters, M. (2000). Limits of exact equilibria for capacity constrained sellers with costly search. *Journal of Economic Theory 95*(2), 139–168.

Pinsky, M. and S. Karlin (2010). *An introduction to stochastic modeling*. Academic press.

Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment vacancies, and real wages. *American Economic Review 75*(4), 676–690.

Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT press.

Porteus, E. L. (2002). *Foundations of stochastic inventory theory*. Stanford University Press.

Prescott, E. C. (1975). Efficiency of the natural rate. *Journal of Political Economy 83*(6), 1229–1236.

Reinganum, J. F. (1979). A simple model of equilibrium price dispersion. *Journal of Political Economy 87*(4), 851–858.

Rhodes, A., M. Watanabe, and J. Zhou (2021). Multiproduct intermediaries. *Journal of Political Economy 129*(2), 000–000.

Rubinstein, A. and A. Wolinsky (1987). Middlemen. *The Quarterly Journal of Economics 102*(3), 581–593.

Rust, J. and G. Hall (2003). Middlemen versus market makers: A theory of competitive exchange. *Journal of Political Economy 111*(2), 353–403.

Salz, T. (2017). Intermediation and competition in search markets: An empirical case study. *working paper, Columbia University*.

Scarf, H. (1960). The optimality of (s, s) policies in the dynamic inventory problem. *Mathematical Methods in the Social Sciences*.

Shevchenko, A. (2004). Middlemen. *International Economic Review 45*(1), 1–24.

Shi, S. (2009). Directed search for equilibrium wage–tenure contracts. *Econometrica 77*(2), 561–584.

Smith, E. (2004). Intermediated search. *Economica 71*(284), 619–636.

Spulber, D. F. (1996). Market microstructure and intermediation. *Journal of Economic perspectives 10*(3), 135–152.

Stahl, D. O. (1989). Oligopolistic pricing with sequential consumer search. *The American Economic Review*, 700–712.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy 69*(3), 213–225.

Stoll, H. R. (1978). The supply of dealer services in securities markets. *The Journal of Finance 33*(4), 1133–1151.

Watanabe, M. (2010). A model of merchants. *Journal of Economic Theory 145*(5), 1865–1889.

Watanabe, M. (2020). Middlemen: A directed search equilibrium approach. *The BE Journal of Macroeconomics 20*(2).

Whitin, T. M. (1955). Inventory control and price theory. *Management science 2*(1), 61–68.

Wright, R., P. Kircher, B. Julien, and V. Guerrieri (2017). Directed search: A guided tour. Technical report, National Bureau of Economic Research.

Wright, R. and Y.-Y. Wong (2014). Buyers, sellers, and middlemen: Variations on search-theoretic themes. *International Economic Review 55*(2), 375–397.

Yang, M. and Y. Zeng (2021). Coordination and fragility in liquidity provision. *Available at SSRN 3265650*.