

A theory of commerce[☆]

Miquel Faig^{a,*} and Belén Jerez^b



^a*Department of Economics, University of Toronto, 150 St. George Street, Toronto, Canada M5S 3G7*

^b*Departamento de Economía, Universidad Carlos III de Madrid, Getafe, Madrid 28903, Spain*

Abstract

The theory of commerce advanced here captures prominent features of retail trade: large employment, congestion effects, anonymous posted prices, and quantity discounts. This theory is built around a directed search model where buyers' preferences are private information. The analytical solution is easily inserted in a Neoclassical growth framework. In this framework, the parameters of retail trade are calibrated using commercial margins and employment. Welfare properties depend on the sellers' ability to charge two-tier prices. With two-tier prices, the directed search equilibrium is efficient. Otherwise, it is not. This contrasts with the full information benchmark, where directed search is always efficient.

JEL classification: D43; D82; D83; E1

Keywords: Retail trade; Directed search; Ex ante price offers; Private information

1. Introduction

Commercial margins constitute a very large fraction of the cost of purchasing a typical good. On average, for every dollar spent before tax in a retail store in the

[☆]Work on this paper started while this author was visiting the Institute of Policy Analysis at the University of Toronto and was affiliated to the Universitat Autònoma de Barcelona. We thank both institutions for their hospitality. Financial support from SSHRC of Canada (Grant 410010028) and DGICYT of Spain (projects BEC2002-00642, BEC2000-0172 and HI2001-0039) is gratefully acknowledged.

*Corresponding author. Fax: +416-978-5519.

E-mail addresses: mfaig@chass.utoronto.ca (M. Faig), mjerez@eco.uc3m.es (B. Jerez).

United States, only 57 cents reach the producers of the merchandise. The remaining 43 cents cover the margins of the merchant wholesaler (15 cents) and the retailer (28 cents).¹ Models with Walrasian markets abstract from commercial activities. This abstraction would be of little importance if these activities were limited to transporting commodities from factories to consumers, for then we could think of commerce as a stage of production with a conventional technology. However, a large fraction of commercial costs are incurred in the process of contacting and dealing with buyers in bilateral meetings. Moreover, the size of these costs is independent of the quantities that buyers end up acquiring and normally cannot be recovered if the buyer chooses to buy nothing. These facts can only be properly modeled with non-Walrasian markets. To develop an appropriate framework of analysis, we advance a directed search model of commerce. This model integrates price competition in an environment where trades are costly and bilateral. In the model, competition is multidimensional in the sense that buyers and sellers care not only about prices but also about the expected time it takes for them to perform a trade. Moreover, to generate pricing mechanisms which are consistent with those we typically observe in retail trade, we assume that buyers' preferences are private information.

Our model seeks to capture the following features of retail trade. The cost of a single transaction is small. However, buyers seek a diversified basket of goods which entails performing many transactions. These transactions then employ part of the scarce labor available in the economy. In addition, the matching of buyers and sellers is frictional, so quite often some buyers are lining up for assistance in front of a seller, while another seller nearby is idle. Sellers post price offers to attract buyers to their outlets. However, buyers care not only about prices, they also care about the expected time it takes for them to make a purchase. Retail prices typically have the following properties: they are anonymous (that is they do not depend on the buyers' preferences), buyers do not pay if they choose not to purchase a good, and buyers quite often receive quantity discounts that can be either explicit or built in the packaging of products.

The center piece of our model is the equilibrium concept of directed search, sometimes referred to as *ex ante* price posting, or competitive search. This concept has been introduced by Peters [14] and Montgomery [11], and it has been further studied by a long list of contributors including Shimer [22], Moen [10], and Mortensen and Wright [13]. Applied to our model of retail trade, the idea of this equilibrium concept is that sellers post a price schedule that specifies the terms at which they commit to trade. Buyers observe all posted price schedules and direct their search towards the subset of sellers announcing a particular price schedule. We refer to the set of sellers announcing a particular price schedule and the buyers that direct their search towards them as a submarket. The assignment of buyers to sellers

¹This break-up assumes that the good goes through one merchant wholesaler and one retailer. In our calculations, we use the gross commercial margins over sales ratios calculated by the Bureau of the Census, Service Sector Statistics Division for 2000 (0.278 for retail and 0.2054 for wholesale). This data was downloaded in July 1, 2002 from <http://www.census.gov/svsd/www/artstbl.html> and <http://www.census.gov/svsd/www/atspur.txt>.

in a submarket is then dictated by an exogenous matching technology as in Mortensen [12] and Pissarides [17]. When a buyer and a seller meet in a submarket, the buyer chooses the quantity to be transacted and pays for it according to the price schedule. As noted by Acemoglu and Shimer [1], this formulation of directed search encompasses many reasonable possibilities. For example, it includes the possibility that buyers randomly select a seller among those who post equivalent price schedules as in the frictional assignment literature [2,14,15]. Another possibility is that submarkets are located in separate places (malls or streets), and that in each submarket all sellers post an identical price schedule and buyers and sellers meet according to an exogenous matching technology.

By adopting directed search we are able to capture that in retail trade sellers use prices to attract customers in an environment where buying and selling activities are labor consuming and take place in bilateral meetings. Moreover, we capture that buyers and sellers care not only about prices, but also about the expected time it takes for them to complete a purchase. Finally, we allow for the possibility of frictional matching between buyers and sellers.

We should clarify here the role of matching frictions in our model. Matching frictions in our model should not be interpreted as buyers not knowing where to buy a shirt, but rather as buyers not knowing how many other buyers are going to visit a particular seller with a limited capacity to serve. Hence, buyers, in general, are uncertain about the time they will spend to make a purchase, which includes the time waiting in line to be served. Similarly, sellers are uncertain about the time they will spend to make a sale, which includes the time they will remain idle while waiting for a customer. Matching frictions are not necessary for our theory to be consistent and relevant. In fact, we show that our main results continue to hold in the limiting special case where the short side of the market is always matched (frictionless matching). We study this special case in some detail.² In general, we allow for matching frictions in order to capture the uncertainty that characterizes retail transactions, as well as to endogenize the congestion of buyers relative to sellers in the market (i.e., the relative amounts of labor employed in buying and selling activities). This congestion determines the trading frequencies of buyers and sellers.

An important innovation we introduce to the concept of directed search is that sellers cannot observe the buyers' willingness to pay (the buyers' type). We assume that when a buyer meets a seller, a random shock affecting the buyer's willingness to pay is realized. Instantly, the buyer becomes aware of his type but the seller does not. All relevant information must then be conveyed through prices, just as in the standard Walrasian benchmark. Introducing this form of private information in the model allows us to generate prices that are consistent with those we typically observe in retail trade. If the seller could observe the buyer's type (full information), prices in a directed search equilibrium would be type dependent. That is, prices would not be anonymous as we typically observe in retail trade.

If the seller cannot observe the buyer's type, the price schedule cannot be type dependent. In this case, the equilibrium price schedule depends on the timing of the

²We thank an anonymous referee for suggesting this simplified version of our model to us.

preference shock. If the buyer and the seller meet prior to the realization of the preference shock, the model predicts two-tier pricing. That is, prior to the realization of the preference shock, the seller charges a flat fee to the buyer. This flat fee covers the labor cost of the seller. After the preference shock is realized, the buyer chooses the quantity to be transacted and pays a price equal to the marginal cost of production. With two-tier pricing, buyers pay the flat fee even if they purchase minimal amounts or nothing. In retail trade, two-tier pricing is rare. Some discount stores charge membership fees to help cover their fixed costs. However, those instances are the exception not the rule. In our baseline model, we assume that the preference shock is realized as soon as the buyer and the seller meet. This assumption, combined with the fact that the buyer's willingness to pay can be arbitrarily low, rules out flat fees. The price paid by the buyer must then be higher than the marginal cost of production in order to remunerate the seller's labor.

The welfare properties of the model depend on the information available to buyers and sellers when they trade, as well as on the timing on the preference shock. With full information (type-dependent prices) and with private information and two-tier pricing, the directed search equilibrium implements the first best allocation. In our baseline model, where the two-tier pricing schedule is not feasible, the directed search equilibrium is not efficient. In the absence of flat fees, sellers must be remunerated through a commercial margin. The commercial margin creates a wedge between the buyer's willingness to pay and the marginal cost of production. This wedge reduces the gains from trade. In addition, as long as matching is frictional, the commercial margin in equilibrium is not large enough to generate the efficient ratio of sellers over buyers. Consequently, the average lineup of buyers in front of a seller is inefficiently long.

Our model can be easily incorporated in a Neoclassical growth framework with two sectors. One of these sectors produces goods combining capital and labor as is typical in the Neoclassical framework, while the other exchanges goods in retail markets. The combined model can be estimated using standard empirical data. In particular, the parameters of the retail sector can be identified and estimated using commercial margins and the allocation of labor in the economy.³

The tractability of the model makes it suitable for further extensions. For example, one can use this framework to introduce money along the lines of the search theoretical approach of Kiyotaki and Wright [6,7]. Faig [4] constitutes a first attempt in this direction in a simpler version of the present model. In this simpler version, sellers are restricted to make offers which consist of a single quantity–payment pair. Our main improvement with respect to Faig [4] is the relaxation of this constraint by allowing sellers to make offers that consist of a price schedule that maps the quantity chosen by a buyer into the payment to be made to the seller. This price schedule serves as a mechanism to reveal the buyers' private information about their preferences. Also, in Faig [4] search is undirected.

³ Shi [20,21] also incorporates a search model in a Neoclassical growth framework. However, the model in Shi assumes full information and random (undirected) search.

Two recent papers, Soller-Curtis and Wright [23] and Camera and Delacroix [3], also study search-theoretic models where the buyers' willingness to pay for a good is private information. In both of these papers, goods are indivisible and search is undirected. Curtis and Wright assume that sellers make take-it-or-leave-it offers to buyers and they focus on the coexistence of two prices for the same good in equilibrium. Camera and Delacroix focus on the endogenous determination of the trading mechanism, that is sellers choose if they want to commit to a posted price or to bargain once they meet a buyer. When buyers have identical preferences and search is undirected, pre-committing to a price is preferable from the sellers point of view because it allows them to extract the whole trading surplus. However, when the preferences of the buyers are heterogeneous, the bargaining process allows the seller to infer information about the buyers' preferences and hence to price discriminate. In our paper, goods are divisible, so sellers can commit to a non-linear price schedule which allows a restricted form of price discrimination even without bargaining.⁴ Also, we incorporate the equilibrium concept of directed search which endogenizes the market power of sellers.

Peters and Severinov [16] also study a model of directed search where the preferences of the buyers are private information. However, in their model the price setting mechanism is an auction among the buyers that meet a seller. Moreover, because a single unit is sold in each auction, their paper does not deal with the price incentives to endogenously determine the size of each purchase.

The paper is organized as follows. Section 2 presents an overview of the model and our equilibrium concept. Section 3 characterizes the optimal behavior of a representative household for given prices. Section 4 studies how prices are endogenously determined under directed search by analyzing the interaction between buyers and sellers in the retail market. Section 5 combines the analyses of the previous two sections into a general equilibrium model where both the behavior of households and retail prices are endogenous. Section 6 studies the welfare properties of a directed search equilibrium. Section 7 incorporates a generalized version of our model in a dynamic Neoclassical growth framework and discusses how to identify the parameters using standard data. Section 8 briefly discusses some of the issues one must confront when extending the present model and concludes. The proofs are gathered in the Appendix.

2. Overview of the model

The economy consists of a continuum of households with measure one who produce and consume differentiated goods. Households do not consume the goods they produce so they need to trade. Trading activities involve some degree of

⁴Admittedly, some transactions in retail trade involve bargaining, and the bargaining process may partially reveal the willingness to pay of a buyer. Yet, the amount of information revealed in the bargaining process is also limited by informational constraints. Moreover, bargaining in retail trade is rare and is inconsistent with ex ante price posting.

idiosyncratic uncertainty to be specified below. To avoid the ex post heterogeneity induced by idiosyncratic uncertainty, which severely limits the tractability of the model, we follow Shi [19] in assuming that each household is composed of a large number of individuals who seek to maximize the utility of the household. These individuals independently perform the production and exchange activities in the household. Thanks to the large household assumption, each household faces no uncertainty, even though the members of the household who perform trading activities are subject to idiosyncratic risks. All members of the household equally share the consumption of goods which is their only source of utility, so there is no conflict of interests among them.

To construct a simple environment where households buy the goods they consume in a retail trade sector, we make the following assumptions. Households are uniformly distributed on the interval $[0, 1]$. All households are identical except that each household produces a different divisible good. Thus, a continuum of differentiated goods on the $[0, 1]$ interval are produced in the economy. Households like to consume all goods but their own production good. Because of physical constraints, the members of the household who perform manufacturing activities (producers) cannot simultaneously sell the good they produce. Likewise, sellers must remain in their retail outlets to sell their products. Therefore, for trade to take place the buyers of a household must go around visiting sellers of other households. Since sellers never meet each other, direct barter is ruled out.

All payments are denominated in an abstract numeraire. In the version of the model analyzed in this paper, buyers do not need to carry money with them. Instead, all traders have access to a central clearing-house that records the credits (payments received by sellers) and debits (payments made by buyers) of all households and ensures that their budget constraint is satisfied.

Trading activities are subject to two kinds of uncertainty. First, buyers sometimes find goods that, because of idiosyncratic factors, fit well the needs of their households while in other occasions they do not. Second, because of matching frictions, trading meetings between buyers and sellers are partially random. Thus, a trader may or may not be able to perform a transaction during a given period.

We model the first type of uncertainty by assuming a preference shock ε that scales the utility that a good brings to the household. Preference shocks are realized once a buyer of the household meets a seller. Shocks are independently and identically distributed across trading meetings, and their distribution is common knowledge. However, the realized value of ε is the buyer's own private information. Therefore, in a trading meeting the seller ignores the willingness to pay of the buyer. We believe that this ignorance is key for understanding retail pricing in the same way that unobservable characteristics of taxpayers are key for understanding income tax schedules.

In order to model both the matching frictions and the price competition that characterize retail markets, we assume the following form of directed search. Prior to the trading period, each seller j simultaneously posts and commits to a price schedule $\mathcal{Z}_j(q)$. This schedule specifies the payment required in a transaction as a function of

the quantity exchanged.⁵ In the next stage, buyers observe the menu of price schedules posted by all sellers and simultaneously choose where to shop. That is, each buyer i decides to trade at a particular price schedule $\mathcal{Z}_i(q)$ in the set of posted schedules $\{\mathcal{Z}_j(q) : \text{for all } j\}$. Buyers can also play mixed strategies and randomize over price schedules for which they are indifferent. If buyer i meets a seller posting $\mathcal{Z}_i(q)$ during the trading period, the buyer privately learns the valuation ε of her household for the seller's good, chooses the quantity to be transacted and pays for it according to the posted price schedule. Buyers can always choose not to buy anything in which case they pay nothing. If no meeting takes place, then there is no trade. We refer to the set of sellers posting $\mathcal{Z}_j(q)$ and the set of buyers that direct their search to this price schedule as submarket j .

Depending on the buyers's search decisions, there may be longer lineups in some submarkets than in some others. To capture this, we let $\theta_j \in [0, \infty]$ be the ratio of buyers to sellers in the submarket with a posted price schedule $\mathcal{Z}_j(q)$:

$$\theta_j = \frac{B_j}{S_j}, \quad (1)$$

where S_j is the measure of sellers posting $\mathcal{Z}_j(q)$ and B_j is the measure of buyers who decide to trade at this price schedule. We refer to θ_j as the congestion in submarket j , or the average queue of buyers in front of a seller in this submarket. For the time being and to facilitate the exposition, we assume that buyers and sellers may perform at most one transaction during the trading period (Section 5.2 relaxes this assumption). In a submarket with congestion θ_j , the probability that a seller meets a buyer is $m^s(\theta_j)$, where $m^s: [0, \infty] \rightarrow [0, 1]$ is continuously differentiable, increasing, and concave. Symmetrically, a buyer meets a seller with probability $m^b(\theta_j)$, where $m^b: [0, \infty] \rightarrow [0, 1]$ is continuously differentiable, decreasing, and convex. If many buyers seek a few sellers (θ_j is high), then it is easy for a seller to find a buyer and hard for a buyer to find a seller. By having m^s and m^b depend only on θ_j , we implicitly assume constant returns in matching, so

$$\mathcal{M}(B_j, S_j) = B_j m^b(\theta_j) = S_j m^s(\theta_j), \quad (2)$$

where \mathcal{M} is a standard matching function that maps the measures of buyers and sellers in submarket j onto the measure of trading meetings in this submarket (see [17]).

An interesting special case of m^s is

$$m^s(\theta_j) = 1 - \exp(-\theta_j). \quad (3)$$

This case arises if buyers use identical mixed strategies to select a seller among those who post equivalent price offers and, because selling is time consuming, each seller

⁵ Equivalently, because each seller j may meet buyers with different private valuations ex post, the seller announces a list of quantity-payment pairs $\{q_{jv}, z_{jv}\}_{v \in V}$. It is not restrictive to assume that the number of items in this list is equal to the number of ex post buyer types. A seller's strategy is summarized by a price schedule $\mathcal{Z}_j: Q_j \rightarrow \mathfrak{R}$ where $Q_j = \{q_{jv}\}_{v \in V}$ and $z_{jv} = \mathcal{Z}_j(q_{jv})$ for all $v \in V$. See Maskin and Riley [9] which analyzes the problem of a monopolistic seller who faces no price competition. In their model, however, there are no matching frictions.

can serve at most one customer (see [15]). This is the typical setup in the frictional assignment literature. Another interesting special case is

$$m^s(\theta_j) = \frac{\theta_j}{1 + \theta_j}. \quad (4)$$

This special case arises if each buyer is randomly matched with a trader (buyer or seller) in the submarket where search is directed. (In this case, $m^s(\theta_j)$ is equal to the fraction of buyers over traders in submarket j). Finally, the matching technology that arises when the search frictions vanish is: $\mathcal{M}(B_j, S_j) = \min\{B_j, S_j\}$. In this case, the short side of the market is always served:

$$m^s(\theta_j) = \min\{\theta_j, 1\}. \quad (5)$$

The concept of directed search equilibrium is still well defined with this frictionless matching technology. Our theory does not include this benchmark as a special case because the matching technology is not differentiable. However, its analysis is similar and in many ways simpler (see Section 5.1).

In the environment we consider, buyers and sellers are ex ante symmetric. Given that buyers are free to choose among different price schedules, in a directed search equilibrium, all buyers must attain the same expected payoff. Furthermore, in equilibrium no seller should have an incentive to deviate by posting a different price schedule. To attract buyers, the offer of the deviating seller must yield at least the common expected payoff buyers attain in equilibrium. Therefore, in a directed search equilibrium the price schedules and the degree of congestion associated with them must maximize the expected payoff of sellers subject to the constraint that the buyers get the common expected payoff. As we show in Section 4, this implies that in our framework there is a single submarket in equilibrium. Therefore, all transactions take place according to the same price schedule. Also, all trades face the same congestion, so the probability of trading is the same for all buyers and all sellers.

We proceed to develop our model with the following steps. In Sections 3–5, we construct a static version of the model. In Section 3, we analyze the optimal behavior of a household who takes as given the price schedule $\mathcal{Z}(q)$ and the degree of congestion θ in the market. Section 4 studies how the price schedule $\mathcal{Z}(q)$ and the congestion θ are endogenously related under directed search. Section 5 collects the results of Sections 3 and 4 in a general equilibrium model where the behavior of the household, the prices, and the congestion are endogenously determined.

3. A representative household

In this section, we describe the behavior of a household whose buyers and sellers participate in a retail market with price schedule $\mathcal{Z}(q)$ and degree of congestion θ . Because there is a continuum of households in the economy, the household takes the price schedule and the degree of congestion as given. We assume $\mathcal{Z}(q)$ to be continuously differentiable and concave with $\mathcal{Z}(0) = 0$. The next section endogen-

eizes $\mathcal{Z}(q)$ and validates these assumptions. Since households are symmetric, we adopt the following notation. Lower-case letters denote the decision variables of the household. Upper-case letters denote the decisions of the other households and hence aggregate quantities, which are also taken as given by our household. In a symmetric equilibrium, lower-case letters are equal to the corresponding upper-case letters.

The household is composed of a countably infinite number of individuals. Each household member is endowed with one unit of labor and can perform one of three activities: production (producers), purchase (buyers), and sale of commodities (sellers). We describe the allocation of labor in the household as the fraction of household members assigned to each of the three activities. These fractions are denoted by n , b , and s , respectively. The assumption that there is an infinite number of members in the household ensures full diversification of trading risks. We assume that this number is countable to ensure that the set of buyers in the household is of measure zero relative to the set of goods for sale by other households. This assumption is technically important because it implies that there are no repeated purchases of the same good with probability one. Also, the standard Law of Large Numbers assumes a countable infinite number of random variables.

The timing of the model is as follows. There is one period. At the beginning of the period, the members of the household are divided into producers, buyers, and sellers. Then, all members perform their assigned activities. Specifically, producers use their labor to generate output, which immediately becomes available for sale. Buyers visit sellers from other households. Upon meeting a seller, they experience a random shock ε that scales the incremental utility that the good brings to the household. Shocks are uniformly distributed on $[0, 1]$ and are independent across trading meetings. The distribution of these shocks is common knowledge, but the realized value of ε is the buyer's own private information. Contingent on the realized value of ε , buyers choose the quantity q_ε they want to acquire from the seller, and pay $z_\varepsilon = \mathcal{Z}(q_\varepsilon)$. This payment is immediately debited from the household's account in the clearing-house. Sellers go to retail outlets and wait for buyers to visit them. When a seller of the household is visited by a buyer who purchases a quantity Q_ε , the seller fetches this quantity from the current household production and collects a payment $Z_\varepsilon = \mathcal{Z}(Q_\varepsilon)$, which is immediately transferred to the household's account in the clearing-house. Finally, once all their tasks are completed, all the individuals of the household get together and equally share the consumption of the goods purchased.

The household's utility is $\mathcal{U}(c)$, where $\mathcal{U} : \mathbb{R}_+ \rightarrow \mathbb{R}$ is continuously differentiable, strictly increasing, and concave. The variable c is a hedonic measure of consumption, which satisfies

$$c = \left[b m^b(\theta) \int_0^1 \varepsilon q_\varepsilon^{1-\sigma} d\varepsilon \right]^{\frac{1}{1-\sigma}}, \quad \sigma \in (0, 1). \quad (6)$$

This hedonic measure of consumption follows from a standard Dixit–Stiglitz aggregator by applying the Law of Large Numbers. The Dixit–Stiglitz aggregator is

defined over the set of goods purchased and consumed by the household.⁶ For completeness, a detailed derivation of (6) is provided in the Appendix. The term $bm^b(\theta)$ represents the fraction of household members who seek to buy goods from other households and are successful in finding a seller. That is, $bm^b(\theta)$ is the number of purchases per capita in the household. The integral $\int_0^1 \varepsilon q_\varepsilon^{1-\sigma} d\varepsilon$ measures the average contribution of a purchase. The weight associated to a purchase is $\varepsilon q_\varepsilon^{1-\sigma}$, where ε is the realization of the preference shock and q_ε is the quantity purchased conditional on such a realization. Since there is a countable number of purchases, the Law of Large Numbers implies that the ex post distribution of the realized preference shocks is uniform on $[0, 1]$. The hedonic measure of consumption in (6) implies that c doubles when q_ε doubles for all ε , and c more than doubles when $bm^b(\theta)$ doubles. Therefore, the household has a preference for variety, which increases with the parameter σ and vanishes when $\sigma \rightarrow 0$. Despite the preference for variety, the household does not buy infinitesimal amounts because it is costly to send extra buyers to the market. To avoid dealing with uninteresting corner solutions, we assume that U is sufficiently concave, so utility is a concave function of b .

Production (per capita) y depends on the fraction of labor employed in manufacturing:

$$y = f(n). \quad (7)$$

The production function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is assumed continuously differentiable, strictly increasing, and concave. The average sales of the household are equal to the number of sales per capita times average quantity sold in a transaction. That is, average sales are $sm^s(\theta)\bar{Q}$ where $\bar{Q} = \int_0^1 Q_\varepsilon d\varepsilon$. Since there is no aggregate uncertainty in a large household, the amount of output both produced and sold is identical:

$$y = sm^s(\theta)\bar{Q}. \quad (8)$$

The household must satisfy the budget constraint, that is, the average expenditure of the household cannot exceed its average sales revenue:

$$sm^s(\theta)\bar{Z} - bm^b(\theta)\bar{z} \geq 0. \quad (9)$$

Here, $\bar{Z} = \int_0^1 Z(Q_\varepsilon) d\varepsilon$ is the average revenue of a seller who is successful in finding a buyer, and $\bar{z} = \int_0^1 Z(q_\varepsilon) d\varepsilon$ is the average expenditure of a buyer who is successful in finding a seller.

The fractions of individuals assigned to the three different activities in the household must add up to one:

$$b + s + n = 1. \quad (10)$$

⁶Note that this set is different from the set of goods for sale by other households. In particular, the set of goods purchased by the household is countable while the set of goods for sale by other households is uncountable.

The household chooses $\{q_\varepsilon\}_{\varepsilon \in [0,1]}$, b , s , n , c and y to maximize \mathcal{U} subject to (6)–(10), and non-negativity constraints for all variables.⁷ Condition (8) can be substituted into (7) to form a single resource constraint. Also, (6) can be substituted into the objective function $\mathcal{U}(c)$. Using Lagrange multipliers μ , λ , and ϑ for the resource constraint, the budget constraint, and the labor allocation constraint respectively, the first-order conditions for an interior maximum are

$$\mathcal{U}'(c)c^\sigma \varepsilon q_\varepsilon^{-\sigma} = \lambda \mathcal{Z}'(q_\varepsilon) \quad \text{for } \varepsilon \in [0, 1], \quad (11)$$

$$m^b(\theta) \int_0^1 \left[\frac{\varepsilon \mathcal{U}'(c)c^\sigma q_\varepsilon^{1-\sigma}}{1-\sigma} - \lambda z_\varepsilon \right] d\varepsilon = \vartheta, \quad (12)$$

$$m^s(\theta) \int_0^1 (\lambda Z_\varepsilon - \mu Q_\varepsilon) d\varepsilon = \vartheta \quad (13)$$

and

$$\mu f'(n) = \vartheta. \quad (14)$$

Condition (11) states that, conditional on each possible realization ε , buyers must equate the marginal utility of purchasing an extra unit of a good with the marginal value of the payment required in return. The other conditions imply that the value of the marginal product of labor in all three occupations must be the same. This common value is ϑ . Condition (12) equates ϑ to the expected consumer surplus obtained by a buyer. Condition (13) equates ϑ to the expected surplus generated by a seller. Finally, condition (14) equates ϑ to the value of the marginal product of labor of a producer.

4. Commercial pricing

In this section, we describe the retail market. We model the interaction between buyers and sellers in the marketplace in a directed search equilibrium where sellers do not know how much buyers are willing to pay for their merchandise. A bilateral trade is described by a pair $(q, z) \in \mathfrak{R}_+^2$ specifying the quantity q supplied by the seller and the payment z given in return by the buyer. In the previous section, the two elements of this vector were linked by an exogenous price schedule $z = \mathcal{Z}(q)$. The purpose of this section is to endogenize \mathcal{Z} .

The revelation principle allows us to reformulate the problem where sellers post price schedules and buyers, with different realizations of the preference shock, self-select along the schedule of the sellers they are matched with. Without loss of generality, we may assume that sellers posts direct revelation mechanisms that induce truth-telling by the buyers (that is, incentive compatible mechanisms). Feasible

⁷We allow b, s and n to be non-negative real numbers even though with a countably infinite number of household members we should limit our attention to rational numbers. Since rational numbers are dense in the reals, the equilibrium we obtain in Section 5 is strictly speaking an approximate equilibrium.

mechanisms must also satisfy the individual rationality constraints of the traders. Buyers observe the mechanisms posted by the sellers and decide to direct their search to those posting a particular mechanism. In a directed search equilibrium, the posted mechanisms and the degrees of congestion associated with them must maximize the expected payoff of the posting seller taking as given the common expected payoff attained by the buyers. This implies that the mechanisms posted by sellers must be optimal in the set of incentive compatible and individually rational mechanisms; that is, any posted mechanism must be incentive efficient. Second, the equilibrium combinations of congestion and posted mechanisms must also be efficient.

4.1. Incentive efficient direct revelation mechanisms

We begin by characterizing the set of incentive efficient direct revelation mechanisms. That is, we characterize the direct revelation mechanisms that maximize a weighted sum of the expected payoffs of the buyer and the seller subject to the incentive compatibility and individual rationality constraints. We then show that the outcome of each incentive efficient direct revelation mechanism is implemented when sellers post a price schedule in a particular class, and buyers who choose to trade at these prices select their most preferred price-quantity combination depending on the realization of the preference shock.

In a meeting between a buyer and a seller, the buyer knows the realized value of ε , but the seller does not. All the seller knows is that ε is a random variable uniformly distributed on the interval $[0, 1]$. We refer to the realization of ε as the buyer's *type*.

All traders maximize the objectives of the households they belong to. Specifically, the incremental utility of a buyer's purchase to the household is the consumer surplus:

$$\mathcal{U}^b(q, z; \varepsilon) = \frac{\varepsilon \psi q^{1-\sigma}}{1-\sigma} - \lambda z, \quad (15)$$

where

$$\psi \equiv \mathcal{U}'(c)c^\sigma, \quad (16)$$

and λ is the Lagrange multiplier associated with the budget constraint (9). The incremental utility of a seller's sale is the (gross) commercial margin. The commercial margin measures the spread between sales receipts and the cost of the merchandise sold:

$$\mathcal{U}^s(q, z) = \lambda z - \mu q, \quad (17)$$

where μ is the multiplier associated with the resource constraint (8). That is, \mathcal{U}^b and \mathcal{U}^s are defined so that they are aligned with the household's first-order conditions (12) and (13). By symmetry, the variables ψ , λ , and μ are identical across households. Because each buyer and each seller are infinitesimal in the household they belong to, they take the values of ψ , λ , and μ as given (i.e., these values are not affected by their individual actions). For ease of exposition, we shall refer to \mathcal{U}^b and \mathcal{U}^s as the utility functions of buyers and sellers, respectively.

A direct revelation mechanism is a schedule of type-contingent trades $\{q_\varepsilon, z_\varepsilon\}_{\varepsilon \in [0,1]}$. A direct revelation mechanism is incentive efficient if it maximizes a weighted sum of the expected utilities of buyers and sellers:

$$\{q_\varepsilon, z_\varepsilon\}_{\varepsilon \in [0,1]} = \arg \max \left[(1 - \omega) \int_0^1 \mathcal{U}^b(q_\varepsilon, z_\varepsilon; \varepsilon) d\varepsilon + \omega \int_0^1 \mathcal{U}^s(q_\varepsilon, z_\varepsilon) d\varepsilon \right], \quad (18)$$

where $\omega \in [0, 1]$, subject to the following constraints:

1. *Incentive compatibility*: Buyers must have no incentive to lie about their type:

$$\varepsilon' \in \arg \max_{\varepsilon \in [0,1]} [\mathcal{U}^b(q_\varepsilon, z_\varepsilon; \varepsilon')] \quad \text{for all } \varepsilon' \in [0, 1]. \quad (19)$$

2. *Individual rationality*: Buyers and sellers must receive non-negative utility in all meetings:⁸

$$\mathcal{U}^b(q_\varepsilon, z_\varepsilon; \varepsilon) \geq 0 \quad \text{for all } \varepsilon \in [0, 1] \quad (20)$$

and

$$\mathcal{U}^s(q_\varepsilon, z_\varepsilon) \geq 0 \quad \text{for all } \varepsilon \in [0, 1]. \quad (21)$$

In the Appendix, we use standard arguments in the mechanism design literature to characterize the set of incentive efficient direct revelation mechanisms. This characterization is summarized in the following proposition:

Proposition 1. *The incentive efficient direct revelation mechanisms which solve program (18) subject to (19)–(21), are the following:*

$$q_\varepsilon = \begin{cases} 0 & \text{for } \varepsilon \in [0, \gamma) \\ \left(\frac{\psi}{\mu} \frac{\varepsilon - \gamma}{1 - \gamma}\right)^{\frac{1}{\sigma}} & \text{for } \varepsilon \in [\gamma, 1], \text{ and} \end{cases} \quad (22)$$

$$z_\varepsilon = \frac{1}{\lambda} \left[\gamma \frac{\psi q_\varepsilon^{1-\sigma}}{1-\sigma} + (1-\gamma)\mu q_\varepsilon \right], \quad (23)$$

where

$$\gamma = \begin{cases} 0 & \text{for } \omega \in [0, 0.5), \\ \frac{2\omega - 1}{3\omega - 1} & \text{for } \omega \in [0.5, 1]. \end{cases} \quad (24)$$

⁸This property follows from the combination of two facts: the buyer observes the variety offered by the seller as soon as they meet, and both buyers and sellers have the option of not trading. If sellers could hide their variety for sale, then they could demand a payment from the buyers to reveal this information. In this case, buyers could end up with negative utility in some meetings.

The following proposition establishes that the outcome of an incentive efficient direct revelation mechanism is implemented when sellers post a simple non-linear price schedule and buyers optimally choose the quantity they want to acquire at these prices.

Proposition 2. *The outcome of an incentive efficient direct revelation mechanism is attained when buyers choose q facing the following price schedule:*

$$\mathcal{Z}(q) = \frac{1}{\lambda} \left[\gamma \frac{\psi q^{1-\sigma}}{1-\sigma} + (1-\gamma)\mu q \right], \quad (25)$$

where $\gamma \in [0, 0.5]$ satisfies (24). We refer to a schedule with parameter γ as the price schedule γ .

As long as commercial margins are positive ($\gamma > 0$), the pricing schedule $\mathcal{Z}(q)$ is strictly concave. This concavity implies that the per unit price of goods declines with q , or equivalently that buyers obtain quantity discounts. This is not unrealistic. In retail trade, sometimes we observe explicit quantity discounts, but most often quantity discounts are implicit in the packaging of products (the larger is the box of nails, the lower is the per unit price).

When buyers have full market power, that is when $\omega = 0$, prices cover only the marginal cost of production ($\gamma = 0$). In this case, buyers capture the whole trading surplus and the individual rationality constraint of the seller binds. This constraint continues to bind as long as $\omega \leq 0.5$. For $\omega > 0.5$, commercial margins are positive and both the buyer and the seller appropriate a fraction of the trading surplus. Even when sellers have full market power, that is when $\omega = 1$ and $\gamma = 0.5$, they are not able to extract the whole trading surplus because they do not know their clients' type.⁹ For $\omega \leq 0.5$, the buyers' incentive compatibility constraints are not binding. Hence, when $\gamma = 0$, the incentive efficient direct revelation mechanism characterized in Proposition 1 maximizes the gains from trade under full information (i.e., the mechanism is first best efficient).¹⁰ However, for $\omega > 0.5$, the buyers' incentive compatibility constraints bind and the gains from trade are lower than under full information. Thus, unless buyers appropriate the whole trading surplus, the trading surplus will be lower than under full information. In our model, buyers cannot

⁹The case where $\gamma = 0.5$ is an interesting benchmark, which we denote monopolistic search. With monopolistic search, sellers post the prices that maximize their expected profits in an environment where their price schedules have no effect on the number of clients visiting their outlets. This might be the relevant equilibrium concept for some tourist areas where buyers have little knowledge about where to shop. However, in most commercial areas, sellers are aware that they can attract buyers to their outlets by posting low prices.

¹⁰Since the preferences of buyers and sellers are quasilinear, the (expected) gains from trade in a match are

$$\int_0^1 \mathcal{U}^b(q_\varepsilon, z_\varepsilon; \varepsilon) d\varepsilon + \int_0^1 \mathcal{U}^s(q_\varepsilon, z_\varepsilon) d\varepsilon = \int_0^1 \left(\frac{\varepsilon \psi q_\varepsilon^{1-\sigma}}{1-\sigma} - \mu q_{q_\varepsilon} \right) d\varepsilon.$$

When $\gamma = 0$, the quantities in (22) maximize this function.

appropriate the whole trading surplus in equilibrium because there would be no sellers (see Section 5).

4.2. Trade-off between prices and congestion

Let $\mathcal{V}^b(\theta_j, \gamma_j)$ and $\mathcal{V}^s(\theta_j, \gamma_j)$ respectively be the expected utility of buyers and sellers in a submarket with congestion θ_j and price schedule γ_j :

$$\mathcal{V}^b(\theta_j, \gamma_j) = m^b(\theta_j) \int_0^1 \left(\frac{\varepsilon \psi q_\varepsilon^{1-\sigma}}{1-\sigma} - \lambda z_\varepsilon \right) d\varepsilon \quad (26)$$

and

$$\mathcal{V}^s(\theta_j, \gamma_j) = m^s(\theta_j) \int_0^1 (\lambda z_\varepsilon - \mu q_\varepsilon) d\varepsilon, \quad (27)$$

where q_ε and z_ε are given by (22) and (23). Integrating these expressions we obtain

$$\mathcal{V}^b(\theta_j, \gamma_j) = m^b(\theta_j) \frac{\sigma^2}{1-\sigma^2} \psi^\frac{1}{\sigma} \mu^{\frac{\sigma-1}{\sigma}} (1-\gamma_j)^2 \quad (28)$$

and

$$\mathcal{V}^s(\theta_j, \gamma_j) = m^s(\theta_j) \frac{2\sigma^2}{1-\sigma^2} \psi^\frac{1}{\sigma} \mu^{\frac{\sigma-1}{\sigma}} \gamma_j (1-\gamma_j). \quad (29)$$

In a directed search equilibrium, \mathcal{V}^b must be common to all active submarkets, because buyers are free to choose the submarket where they direct their search. Also, the pair (θ_j, γ_j) must maximize $\mathcal{V}^s(\theta_j, \gamma_j)$ subject to the constraint that $\mathcal{V}^b(\theta_j, \gamma_j)$ is equal to the common expected utility attained by buyers in all active submarkets. The solution to this maximization problem, if one exists, is unique.¹¹ Hence, there is a unique submarket in a directed search equilibrium, and the subscript j can be omitted. This is consistent with our assumptions in Section 3: when all households behave identically, all traders face the same prices and the same congestion. Also, as long as sellers capture a fraction of the expected trading surplus, the equilibrium pair (θ, γ) is interior and satisfies the following first-order condition:

$$\frac{\frac{\partial \mathcal{V}^b(\theta, \gamma)}{\partial \theta}}{\frac{\partial \mathcal{V}^b(\theta, \gamma)}{\partial \gamma}} = \frac{\partial \mathcal{V}^s(\theta, \gamma)}{\partial \gamma}. \quad (30)$$

That is, buyers and sellers must have a common marginal rate of substitution of θ for γ . Differentiating (28) and (29) and substituting into (30), we obtain

$$\gamma = \frac{1 - \eta(\theta)}{2}, \quad (31)$$

¹¹ Substituting the constraint in the objective function and dropping unnecessary constants, θ_j must maximize the strictly concave function: $m^s(\theta_j) \{1 - [vm^b(\theta_j)]^{-0.5}\} [vm^b(\theta_j)]^{-0.5}$, where v is a constant inversely proportional to the given value for \mathcal{V}^b .

where $\eta(\theta)$ is the elasticity of the function m^s :

$$\eta(\theta) = \frac{m^{s'}(\theta)\theta}{m^s(\theta)}. \quad (32)$$

Eq. (31) shows that the price schedule and the congestion level are endogenously related in a directed search equilibrium. The term $1 - \eta(\theta)$ measures the contribution of sellers in the matching process. As long as the elasticity function $\eta(\theta)$ is decreasing, this contribution increases with the congestion θ . Eq. (31) says that the higher this contribution, the higher the fraction of the surplus appropriated by the seller in a directed search equilibrium. In the next section, we pinpoint the equilibrium pair (θ, γ) by combining the endogenous relation in Eq. (31) with the optimal choice of the fractions of buyers and sellers in each household.

5. General equilibrium

In this section, we characterize a directed search equilibrium by combining the optimal behavior of households in Section 3 with the endogenous determination of the price schedule in Section 4.

Definition. The tuple $\{n, b, s, c, y, q_e, z_e, \psi, \mu, \lambda, \mathcal{Z}(q), \theta\}$ is a symmetric directed search equilibrium if

1. All households choose $\{n, b, s, c, y, q_e\}$ to solve their optimization problem taking as given the price schedule $\mathcal{Z}(q)$, the market congestion ratio θ and the decisions of other households, and the households' implied payments are $z_e = \mathcal{Z}(q_e)$.
2. The payment schedule $\mathcal{Z}(q)$ satisfies (25) when ψ satisfies (16) and μ and λ are the Lagrange multipliers associated with the resource constraint and the budget constraint in the household's optimization problem.
3. The pair (γ, θ) satisfies condition (31).
4. The market congestion ratio is consistent with individual behavior: $\theta = b/s$.

Since in a symmetric directed search equilibrium all households behave identically, lower-case and upper-case letters coincide. Symmetry and the equilibrium condition 4, together with constraints (2) and (8), imply that the amount of each good that is produced and traded in equilibrium satisfies:

$$y = f(n) = sm^s(b/s)\bar{q} = bm^b(b/s)\bar{q}, \quad (33)$$

where $sm^s(b/s)$ (equivalently, $bm^b(b/s)$) is the number of sales per capita in each household, and $\bar{q} = \int_0^1 q_e de$ is the average quantity of output sold in a transaction.

In equilibrium, a household must obtain the same utility from allocating an individual to any of the three activities, as it is implied by the first-order conditions (12)–(14). Therefore, $\mathcal{V}^b(\theta, \gamma)$ and $\mathcal{V}^s(\theta, \gamma)$ must be equal. Using (28) and (29)

together with (2), this equality implies

$$\theta = \frac{1-\gamma}{2\gamma}. \quad (34)$$

According to (34), the congestion of buyers in the market θ is inversely related to γ . As γ increases buyers capture a smaller fraction of the trade surplus (see Proposition 2), so households respond by sending fewer buyers and more sellers to the marketplace. Combining (34) with condition (31), we obtain the pair (θ, γ) . In particular, congestion θ is implicitly determined by the following equation:

$$\frac{1}{\frac{1}{2} + \theta} = 1 - \eta(\theta). \quad (35)$$

Consequently, the equality of expected payoffs of buyers and sellers, $V^b(\theta, \gamma) = V^s(\theta, \gamma)$, together with the common marginal rate of substitution of θ for γ by buyers and sellers determine the two key variables in the retail sector: the congestion θ and the price schedule γ .

To determine the remaining variables in the model, we must combine the resource constraints, the household's first-order conditions, and the definitions of θ , c , and ψ . Equating $V^s(\theta, \gamma)$ to the marginal payoff of a producer, $\mu f'(n)$, we obtain

$$f'(n) = m^s(\theta) \frac{2\sigma^2}{1 - \sigma^2} \left(\frac{\psi}{\mu} \right)^{\frac{1}{\sigma}} \gamma (1 - \gamma). \quad (36)$$

Integrating the quantities traded when buyers face a price schedule γ which are given by (22), we obtain the average quantity sold in a trading meeting

$$\bar{q} = \frac{\sigma}{1 + \sigma} \left(\frac{\psi}{\mu} \right)^{\frac{1}{\sigma}} (1 - \gamma). \quad (37)$$

Combining (7), (8), (36), and (37) gives

$$\frac{f'(n)}{f(n)} = \frac{2\sigma}{1 - \sigma} \frac{\gamma}{s}. \quad (38)$$

Using the equilibrium condition $\theta = b/s$, and the labor resource constraint (10), we have

$$s = \frac{1 - n}{1 + \theta}. \quad (39)$$

Finally, combining (38) with (39) and using (34), we obtain

$$\frac{(1 - n)f'(n)}{f(n)} = \frac{\sigma(1 + \gamma)}{1 - \sigma}. \quad (40)$$

Eq. (40) determines n for a given value of γ . In a directed search equilibrium, γ and θ are obtained from (31) and (35). Given θ , γ , and n , the equilibrium values for s , b , ψ/μ , q_s , c , y , ψ , and μ are recursively determined by (38), (1), (36), (22), (6), (7), (16), and (36). The utility value of the payment schedule $\lambda \mathcal{Z}(q)$ is determined by (25), but the precise values of λ and $\mathcal{Z}(q)$ are indeterminate because they depend on the units in which payments are measured.

The following proposition summarizes the conditions for the existence and uniqueness of a directed search equilibrium (the proof is in the Appendix):

Proposition 3. *If the elasticity $\eta(\theta)$ is non-increasing and belongs to the interval $(0, 1)$, and the following terminal conditions are satisfied: $f(0) = 0$ and $f'(0) > 0$, then a directed search equilibrium exists and is unique.¹²*

Under the assumptions in Proposition 3, the left-hand side of (40) is a decreasing function of n . Therefore, the equilibrium amount of labor allocated to production is inversely related to the preference for diversity, measured by σ , and prices, measured by γ . A higher preference for diversity means that households are more willing to sacrifice labor to shop for a diverse basket of goods. Likewise, higher prices mean that households find it in their interest to allocate more labor to sell goods. Therefore, both a higher σ and a higher γ divert labor from production to trading activities. In our model, the value of γ is endogenous. As stated in (31) γ is increasing with $1 - \eta(\theta)$ with θ determined in (35). Therefore, the value of n is inversely related to the equilibrium contribution of sellers in the matching process measured by the term $1 - \eta(\theta)$.

To calculate the average commercial margin in a directed search equilibrium, we use (22) to obtain the following relationship:

$$\psi \int_0^1 \varepsilon q_e^{1-\sigma} d\varepsilon = (1 + \gamma\sigma)\mu\bar{q}. \quad (41)$$

Substituting (41) and (34) into (12) and (13), we get

$$\frac{2\gamma}{1-\gamma} \left(\frac{1+\gamma\sigma}{1-\sigma} \mu\bar{q} - \lambda\bar{z} \right) = \lambda\bar{z} - \mu\bar{q}. \quad (42)$$

Using (42), we obtain that the average commercial mark-up (ratio of the average commercial margin in the retail sector over the production cost of the merchandise):

$$\frac{\lambda\bar{z} - \mu\bar{q}}{\mu\bar{q}} = \frac{\sigma}{1-\sigma} [1 - \eta(\theta)]. \quad (43)$$

The average commercial mark-up provides a measure of the size of retailing relative to manufacturing because it is the ratio of the value added in these two sectors. Eq. (43) says that the value added of retailing relative to manufacturing increases with both the contribution of sellers in the matching process measured by $1 - \eta(\theta)$, and the preference for diversity σ . The higher the sellers' contribution in the matching process $1 - \eta(\theta)$, the higher the value added by any one seller in the retail sector. A higher preference for diversity σ raises the households' willingness to allocate their labor to buying in order to acquire a diverse basket of goods. An increase in the number of buyers reduces the search costs of sellers and raises the value added of a seller in the retail sector.

¹²This proof also uses the continuity and concavity of m^s and f , and the continuity and convexity of m^b assumed throughout the paper.

5.1. Special case: frictionless matching

With a frictionless matching technology $\mathcal{M}(B, S) = \min\{B, S\}$, the characterization of the directed search equilibrium is identical except for the trade-off between the congestion and the price schedule. To maximize the expected utility of the seller in Eq. (29) subject to the constraint that the expected utility of the buyer in Eq. (28) is constant, we must set $\theta = 1$. Therefore, in equilibrium $B = S$, so all buyers and all sellers are matched with probability one. Eq. (31) is no longer relevant, and Eq. (34) implies $\gamma = 1/3$. Eq. (40) still determines the value of n once we set $\gamma = 1/3$. Likewise, given θ , γ , and n , the equilibrium values for s and b (which are identical), ψ/μ , q_e , c , y , ψ , and μ are still determined recursively using (39), (36), (22), (6), (7), (16), and (36). Moreover, the exact same argument used in the proof of Proposition 3 implies that a directed search equilibrium exists and is unique if $f(0) = 0$ and $f'(0) > 0$.

If matching technology is frictionless, the average commercial mark-up is

$$\frac{\lambda\bar{z} - \mu\bar{q}}{\mu\bar{q}} = \frac{2}{3} \frac{\sigma}{1 - \sigma}. \quad (44)$$

As in the previous section, the average commercial mark-up increases with the preference for diversity σ . However, the average commercial mark-up is not affected by search costs since sellers trade with probability one.

In Section 6, we study the efficiency properties of the directed search equilibrium allocation. Most qualitative results in Section 6 hold whether or not matching frictions are present. Our main result is that the directed search equilibrium is inefficient. The effect of matching frictions is to distort the directed search equilibrium allocation further away from the first best allocation.

5.2. Multiple matches

In our model, there is no logical reason to restrict buyers and sellers to at most one match in the trading period. The extension to multiple matches is straightforward for the following reasons. Individual traders (buyers or sellers) are infinitesimal in the household where they belong, so their individual actions have no effect on the shadow valuations of goods, labor, and payments in the household: ψ , μ , and λ . Moreover, thanks to the clearing-payments mechanism, an individual buyer is not restricted by the money he or she carries. Likewise, thanks to the immediate access to current production, an individual seller is not restricted by available commercial inventories. Therefore, the optimal strategies of buyers and sellers in successive trading meetings do not depend on their individual histories. Our analysis and results apply with almost no modification to a generalized model where buyers and sellers can perform multiple matches during the trading period. In that model, we reinterpret $m^s(\theta)$ and $m^b(\theta)$ as the expected number of trading matches that a seller and a buyer perform during the trading period, respectively. With this reinterpretation, the image of these two functions is $[0, \infty)$ instead of

$[0, 1]$.¹³ Conveniently, this allows for a Cobb–Douglas matching technology for which the elasticity of $m^s(\theta)$ is a constant.

6. Welfare

This section studies the welfare properties of the directed search equilibrium. To this end, it characterizes the optimal allocation that a benevolent central planner would choose in order to maximize the utility of a representative household. Following standard practice, the central planner is not only bound by the resources available in the economy, but also by the bilateral matching among traders. In addition, it is sensible that we restrict the central planner to information that is publicly available. However, imposing this restriction is much more subtle than it first appears. The incentive compatibility constraint (19) arising from the fact that the buyers’ types are private information is only binding when selling costs must be financed with the revenue from sales (see below). Moreover, the central planner can easily affect this revenue with policy tools that, in principle, are respectful to private information, for example, a sales tax or a sales subsidy. Due to the subtleties of restricting the central planner to public information, we start by characterizing the first best allocation, in which the central planner is not bound by the incentive compatibility constraints imposed by private information. This allocation is an interesting benchmark in itself and is useful to evaluate the welfare costs of private information. Later, we introduce private information with specific assumptions about the policy tools at the disposal of the central planner.

6.1. The optimum with complete information

The first best allocation is one in which the central planner maximizes the utility of a representative household subject to the resources available in the economy and the bilateral matching among traders. Let $\mathcal{M}(b, s)$ be the aggregate matching function, that is $\mathcal{M}(b, s) = bm^b(b/s) = sm^s(b/s)$. The planner must solve the following program:

$$\max_{b, s, n, q_e} \mathcal{U}(c), \text{ where } c = \left[\mathcal{M}(b, s) \int_0^1 \varepsilon q_\varepsilon^{1-\sigma} d\varepsilon \right]^{\frac{1}{1-\sigma}}, \quad (45)$$

where subject to (10) and

$$\mathcal{M}(b, s) \int_0^1 q_\varepsilon d\varepsilon = f(n). \quad (46)$$

¹³With multiple matches, it is important that buyers can play mixed strategies when they choose which prices to search. Otherwise, a deviating seller would have a hard time attracting prospective buyers because these would dislike being in a submarket where there is a single good.

The first-order conditions of this problem are

$$\varepsilon\psi q_\varepsilon^{-\sigma} = \mu, \quad (47)$$

$$\mathcal{M}_b(b, s) \int_0^1 \frac{\varepsilon\psi q_\varepsilon^{1-\sigma}}{1-\sigma} d\varepsilon - \mu \mathcal{M}_b(b, s) \int_0^1 q_\varepsilon d\varepsilon = \mu f'(n), \quad (48)$$

$$\mathcal{M}_s(b, s) \int_0^1 \frac{\varepsilon\psi q_\varepsilon^{1-\sigma}}{1-\sigma} d\varepsilon - \mu \mathcal{M}_s(b, s) \int_0^1 q_\varepsilon d\varepsilon = \mu f'(n), \quad (49)$$

where μ is the Lagrange multiplier of (46) and $\psi = \mathcal{U}'(c)c^\sigma$. In the first best, the marginal utility of consuming each good must be equal to the marginal cost of production. Also, the marginal social benefit of employment in all three activities must be the same. Comparing (48) and (49), we obtain

$$\mathcal{M}_b(b, s) = \mathcal{M}_s(b, s). \quad (50)$$

Equality (50) implies that for a given number of traders the number of trading meetings is maximized. Eqs. (10) and (46) to (49) can be easily manipulated to obtain an almost explicit solution of the first best allocation (see the Appendix for details):

Proposition 4. *The first best allocation, in which the central planner has both complete information and control over all variables, is characterized by the following equations:*

$$\frac{1}{1+\theta} = 1 - \eta(\theta), \quad (51)$$

$$\frac{(1-n)f'(n)}{f(n)} = \frac{\sigma}{1-\sigma}, \quad (52)$$

$$s = \frac{1-n}{1+\theta}, \quad (53)$$

$$b = \theta s \quad (54)$$

and

$$q_\varepsilon = \frac{1+\sigma}{\sigma} \frac{f(n)}{\mathcal{M}(b, s)} \varepsilon^{\frac{1}{\sigma}}. \quad (55)$$

The following proposition compares the allocations in the first best and in the directed search equilibrium. The proposition shows that congestion is higher and production is lower in equilibrium than in the first best. To facilitate comparison of other variables, the proposition specializes the production and matching technologies to standard functional forms:

Proposition 5. *Suppose that all the assumptions used in Proposition 3 to ensure existence and uniqueness of an equilibrium hold. Let an asterisk denote first best and no*

asterisk denote directed search equilibrium. The following relations hold:

$$\theta > \theta^*, \quad (56)$$

$$n < n^*. \quad (57)$$

In addition, if f is isoelastic,

$$\frac{s}{n} \geq \frac{s^*}{n^*}, \text{ with equality if } \eta(\theta) \text{ is constant.} \quad (58)$$

Finally, if f is constant returns to scale,

$$\bar{q} < \bar{q}^*. \quad (59)$$

In a directed search equilibrium, selling costs are financed with commercial margins that create a wedge between the marginal cost of production of merchandises and the price paid by buyers. This wedge reduces the quantities that buyers purchase in each transaction for given valuations of goods and labor ψ and μ (compare (22) and (47)). However, ex ante price competition among sellers narrows commercial margins so the congestion of buyers to sellers in equilibrium is higher than in the first best. Intuitively, the equilibrium price schedule must play two conflicting allocational roles: It must signal buyers the opportunity cost of the goods they intend to purchase, and it must finance retail costs in order to induce an efficient ratio of sellers over buyers. The equilibrium price schedule settles on a compromise between these two roles. Prices are higher than the social opportunity cost of goods but not high enough to finance the first best ratio of sellers over buyers. Even though, commercial margins induce buyers to purchase smaller quantities, the comparison between q_e and q_e^* is ambiguous because the equilibrium and the first best assign a different shadow value to the cost of production relative to the utility of consumption (μ/ψ). In the special case where f is constant returns to scale, we can prove that on average buyers purchase an inefficiently low quantity when they meet a seller.

In the baseline case where both f and m^s are isoelastic, the ratio (s/n) is identical in the equilibrium and the first best allocations. Therefore, in this baseline case the inefficiently high (b/s) ratio implies that in equilibrium the measure of buyers is inefficiently high and the measures of sellers and producers are inefficiently low. In general, the equilibrium allocation of labor relative to the first best depends on whether the elasticities of f and m^s are increasing or decreasing with respect to n and θ , respectively.

The following proposition compares the allocations in the first best and in the directed search equilibrium when the matching technology is frictionless. With this matching technology, the planner always chooses $\theta = 1$. Eqs. (52)–(55) still characterize the first best allocation once we set $\theta = 1$ and $\mathcal{M}(b, s) = b = s$. Proposition 6 shows that in equilibrium the congestion is efficient, but the production is still lower than in the first best.

Proposition 6. Suppose that the matching function is efficient and that the assumptions on f used in Proposition 3 hold (so a unique directed search equilibrium exists). Let an asterisk denote first best and no asterisk denote directed search equilibrium. The following relations hold:

$$\theta = \theta^* = 1, \quad (60)$$

$$n < n^*, \quad (61)$$

$$b = s > b^* = s^*, \quad (62)$$

$$\bar{q} < \bar{q}^*. \quad (63)$$

When the matching technology is efficient, the equilibrium price schedule induces an efficient ratio of sellers over buyers, but prices are still higher than the marginal cost of production. On average buyers purchase an inefficiently low quantity when they meet a seller. Even though the ratio (b/s) is efficient, in equilibrium there are too many buyers and sellers performing transactions that are too small. As a result, whether or not matching frictions are present, the relative amount of labor allocated to production relative to selling is inefficiently low.

As in Moen [10] and Shimer [22], the first best allocation can be implemented as a directed search equilibrium when buyers and sellers have complete information. In this case, a bilateral trade must maximize the joint surplus of buyers and sellers. Moreover, the ex ante price competition among sellers leads to sharing this surplus according to Hosios's [5] rule, that is buyers get a fraction $\eta(\theta)$ of the surplus and sellers get $1 - \eta(\theta)$, and θ is the first best level of congestion. With private information about buyers' types, this equilibrium breaks down because the incentive compatibility constraint (19) is violated.

The first best allocation can be decentralized as a directed search equilibrium where buyers' types are private information if sellers can charge a flat fee to prospective buyers prior to the realization of the preference shock ε . In this case, the flat fee covers selling costs without having to add a positive commercial margin on the price of merchandises. In our model, such a fee is prevented by the assumption that buyers get the preference shock as soon as they meet a seller, so a buyer is not willing to pay the fee if the realization of ε is low. That is, flat fees violate the individual rationality constraint of buyers. More generally, one could realistically assume that the buyers' satisfaction from a commercial transaction depends on the service effort provided by the seller. In this more complicated model, the flat unconditional fee is also discouraged by the moral hazard problem it generates on the effort exercised by sellers. In reality, we find flat fees in warehouse clubs. However, sales in warehouse clubs are a small fraction of the economy-wide retail sales, and even these clubs charge fees that cover a small fraction of their commercial costs.

The following proposition summarizes these two ways of decentralizing the first best allocation:

Proposition 7. *The first best allocation where the planner has complete information can be implemented as a directed search equilibrium if buyers and sellers have complete information or if sellers can charge a flat fee to prospective buyers prior to the realization of the preference shock ε .*

6.2. The optimum with private information

This subsection characterizes second best allocations when the central planner has limited information. Given that buyers' types are private information it is natural to assume that the central planner cannot directly observe them. However, if the central planner can monitor the allocation of labor in each household, then the first best allocation can be easily implemented by dictating to households the allocation of labor, charging buyers the marginal cost of producing merchandise, and transferring the proceeds of these payments to sellers. With this mechanism, buyers truthfully reveal their types without any efficiency loss. To make our analysis more realistic and more interesting than this simple result, we assume that the central planner cannot directly observe how households allocate their labor. This assumption is in line with the unobservability of leisure (and so the allocation of time when there are two activities) in the standard theory of taxation.

Specifically, we assume that the planner can only observe market transactions and only has control of the direct revelation mechanisms by which transactions are conducted, or equivalently the price schedules faced by buyers and sellers. Given these price schedules, buyers choose the quantities they want to purchase, which sellers supply as long as prices are above the marginal costs of production. Households decide the allocation of labor taking into account the expected returns from each activity.

In principle, buyers and sellers could face different price schedules. If so, the gap between the two schedules is a sales tax (or a sales subsidy) to be collected (or distributed) by a government who balances its budget with a lump-sum subsidy (or a lump-sum tax) on households. In the two propositions that follow, we make alternative assumptions on the central planner's ability to control the pricing mechanism and to impose lump-sum taxes on households.

In the absence of any restriction on these policy instruments, the central planner can implement the first best allocation despite the presence of private information:

Proposition 8. *If the central planner can resort to lump-sum taxes on households to finance a linear sales subsidy, then the first best allocation characterized in Proposition 4 can be decentralized with the following price schedule faced by buyers:*

$$\mathcal{Z}(q) = \frac{\eta(\theta)\mu q}{\lambda}, \quad (64)$$

and a linear sales subsidy at the gross rate:

$$T = \frac{1 - \sigma\eta(\theta)}{(1 - \sigma)\eta(\theta)}, \quad (65)$$

where θ is the solution to (51). (The price schedule faced by sellers is $T\mathcal{Z}(q)$).

This proposition implies that the planner could not do better by introducing additional control instruments. Intuitively, the proposition holds because the planner picks the tab for the selling costs with the sales subsidy. Hence, the prices faced by the buyers can be equated to the marginal social cost of production of merchandises.

The implementation of the first best allocation in Proposition 8 depends on the existence of lump-sum taxes. In the model, this is not problematic because all households are identical. However, this homogeneity is unrealistic and has been assumed only for simplicity. The next proposition assumes the absence of lump-sum taxes.

Proposition 9. *If the only policy instrument is to set the retail trade price schedule and the marginal product of labor is constant (the production function is affine), a directed search equilibrium is efficient.*

If the production function is affine, a central planner without recourse to lump-sum taxes cannot improve upon the directed search equilibrium. When the production function is not affine, regulation of the price schedule can be welfare enhancing because it has an indirect effect on the marginal product of labor which differs from its first best value. However, as long as the production function is differentiable, it can always be approximated with an affine production function, and so the welfare gains of deviating from the directed search equilibrium are second order of magnitude. In the next section, we use numerical simulations to check that these welfare gains are very small.

7. Commerce in a Neoclassical growth framework

This section sketches how to embed the model developed in the previous sections in a Neoclassical growth framework. In the extended model, time is infinite and discrete. Each period in the life of a household proceeds as in the static model. Also, in the resulting synthesis, the economy has two sectors: one produces goods combining capital and labor as in the Neoclassical growth model, the other exchanges goods in directed search markets where buyers' types are private information. To make our model as close as possible to the basic Neoclassical growth model, we assume that investment does not require installation or commercial costs. However, future work could incorporate these features. For brevity, we omit all proofs, which are either standard in the Neoclassical growth theory or parallel the arguments in the previous sections.

The household's intertemporal utility function is

$$\sum_{t=0}^{\infty} \beta^t \mathcal{U}(c_t), \quad (66)$$

where $\beta \in (0, 1)$ is the household's discount factor.

Also, the production of goods requires not only labor but also capital:

$$y_t = \mathcal{F}(k_t, n_t). \quad (67)$$

The production function $\mathcal{F}: \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ maps capital and labor into output. This function is continuously differentiable, increasing in both arguments, concave, and homogeneous of degree one. Also, the Inada conditions for an interior solution are assumed to apply.

Goods can be used for both consumption and investment. When goods are used for consumption, they are exchanged in the same type of markets as those described in the previous sections. Goods used for consumption are perishable and must be sold in the same period in which they are produced. Goods are used for investment are perfect substitutes for one another. To save commercial costs, households use part of their own output to increase their capital stock. Therefore, households allocate a fraction of their output to be sold and another fraction to investment:

$$y_t = k_{t+1} - k_t(1 - \delta) + s_t m^s(\theta_t) \bar{q}_t, \quad (68)$$

where $\delta \in (0, 1)$ is the depreciation rate. The fraction destined for sale is consumed by the purchasing households.

Each period t the household is subject to constraints (6), (9) to (10), (67) and (68).¹⁴ The problem of the household is to choose $\{q_{et}\}_{e \in [0,1]}, b_t, s_t, n_t, c_t, y_t$ and k_{t+1} to maximize (66) subject to this system of constraints. We use Lagrange multipliers $\beta^t \mu_t, \beta^t \lambda_t$, and $\beta^t \vartheta_t$ for the resource constraint, the budget constraint, and the labor allocation constraint, respectively. Then, the optimal behavior of a household is characterized by the same first-order conditions as those in Section 3, that is (11)–(14), with all variables indexed by the corresponding time subscript, and the obvious modification that output and the marginal product of labor now depend on the capital stock. Moreover, the following two conditions must hold:

$$\mu_t = \mu_{t+1} [\mathcal{F}_k(k_{t+1}, n_{t+1}) + 1 - \delta] \beta, \quad (69)$$

and

$$\lim_{t \rightarrow \infty} \beta^{-t} \mu_t k_t = 0. \quad (70)$$

Eq. (69) states that the value of one unit of output today is equal to the present discounted value of the gross marginal product of capital, while Eq. (70) is a standard transversality condition.

A directed search equilibrium is a set $\{n_t, b_t, s_t, c_t, y_t, k_{t+1}, q_{et}, z_{et}, \psi_t, \mu_t, \lambda_t, \mathcal{Z}_t(q), \theta_t\}_{t=0}^\infty$ satisfying conditions analogous to the ones in the definition in Section 5. This equilibrium is now described by a system of difference equations for the

¹⁴Regarding condition (9), we could allow the clearing-house to pay or charge interest on the balance of credits and debits imposing only an intertemporal budget constraint on the households. However, since all households behave identically, all balances must be zero in equilibrium. Therefore, we can simplify the exposition by assuming that the balance of credits and debits for each household must be zero in each period.

variables k_t and μ_t : (68) and (69) together with the time-indexed versions of conditions (1), (10), (16), (22), (31), (34), (36),¹⁵ (37), (39), and (67). This system together with the initial condition k_0 and the terminal condition (70) determines the equilibrium path. For all capital stocks, Eqs. (31) and (34) still determine the pricing parameter γ_t and the congestion θ_t of buyers over sellers in the market for consumption goods. Therefore, these two variables are constant along an equilibrium path.

Qualitatively, the dynamics of capital accumulation are identical to those of the Neoclassical growth model. Capital converges monotonically to a steady state where k_t and μ_t are constant. For low capital stocks, both the marginal product of capital and the utility price of capital μ_t are high relative to the steady state. High levels of μ_t induce low consumption and as a result high saving. As capital is accumulated, households not only increase the fraction of output allocated to consumption, but also the fraction of labor allocated to the exchange of commodities.

In the steady state, Eq. (69) implies that the net marginal product of capital is equal to the subjective discount rate:

$$\mathcal{F}_k\left(\frac{k}{n}, 1\right) - \delta = \frac{1}{\beta} - 1. \quad (71)$$

When net investment is zero, analogous steps to the derivation of (40) yield:

$$\frac{\mathcal{F}_n\left(\frac{k}{n}, 1\right)}{\mathcal{F}\left(\frac{k}{n}, 1\right)} - \delta \frac{k}{n} = \frac{\sigma(1 + \gamma)}{1 - \sigma}. \quad (72)$$

Eqs. (71) and (72) determine the steady-state capital stock and labor allocated into production. The steady-state values of the remaining variables (b , s , q_e , and μ) are obtained from equations analogous to those in Section 5.

As in the version of the model without capital, a central planner who faces the same informational constraints as the market and who only has control of the direct revelation mechanism by which transactions are conducted cannot improve upon the directed search equilibrium allocation when the production technology is linear. With the existence of capital, constant returns to scale does not imply that the production technology is linear. However, with constant returns to scale the difference in welfare between a directed search equilibrium and the allocation that is attained when the central planner chooses the direct revelation mechanism are negligible for reasonable parameters (see the numerical example in the following subsection).

As in the Neoclassical growth model, output and capital converge to a steady growth path if the utility function \mathcal{U} is isoelastic and the efficiency of labor in the production of goods grows at a constant rate. Also, the utility function can be easily extended to include leisure or home services. In this case, the restrictions on \mathcal{U} for convergence to a steady growth path are the standard ones.

¹⁵Taking into account that the marginal product of labor depends on both capital and labor.

7.1. Numerical calibration

In the context of the Neoclassical growth framework, our model of commerce can be estimated using standard economic data. In this subsection, we discuss how to identify the parameters of the model, and we provide a numerical calibration of the model. For this purpose, we assume logarithmic preferences: $U(c_t) = \ln c_t$,¹⁶ and Cobb–Douglas matching and production technologies: $\mathcal{M}(b_t, s_t) = A_0 b_t^\eta s_t^{1-\eta}$ and $\mathcal{F}(k_t, e_t n_t) = A_1 k_t^\alpha (e_t n_t)^{1-\alpha}$. (The functional form of the matching technology could be estimated with data on the distributions of idle serving spells by sellers and of waiting times to be served by buyers, but we leave this empirical work for future research.) We also assume that the efficiency of labor in the production of goods (e_t) grows at a constant rate g .

As is standard in the Neoclassical growth model, the parameters α , β , δ , and g can be estimated using capital and labor income shares in the sector producing goods, the real return on capital, the durability of capital, and the average growth rate of output. In our numerical example in [Table 1](#), we pick standard estimates for these parameters. The two remaining parameters of the model to estimate are the preference for diversity (σ) and the elasticity of matches with respect to the number of sellers (η). (The values of A_0 and A_1 just reflect the units of measurement and are irrelevant for all the calculations that follow.) These two parameters can be identified with empirical estimates of the average commercial mark-up and the congestion in the retail sector. In a directed search equilibrium, Eqs. (31), (35), and (43) imply:

$$\text{Average commercial mark-up} \left(\frac{\lambda_t \bar{z}_t - \mu_t \bar{q}_t}{\mu_t \bar{q}_t} \right) = \frac{\sigma(1-\eta)}{1-\sigma} \quad (73)$$

and

$$\text{Congestion} \left(\frac{b_t}{s_t} \right) = \frac{1+\eta}{2(1-\eta)}. \quad (74)$$

Therefore, with estimates of the average commercial mark-up and market congestion we can solve for σ and η .

The Bureau of the Census of the United States reports that the average commercial margin over sales for retail trade has quite stable around 0.28 during the last decade.¹⁷ This implies that the average commercial mark-up is approximately 0.39 (i.e., 0.39 is approximately equal to $0.28/(1-0.28)$). The Bureau of Labor Statistics measures the number of production employees in retail trade and the average weekly hours that these employees work. The product of these two measures

¹⁶When preferences are logarithmic, the system of difference equations characterizing a directed search equilibrium is greatly simplified because then $s_t m^s(\theta_t) \bar{q}_t = (1 + \sigma\gamma)^{-1} \mu_t^{-1}$.

¹⁷See <http://www.census.gov/svsd/www/artstbl.html>. The commercial margin over sales varies widely by the type of business of the retailers. For example, in 2000, at the lower end, we find the commercial margins for Warehouse Clubs and Superstores (0.167), Automotive Dealers (0.175), and Gasoline Stations (0.208). At the upper end, we have the margins for Specialty Food Stores (0.419), Clothing and Footwear (0.426), and Furniture (0.441). In this paper we abstract from the reasons why different goods may trade with different margins, although this is an interesting topic for future research.

Table 1
Numerical example

Production function: $\mathcal{F}(k, en) = A_1 k^\alpha (en)^{1-\alpha}$, $\alpha = 0.36$

Depreciation rate: 0.1

Rate of interest: $r = 0.04$

Rate of growth of e : $g = 0.018$

Discount factor: $\beta = (1 + g)/(1 + r)$

One period utility: $\ln(c)$

Preference for diversity: $\sigma = 0.42$

Matching technology: $M(b, s) = A_0 b^\eta s^{1-\eta}$, $\eta = 0.46$.

	Directed search equilibrium	First best allocation
Balanced growth path		
Producers (n)	0.500	0.559
Sellers (s)	0.213	0.237
Buyers (b)	0.287	0.203
Capital/labor ratio (k/n)	4.374	4.374
Pricing weight parameter (γ)	0.270	0
Quantity purchased when $\varepsilon \geq \gamma(q_\varepsilon)$	$11.23(1.37\varepsilon - 0.37)^{2.38}$	$10.13\varepsilon^{2.38}$
Welfare relative to the first best (equivalent percentage change in consumption)		
Comparison across balanced paths	-0.0794	0
With same initial capital	-0.0787	0

is the empirical counterpart of s_t (465 million hours/week in 1986).¹⁸ To calculate b_t , we multiply the average time spent shopping by an adult (3.4 h/week in 1986) reported in Robinson et al. [18, p. 84]¹⁹ by the number of shoppers in the economy (United States population 16 and over). This product measures the empirical counterpart of b_t (630 million hours/week in 1986). This implies that congestion is 1.35 ($\simeq 630/465$). Applying these estimates to the system (73) and (74), we obtain $\eta = 0.46$ and $\sigma = 0.42$.

Using the estimated parameters, Table 1 compares the directed search equilibrium and the first best allocation. These calculations use the formulas described in previous sections for the balanced growth paths and use numerical methods for the transitional dynamics necessary to calculate the last row of the Table. In equilibrium, households spend less time producing and selling goods and more time shopping than in the first best. Also, in equilibrium buyers leave empty handed from 27 percent of the trading meetings while they always acquire a positive amount of goods in the first best. Welfare in the two allocations is quite different. Changing from the first best balanced path to the equilibrium balanced path is equivalent to a 7.94

¹⁸We use 1986 to estimate congestion because the survey on the time individuals spend shopping that we have on hand refers to that year. The data was downloaded in July 1, 2002 from <http://stats.bls.gov/data/home.htm> (production employees in retail trade: 15.924 million, average hours worked by production employees: 29.2).

¹⁹This is the most recent estimate of average time allocated to shopping that we could find. We do not expect that indices of congestion vary dramatically over time.

percent drop in consumption. When the transitional costs of changing the capital stock are taken into account this percentage drops slightly to 7.87.²⁰ This large welfare cost is not easy to avoid with correcting policies. For example, regulating the price schedule in a directed search equilibrium leads only to a negligible welfare improvement equivalent to less than a 10^{-5} percent increase in consumption. (This improvement would be zero if the production technology were linear instead of Cobb–Douglas). The large difference in welfare between a directed search equilibrium and the first best allocation is due to the necessity of financing retail costs with large commercial mark-ups. This inefficiency is unavoidable in the absence of some form of lump-sum taxes which would allow to subsidize retail trade.

8. Conclusion and extensions

Search models have been used to study decentralized markets where traders meet bilaterally. These models have been useful to analyze the labor market. Also, with the work of Kiyotaki and Wright [6,7] they have become the dominant paradigm for the theoretical microfoundations of money. Our paper uses a search model to capture some important features of the retail sector.

We adopt the concept of directed search equilibrium in an environment with private information in order to capture the following features of retail trade. The cost of a single transaction is small, yet buyers and sellers spend large amounts of labor performing transactions. The matching of buyers and sellers is in general frictional. Sellers post prices to attract buyers. Both buyers and sellers care about the probability of realizing a trade in a given period. Finally, retail prices are typically anonymous, imply no payments if there are no purchases, and involve quantity discounts.

In this framework, we study the welfare properties of a directed search equilibrium allocation by comparing it to the allocation chosen by a central planner. If the planner faces the same informational constraints as the market and only has control of the market prices, then the directed search equilibrium allocation coincides with the choice of the planner when the production technology is linear. However, if the planner can use lump-sum taxes to subsidize sales, the planner can improve upon the equilibrium allocation. In fact, the planner can achieve the first best allocation by introducing a linear subsidy on sales.

Even if large commercial margins are unavoidable in the absence of lump-sum taxation, their existence is important for policy design. For example, large commercial margins have a profound effect on the welfare cost of sales taxes, because a zero tax rate on sales is already a large departure from the first best, the standard result that small taxes cause negligible dead-weight-losses does not apply. Moreover, because of standard tax equivalence results, the welfare costs of income taxes must also be much larger when commercial margins are taken into account.

²⁰In this comparison, we assume that both the equilibrium path and the first best path start with the same capital stock (the one in the first best).

The informational structure in our model allows us to generate retail prices that are anonymous and include no flat fees. For some applications, one may wish to abstract from the specific form the equilibrium pricing schedules take. In this case, the model can be drastically simplified by either eliminating the preference shocks, or assuming full information, or allowing two-tier pricing. Likewise, one may wish to abstract from waiting times by buyers and idle spells by sellers. In this case, a further simplification can be achieved assuming a frictionless matching technology.

For simplicity, we have abstracted from the existence of money by assuming that payments can be made through a central clearing-house. In the absence of this centralized system of payments, money is a useful device that facilitates trade as in Kiyotaki and Wright [6,7]. Faig [4] introduces money in a simpler version of the present model where sellers are restricted to make offers that consist of a single quantity–payment pair. The main complication of introducing money in the present set-up is that in equilibrium when buyers are lucky to find a good for which they have a high valuation they would like to spend more money than they carry. Because this is not possible, they are liquidity constrained. Moreover, these liquidity constraints affect equilibrium price schedules. Despite this complication, the model remains analytically tractable, but the algebraic expressions are longer and harder to interpret than in the present contribution. For this reason, we plan to study the monetary version of the present model in a separate paper.

Appendix

A.1. Derivation of (6)

We derive the hedonic measure of consumption in (6) from a standard Dixit–Stiglitz aggregator by using the Law of Large Numbers. Let q_j be the quantity purchased and ε_j be the preference shock in each match j between a buyer of the household and a seller. Because there is a countable number of such matches and a continuum of goods for sale, with probability one an additional purchase brings a new good to the set of goods consumed by the household. Therefore, the number of goods the household can consume is equal to the number of matches.

Take the case where the set of household members is finite. In this case, the number of goods the household can consume is also finite. We assume the following Dixit–Stiglitz aggregator:

$$c = \left(\frac{1}{N} \sum_{j=1}^J \varepsilon_j q_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \quad (\text{A.1})$$

where N denotes the number of household members and J denotes the number of goods the household can consume. The variable c measures the per capita contribution of the purchases of the household. The goods acquired have different valuations for the household. For a given match j and a given quantity purchased q_j ,

the valuation is maximal when the realization of the shock is $\varepsilon_j = 1$, and it declines linearly to 0 as ε_j decreases to 0.

Taking limits as the number of household members goes to infinity (so the number of goods the household can consume also goes to infinity) we have that

$$\begin{aligned} c &= \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{j=1}^J \varepsilon_j q_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}} = \lim_{N \rightarrow \infty} \left(\frac{J}{N} \sum_{j=1}^J \frac{\varepsilon_j q_j^{1-\sigma}}{J} \right)^{\frac{1}{1-\sigma}} \\ &= \left(b m^b(\theta) \lim_{J \rightarrow \infty} \sum_{j=1}^J \frac{\varepsilon_j q_j^{1-\sigma}}{J} \right)^{\frac{1}{1-\sigma}}. \end{aligned} \quad (\text{A.2})$$

When the number of household members is large, the number of goods purchased per capita is $\lim_{N \rightarrow \infty} J/N = b m^b(\theta)$. The average contribution of a purchase is $\lim_{J \rightarrow \infty} \sum_{j=1}^J \frac{\varepsilon_j q_j^{1-\sigma}}{J}$. The stochastic variables ε_j are uniformly distributed on the interval $[0, 1]$ and are independent across matches. Let q_ε denote the quantity purchased by a buyer when the realization of the preference shock is ε . Applying the Law of Large Numbers to expression (A.2) implies the formula in (6).

Proof of Proposition 1. Let the indirect utility of a type- ε buyer be defined as

$$v_\varepsilon \equiv \mathcal{U}^b(q_\varepsilon, z_\varepsilon; \varepsilon). \quad (\text{A.3})$$

Using v_ε , the incentive compatibility constraint (19) can be restated with the help of the following standard result (see [8, Proposition 23.D.2]):

Lemma. *A direct revelation mechanism satisfies the incentive compatibility constraint (19) if and only if q_ε is non-decreasing in ε and the indirect utility function satisfies*

$$v_\varepsilon = v_0 + \int_0^\varepsilon \frac{\partial}{\partial x} \mathcal{U}^b(q_x, z_x; x) dx = v_0 + \int_0^\varepsilon \frac{\psi q_x^{1-\sigma}}{1-\sigma} dx, \text{ for all } \varepsilon \in [0, 1]. \quad (\text{A.4})$$

Using this lemma and constraints (20) and (21) together with the definitions (15), (17), and (A.3), the incentive efficient direct revelation mechanisms can be characterized as the solution to the following program:

$$\max_{\{q_\varepsilon, v_\varepsilon\}_{\varepsilon \in [0,1]}} \int_0^1 \left[(1 - \omega) v_\varepsilon + \omega \left(\frac{\varepsilon \psi q_\varepsilon^{1-\sigma}}{1-\sigma} - \mu q_\varepsilon - v_\varepsilon \right) \right] d\varepsilon \quad (\text{A.5})$$

subject to²¹

$$\dot{v}_\varepsilon = \frac{\psi q_\varepsilon^{1-\sigma}}{1-\sigma}, \quad (\text{A.6})$$

$$q_\varepsilon \geq 0, \quad (\text{A.7})$$

$$v_\varepsilon \geq 0, \quad (\text{A.8})$$

²¹ \dot{v}_ε denotes the derivative of v with respect to ε evaluated at ε .

$$\pi_\varepsilon \equiv \frac{\varepsilon\psi q_\varepsilon^{1-\sigma}}{1-\sigma} - v_\varepsilon - \mu q_\varepsilon \geq 0, \quad (\text{A.9})$$

q_ε is non-decreasing, (A.10)

$$q_0 = 0, \quad \text{and} \quad v_0 = 0. \quad (\text{A.11})$$

Here, π_ε denotes the seller's surplus upon meeting a buyer of type ε . The payment in the transaction satisfies:

$$\lambda z_\varepsilon = \frac{\varepsilon\psi q_\varepsilon^{1-\sigma}}{1-\sigma} - v_\varepsilon = \pi_\varepsilon + \mu q_\varepsilon. \quad (\text{A.12})$$

This program can be solved with a standard application of the Pontryagin's Maximum Principle. We organize the solution to the program in three steps:

(a) *Constraints (A.6)–(A.11) imply that there is a $\gamma \in [0, 1]$ such that the solution $\{q_\varepsilon, v_\varepsilon\}_{\varepsilon \in [0, 1]}$ to program (A.5) obeys: $q_\varepsilon = 0$ and $v_\varepsilon = 0$ for $\varepsilon \in [0, \gamma]$, and $q_\varepsilon > 0$ and $v_\varepsilon > 0$ for $\varepsilon > \gamma$.*

Constraints (A.10) and (A.11) immediately imply that there is $\gamma \in [0, 1]$ such that $q_\varepsilon = 0$ for $\varepsilon \in [0, \gamma]$, and $q_\varepsilon > 0$ for $\varepsilon > \gamma$. When $q_\varepsilon = 0$, constraints (A.8) and (A.9) imply that $v_\varepsilon = 0$. With these results, constraint (A.6) implies that $q_\varepsilon > 0$ if and only if $v_\varepsilon > 0$. Finally, if γ were 1, the optimized value of (A.5) would be zero, which cannot be a solution to the maximization program because there are many feasible direct revelation mechanisms that achieve a positive value. For example, $q_\varepsilon = v_\varepsilon = 0$ if $\varepsilon < 0.5$, and $q_\varepsilon = \left(\frac{\psi(2\varepsilon-1)}{\mu}\right)^{\frac{1}{\sigma}}$ and $v_\varepsilon = \frac{1}{2}\frac{\sigma}{1-\sigma}[\psi(2\varepsilon-1)]^{\frac{1}{\sigma}}\mu^{\frac{\sigma-1}{\sigma}}$ otherwise.²² \square

(b) *For $\omega \in (0.5, 1]$, the solution to program (A.5) subject to (A.6)–(A.11), is the following:*

$$q_\varepsilon = \begin{cases} 0 & \text{for } 0 \leq \varepsilon < \gamma, \\ \left(\frac{\psi}{\mu} \frac{\varepsilon - \gamma}{1 - \gamma}\right)^{\frac{1}{\sigma}} & \text{for } \gamma \leq \varepsilon \leq 1, \end{cases} \quad (\text{A.13})$$

$$v_\varepsilon = \begin{cases} 0 & \text{for } 0 \leq \varepsilon < \gamma, \\ \frac{\sigma\mu(1-\gamma)}{1-\sigma} \left(\frac{\psi}{\mu} \frac{\varepsilon - \gamma}{1 - \gamma}\right)^{\frac{1}{\sigma}} & \text{for } \gamma \leq \varepsilon \leq 1, \end{cases} \quad (\text{A.14})$$

where

$$\gamma = \frac{2\omega - 1}{3\omega - 1}. \quad (\text{A.15})$$

²²This is the optimal direct mechanism for $\omega = 1$.

Let ζ_ε denote the co-state variable associated with the differential equation (A.6). The current-value Hamiltonian of the program is

$$\mathcal{H}(q_\varepsilon, v_\varepsilon, \zeta_\varepsilon, \varepsilon) = (1 - \omega)v_\varepsilon + \omega \left(\frac{\varepsilon\psi q_\varepsilon^{1-\sigma}}{1-\sigma} - \mu q_\varepsilon - v_\varepsilon \right) + \zeta_\varepsilon \frac{\psi q_\varepsilon^{1-\sigma}}{1-\sigma}. \quad (\text{A.16})$$

The first-order necessary condition with respect to the control variable q_ε is ($\mathcal{H}_{q_\varepsilon} = 0$):

$$\omega(\psi\varepsilon q_\varepsilon^{-\sigma} - \mu) + \zeta_\varepsilon \psi q_\varepsilon^{-\sigma} = 0. \quad (\text{A.17})$$

The co-state variable must obey ($\dot{\zeta}_\varepsilon = -\mathcal{H}_{v_\varepsilon}$):

$$\dot{\zeta}_\varepsilon = -(1 - 2\omega). \quad (\text{A.18})$$

Finally, the transversality implies²³

$$\zeta_1 = 0. \quad (\text{A.19})$$

The value of the co-state variable ζ_ε can be solved for explicitly using conditions (A.18) and (A.19)

$$\zeta_\varepsilon = (2\omega - 1)(\varepsilon - 1). \quad (\text{A.20})$$

Substituting (A.20) into (A.17), solving for q_ε , and using (A.15), we obtain

$$q_\varepsilon = \left(\frac{\psi}{\mu} \frac{\varepsilon - \gamma}{1 - \gamma} \right)^{\frac{1}{\sigma}}. \quad (\text{A.21})$$

Integrating (A.6) from γ to ε and noting that $v_\gamma = 0$, we obtain

$$v_\varepsilon = \frac{\sigma}{1 - \sigma}(1 - \gamma)\mu q_\varepsilon. \quad (\text{A.22})$$

Finally, using the definition of π_ε in (A.9), we obtain

$$\pi_\varepsilon = \gamma \frac{\mu q_\varepsilon}{1 - \sigma} \left(\sigma + \frac{1 - \varepsilon}{\varepsilon - \gamma} \right). \quad (\text{A.23})$$

So far, we have disregarded constraints (A.7) to (A.11). However, for $\varepsilon \geq \gamma$, the values of q_ε , v_ε , and π_ε in (A.21)–(A.23) are well defined and satisfy these constraints. Moreover, at $\varepsilon = \gamma$, $q_\varepsilon = v_\varepsilon = 0$. Hence, (a) implies $q_\varepsilon = v_\varepsilon = \pi_\varepsilon = 0$ for $\varepsilon \leq \gamma$. The

²³The transversality condition is $\zeta_1 v_1 = 0$. However, $v_1 > 0$ because of step (a).

Hamiltonian is strictly concave with respect to q_ε , and when the Hamiltonian is evaluated at the optimal choice of q_ε it is concave with respect to v_ε . Therefore, (A.13)–(A.15) characterize the unique maximum of the program. \square

(c) For $\omega \in [0, 0.5]$, the solution to program (A.5) subject to (A.6)–(A.11), is the following:

$$q_\varepsilon = \left(\frac{\psi}{\mu}\varepsilon\right)^{\frac{1}{\sigma}} \quad (\text{A.24})$$

and

$$v_\varepsilon = \frac{\sigma}{1 - \sigma} \mu q_\varepsilon. \quad (\text{A.25})$$

When $\omega = 0.5$, the variable v_ε cancels in the objective (A.5) so the problem becomes separable across types. Ignoring constraints (A.9) and (A.10), the first-order conditions of the problem yield (A.24). Integrating (A.6) from 0 to ε and noting $v_0 = 0$, we obtain (A.25). Using the definition in (A.9), we obtain $\pi_\varepsilon = 0$ for all $\varepsilon \in [0, 1]$. The values of q_ε and v_ε in (A.24) and (A.25) satisfy constraints (A.9) and (A.10), so they solve program (A.5) subject to (A.6)–(A.11). Furthermore, the solution for $\omega = 0.5$ maximizes the expected payoff of the buyer subject to a zero expected payoff for the seller. A fortiori, given constraint (A.9), the same solution must apply for $\omega \in [0, 0.5)$, which assigns a lower weight to the seller in the maximized welfare function (A.5). \square

Finally, Proposition 1 results from combining steps (b) and (c) together with (A.12).

Proof of Proposition 2. A buyer facing a price schedule γ purchases a quantity that satisfies the following specialization of (11)

$$\varepsilon\psi q_\varepsilon^{-\sigma} = \gamma\psi q_\varepsilon^{-\sigma} + (1 - \gamma)\mu. \quad (\text{A.26})$$

Solving this equation for q_ε , we obtain (22). Using (25), the implied payments are given by (23). Finally, γ taking values in the interval $[0, 0.5]$ spans all values of γ attained in (24) for $\omega \in [0, 1]$. Thus, Proposition 2 follows. \square

Proof of Proposition 3. Given that the equilibrium variables are recursively determined by the set of equations described in the main text, existence and uniqueness is implied if Eqs. (35) and (40) have a unique admissible solution. The right-hand side of (35) is a continuous and decreasing function of θ , which image spans the interval $(0, 2]$. The left-hand side of (35) is a continuous and non-decreasing function of θ , which is bounded away from 0 and 1. Therefore, Eq. (35) has a unique solution.

The left-hand side of (40) is a decreasing function of n that maps $[0, 1]$ onto $[0, \infty)$, while the right-hand side is positive and independent of n . Hence, Eq. (40) has a unique solution for n in the interval $(0, 1)$. \square

Proof of Proposition 4. Eq. (50) together with the definitions of θ , m^b , and η in (1), (2), and (32), implies (51) and (54). The labor allocation constraint together with (51) and (54) yields (53). Using (47) to solve the integrals in (49) and (46), we obtain

$$f'(n) = \mathcal{M}_s(b, s) \frac{\sigma^2}{1 - \sigma^2} \left(\frac{\psi}{\mu} \right)^{\frac{1}{\sigma}} \quad (\text{A.27})$$

and

$$f(n) = \mathcal{M}(b, s) \frac{\sigma}{1 + \sigma} \left(\frac{\psi}{\mu} \right)^{\frac{1}{\sigma}}. \quad (\text{A.28})$$

Definitions (1), (2), and (32) imply

$$\frac{\mathcal{M}_s(b, s)}{\mathcal{M}(b, s)} = \frac{1 - \eta(\theta)}{s}. \quad (\text{A.29})$$

Eq. (52) results from combining (53) with (A.27)–(A.29). Finally, combining (47) and (A.28), we obtain (55). \square

Proof of Proposition 5. The equations that determine θ and n in a directed search equilibrium, (31) and (35), (40), and in the first best allocation, (51) and (52), can be written as follows:

$$\frac{1}{a_1 + \theta} = 1 - \eta(\theta) \quad (\text{A.30})$$

and

$$\frac{(1 - n)f'(n)}{f(n)} = \frac{\sigma}{1 - \sigma} a_2, \quad (\text{A.31})$$

where $a_1 = a_2 = 1$ in the first best, and $a_1 = 0.5$, $a_2 = (1 + \theta)/(0.5 + \theta) > 1$ in equilibrium. The Implicit Function Theorem applied to (A.30) and (A.31) yields $d\theta/da_1 < 0$ and $dn/da_2 > 0$. Hence, we obtain (57) and (56).

If f has constant elasticity α , equations (39)–(40) and (52)–(53) imply

$$\frac{s}{n} = \frac{\sigma[1 - \eta(\theta)]}{\alpha(1 - \sigma)}, \quad (\text{A.32})$$

for both the first best and the equilibrium. Therefore, (58) follows from (A.32), (56), and η being a non-increasing function of θ .

The resource constraint (8) together with f being constant returns to scale implies that

$$\frac{\bar{q}}{\bar{q}^*} = \frac{\frac{s^*}{n^*} m^s(\theta^*)}{\frac{s}{n} m^s(\theta)}. \quad (\text{A.33})$$

Therefore, (59) follows from (58), (56), and m^s being increasing. \square

Proof of Proposition 6. The proof is a simple variation of that of Proposition 5 and is thus omitted.

Proof of Proposition 7. (a) *When buyer types are public information the directed search equilibrium and the first best allocations coincide.*

As in Section 4, all bilateral trades must be pairwise efficient in equilibrium and the marginal rates of substitution between θ and ξ of buyers and sellers must coincide. With full information, the efficient quantities q_e are calculated by maximizing the trade surplus, $\mathcal{U}^b + \mathcal{U}^s$:

$$q_e = \arg \max \left(\frac{\varepsilon \psi q_e^{1-\sigma}}{1-\sigma} - \mu q_e \right). \quad (\text{A.34})$$

The solution to (A.34) is

$$q_e = \left(\frac{\psi \varepsilon}{\mu} \right)^{\frac{1}{\sigma}}. \quad (\text{A.35})$$

The expected trade surplus is

$$\int_0^1 \left(\frac{\varepsilon \psi q_e^{1-\sigma}}{1-\sigma} - \mu q_e \right) d\varepsilon = \frac{\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}}. \quad (\text{A.36})$$

When the market tightness is θ and buyers receive a fraction ξ of the trade surplus, the expected utilities of buyers and sellers are

$$\mathcal{V}^b(\theta, \xi) = \xi m^b(\theta) \frac{\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}}, \quad (\text{A.37})$$

and

$$\mathcal{V}^s(\theta, \xi) = (1-\xi) m^s(\theta) \frac{\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}}. \quad (\text{A.38})$$

In a directed search equilibrium, the marginal rates of substitution between θ and ξ of buyers and sellers must coincide. Differentiating (A.37) and (A.38), this implies

$$\xi = \eta(\theta). \quad (\text{A.39})$$

Households allocate b and s so $\mathcal{V}^b(\theta, \xi) = \mathcal{V}^s(\theta, \xi)$. Using (A.37) and (A.38), this equality yields (51). Also, households allocate s and n so $\mathcal{V}^s(\theta, \xi) = \mu f'(n)$. Using (A.38) and (A.39) this equality implies

$$\mu f'(n) = m^s(\theta) [1 - \eta(\theta)] \frac{\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}}. \quad (\text{A.40})$$

Using (A.35) to calculate the average sales, the resource constraint becomes

$$f(n) = s m^s(\theta) \frac{\sigma}{1+\sigma} \left(\frac{\psi}{\mu} \right)^{\frac{1}{\sigma}}. \quad (\text{A.41})$$

Combining (A.40) and (A.41), and using (1), (51), and (10), we obtain (52) and (53). From (A.40),

$$\left(\frac{\psi}{\mu}\right)^{\frac{1}{\sigma}} = \frac{f(n)}{sm^s(\theta)} \frac{1+\sigma}{\sigma}, \quad (\text{A.42})$$

which combined with (2) and (A.35) gives (55). This completes the proof of (a).

(b) *If buyer types are private information but sellers charge a lump-sum fee to prospective buyers prior to the realization of ε , then the directed search equilibrium and the first best allocations coincide.*

Let p be the fee a seller charges to prospective buyers. After the fee p has been paid and ε is realized, the trading game between a buyer and a seller is identical to the one in Section 4. Hence, the payment schedule net of the fee p that implements incentive efficient direct revelation mechanisms has still the functional form (25) and the quantities purchased by the buyer are given by (22). The expected utilities of the traders in a market with congestion θ , price schedule γ , and fee p are

$$\mathcal{V}^b(\theta, \gamma, p) = m^b(\theta) \left(\frac{\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}} \gamma^2 - \lambda p \right) \quad (\text{A.43})$$

and

$$\mathcal{V}^s(\theta, \gamma, p) = m^s(\theta) \left[\frac{2\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}} \gamma (1-\gamma) + \lambda p \right]. \quad (\text{A.44})$$

In a directed search equilibrium, the marginal rates of substitution of θ for γ , and of γ for p must be equal for buyers and sellers, so

$$\frac{\eta(\theta) - 1}{\eta(\theta)} = \frac{\gamma}{1-2\gamma} \frac{\frac{2\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}} \gamma (1-\gamma) + \lambda p}{\frac{\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}} \gamma^2 - \lambda p} \quad (\text{A.45})$$

$$\gamma = 1. \quad (\text{A.46})$$

Substituting (A.46) in (A.45) and solving for p ,

$$p = \frac{1}{\lambda} \frac{\sigma^2}{1-\sigma^2} \psi^{\frac{1}{\sigma}} \mu^{\frac{\sigma-1}{\sigma}} [1 - \eta(\theta)]. \quad (\text{A.47})$$

Note that this fee implies $\mathcal{V}^b(\theta, \gamma, p) > 0$, so buyers are willing to pay the fee.

Households allocate b and s so $\mathcal{V}^b(\theta, \gamma, p) = \mathcal{V}^s(\theta, \gamma, p)$. Using (A.46) and (A.47), this equality yields (51). Also, households allocate s and n so $\mathcal{V}^s(\theta, \gamma, p) = \mu f'(n)$. Using (A.46) and (A.47), this equality yields (A.40). To show that (52), (53) and (55) hold, we use the exact same steps used in the proof of (a). \square

Proof of Proposition 8. A type- ε buyer chooses

$$q_\varepsilon = \arg \max \left\{ \frac{\varepsilon \psi q^{1-\sigma}}{1-\sigma} - \lambda \mathcal{Z}(q) \right\} = \left(\frac{\psi \varepsilon}{\mu \eta} \right)^{\frac{1}{\sigma}}. \quad (\text{A.48})$$

Hence, the return of a seller when matched with a type- ε buyer is

$$\lambda T\mathcal{Z}(q_\varepsilon) - \mu q_\varepsilon = (T\eta - 1)\mu q_\varepsilon = \frac{\sigma(1-\eta)}{1-\sigma} \left(\frac{\psi\varepsilon}{\mu\eta}\right)^{\frac{1}{\sigma}} \mu, \quad (\text{A.49})$$

where $\eta = \eta(\theta)$ and θ is the first-best level of congestion given by (51). By (51), $\eta \leq 1$. Thus, the seller's individual rationality constraint is satisfied.

When the market tightness is θ , the expected utilities of the traders are

$$\mathcal{V}^b(\theta) = m^b(\theta)\eta \frac{\sigma^2}{1-\sigma^2} \left(\frac{\psi}{\mu\eta}\right)^{\frac{1}{\sigma}} \mu, \quad (\text{A.50})$$

$$\mathcal{V}^s(\theta) = m^s(\theta)(1-\eta) \frac{\sigma^2}{1-\sigma^2} \left(\frac{\psi}{\mu\eta}\right)^{\frac{1}{\sigma}} \mu. \quad (\text{A.51})$$

The household chooses its labor allocation so that $\mathcal{V}^b(\theta) = \mathcal{V}^s(\theta) = \mu f'(n)$. The first equality implies that θ satisfies (51) so $\eta = \eta(\theta)$. The second one implies

$$f'(n) = m^s(\theta)(1-\eta) \frac{\sigma^2}{1-\sigma^2} \left(\frac{\psi}{\mu\eta}\right)^{\frac{1}{\sigma}}. \quad (\text{A.52})$$

Using (A.48) to calculate the average sales, the resource constraint becomes

$$f(n) = sm^s(\theta) \frac{\sigma}{1+\sigma} \left(\frac{\psi}{\mu\eta}\right)^{\frac{1}{\sigma}}. \quad (\text{A.53})$$

Combining (A.52) and (A.53), and using (1), (51), and (10), we obtain (52) and (53). From (A.53),

$$\left(\frac{\psi}{\mu}\right)^{\frac{1}{\sigma}} = \frac{f(n)}{sm^s(\theta)} \frac{1+\sigma}{\sigma} \eta^{\frac{1}{\sigma}}, \quad (\text{A.54})$$

which combined with (2) and (A.48) gives (55). \square

Proof of Proposition 9. The direct revelation mechanism chosen by the planner must be incentive efficient, so we can restrict our search to price schedules in (25) for some unknown parameter γ . Substituting (22) into (6), solving for the integral, and using (2) we obtain

$$c = \left\{ \frac{sm^s(\theta)\sigma}{1+\sigma} (1-\gamma)(1+\sigma\gamma) \right\}^{\frac{1}{1-\sigma}} \left(\frac{\psi}{\mu}\right)^{\frac{1}{\sigma}}. \quad (\text{A.55})$$

Denoting $f(n) = a + wn$, and using (7), (8), (34), (37), (39), and (40), Eq. (A.55) simplifies to

$$c = \left(\frac{2w\sigma^2}{1+\sigma}\right)^{\frac{\sigma}{1-\sigma}} (1-\sigma)(a+w)^{\frac{1}{1-\sigma}} [\gamma(1-\gamma)m^s(\theta)]^{\frac{\sigma}{1-\sigma}}. \quad (\text{A.56})$$

Since (34) holds, the maximization of c implies:

$$(2\gamma - 1)m^s(\theta) + \gamma(1 - \gamma)(m^s)'(\theta)\frac{1}{2\gamma^2} = 0. \quad (\text{A.57})$$

Given (32) and (34), (A.57) implies (31). \square

References

- [1] D. Acemoglu, R. Shimer, Efficient unemployment insurance, *J. Polit. Economy* 107 (1999) 893–928.
- [2] K. Burdett, S. Shi, R. Wright, Pricing with frictions, manuscript, 2001.
- [3] G. Camera, A. Delacroix, Endogenous selection of a trade mechanism in a search-theoretic environment, manuscript, 2001.
- [4] M. Faig, A search theory of money and commerce with neoclassical production, manuscript, 2001.
- [5] A.J. Hosios, On the efficiency of matching and related models of search and unemployment, *Rev. Econ. Stud.* 57 (1990) 279–298.
- [6] N. Kiyotaki, R. Wright, On money as a medium of exchange, *J. Polit. Economy* 97 (1989) 927–954.
- [7] N. Kiyotaki, R. Wright, A search-theoretic approach to monetary economics, *Amer. Econ. Rev.* 83 (1) (1993) 63–77.
- [8] A. Mas-Colell, M. Winston, J. Green, *Microeconomic Theory*, Oxford University Press, New York, 1995.
- [9] E. Maskin, J. Riley, Monopoly with incomplete information, *RAND J. Econ.* 15 (1984) 171–196.
- [10] E. Moen, Competitive search equilibrium, *J. Polit. Economy* 105 (2) (1997) 385–411.
- [11] J. Montgomery, Equilibrium wage dispersion and interindustry wage differentials, *Quart. J. Econ.* 106 (1) (1991) 163–179.
- [12] D. Mortensen, Property rights and efficiency in mating, racing and related games, *Amer. Econ. Rev.* 72 (1982) 968–979.
- [13] D. Mortensen, R. Wright, Competitive pricing and efficiency in search equilibrium, *Int. Econ. Rev.* 43 (1) (2002) 1–20.
- [14] M. Peters, Ex-ante price offers in matching games: non-steady states, *Econometrica* 59 (5) (1991) 1425–1454.
- [15] M. Peters, Limits of exact equilibria for capacity constrained sellers with costly search, *J. Econ. Theory* 95 (2) (2000) 139–168.
- [16] M. Peters, S. Severinov, Competition among sellers who offer auctions instead of prices, *J. Econ. Theory* 75 (1) (1997) 141–179.
- [17] C. Pissarides, *Equilibrium Unemployment Theory*, Basil Blackwell, Oxford, 1990.
- [18] J.P. Robinson, V.G. Andreyenkov, V.D. Patrushev, *The Rytm of Everyday Life: How Soviet and American Citizens Use Time*, Westview Press, Boulder, CO, 1989.
- [19] S. Shi, A divisible search model of fiat money, *Econometrica* 65 (1997) 75–102.
- [20] S. Shi, Search, inflation, and capital accumulation, *J. Monet. Econ.* 44 (1999) 81–103.
- [21] S. Shi, Liquidity, bargaining, and multiple equilibria in a search monetary model, *Ann. Econ. Finance* 2 (2001) 191–197.
- [22] R. Shimer, Contracts in a frictional labour market, MIT manuscript, 1996.
- [23] E. Soller-Curtis, R. Wright, Price setting and price dispersion in a Monetary economy; or, The law of two prices, manuscript, 2000.