An Integrated Theory of Intermediation and Payments

Marshall Urias *

Abstract

This paper develops an integrated theory of intermediation and payments in wholesale and retail goods markets. The model synthesizes the search-theoretic approach to intermediation with the New Monetarist approach to payments. Under perfect credit consumption falls short of the efficient allocation due to a hold-up problem. Improving the bargaining position of intermediaries increases consumption and the measure of middlemen. When credit is not feasible there is a two-sided hold-up problem associated with middlemen’s inventory choice and consumers’ portfolio choice. This results in multiple steady state equilibria and two regimes which determine the response of consumption and the measure of middlemen with respect to fundamentals. Which regime the economy is in depends on the bargaining power of middlemen in the wholesale market and cost of entry. There exists a threshold nominal interest rate below which monetary policy is ineffective.

*Correspondence: Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA. Email: murias@uci.edu.
1 Introduction

Two institutions that facilitate exchange are intermediaries and money. Both are ubiquitous and provide explanations for how agents realize gains from trade in the presence of frictions. Progress has been made toward understanding why these institutions emerge by accounting for the frictions that inhibit trade, and then showing how these institutions alleviate these frictions. Rubinstein and Wolinsky (1987) advocated “a basic model that describes explicitly the trade frictions that give rise to the function of middlemen.” Karaken and Wallace (1980) argued that “progress can be made in monetary theory ...only by modeling monetary arrangements explicitly.” In this paper, I unify these two institutions into a single framework and characterize how their interaction affects the gains and distribution from trade.

I develop a model of intermediation to study the effects that different payment instruments have on inventory, entry, consumption and monetary policy. I characterize how intermediation interacts with money and credit toward achieving desirable outcomes, the strategic complementarities that exist between middlemens’ inventory choices and consumers’ portfolio choices, and the efficacy of monetary policy.

I work from the premise that any model of intermediation should include an explicit description of the trading process, and to study the impact of different payment instruments the model should take seriously the frictions that generate a need for such instruments. Thus, I use the New Monetarist framework which provides micro-foundations for the emergence of mediums of exchange (e.g. money) while naturally incorporating a search-theoretic approach to intermediation. The model environment is an extension of Lagos and Wright (2005) including a wholesale and retail market, varying degrees of commitment, endogenous inventory decisions, and free entry of middlemen.

Incorporating middlemen into a general equilibrium model has several implications for the functioning of markets. Middlemen are not simply a technology, but rather a strategic agent that possesses market power in output and input markets. Since prices are not taken as given in this case, middlemen set bid and ask prices to maximize expected profits subject to the willingness of producers to supply goods and consumers to purchase them. Not only does the model address
how prices and quantities are set by agents with market power, but also examines how competition among middlemen affect equilibrium allocations.

The form of exchange considered here is decentralized trade with pairwise meetings of agents where middlemen intermediate between consumers and producers. Producers have access to technology allowing them to produce a divisible retail good. Middlemen have access to search technology enabling them to find a producer from whom they buy the good, and then find a consumer to whom they resell the good. Bid and ask prices are determined by sequential bilateral bargaining and depend on the availability of enforceable contracts. The equilibrium is distorted by holdup problems under incomplete contracts and also by liquidity constraints from imperfect enforcement. The inherent frictions of the environment generate a two-sided holdup problem: one associated with inventory choices by middlemen, and the other from portfolio choices by consumers.

The environment supposes infinite search and matching costs for direct trade, thereby generating an essential role for middlemen. This follows in the tradition of Rubinstein and Wolinsky (1987) where middlemen are active in equilibrium only if they can match with consumers at a faster rate than producers. In RW, these matching rates are exogenous and so intermediated trade exists only if middlemen are endowed with a superior matching technology. In my model middlemen are indeed endowed with superior matching technology, but the rate at which intermediated trading opportunities arrive is endogenous and depends on the strategic choices of all agents. More specifically, a middleman's decision to enter the market depends on its relative bargaining position and likelihood of engaging in trade, which in turn is affected by the aggregate measure of active of middlemen.

The framework expositions varying degrees of commitment on the part of agents. First, the environment is such that agents can fully commit to pay for current trades at a future date, providing a benchmark case used to understand the interaction between wholesale and retail markets and the inherent frictions in the environment. Second, I consider the case where agents can strategically default which endogenizes borrowing capacity and alters the gains from intermediated trade. Finally, I consider the case where agents cannot credibly commit and instead must use a liquid asset to trade quid pro quo. This introduces a portfolio choice for agents that require money balances to purchase goods, and yields interesting strategic complementarities between consumer’s portfolio
choice and middlemen’ inventory choice. I also consider various payments in wholesale transactions.

I find that even when there is perfect enforcement in both wholesale and retail markets, consumption falls short of the first-best due to a holdup problem. Since middlemen must purchase inventory prior to meeting a consumer, this cost is sunk in the retail market leading to under investment in inventory. Improving the bargaining position of middlemen—increasing bargaining power in wholesale or retail markets or increasing its outside option—increases the number of operative middlemen and hence improves the extensive margin of trade. The amount of consumption per trade, the intensive margin, increases with retail bargaining power, decreases with wholesale bargaining power due to a thick market externality, and increases with the cost of entry. The bid-ask spread is strictly positive, bounded above by the joint surplus from direct trade, and increasing in the bargaining position of middlemen.

Sequential bargaining in wholesale and retail markets allow the cost associated with the holdup problem to be divided between middlemen and producers. More specifically, the wholesale transaction internalizes the downstream search costs and distributes it between middlemen and producers. If a middleman has complete bargaining power in wholesale transactions, then the cost associated with the holdup problem is completely borne by the producer.

When credit is not feasible in retail transactions, the strategic interactions between consumers’ portfolio choice and middlemens’ inventory holdings generate a two-sided holdup problem. One consequence is that there exist multiple steady state equilibria. Moreover, there exists a non-monotone relationship between the quantity traded and the degree of entry. This is counter to a monetary economy without intermediation. In short, even when consumers choose to hold very large real balances, their consumption opportunities are still constrained by middlemens’ inventory holdings. When there are few middlemen, entry incentivizes consumers to hold more real balances and results in more trade. When there are many middlemen, congestion effects cause middlemen to purchase fewer inventories which constrains consumption opportunities and results in less trade.

As a consequence of this two-sided holdup problem, there exist two regimes that dictate the response of the equilibrium to changes in fundamentals. Which regime the economy is in depends on the
bargaining power of middlemen in the wholesale market and the cost of entry. Low wholesale bargaining power or high entry costs results in a regime characterized by few middlemen and hence consumer’s liquidity constrains the allocation. High wholesale bargaining power or low entry costs results in a regime with many middlemen and hence inventory constrains the solution. If the economy is in the former regime (liquidity constrained) then monetary policy behaves as usual: lower nominal rates increase consumers’ real balances resulting in more consumption and entry. If, however, the economy is in the latter regime (inventory constrained), monetary policy is ineffective at some nominal interest rate $\bar{\hat{r}} > 0$: reducing the nominal interest rate does not affect real balances because consumers anticipate that they will not be able to use them given inventory is constrained. Monetary policy has no effect on the quantity of trade or the entry of middlemen.

The model describes the environment in which middlemen operate, the different payment instruments available to agents, and the resulting equilibria. Section 4 discusses a pure credit economy as a benchmark case to highlight the inherent frictions in the environment. Section 5 considers the case of strategic default and endogenous debt limits. Section 6 relaxes the notion of perfect commitment in retail trades and therefore provides an endogenous need for consumers to hold money. Section 7 introduces the need for money in wholesale trades and shows that multiplicity and non-monotonicities vanish when money is inessential in retail trades. Section 8 concludes and discusses how the framework can contribute an intermediation theory of the firm.

2 Motivating Data

Apart from being anecdotally ubiquitous, intermediation activities constitute a non-trivial share of the U.S. economy. As a rough estimate, shares of GDP attributed to intermediation include retail trade (5.9%), wholesale trade (5.9%), finance and insurance (7.3%), transportation and warehousing (3%). This estimate, a conservative one at that since it assumes zero value-added attributed to intermediation for all other industries, suggests intermediation activities account for approximately 22.1% of 2016 U.S. GDP. Moreover, disintermediation has not occurred. Figure 1 shows that my rough estimate of intermediation have been relatively stable, just shy of one-quarter of the U.S.
Middlemen must exist because the benefits from intermediated trade are greater than those from direct trade. One explanation is that direct trade is very costly. Evidence by McGraw Hill (2013) estimates it takes on average 4.3 phone calls before a manufacturer finds a consumer at a total cost of just over $589. Although producers often pine for “cutting out the middleman,” too often they forget that they still have to provide their function. If a producer bypasses intermediaries, it must incur the costs associated with distributing the good to end consumers. A survey of 200 local food producers in California by Brimlow (2016) found that 72 percent were selling out of the area to wholesale brokers. Producers surveyed claimed it was difficult to find local buyers or that marketing and advertising costs to sell locally would be too high. Middlemen provided a competitive advantage in delivering goods to final consumers.

Many intermediation activities exist but are difficult to discern in aggregate data. Manufacturing firms, for example, spend considerable resources on marketing, transportation, inventory management and other activities that are conventionally described as intermediation. To give some perspective, the proportion of manufacturing employees classified as non-production averaged 29 percent over the past two decades with minor variation. Although not all of these non-production employees engage in intermediation activities like sales, inventory management, warehousing, logistics etc., surely some do.
The benefits of intermediated trade are amplified when goods must be transported internationally. Intermediation has been a major driver of globalization bringing goods and services from local producers to international consumers. For example, in the early 1980s only 300 Japanese trading firms accounted for 80 percent of total Japanese trade and the ten largest of these firms were responsible for 30 percent of Japan’s GNP (Rossman, 1998). Hong Kong and Singapore are frequently cited examples of entrepot economies where intermediation activities account for a sizable portion of GDP growth. Feenstra (2003) finds that in 1998, total trade was 259 percent of GDP in Hong Kong and 269 percent in Singapore.

The process by which middlemen match with producers and consumers is itself a productive process that uses resources. One of the most closely watched metrics to gauge retail performance is average inventory turnover; the reciprocal of which is days inventory outstanding (DOI). In a frictionless environment, middlemen could stock and sell inventories instantaneously and the DOI would be zero. However, in reality it takes time to clear goods markets resulting in varying DOIs across industries and firms. There is a large degree of heterogeneity both across and within industries, but the average DIO in 2015 was 87.4 days. A model of intermediation in goods markets should take seriously the frictions that preclude the instantaneous purchase and resale of inventory.

Although technology has not led to disintermediation, as many tech guru prophets predicted, it has revolutionized retail payment technology. The means of executing quid pro trades using currency or digital substitutes is experiencing a rapid evolution and it is important to understand how the role of various retail payment instruments affect merchant behavior and vice versa. The 2015 Survey of Consumer Payment Choice (SCPC) finds that while there are nine identified payment instruments, consumers still predominantly use debit cards (32.5 percent of monthly payments), cash (27.1 percent), and credit cards (21.3 percent).

3 Literature Review

The study of middlemen and how their presence influences market allocations dates to Rubenstein and Wolinsky (RW) who advocated “a basic model that describes explicitly the trade frictions that...
give rise to the function of middlemen.” RW modeled the exchange process and trading frictions between sellers, middlemen, and buyers thereby providing a framework for endogenizing the extent of intermediation and its effect on the distribution of gains from trade. Subsequent models, such as those by Nosal, Wong, and Wright (2011, 2014, 2016) (NWW), expand on RW to include production, search costs, Nash bargaining, and occupational choice. The present paper seeks to contribute the study of middlemen in the spirit of RW and NWW.

There are various explanations for why middlemen are valuable. RW suggest that intermediation is a way of alleviating search frictions by providing more frequent consumption opportunities for consumers. Although, in RW the meeting rates are taken as given so that middlemen are simply agents with exogenously endowed technological advantage. Given some exogenous meeting process, some papers focus on the role of middlemen as guarantors of quality (Biglaiser 1993, Li 1998) while others suggest that middlemen are able to satisfy consumers’ demand for varieties of goods whereas individual producers cannot (Johri and Leach, 2002) (Shevchenko, 2004). Watanabe (2010) argues that middlemen have the advantage of inventory capacity relative to producers and that capacity constraints are an important determinant for the endogenous meeting rates.

Lotz and Zhang consider payment instruments used by consumers. Rocheteau and Zhang consider payment instruments used by firms.

4 Middlemen in a Pure Credit Economy

4.1 Environment

Time is discrete and continues forever. Each period is divided into three subperiods where different transactions take place. During the first two subperiods trades occur in decentralized markets (DM) where agents’ trading opportunities arrive according to a random bilateral matching process. The arrival rate of a trading opportunity for an agent of type $i$ with an agent of type $j$ will be denoted $\alpha_{ij}$. That is, $\alpha_{ij}$ captures the severity of trading frictions present in the decentralized market. There exists a unique perishable retail good traded in the DMs denoted by $q$. During the
third subperiod all agents meet in a centralized market (CM) where they produce, consume and exchange the numeraire good, denoted \( x \), without impediments to trade.

There are three types of agents: producers (P), consumers (C), and middlemen (M). Each type of agent is completely characterized by idiosyncratic preferences and technology. Producers have no desire to consume in the DM but are able to produce the retail good. Consumers desire the retail good in the DM, but are unable to produce it. Middlemen do not desire the retail good, but are able to purchase it from producers in the first DM and resell it to consumers in the second DM. Any unsold inventory can be transformed at rate \( R \) into the numeraire during the CM.

In the first subperiod, there is a wholesale market where middlemen can meet producers and acquire the retail good. Matches occur randomly according the arrival rates \( \alpha_{pm}, \alpha_{mp} \) such that a subset \( \tilde{P} \subset P \) of producers are matched with \( \tilde{M}_w \subset M \) middlemen. When a producer and a middleman meet they agree on terms of trade \((q^w, b^w)\) where \( q^w \) denotes the amount a producer sells to a middleman and \( b^w \) the amount of debt that a middleman incurs in the exchange. In the second superiod, there is a retail market where consumers purchase the retail good. Matches arrive according to \( \alpha_{cm}, \alpha_{mc} \) such that a subset \( \tilde{C} \subset C \) of consumers are matched with \( \tilde{M}_r \subset M \) middlemen. When a middleman and consumer meet they agree on terms of trade \((q^r, b^r)\) where \( q^r \) denotes the quantity of the good transferred to the consumer and \( b^r \) the amount of debt issued by the consumer. In the centralized market, middlemen can transform the retail good into the numeraire at rate \( R \) and all debts are settled by producing and transferring the numeraire good.

The above specification assumes that credit is a feasible means of payment. This assumption can be justified by positing the existence of a record keeping technology and enforcement mechanism which gives agents perfect memory (Kocherlakota 1998). The pure credit economy is a useful benchmark for analyzing the efficiency properties of middlemen and how terms of trade in the retail and wholesale market are linked. I will also consider the case where credit is infeasible, either in the retail or wholesale market, which necessitates agents to make a decision to accumulate liquid assets. This portfolio problem has implications for the equilibrium allocation and allows for a discussion of monetary policy and how the transmission of monetary policy is affected by the presence of an intermediated sector.
The period utility function of a consumer, producer, and middleman are given by,

\[ U^C(q, x) = u(q) + x \]
\[ U^P(q, x) = -c(q) + x \]
\[ U^M(x) = x \]

In the DM, consumers derive utility \( u(q) \) from the retail good and producers incur cost \( c(q) \) to produce the retail good. In the CM all agents enjoy linear utility in the numeriare good, where \( x < 0 \) is interpreted as the disutility of producing the numeraire. Middlemen derive utility only from the numeraire (and receive no utility from producing it for themselves since CM production costs are linear). Middlemen must purchase the retail good from producers in the wholesale market and resell it to consumers in the retail market. All goods are nonstorable between time periods. All agents discount future utility by a factor \( \beta \in (0, 1) \).

**ASSUMPTION 1.** Utility \( u(\cdot) \) and costs \( c(\cdot) \) are \( C^2 \) functions defined on \( \mathbb{R}_+ \) and obey the usual properties: \( u' > 0, u'' < 0, u(0) = 0, u'(0) = \infty, c' > 0, c'' > 0, c(0) = 0, c'(0) = 0 \). Additionally, \( \hat{q} < \tilde{q} \) where \( u'(\hat{q}) = c'(\hat{q}) \) and \( u'(\tilde{q}) = R \).

Differentiating types ex-ante makes it simple to introduce an extensive margin of trade. In particular, a subset of middlemen with measure \( n_t \) enter the wholesale market each period \( t \) at cost \( k \). Assuming equal measure of agents on either side of a market, and normalizing the measure of buyers and sellers to one, I have that

\[ \mu(n) = \alpha_{cm}(n) = n\alpha_{mc}(n) \]
\[ \gamma(n) = \alpha_{pm}(n) = n\alpha_{mp}(n) \]

This specification allows search externalities in both the wholesale and retail markets where trading opportunities depend on the ratio of middlemen to sellers and buyers.\(^1\)

\(^1\)Implicit in the description of the environment is that sellers are never matched directly with consumers. This can be interpreted as extreme matching frictions such that \( \alpha_{cp} = \alpha_{pc} = 0 \). In this sense, middlemen are essential in order
ASSUMPTION 2. Matching functions are homogeneous of degree one and exhibit standard properties: \( \mu'(n) > 0, \mu''(n) < 0, \mu(n) \leq \min(1, n), \mu(0) = 0, \mu'(0) = 1, \mu(\infty) = 1 \) and identical conditions on \( \gamma(\cdot) \).

4.2 Planner’s Problem

I consider the problem of a social planner who each period chooses the measure \( n_t \) of active middlemen and an allocation \( \{q_t^r(i), q_t^w(i)\} \) for all matched agents, \( i \in \tilde{P} \cup \tilde{M}_w \cup \tilde{M}_r \cup \tilde{C} \), and \( \{h_t(i)\} \) for all agents. The planner is constrained by the environment in the sense that he cannot choose the set of matched agents but only \( n_t \), and then the sets are determined randomly in accordance with the matching technology. If the planner treats all agents identically, and confining attention to stationary allocations, the relevant period welfare function is given by,

\[
W_t = (2 + n)x + (\gamma(n)\mu(n)/n)u(q^r) - \gamma(n)c(q^w).
\]

The first term is net consumption in the CM enjoyed by all agents, the second term is the utility of a unit measure of consumers in the retail market who find a middleman holding inventory, and the third term is the cost incurred by a unit measure of producers who find a middleman. The planner wishes to maximize \( \sum_{t=0}^{\infty} \beta^t W_t \) subject to the following feasibility constraints,

\[
(2 + n)x \leq (\gamma(n)\mu(n)/n)R(q^w - q^r) + \gamma(n)(1 - \mu(n)/n)Rq^w \\
q^r \leq q^w
\]

The first constraint states that net consumption of the numeraire can be no greater than unsold inventory. The second constraint requires that an individual buyer can never purchase more than a middleman carries in inventory.

PROPOSITION 1. The constrained efficient allocation \( (q^*, n^*) \in \mathbb{R}_+ \) solves the planner’s problem to alleviate the extreme frictions placed on trade. Although stark, the purpose of this paper is to examine allocations in an environment with essential middlemen rather than derive the endogenous emergence of an intermediated sector.
lem and is given by,

\[ q^* = q^f = q^w \]

(1)

\[(\mu(n^*)/n^*)(u'(q^f) - R) = c'(q^w) - R \]

(2)

\[ k = (\gamma(n^*)\mu(n^*)/n^*)(u(q^f) - Rq^f) + \gamma'(n^*)(Rq^w - c(q^w)) \]

(3)

**Proof.** Maximizing \( W_t \) at each date, first order conditions for the planner’s problem reveal that there are two potential solutions: one where the feasibility constraint \( q^f \leq q^w \) binds and one where it does not. Assumption 1 rules out the non-binding case. \( \square \)

Having described the efficient allocation, I now turn to the decentralized pure credit economy and characterize equilibria. For now, I maintain that agents are able to fully commit to pay for current trades at a future date. Although stark, this assumption provides a benchmark case against which I can measure the equilibrium outcomes of economies wherein agents are unable to fully commit and therefore require some liquid means of payment. I return to this issue in Section 4 and Section 5.

### 4.3 Centralized Market

A consumer enters the CM with some amount of debt \( b \) incurred in the retail market and optimally chooses production of the numeraire in order finance consumption and repay debt. He then enters the following period’s DM which yields expected utility \( V^C \).

\[ W^C(b) = x + \beta V^C \]

s.t. \( x = -b \)

A middleman enters the CM with net wealth (credit from consumers less debt owed to producers) \( b \) and unsold inventory \( q \). He finances his consumption of the numeraire using net wealth, transforming unsold inventory at rate \( R \), and producing the numeraire at unit cost. He then chooses
whether or not to enter the following period’s DM with expected utility $V^M_1$.

$$W^M(q, b) = x + \beta \max\{V^M_1, W^M\} \quad s.t. \quad x = b + Rq$$

A producer enters the CM with credit owed by middlemen which it uses to finance its consumption of the numeraire.

$$W^P(b) = x + \beta V^P \quad s.t. \quad x = b$$

Substituting the budget constraints into their respective objective functions, the CM value functions for agents are given by the following:\(^2\)

$$W^C(b) = -b + \beta V^C$$
$$W^M(q, b) = b + Rq + \beta \max\{V^M_1, W^M\}$$
$$W^P(b) = b + \beta V^P$$

### 4.4 Decentralized Markets

Having characterized the CM value functions, I move back one sub-period to the retail market where consumers and middlemen are randomly matched. A consumer entering the retail market finds a middleman with probability $\mu(n)$ upon which terms of trade $(q^r, b^r)$ are executed and then enters the CM.

$$V^C = \mu(n)[u(q^r) - b^r] + \beta V^C$$

A middleman enters the retail market with some amount of inventory purchased from a producer and the corresponding debt $(q, b)$. With probability $\mu(n)/n$ he finds a consumer upon which terms of trade $(q^r, b^r)$ are executed and then enters the CM. If the middlemen does not find a consumer, he carries all unsold inventory and debt into the CM.

---

\(^2\)Notice that all agents’ CM value function are linear in net wealth. When agents choose to acquire liquid assets, the portfolio decisions are history independent so that there is a degenerate distribution of asset holdings. This result is an artifact of quasi-linear preferences and delivers tractable results without sacrificing economic insight.
\[ V_2^M(q, b) = (\mu(n)/n)W^M(q - q^r(q), -b + b^r(q)) + (1 - \mu(n)/n)W^M(q, b) \]

Notice that terms of trade in the retail market depend on what occurred in the wholesale market. If a middleman did not purchase any inventory in the wholesale market \((q = 0)\) then he surely cannot sell anything in the retail market. More generally, the amount of inventory \(q\) constrains the set of feasible allocations in the retail market. Using CM value function (5) I have that,

\[ V_2^M(q, b) = (\mu(n)/n)(-Rq^r(q) + b^r(q)) + Rq - b + \beta \max\{V_1^M, W^M\} \]  

\[ (8) \]

A middleman’s expected value in the retail market is the probability he finds a consumer times the value of the match, plus the guaranteed value of transforming unsold inventory in the CM, plus the continuation value of entering next period’s wholesale market.

I now move back one sub-period to the wholesale market where middlemen purchase inventory from producers. When a middleman enters the wholesale market he incurs entry cost \(k\), meets a producer with probability \(\gamma(n)/n\), executes terms of trade \((q^w, b^w)\) and then enters the retail market. Otherwise he enters the retail market with zero inventory and zero debt.

\[ V_1^M = (\gamma(n)/n)V_2^M(q^w, -b^w) + (1 - \gamma(n)/n)V_2^M(0, 0) - k \]

Using the retail value function (8) I have that,

\[ V_1^M = (\gamma(n)/n)((\mu(n)/n)[-Rq^r(q^w) + b^r(q^w)] + Rq^w - b^w) - k + \beta \max\{V_1^M, W^M\}. \]  

\[ (9) \]

A middleman’s expected value in the wholesale market is the expected value of acquiring inventory \(q^w\) with probability \(\gamma(n)/n\). The value of holding inventory includes the guaranteed value of transforming it at rate \(R\) in the CM and repaying debts plus the value of carrying inventory into the retail market. Of course, the value of inventory in the retail market depends on the terms of trade \((q^r, b^r)\). Suppose, for example, that the terms of trade are such that the consumer receives
the entire value of surplus from a retail match. In this case, a middleman would only receive utility
from transforming inventory in the CM and would never choose to operate in the retail market.
This is an uninteresting equilibrium. In order to generate an equilibrium where consumers have
an opportunity to consume and middlemen to actually behave as intermediaries (buy and resell)
it will be necessary to implement a bargaining protocol that gives the middleman some bargaining
power in the retail market.

A producer entering the wholesale market finds a middleman with probability \( \gamma(n) \), produces the
good at cost \( c(q) \), and receives credit to be settled in the CM.

\[
V^P = \gamma(n) (-c(q^w) + b^w) + \beta V^P
\] (10)

4.5 Bargaining Sets

In this section I characterize the set of allocations that are incentive feasible—the terms of trade
which satisfy agents’ participation constraints. First, I characterize the bargaining set that exists
between a middleman and a consumer in the retail market. If an agreement is reached, a consumer’s
utility level is \( u^C = u(q^r) + W^C(-b^r) \) and a middleman’s utility level is \( u^M = W^M(q - q^r, -b + b^r) \).
If there is no agreement then a consumer receives utility \( u^C_0 = W^C(0) \) and a middleman receives
utility \( u^M_0 = W^M(q, -b) \). Using the CM value functions we can write the value of the surplus from
a match as follows:

\[
\begin{align*}
    u^C - u^C_0 &= u(q^r) - b^r \\
    u^M - u^M_0 &= b^r - R(q^r)
\end{align*}
\]

A proposed trade is incentive feasible only if both agents earn non-negative surpluses from the
agreement. The set of incentive feasible allocations is defined as \( \Omega_r = \{(q^r, b^r) : R(q^r) \leq b^r \leq u(q^r, q^r \leq q^w) \} \) and may be constrained by the amount of inventory a middleman carries into a
retail match. In particular, a middleman can never sell more of the consumption good than it has
in inventory. Formally, we can define the Pareto frontier of bargaining set as follows,

\[
\max_{q^r, b^r} u^C = u(q^r) - b^r + u^C_0 \\
\text{s.t. } b^r - Rq^r + u^M_0 \geq u^M \\
\text{s.t. } 0 \leq q^r \leq q^w
\]

The jointly efficient outcome is \(u'(\tilde{q}) - R = 0\) and \(\tilde{b} = R\tilde{q} + u^M - u^M_0\). If a middleman is holding sufficient inventory \(q^w \geq \tilde{q}\), then the jointly efficient amount will be purchased. If, however, a middleman holds insufficient inventory \(q^w < \tilde{q}\) then all inventory will be sold and debt will be issued to compensate the middleman for the opportunity cost of selling the good and provide some surplus \(b^r = Rq + u^M - u^M_0\). The equation for the frontier is given by,

\[
u^M - u^M_0 = u(q^*) - Rq^* - (u^C - u^C_0) \quad \text{if } q^w \geq \tilde{q} \\
= u(q^w) - (u^C - u^C_0) - Rq^w \quad \text{if } q^w < \tilde{q}
\]

Notice that \(\partial^2 u^M / \partial (u^C)^2 = 0\) for all \(q\) so that the frontier is always linear. However, the bargaining set monotonically shrinks as \(q^w\) declines.

I now characterize the bargaining set that exists between a middleman and a producer in the
wholesale market. If an agreement is reached, then a middleman receives utility $u^M = V^M_2(q^w, b^w)$ and the producer receives $u^P = -c(q^w) + W^P(b^w)$. If there is no agreement then the middleman gets $u^M_0 = V^M_2(0, 0)$ and the producer gets $u^P_0 = W^P(0)$. Using the CM value functions we can write the value of the surplus from a match as follows:

$$u^M - u^M_0 = \pi(q^w) - b^w \equiv \frac{\mu(n)}{n}[-Rq^r(q^w) + b^r(q^w)] + Rq^w - b^w$$

$$u^P - u^P_0 = -c(q^w) + b^w$$

The set of incentive feasible allocations is thus defined as $\Omega_w = \{(q^w, b^w) : c(q^w) \leq b^w \leq \pi(q^w)\}$ and depends on the division of surplus in the retail market. In particular, the size of the bargaining set is increasing in the amount of surplus a middleman receives in a retail match. Once the amount of inventory exceeds the first best $q^w > \tilde{q}$ the marginal benefit of inventory is simply its value in the CM. That is, the upper contour of the bargaining set is linear with slope $R$ for all $q^w > \tilde{q}$.

Also note that greater entry induces a congestion effect in the retail market shrinking the set of incentive feasible trades. As $n \to \infty$ the feasible trades must satisfy $c(q^w) \leq b^w \leq Rq^w$ indicating that middlemen only realize value from transforming inventory into the numeraire. Alternatively, as $-Rq^r(q^w) + b^r(q^w) \to 0$ we again have that $c(q^w) \leq b^w \leq Rq^w$. Whether the efficient quantity traded is incentive feasible $q^* \in \Omega_w$ depends on the share of retail trade surplus a middlemen receives. Of course, if $c(q^*) \leq Rq^*$ then the efficient quantity is incentive feasible even when the middleman’s share of retail surplus approaches zero; although this is not true in general and largely depends on the rate $R$ at which a middleman can transform unsold inventory.

The Pareto frontier of the bargaining set is defined by the following:

$$\max_{q^w, b^w} u^M = \pi(q^w) - b^w + u^M_0$$

$$s.t. \quad -c(q^w) + b^w + u^P_0 \geq u^P$$
The jointly efficient outcome is given by the solution to the following,

\[
\left(\frac{\mu(n)}{n}\right) \frac{\partial(-Rq^w(q^w) + b^r(q^w))}{\partial q^w} + R = c'(q^w)
\]

\[
b^w = u^P - u^P_0 + c(q^w)
\]

The jointly efficient allocation equates the marginal benefit to a middleman of acquiring inventory to the cost of producing that inventory. The marginal benefit to a middleman is the surplus received in the retail market with probability \(\mu(n)/n\) plus the ability to transform any unsold inventory at rate \(R\) with probability one. With the optimal quantity determined, debt is issued by the middleman to compensate the producer for its cost of production and provide some surplus.

Notice that the jointly efficient allocation in the wholesale market depends on how the surplus from the retail market is split. Suppose, for example, that middlemen received zero share of the surplus in the retail market. Then the jointly efficient quantity of the consumption good would reduce to \(R = c'(q^w)\). Of course, this corresponds to middlemen choosing not to enter the retail market and simply transforming inventory into the numeraire. Any equilibrium with active middlemen in the retail market requires a non-zero share of retail surplus for middleman and this value will be taken into consideration when purchasing inventory in the wholesale market. The Pareto frontier is linear.
and strictly decreasing in the share of surplus received by middlemen in the retail market,

\[ u^M - u_0^M = (\mu(n)/n)S_r + Rq^w - c(q^w) - (u^P - u_0^P) \]

### 4.6 Entry by Middlemen

A middleman participates in the retail market if

\[ V_2^M(q^w, -b^w) \geq W^M(q^w, -b^w). \]

This is equivalent to \(-Rq^r + b^r \geq 0\) which simply says that a middleman must receive a non-negative surplus from a match with a consumer. The decision to search in the retail market is equivalent to the participation constraint in any retail match.

A middleman chooses to search in the wholesale market if the value of doing so is at least a great as the cost of entry, \(V_1^M \geq 0\). Using the value functions this is equivalent to the following,

\[ (\gamma(n)/n)(\mu(n)/n)[-Rq^r(q^w) + b^r(q^w)] + Rq^w - b^w) \geq k. \]

The value of searching in the wholesale market is the value of selling inventory in the retail market with probability \((\mu(n)/n)\) plus the value of transforming inventory in the centralized market with probability one. Notice that it may be profitable for a middleman to acquire inventory in the wholesale market even if it expects no surplus from the retail market. This occurs when \((\gamma(n)/n)(Rq^w - b^w) \geq k\) – inventory holding costs must be sufficiently low to induce entry in the wholesale market if there is no surplus from the retail market. Any surplus from retail trades relaxes the condition for entry.

### 4.7 Equilibrium

An equilibrium is defined as \(\{q^w, q^r, b^w, b^r, n\}\) where a bargaining protocol determines the terms of trade and free entry determines the measure of active middlemen. As discussed earlier, the terms of trade settled in the retail market affect the terms of trade in the wholesale market. In particular, the middleman must receive a non-zero share of the retail surplus in order to incentivize them
to participate in retail trades. For this reason, certain bargaining protocols such as a take-it-or-
leave-it offer by the consumer yield equilibrium with no intermediation. I restrict my attention to
stationary equilibria with active middlemen and adopt bargaining protocols appropriately.

I begin with the retail market where a consumer meets a middleman. The terms of trade will be
determined by Kalai proportional bargaining,

\[
\max_{q^r, b^r} u(q^r) - b^r \\
\text{s.t. } u(q^r) - b^r = \frac{\theta_r}{1 - \theta_r} (-Rq^r + b^r) \\
\text{s.t. } q^r \leq q^w
\]

where \( \theta_r \) denotes the bargaining power of a consumer. The unconstrained solution obtains where
the marginal benefit of consumption equals the marginal opportunity cost of the sale \( u'(\tilde{q}) = R \)
and the corresponding transfer is \( \tilde{b} = \theta_r R\tilde{q} + (1 - \theta_r)u(\tilde{q}) \). If a middleman holds too little inventory
\( q^w < \tilde{q} \) however, then the solution is constrained such that a consumer purchases all inventory and
issues the corresponding amount of debt,

\[
q^r = q^w \tag{11} \\
b^r = \theta_r Rq^r + (1 - \theta_r)u(q^r) \tag{12}
\]

Under proportional bargaining, the surplus received by either agent monotonically increases as the
bargaining set expands. An extra unit of inventory held by a middleman (relaxing the constraint)
generates extra surplus up to the jointly efficient allocation. That is, \( \partial S_r / \partial q^w = u'(q^w) - R \) if
\( q^w < \tilde{q} \) and is zero otherwise. The terms of trade in the wholesale market, where a middleman
purchases inventory from a producer, are settled according to,

\[
\max_{q^w, b^w} (\mu(n)/n)[-Rq^r(q^w) + b^r(q^w)] + Rq^w - b^w \\
\text{s.t. } (\mu(n)/n)[-Rq^r(q^w) + b^r(q^w)] + Rq^w - b^w = \frac{\theta_w}{1 - \theta_w} (-c(q^w) + b^w)
\]
where $\theta_w$ denotes the bargaining power of a middleman. The solution to the above program is

$$
(\mu(n)/n)(1 - \theta_r)(u'(q^w) - R)^+ + R = c'(q^w) \quad (13)
$$

$$
b^w = (1 - \theta_w)((\mu(n)/n)(1 - \theta_r)(u(q^w) - Rq^w) + Rq^w) + \theta_w c(q^w) \quad (14)
$$

where $x^+ = \max\{0, x\}$. The marginal benefit from an extra unit of inventory is the expected surplus it brings to a retail match plus its guaranteed recycle value. The optimal amount of inventory equates this marginal benefit to the marginal cost borne by a producer.

**PROPOSITION 2.** For a given level of entry $n$, the amount of inventory purchased in the decentralized equilibrium with credit is less than the jointly efficient quantity in retail trades, $q^w < \tilde{q}$ and less than the first-best allocation, $q^w < q^*$, for all $\theta_r > 0$.

The first result is due to a hold-up problem: since a middleman is required to purchase inventory prior to meeting a consumer, this cost is sunk during bargaining in the retail market. Consequently, a middleman will never purchase enough inventory to satiate consumer demand. The second result follows because a middleman does not receive the full value of his investment in inventory in the subsequent retail market so long as $\theta_r > 0$. A graphical description is provided in Figure 6.

Middlemen endogenously choose to participate in the wholesale market at cost $k$. Free entry implies...
\[ V_1^M = W^M(0) = 0. \] Using (9) I have that,

\[
\frac{\gamma(n)}{n} \theta_w \left( \frac{\mu(n)}{n} (1 - \theta_r) (u(q^w) - Rq^w) + Rq^w - c(q^w) \right) = k \tag{15}
\]

Given the quantity traded in the wholesale market, the measure of middlemen \( n \) will adjust such that the value of entering the wholesale market and the value of not entering are equated to zero. Notice that a necessary assumption to guarantee \( n > 0 \) is that \( \theta_w ((1 - \theta_r) (u(q^w) - Rq^w) + Rq^w - c(q^w)) \geq k \) where \( q^w \) is determined by (13). It is clear that this requires \( \theta_w > 0 \) and the constraint is most slack when \( \theta_r = 0 \).

An equilibrium jointly describes the terms of trade and measure of operative middlemen using (11)-(15). Notice that there exist complementarities between the extensive and intensive margins. The probability of a trading opportunity depends on the measure of active middlemen which in turn affects the quantity of inventory purchased in the wholesale market. Observing (13), it is clear that greater entry induces lower expected utility in the retail market which exacerbates under-investment and hence reduces retail consumption.

An equilibrium can be summarized by equations (13) and (15) which jointly determine \((q^w, n)\). By (13) there is an inverse relationship between the measure of active middlemen and the quantity of retail consumption. Higher entry causes a congestion effect which lowers the expected surplus in the retail market and leads to under-investment of inventory. Define \( \bar{q} \) such that \((1 - \theta_r)(u'(q) - R) = c'(q) - R\) which corresponds to the case where \( n \to 0 \). As the probability of finding a trading partner in the retail market approaches one the quantity traded increases up to \( \bar{q} \) which is still less than the first best because middlemen do not realize the full return of their investment in inventory as long as \( \theta_r > 0 \). Define \( q \) such that \( c'(q) = R \) which corresponds to the case where \( n \to \infty \). As entry becomes unbounded the probability of finding a match approaches zero so that a middleman only realizes return on inventory by transforming it into the numeraire. Free entry condition (15) shows a positive relationship between entry and the quantity of the consumption good traded. To induce entry, a middleman must be compensated for the lower probability of a match with a greater quantity traded per match. Equilibrium \((q^w, n)\) occurs at the intersection of (13) and (15). The
quantity of debt issued is then determined by (12) and (14) where it is clear from (11) that $q^r = q^w$.

**PROPOSITION 3.** The decentralized equilibrium can never reach the socially efficient allocation.

Proof. The first best quantity traded obtains only where $\theta_r = 0$ so that middlemen realize the full return on their inventory. However, optimal entry occurs for a version of the Hosios condition where $\theta_r = -\left(\frac{\gamma(n)}{\gamma'(n)}\right)\left(\frac{\mu(n)/n}{\mu(n)/n'}\right)$ and $\theta_w = n\gamma'(n)/\gamma(n)$ obtained from comparing (15) to (3). These two optimality conditions are contradictory therefore the socially efficient $(q, n)$ cannot be reached. □

**PROPOSITION 4.** Decreasing the bargaining power of middlemen in the retail market results in less entry and less trade, $\partial n/\partial \theta_r < 0, \partial q/\partial \theta_r < 0$. Increasing the bargaining power of middlemen in the wholesale market results in more entry but less trade, $\partial n/\partial \theta_w > 0, \partial q/\partial \theta_w < 0$. Increasing the cost of entry results in less entry but more trade, $\partial n/\partial k < 0, \partial q/\partial k > 0$. A decrease in inventory holding costs increases entry and has an ambiguous effect on quantity traded, $\partial n/\partial R > 0$.

If the extensive margin were shut down (exogenous $n$) then altering the bargaining power in the wholesale market would have no effect on the quantity traded; it would simply adjust the division of surplus between producers and middlemen. However, when entry is endogenous middlemen
internalize the higher share of expected surplus resulting in more entrants. More entrants decrease the probability of a retail match which incentivizes middlemen to purchase less inventory and thus engage in less trade. Also interesting is that higher entry costs can actually increase the quantity traded. Entry costs have no effect on the division of surplus as can be seen from (13); however, they do affect ex-ante profits seen from (15). To compensate for the lower probability of a retail match due to fewer middlemen, a greater quantity must be traded in equilibrium. The effect on entry from a decrease in on inventory holding costs is ambiguous. Lower inventory costs improves a middleman’s outside option resulting in more trade in retail matches; (13) shifts northeast. Concurrently, an improved outside option encourages more entry which congests the wholesale market and decreases inventory purchases; (15) shifts northwest.

**PROPOSITION 5.** The intermediation spread for middlemen is strictly positive and given by,

\[
(b^r - b^w)(q) = (u(q) - Rq)(1 - \theta_r - (1 - \theta_r)(1 - \theta_w)\mu(n)/n) + \theta_w(Rq - c(q)).
\]

The spread is increasing in the bargaining power of middlemen \(\partial(b^r - b^w)/\partial\theta_w > 0\), \(\partial(b^r - b^w)/\partial\theta_r < 0\), the amount traded \(\partial(b^r - b^w)/\partial q > 0\), inventory holding costs \(\partial(b^r - b^w)/\partial R > 0\) and the measure of middlemen \(\partial(b^r - b^w)/\partial n > 0\).

Of interest is the proportion of retail surplus captured by a middleman: \((1 - \theta_r - (1 - \theta_r)(1 - \theta_w)\mu(n)/n)\). The first term captures the primitive bargaining power of middlemen in retail trade, whereas the second term reveals the interaction between wholesale and retail trade. Suppose, for example, that a middleman has all bargaining power in wholesale trades \(\theta_w = 1\). In this case, the producer is forced to internalize not only its own production costs, but also the search costs associated with the retail market. That is, the wholesale transaction internalizes the downstream search costs and distributes it between middlemen and producers. This mechanism underlies the intuition for why \(\partial(b^r - b^w)/\partial n > 0\). More entry decreases the expected value of a retail match, and therefore requires a smaller payment in wholesale trades. Concurrently, more entry does not effect the terms of trade in retail matches. A middleman can extract up to the full surplus \(u(q) - c(q)\) when \(\theta_r = 0, \theta_w = 1\).
5 Limited Commitment

Thus far I have assumed that credit is perfect. That is, there exists a record keeping technology and enforcement mechanism that replicates perfect memory and ensures debt repayment. Now suppose that such an enforcement technology does not exist, and so repayment of debt must be self-enforcing. Buyers will be allowed the possibility of strategic default, but understand that their actions are publicly recorded and punishment for default is exclusion from all future credit trades.

I begin with the retail market, and denote $\bar{b}^r$ the consumer’s debt limit which is the maximum amount that a buyer is willing to repay. The consumer will have an incentive to repay his debt in the CM if and only if $-b^r + \beta V^C \geq 0$. The sum of the buyer’s current and continuation payoffs if he repays his debt must be greater than the continuation (autarkic) payoff of zero if he defaults.

The debt limit is thus defined as,

$$\bar{b}^r = \frac{\mu(n)}{\mu(n) + r} u(q^r)$$  \hspace{1cm} (16)

and the set of incentive feasible allocations in the retail market is given by $\Omega^{lc}_r = \{(q^r, b^r) : Rq^r(q^w) \leq b^r \leq \bar{b}^r\}$.\footnote{There exist a continuum of stationary credit equilibria indexed by debt limits $\bar{b} < \bar{b}^r$ supported by self-fulfilling beliefs. I restrict my attention to the “not-too-tight” borrowing constraint which is sufficiently tight to prevent default but not too tight so as to leave unexploited gains from trade.} Compared to full commitment, the set of feasible trades is strictly smaller. Moreover, the debt limit is increasing with the measure of active middlemen. More middlemen increase the frequency of trading opportunities which makes having access to credit more valuable.

Whether or not the jointly efficient quantity lies in the incentive feasible set, $q^r \in \Omega^{lc}_r$, depends on $(q^w, n, \beta)$. That is, for any given discount factor there exists a threshold level of entry such that if there are too few middlemen then the jointly efficient quantity is not incentive feasible. The intuition is that if there are too few middlemen in the retail market, then exclusion from future retail trades is not punishing enough to induce debt repayment.

I now consider the wholesale market, and denote $\bar{b}^w$ the middleman’s debt limit defined by $-\bar{b}^w + \beta V^M_1 = 0$, or written explicitly,

$$\bar{b}^w = \frac{(\gamma(n)/n)(\mu(n)/n)(-Rq^r + b^r) + (\gamma(n)/n)Rq^w - k}{r + \gamma(n)/n}$$  \hspace{1cm} (17)
and the set of incentive feasible allocations in the wholesale market is given by \( \Omega_{lc}^w = \{ \bar{b}^w \leq b^w \leq c(q^w) \} \). Whether or not the efficient amount of inventory is incentive compatible depends on \((n, \beta, k)\). The debt limit is decreasing in the measure of operative middlemen \(n\), increasing in their patience \(\beta\), and decreasing in the cost of entry \(k\). For a given \((\beta, k)\) there exists a threshold level \(\bar{n}_w\) such that for all \(n \geq \bar{n}_w\) there is no \(b^w\) that can support the efficient quantity trade. Intuitively, too many middlemen congests the market which reduces the benefit of avoiding autarky.

I continue to assume that terms of trade are settled by proportional bargaining. In a retail match, the jointly efficient quantity \(\bar{q}\) is purchased if \(\bar{b}^r > \theta_r R\bar{q} + (1 - \theta_r)u(q)\). Otherwise, the consumer borrows up to the debt limit and purchases the maximum amount that a middleman is willing to sell, \(\bar{b}^r = \theta_r Rq^r + (1 - \theta_r)u(q^r)\). In a wholesale match, a middleman will purchase the jointly efficiency quantity given by (13) if \(\bar{b}^w > b^w\) where \(b^w\) is given by (14). Otherwise, the middleman borrows up to the debt limit and purchases as much as a producer is willing to sell in exchange for \(\bar{b}^w\).

In the following sections, I further relax the notion that agents are able to commit and introduce a role for a medium of exchange.
6 Middlemen in a Monetary Retail Market

In this section, I investigate the role that money plays in facilitating trade within an intermediary sector. I assume that money is necessary in retail market transactions due to anonymity and lack of record keeping, and that credit is feasible in the wholesale market for simplicity. Money is modeled as a perfectly divisible, intrinsically useless asset. Agents endogenously select to hold any non-negative amount of money allowing them to purchase the consumption good in the retail market. I assume that the quantity of money grows at a constant rate $M_{t+1} = \nu M_t$ and is injected by lump-sum transfers $T$ to buyers. One unit of money $m$ purchases $\phi$ units of the numeraire good in the centralized market. I call $\phi$ the value of money.

The critical difference is the terms of trade in the retail market. Since credit is not feasible between middlemen and consumers, the terms of trade in the retail market $(q^r, d^r)$ indicate a quantity of good exchanged for some amount of fiat money $d^r$. In the CM all agents exchange money and goods. In principle, any type of agent can choose to accumulate money in the CM. As we will see, however, only consumers realize liquidity value from holding money in the retail market.

6.1 Centralized Market

In the centralized market debts are settled and the numeraire good is exchanged for money. Any agent is allowed to accumulate money balances and carry them into the following period. The value functions are analogous to (4)-(6) except that now each agent finances current consumption, debt repayment, and future money holdings with current money holdings.

$$W^C_t(m) = \phi_t m + T + \max_{m'} \left[ -\phi_t m' + \beta V^C_{t+1}(m') \right]$$  \hspace{1cm} (18)

$$W^M_t(q, \tilde{m}, b) = Rq + \phi_t \tilde{m} - b + \max_{\tilde{m}'} \left[ -\phi_t \tilde{m}' + \beta \max\{V^M_{1,t+1}(\tilde{m}'), W^M_{t+1}(\tilde{m}')\} \right]$$  \hspace{1cm} (19)

$$W^P_t(\hat{m}, b) = b + \phi_t \hat{m} + \max_{\hat{m}'} \left[ -\phi_t \hat{m}' + V^P_{t+1}(\hat{m}') \right]$$  \hspace{1cm} (20)

\textsuperscript{4}We may imagine that producers are sophisticated in the sense that they are able to record and recognize members of the intermediary sector. That is, each producer has technology which assigns a name to each middleman and can find said middleman in the CM in order to collect on debts.
Inventory, net wealth, and current money balances enter as state variables while future money balances are the choice variable. Notice that all three CM value functions are linear in the state variables (e.g. $W^M(q, \tilde{m}, b) = Rq + \phi \tilde{m} - b + W(0,0,0)$) which implies that current wealth does not affect agents’ choice of future money holdings. Wealth effects disappear due to quasi-linear preferences and greatly simplifies the model analytically since the distribution of money holdings across agents is degenerate.

### 6.2 Decentralized Market

The value functions for the wholesale and retail markets follow precisely the same logic as in the pure credit economy. The only difference is that retail market transactions require money. When a consumer and middleman meet, terms of trade $(q^r, d^r)$ are executed where $d^r$ denotes the amount of money transferred. The retail and wholesale value functions are thus analogous to (7)-(10),

\[ V^C_t(m) = \mu(n)[u(q^r) + W^C_t(m - d^r)] + (1 - \mu(n))W^C_t(m) \]  \hfill (21)

\[ V^M_{2,t}(q, b, \tilde{m}) = (\mu(n)/n)W^M_t(q - q^r, b, \tilde{m} + d^r) + (1 - \mu(n)/n)W^M_t(q, b, \tilde{m}) \]  \hfill (22)

\[ V^M_{1,t}(\tilde{m}) = (\gamma(n)/n)V^M_{2,t}(q^w, b^w, \tilde{m}) + (1 - \gamma(n)/n)V^M_{2,t}(0,0,\tilde{m}) \]  \hfill (23)

\[ V^P_t(\tilde{m}) = \gamma(n)[-c(q^w) + W^P_t(b^r, \tilde{m})] + (1 - \gamma(n))W^P_t(0,\tilde{m}) \]  \hfill (24)

### 6.3 Bargaining Sets

I now characterize the bargaining set that exists between middlemen and consumers in the retail market when credit is not feasible. The value of the surplus from a match for a consumer and middlemen is given by,

\[ u^C - u^C_0 = u(q^r) - \phi d^r \]

\[ u^M - u^M_0 = \phi d^r - Rq^r \]

Contrary to the pure credit economy, a monetary transfer with real value $\phi d^r$ is used to split the
total surplus. The set of incentive compatible trades is thus \( \Omega_r = \{(q^r, d^r) : Rq^r \leq \phi d^r \leq u(q^r) \} \).

Notice that the set of feasible allocations now depends not only on a middleman’s inventory holdings but also on a consumer’s money holdings. For example, it may be the case that a middleman has sufficient inventory to achieve the pairwise efficient outcome \((q^w > \tilde{q})\), but a consumer does not hold enough money to purchase \(\tilde{q}\). The set of feasible utility levels is constrained by \((q^r, d^r) \in [0, q^u] \times [0, m]\). The Pareto frontier of the bargaining set is described taking into account these constraints,

\[
\max_{q^r, d^r} u^c = u(q^r) - \phi d^r + u^c_0 \\
\text{s.t.} \quad \phi d^r - Rq^r + u^M_0 \geq u^M \\
\text{s.t.} \quad 0 \leq q^r \leq q, \quad 0 \leq d^r \leq m
\]

Clearly, the pairwise efficient allocation is the same as the pure credit economy, \(u'(\tilde{q}) = R\), requiring an exchange of \(\phi d^r = R\tilde{q} + u^M - u^M_0\) units of real money. Now, however, there are two constrained solutions. If a middleman carries \(q^w < \tilde{q}\) then a consumer will purchase all inventory with existing money balances, \(q^r = q^w\), and provide the monetary transfer \(\phi d^r = Rq^r + u^M - u^M_0\). If a consumer carries too little money \(\phi m < \min\{Rq^r + u^M + u^M_0, R\tilde{q} + u^M + u^M_0\}\) then the consumer spends all money balances to acquire as much of the search good as possible, \(\phi m = Rq^r + u^M - u^M_0\).

The Pareto frontier is defined as follows,

\[
u^M - u^M_0 = u(\tilde{q}) - R\tilde{q}^* - (u^C - u^C_0) \text{ if } q \geq \tilde{q} \quad \text{and} \quad \phi m \geq \min\{Rq + u^M + u^M_0, R\tilde{q} + u^M + u^M_0\} \\
= u(q) - Rq - (u^C - u^C_0) \text{ if } q < \tilde{q} \quad \text{and} \quad \phi m \geq \min\{Rq + u^M + u^M_0, R\tilde{q} + u^M + u^M_0\} \\
= \phi m - Ru^{-1}(u^C - u^C_0 + \phi m) \text{ if } q < \tilde{q} \quad \text{and} \quad \phi m < \min\{Rq + u^M + u^M_0, R\tilde{q} + u^M + u^M_0\}
\]

The shape of the bargaining set depends on whether inventory or money holdings constrain the solution. If inventory is binding then the Pareto frontier is linear, as was the case in the credit economy. If money is binding, however, then the frontier is concave: \(\frac{\partial^2 u^M}{\partial(u^C)^2} < 0\) if \(\phi m - R\tilde{q} - (u^M - u^M_0) < 0\). Figure 9 depicts the two possible shapes of the frontier.
I now characterize the bargaining set that exists between a middleman and a producer in the wholesale market. The value of the surpluses are given by,

\[ u^M - u^M_0 = \left( \mu(n)/n \right) [\phi d^r(q^w, m) - Rq^r(q^w, m)] + Rq^w - b^w \]

\[ u^P - u^P_0 = -c(q^w) + b^w \]

which are identical to the pure credit economy except that now the expected surplus received by a middleman depends on a consumer’s choice of money holdings. The set of incentive feasible allocations looks the same as under credit, except that the size of the set depends on both a consumer’s money holdings and the bargaining protocol used in the retail market. A consumer’s portfolio choice will affect the size of the total surplus to be divided in the retail market, and the bargaining protocol will dictate how that surplus is divided. For any given level of inventory \( q^w \) the size of the bargaining set in the wholesale market is weakly decreasing in consumer money holdings \( m \). Formally, the set of incentive feasible allocations is given by \( \Omega_w = \{(q^w, b^w) : c(q^w) \leq b^w \leq (\mu(n)/n)[-Rq^r(q^w, m) + d^r(q^w, m)] + Rq^w \} \). Figure 10 depicts how the set of incentive feasible trades shrinks as consumer’s portfolio of real money balances shrinks.
The Pareto frontier is similarly derived as in the credit economy and given by,

$$\frac{\mu(n)}{n} \partial \left[ -Rq^r(q^w, m) + \phi d^r(q^w, m) \right] \partial q^w + R = c'(q^w)$$

$$b^w = u^P - u^P_0 + c(q^w)$$

where the benefit from carrying a marginal unit of inventory into the retail market depends on a consumer’s choice of money holdings. Recall that if a consumer holds $\phi m < \min \{Rq^w + u^M + u^M_0, R\bar{q} + u^M + u^M_0\}$ then he spends all money balances to acquire as much of the search good as possible. In this case, a middleman realizes zero marginal benefit from bringing extra inventory since a consumer holds too little money.

### 6.4 Entry by Middlemen

A middleman chooses to search in the retail market if he captures a non-zero surplus from retail trade, $-Rq^r(q, m) + \phi d^r(q, m) > 0$. A middleman chooses to search the wholesale market if

$$\left( \gamma(n)/n \right) \left( \mu(n)/n \right) \left[ -Rq^r(q^w, m) + \phi d^r(q^w, m) \right] + Rq^w - b^w \geq k.$$  \hspace{1cm} (25)
Notice that if a consumer’s money holdings constrain the bargaining outcome then the entry condition for a middleman becomes more stringent relative to the credit economy.

6.5 Equilibrium

An equilibrium is defined as \{q^w, q^r, b^w, d^r, n\} and money holdings \{m, \tilde{m}, \hat{m}\} determined by a bargaining protocol, free entry of middlemen, and portfolio decisions. For comparability with the pure credit economy, I continue to settle the terms of trade according to proportional bargaining.

In the retail market we have,

$$\max_{q^r, d^r} u(q^r) - \phi d^r \quad s.t. \quad u(q^r) - \phi d^r = \frac{\theta_r}{1 - \theta_r} (-R q^r + \phi d^r)$$

$$s.t. \quad q^r \leq q, \quad d^r \leq m$$

As before, the unconstrained solution is such that

$$u'(\tilde{q}) = R$$

$$\phi \tilde{d} = \theta_r R \tilde{q} + (1 - \theta_r) u(\tilde{q})$$

Now there are two constrained solutions. If inventory is insufficient we have that,

$$q^r = q^w$$

$$\phi d^r = (1 - \theta_r) u(q^r) + \theta_r R q^r$$

If money holdings are insufficient we have that,

$$\phi m = (1 - \theta_r) u(q^r) + \theta_r R q^r$$ \hspace{1cm} (26)

I now move the wholesale market. Since credit is still feasible in the wholesale market, the terms of trade are identical as in the pure credit economy. The amount of inventory purchased is given
by,
\[ \max_{q^w} \left( \mu(n)/n(1-\theta_r)S_r(q^r; q^w, m) + Rq^w - c(q^w) \right) \]
where the size of the surplus in the retail market \( S_r(q^r; q^w, m) = u(q^r(q^w, m)) + -Rq^r(q^w, m) \) now depends on the portfolio choice of a consumer. Thus, the amount of inventory purchased depends on consumers’ portfolio choices made at the end of the CM. A consumer’s choice of money holdings is given by (16) where I substitute out \( V^C \) using (19),
\[ \max_m - (\phi_t - \beta \phi_{t+1})m + \beta \mu(n) [u(q^r(q^w, m))] - \phi d^r(q^w, m) \] (27)
Notice that if \( \phi_t/\phi_{t+1} < \beta \) then there is no solution to (25) since consumers would demand infinite money balances. If \( \phi_t/\phi_{t+1} = \beta \) then the cost of holding money is equated to the rate of time preference so that agents choice of money holdings is enough to purchase \( \tilde{q} \) and is not unique. Finally, if \( \phi_t/\phi_{t+1} > \beta \) then money is costly to hold and buyers do not carry more money balances than they expect to spend in the retail market and the solution is unique.
I restrict my attention to stationary equilibria (i.e. \( q_t = q_{t+1} = q \)) which requires that aggregate real money balances are constant over time: \( \phi_t M_t = \phi_{t+1} M_{t+1} \). There is thus a one-for-one mapping between the rate of money growth and the rate of inflation: \( \phi_{t+1}/\phi_t = 1/\gamma \). Considering stationary equilibria, and using the proportional bargaining outcome, consumers’ choice of real money balances follow,
\[ \max_z -iz + \mu(n)\theta_r S_r(q^r; q^w, z) \] (28)
where \( (1+i) = (1+r)\gamma \) and \( z = \phi m \). A consumer weighs the cost of holding money \( -iz \) against the liquidity value it brings in the retail market \( \mu(n)\theta_r S_r(q^r; q^w, z) \). In order to guarantee that the problem remains concave and admits a unique maximum it must be the case that \( \theta_r/(1-\theta_r) > i/\mu(n) \). For a given level of entry, the buyer must have enough bargaining power in the retail market for money to be valued in equilibrium. The consumer’s problem is represented graphically in Figure 11. Consumers realize positive marginal benefit from holding additional money balances up to the threshold \( q^w \) representing a middleman’s inventory. If the cost of holding money is sufficiently
Figure 11: Consumer’s Portfolio Decision

If, however, the cost of holding money is low enough then the solution occurs at the boundary where consumers carry enough money to purchase the entire amount of inventory. That is, there is zero liquidity value for real balances $z > z(q^w)$.

I now return to a middleman’s inventory decision and discuss similarities to the consumer’s portfolio problem. Recall that inventory purchases are given by,

$$\max_{q^w} \left( \frac{\mu(n)}{n} (1 - \theta_r) S_r(q^r; q^w, z) + R q^w - c(q^w) \right) \quad (29)$$

I represent the middleman’s problem graphically in Figure 12. A middleman weighs the cost of acquiring inventory against its expected value in the retail market. A middleman realizes positive marginal benefit from carrying extra inventory into the retail market up to some threshold $q^{-1}(z)$ which describes the amount of inventory that a consumer can purchase holding $z$ real balances. Any additional inventory in excess of this threshold yields zero marginal benefit to a middleman. If this threshold is sufficiently high there is an interior solution and the middleman buys less inventory than a consumer is able to purchase with $z$ real balances. If, however, the threshold is sufficiently low the middleman has a boundary solution where he buys exactly the amount of inventory that a consumer can purchase.

I now combine the consumer’s portfolio problem and the middleman’s inventory problem to char-
acterize the equilibrium. Recall that a middleman will purchase inventory up to an amount $q^w$ given by 
$$\frac{\mu(n)}{n}(1 - \theta_r)\frac{\partial S_r}{\partial q^w} + R = c'(q^w).$$
This level of inventory corresponds to the pure credit economy. A consumer will therefore never be able to buy more inventory than $q^w$. Thus, we can restrict the domain of the consumer’s portfolio problem to $[0, q^w]$. The nominal interest rate then determines whether or not the solution is interior or at the boundary $q^w$. If the solution is interior, then the middleman’s state space is restricted to $[0, q^{-1}(z)]$ and the optimal inventory is at the boundary $q^{-1}(z)$. If a consumer’s solution is at the boundary, then the middleman simply purchases the pure credit amount $q^w$. This is represented graphically in Figure 13.

An equilibrium is thus defined by (26),(27),(23). Relative to the pure credit economy, the amount of inventory can be no greater. The underinvestment problem is weakly worse. Weak in the sense that if consumers hold enough real balances, then the amount of inventory is the same as under the pure credit economy; however, if consumers hold too few real balances then there is more underinvestment in inventory. Making credit infeasible in the retail market (and so long as money is costly to hold) necessitates a weakly smaller surplus in the retail market. This decreases the value of holding inventory for a middleman.

Note the effect of nominal interest rates on the quantity of inventory. Conventionally, a higher nominal interest rate increases the opportunity cost of holding money which leads to fewer real balances and less trade. Consider, however, the equilibrium depicted in Figure 13 at $i_2$ where the
Figure 13: Equilibrium in $q^w$
consumer is at a boundary solution. In this case, the choice of real money balances is unaffected by a small change in the nominal interest rate. Even though the cost of holding money decreases, agents will not accumulate more money because they know such extra balances will be useless given that middlemen do not carry enough inventory. The quantity traded will only respond to the nominal interest rate along the set of interior solutions to the consumer’s portfolio problem.

Essential money in retail transactions yields qualitatively different effects than under the pure credit economy. Consider the relationship between the measure of middlemen \( n \) and the amount of inventory purchased \( q^w \) depicted in Figure 13. Suppose, initially, that consumer’s portfolio decision has an interior solution and thus a middleman is at a boundary solution. This corresponds to the equilibrium depicted in Figure 13 at \( i_1 \). Now suppose that more middlemen enter the market \( \uparrow n \). This decreases the expected value of retail trade for middlemen resulting in a leftward shift of the inventory demand curve. Lower inventory demand reduces \( q^w \) and thus shifts the boundary condition for consumers to the left. Concurrently, greater entry increases the expected value of retail trade for consumers causing a rightward shift in the money demand curve. This causes the boundary condition for middlemen to shift right \( \uparrow q^{-1}(z) \). If the increase in \( n \) is small, then the consumer is still at an interior solution, the middleman at a boundary solution, and the quantity of inventory increases. However, for a large increase in \( n \), inventory demand shifts so far to the left that the consumer is at its boundary solution while the middleman is at an interior solution. This implies a lower amount of inventory.

**PROPOSITION 6.** When money is essential in retail trades, the response of inventory to the measure of active middlemen is non-monotone. Define \( q = \min\{q^r, q^w\} \) to be the quantity traded given by (26),(27). We have that

\[
\frac{\partial q}{\partial n} > 0 \quad \text{for} \quad (0, \bar{n})
\]

\[
\frac{\partial q}{\partial n} < 0 \quad \text{for} \quad (\bar{n}, \infty)
\]

where \( \bar{n} \) is such that \( q^r = q^w \).

This is substantively different from the pure credit case due to the portfolio decision of consumers.

37
For \((0, \bar{n})\) an increase in \(n\) incentivizes consumers to hold more money balances and middlemen rationally respond by increasing their purchase of inventory. For \((\bar{n}, \infty)\) an increase in \(n\) incentivizes middlemen to purchase few inventories and consumers rationally respond by holding fewer real balances.

An equilibrium in \((26),(27),(23)\) is represented in Figure 14. Notice that the strategic complementarities between portfolio decisions and entry generate multiple equilibria. I denote the “high” equilibrium as \((q_H, n_H)\) and the “low” equilibrium as \((q_L, n_L)\). Both equilibria are supported by consistent and validated beliefs of agents. Consider the high equilibrium as an example. Suppose that middlemen anticipate that consumers will hold large real balances, and therefore anticipate a large surplus in retail trades. This incentivizes a large measure of entrants which increases the frequency of consumption opportunities making it advantageous for consumer’s to hold large real balances, which supports firms’ beliefs. Similarly, if firms believe consumers will hold few real balances, then entry is low, consumption opportunities are rare, and consumers hold few real balances which validates firms’ beliefs. For the comparative statics that follow I focus on the high equilibrium.

**PROPOSITION 7.** When money is essential in retail trades, the comparative statics in \((q_H, n_H)\) depend on the location of the initial equilibrium. The comparative statics for an initial equilibrium
\( n_0 > \bar{n} \) are as follows: \( \partial q_H / \partial \theta_r < 0, \partial n_H / \partial \theta_r < 0, \partial q_H / \partial \theta_w < 0, \partial n_H / \partial \theta_w > 0, \partial q_H / \partial i = 0, \partial n_H / \partial i = 0 \). The comparative statics for an initial equilibrium \( n_0 < \bar{n} \) are as follows: \( \partial q_H / \partial \theta_r > 0, \partial n_H / \partial \theta_r > 0, \partial q_H / \partial \theta_w > 0, \partial n_H / \partial \theta_w > 0, \partial q_H / \partial i > 0, \partial n_H / \partial i > 0 \).

Increasing consumers’ bargaining power \( \uparrow \theta_r \) increases retail demand \( \uparrow q^r \) and decreases inventory demand \( \downarrow q^w \) for any given level of entry \( n \). Consumers are willing to hold more real money balances given they expect higher surplus in the retail market, and middlemen are less willing to hold inventory. This results in a leftward shift of the \( q \)-curve shown in Figure 15.\(^5\) Concurrently, the \( n \)-curve will rotate counterclockwise about \( q^w \). The left diagram in Figure 15 shows the partial effect of \( \uparrow \theta_r \) on the equilibrium due to the \( q \)-curve. The right diagram in Figure 15 shows the general equilibrium effects accounting for free entry via the \( n \)-curve.

The comparative statics depend on where the initial equilibrium is located. If initial equilibrium is at a level of entry \( n > \bar{n} \) (where middlemen inventory demand is binding) then increasing consumers’ bargaining power will result in fewer entrants and less quantity traded. Middlemen, facing a worse bargaining position, demand fewer inventory and consumers rationally respond by holding fewer real balances. If, however, the initial equilibrium is at some \( n < \bar{n} \) (where consumers’ demand is binding) then there will be more entrants and more quantity traded. Greater bargaining power incentivizes consumers to hold more real balances and middlemen rationally respond by purchasing more inventory. In the extreme case where \( \theta_r = 1 \), the \( q \)-curve shifts far to the left and has a horizontal portion corresponding to \( q^w: c'(q^w) = R \) indicating that middlemen do not realize any value from holding inventory in the retail market. Concurrently, the \( n \)-curve rotates counterclockwise.

Changes in bargaining power in the wholesale market have no effect on the \( q \)-curve but effect entry through the \( n \)-curve. More bargaining power for middlemen generates a larger expected surplus from entry which rotates the \( n \)-curve clockwise indicating more entry for any given level of trade. If the initial equilibrium is at \( n > \bar{n} \) then \( \partial n / \partial \theta_w > 0 \) and \( \partial q / \partial \theta_w < 0 \). More entry induces congestion in the retail market resulting in downward movement along the inventory demand curve. If the initial equilibrium is \( n < \bar{n} \) then \( \partial n / \partial \theta_w > 0 \) and \( \partial q / \partial \theta_w > 0 \). More entry increases the expected

\(^5\)It can also be shown analytically that \( \partial n / \partial \theta_r < 0 \) which verifies the leftward shift of the \( q \)-curve.
value of retail trade for consumers who respond by holding more real balances. These comparative statics are represented in Figure 16.

The above comparative statics suggest that the value of $\theta_w$ can determine where the initial equilibrium lies. If $\theta_w$ is large, middlemen will receive a large fraction of its expected surplus which incentivizes a large measure of entrants for any given quantity traded and the resulting equilibrium will be at $(q_H, n_H) : n_H > \bar{n}$. If however, $\theta_w$ is small, then there will be few entrants and the equilibrium will be at $(q_H, n_H) : n_H < \bar{n}$.

Finally, I consider the effects of monetary policy. As suggested by Figure 13, there is a region over
which monetary policy is ineffective. This occurs when inventory demand is a binding constraint for
the consumer. For any small change in the nominal interest rate, the quantity traded is unchanged
because consumer’s realize zero liquidity value from holding extra money. Figure 17 shows that
this is the case for all initial equilibria with $n > \bar{n}$. If the initial equilibrium is $n < \bar{n}$ then a
lower nominal interest rate increases money demand and middlemen respond by holding greater
inventory, and more trade in the retail market attracts new entrants.  

The efficacy of monetary policy may then crucially depend on how much bargaining power mid-
dlemen possess. If $\theta_w$ is large, then the high equilibrium will be such that monetary policy has
no effect. The intuition is as follows. When middlemen receive a large share of future surpluses,
there is a large measure of entry which increases competitive pressures in the wholesale market
and results in less inventory acquisition. Since middlemen are purchasing too little inventory, con-
sumers’ will not realize the full value of their real balances and so rationally choose to hold only
enough to purchase all inventory. The high equilibrium is thus characterized by too little inventory
and consumers’ facing a boundary solution which leaves monetary policy ineffective. Conversely,
suppose that $\theta_w$ is small, so that there are relatively few entrants. Fewer entrants increases the
probability of a retail trade which incentivizes middlemen to purchase more inventory. Consumers’
are now able to realize the full return on their real balances since middlemen hold large inventories,
and monetary policy is effective.

---

6There exists some threshold nominal interest rate $\bar{i}$ which defines the effective lower bound on interest rates,
below which monetary policy is ineffective. This threshold is implicitly defined as follows: let $(q_i, n_i)$ solve (23) and
(26), then $\bar{i}$ is such that (25) holds at $(q_i, n_i)$.  

41
I now consider the case where money is essential in the wholesale market and assume that credit is feasible in the retail market.

### 7.1 Centralized Market

I express the CM value functions generally in that I allow all agents to accumulate real money balances. Although money appears as a state variable, only middlemen will hold money in equilibrium. Since credit is feasible in the retail market, debt enters as a state variable for middlemen and consumers.

\[
W^C_t(m, b) = \phi_t m - b + T + \max_{m'}[-\phi_t m' + \beta V^C_{t+1}(m')] \quad (30)
\]

\[
W^M_t(q, \tilde{m}, b) = Rq + \phi_t \tilde{m} + b + \max_{\tilde{m}'}[-\phi_t \tilde{m}' + \beta V^M_{1,t+1}(\tilde{m})] \quad (31)
\]

\[
W^P(\tilde{m}) = \phi_t \tilde{m} + \max_{\tilde{m}'}[-\phi_t \tilde{m}' + V^P_{t+1}] \quad (32)
\]
7.2 Decentralized Market

The value functions for the DM follow the same logic and before. Now, however, transactions in the wholesale market must be conducted with money while consumers purchase the search good in the retail market with credit.

\[ V^C_t(m) = \mu(n)[u(q^r) + W^C_t(m, -b^r)] + (1 - \mu(n))W^C_t(m, 0) \]  \hspace{1cm} (33)

\[ V^M_{2,t}(q, \tilde{m}) = \left(\frac{\mu(n)}{n}\right)W^M_t(q - q^r, b^r, \tilde{m}) + (1 - \left(\frac{\mu(n)}{n}\right))W^M_t(q, 0, \tilde{m}) \]  \hspace{1cm} (34)

\[ V^M_{1,t}(\tilde{m}) = \left(\frac{\gamma(n)}{n}\right)V^M_{2,t}(q^w, \tilde{m} - d^w) + (1 - \left(\frac{\gamma(n)}{n}\right))V^M_{2,t}(0, \tilde{m}) \]  \hspace{1cm} (35)

\[ V^P_t(\hat{m}) = \gamma(n)[-c(q^w) + W^P_t(\hat{m} + d^w)] + (1 - \gamma(n))W^P_t(\hat{m}) \]  \hspace{1cm} (36)

7.3 Bargaining Sets

Since credit is feasible in the retail market, the bargaining set is exactly as it appears in the pure credit economy. The frontier is reproduced here for convenience,

\[ u^M - u^M_0 = u(q^*) - (u^C - u^C_0) \hspace{1cm} \text{if} \hspace{0.5cm} q^w \geq q^* \]

\[ = u(q^w) - (u^C - u^C_0) - Rq^w \hspace{1cm} \text{if} \hspace{0.5cm} q^w < q^* \]

The size of the bargaining set in the wholesale market now depends on the money holdings of a middleman following the same logic used in the monetary economy where consumers held money balances. The frontier of the wholesale bargaining set is given by,

\[ u^M - u^M_0 = \frac{\mu(n)}{n}\left[-Rq^r(q^w) + b^r(q^w)\right] + Rq^w - c(q^w) - (u^P - u^P_0) \hspace{1cm} \text{if} \hspace{0.5cm} \phi \tilde{m} \geq c(q^w) + u^P - u^P_0 \]

\[ = \frac{\mu(n)}{n}\left[-q^r(c^{-1}(\phi \tilde{m} - (u^P - u^P_0))) + b^r(c^{-1}(\phi \tilde{m} - (u^P - u^P_0)))\right] + Rc^{-1}(\phi \tilde{m} - (u^P - u^P_0)) - \phi \tilde{m} \hspace{1cm} \text{if} \hspace{0.5cm} \tilde{m} \]

In this case, if the middleman is liquidity constrained, the bargaining set is strictly concave.
7.4 Entry by Middlemen

Middlemen will choose to search the retail market so long as they receive a nonzero surplus from a trade $b^r - Rq^r \geq 0$. A middleman will choose to search in the wholesale market as long as,

$$\frac{\gamma(n)}{n} \left[ \frac{\mu(n)}{n} [-Rq^r(q^w, \bar{m}) + b^r(q^w, \bar{m})] + Rq^w - b^w \right] \geq k$$

where a middleman’s choice of money holdings can constrain both the surplus received in the wholesale and retail market since inventory acquisition depends on money holdings $q^w = q^w(\bar{m})$.

7.5 Equilibrium

An equilibrium is $\{q^w, q^r, b^r, d^w, n\}$ and money holdings $\{m, \bar{m}, \hat{m}\}$. Recall that even in the pure credit economy, there always exists underinvestment in inventory $q^w < q^*$ due to the hold up problem. Therefore, the solution to retail bargaining will always be where the inventory constraint binds,

$$q^r = q^w$$  \hspace{1cm} (37)

$$b^r = \theta_r Rq^r + (1 - \theta_r) u(q^r)$$  \hspace{1cm} (38)

The bargaining solution in the wholesale market may be constrained by a middleman’s money holdings. Formally we have,

$$\max_{q^w, b^w} \left( \frac{\mu(n)}{n} [-Rq^r(q^w) + b^r(q^w)] + Rq^w - \phi d^w \right)$$

$$s.t. \quad \left( \frac{\mu(n)}{n} [-Rq^r(q^w) + b^r(q^w)] + Rq^w - \phi d^w \right) = \frac{\theta_w}{1 - \theta_w} (-c(q^w) + b^w)$$

$$s.t. \quad d^w \leq \bar{m}$$

The unconstrained solution is identical to the pure credit economy, but now there is also a
constrained solution

$$\phi \tilde{m} = \theta_w c(q^w) + (1 - \theta_w) \left[ \frac{\mu(n)}{n} (-Rq^w(\tilde{m}) + b'(\tilde{m})) + Rq^w \right]$$

where it is made explicit that the value received in a retail match depends on the a middleman’s money holdings. Given that credit is feasible in the retail market and the inventory constraint always binds, I can write,

$$z(q^w) = \phi \tilde{m} = \theta_w c(q^w) + (1 - \theta_w) \left[ \frac{\mu(n)}{n} (1 - \theta_r) (u(q^w) - Rq^w) + Rq^w \right]$$  (39)

which makes clear the relationship between money holdings and the amount of inventory purchased.

The amount of real balances carried by producers and consumers never affect the terms of trade and so long as money is costly to hold neither will choose to hold any money. Middlemen, on the other hand, do realize liquidity value from their holdings and choose to hold a quantity which equalizes the cost of holding money to its liquidity value. A middleman solves,

$$\max_{q^w} -i z(q^w) + (\gamma(n)/n) \theta_w \left[ (\mu(n)/n)(1 - \theta_r)(u(q^w) - Rq^w) + (Rq^w - c(q^w)) \right]$$

where $z(q^w)$ denote the amount of inventory that can be bought with $z$ real balances given by (37). Note that in order for the objective function to be concave and admit a unique maximum, it must be the case that $\theta_w/(1 - \theta_w) > i/(\gamma(n)/n)$. For a given level of entry, a middleman must have enough bargaining power when acquiring inventory to incentivize him to hold money. The first order condition for the above problem pins down the quantity of inventory purchased,

$$i = \frac{(\gamma(n)/n) \theta_w \left[ (\mu(n)/n)(1 - \theta_r)(u(q^w) - R) + (R - c'(q^w)) \right]}{\theta_w c'(q^w) + (1 - \theta_w) \left[ (\mu(n)/n)(1 - \theta_r)(u'(q^w) - R) + R \right]}$$  (40)

Relative to the pure credit case, inventory demand is less for all $i > 0$ and coincides precisely with the pure credit case at the Friedman rule ($i = 0$).

Free entry requires that middlemen will continue to enter the wholesale market until expected
Profits are driven to zero.

\[
(\gamma(n)/n) \left[ (\mu(n)/n)(1 - \theta_r)(u(q_r) - Rq_r) + \theta_w(Rq_w - c(q_w)) \right] = (1 - \beta)k
\]

An equilibrium is given by (35)-(39). Figure 18 displays the equilibrium. Relative to the monetary retail market, the equilibrium is simpler because we do not need to consider discontinuities arising from boundary solutions. That is, since credit is available to consumers the amount purchased in the retail market equals the amount of inventory regardless of portfolio choices. Thus, an equilibrium is given by (38) which describes a negative relationship between \( q \) and \( n \) and (39) which describes a positive relationship. This immediately reveals that multiplicity of equilibria is the consequence of essential money in the retail market, as opposed to the wholesale market.

8 Conclusion

The framework lends itself to developing an intermediation theory of the firm, first articulated by Coase (1937) and later refined by Spulber (1999), while including micro-foundations for the role of liquid assets. Stated simply, firms act as a conduit between suppliers and customers when the
gains from intermediated trade are greater than the gains from direct trade. The conditions under which this happens are many, and always depend on the environment described by the modeler.

Presently, middlemen are merchants who buy and resell goods without engaging in any productive activity; while producers are a technology allowing for the manufacture of retail goods. The model can be amended so that middlemen more closely resemble firms in the conventional sense. Producers are reinterpreted as entrepreneurs who have an idea or ability to generate some input into a larger production process. Middlemen are reinterpreted as firms who purchase inputs from entrepreneurs, transform inputs into final consumption goods, and sell to consumers. The value-adding productive process employed by middlemen/firms can be formalized by positing a concave technology $Q=G(q)$. Consumers then enjoy utility $u(Q)$.

The reinterpreted framework places middlemen as the creators and operators of markets. They form bid and ask prices, conduct transactions, and allocate goods. The theory offers an explicit mechanism by which markets clear and equilibrium prices obtain rather than resorting to the theoretical construct of a Walrasian auctioneer.

It is worthwhile to consider alternative market structures while retaining middlemen as an explicit mechanism by which prices are set and quantities are determined. The obvious market structure to explore would be competitive price posting which allows one to price congestion in the market. This market structure may also be a more realistic representation of middlemen as market-makers rather than merchants. Watanabe (2017) considers the case of a monopolistic middleman who can choose whether to act as a merchant or a market-maker.

References


Xun Gong. Middlemen: Intensive and extensive margins with endogenous meeting technology.


9 Appendix

9.1 Planner’s Problem

In order to study efficiency, I consider the problem of a social planner who each period chooses the measure \( n_t \) of active middlemen and an allocation \( \{ q^w(i), q^r(i), q^p(i) \} \) for all matched agents, \( i \in \tilde{P} \cup \tilde{M} \cup \tilde{C} \), and \( \{ x_t(i), y_t(i) \} \) for all agents \( i \in P \cup C \cup M \). The planner is constrained by the environment in the sense that he cannot choose the set of matched agents \( \tilde{P} \cup \tilde{M} \cup \tilde{C} \) but only \( n_t \), and then the sets are determined randomly such that \( \tilde{C} \) has measure \( \mu(n) \), \( \tilde{M} \) has measure \( \mu(n)/n \).

I consider the following period welfare function,

\[
W_t = \int_{C \cup M \cup P} x_t(i) - y_t(i) di + \int_{\tilde{C}} u(q^w_t(i)) di - \int_P c(q^w_t(i)) di - kn_t
\]

The planner wishes to maximize \( \sum_{t=0}^{\infty} \beta^t W_t \) subject to the following feasibility constraints,

\[
\int_{C \cup M \cup P} x_t(i) di \leq \int_{C \cup M \cup P} y_t(i) + \int_{M_w \cap M_r} R(q^w_t(i) - q^r_t(i)) di + \int_{\tilde{M}_w \cap \tilde{M}_r} Rq^w_t(i)
\]
\[
\int_{\tilde{M}_w} q^w_t(i) di \leq \int_{\tilde{P}} q^P_t(i) di
\]
\[
\int_{\tilde{C}} q^r_t(i) di \leq \int_{\tilde{M}_w \cap \tilde{M}_r} q^w_t(i) di
\]

\( q^r_t(i) \leq q^w_t(j) \) for each \((i,j)\) trading pair

\( q^w_t(i) \leq q^P_t(j) \) for each \((i,j)\) trading pair

The first constraint states that the amount of numeraire consumed in the CM can be no greater than the sum of labor and unsold inventory carried by middlemen. The second constraint demands that total inventory can be no greater than total production. The third constraint guarantees that total consumption of the search good is less than inventory holdings of matched middlemen. The fourth and fifth constraints require that an individual buyer can never purchase more than an individual seller can produce or carry in inventory.
Form the Lagrangian,

\[ \mathcal{L} = \int_{\mathcal{C} \cup \mathcal{M} \cup \mathcal{P}} x(i) - y(i) \, di + \int_{\mathcal{C}} u(q^r(i)) \, di - \int_{\tilde{\mathcal{P}}} c(q^w(i)) \, di - kn_t - \lambda_1 \left[ \int_{\tilde{\mathcal{M}}} q^w(i) \, di - \int_{\tilde{\mathcal{P}}} q^p(i) \, di \right] - \lambda_2 \left[ \int_{\tilde{\mathcal{M}}} q^w(i) \, di - \int_{\tilde{\mathcal{P}}} q^p(i) \, di \right] - \lambda_3 \left[ \int_{\tilde{\mathcal{C}}} q^r(i) \, di - \int_{\tilde{\mathcal{M}}} q^w(i) \, di \right] \]

For local maximizers \( \{q^r(i), q^w(i), q^p(i), n\} \) there exist multipliers \( \{\lambda_1, \lambda_2, \lambda_3\} \) such that:

1. \( 0 = \frac{\partial \mathcal{L}}{\partial x(i)} = \frac{\partial \mathcal{L}}{\partial y(i)} = \frac{\partial \mathcal{L}}{\partial q^r(i)} = \frac{\partial \mathcal{L}}{\partial q^w(i)} = \frac{\partial \mathcal{L}}{\partial n} \)

2. \( 0 = \lambda_1 \left[ \int_{\mathcal{C} \cup \mathcal{M} \cup \mathcal{P}} x_t(i) \, di - \int_{\mathcal{C} \cup \mathcal{M} \cup \mathcal{P}} y_t(i) \, di - \int_{\tilde{\mathcal{M}}} R(q^w_t(i) - q^r_t(i)) \, di + \int_{\tilde{\mathcal{M}}} Rq^w_t(i) \right] \)
   \( 0 = \lambda_2 \left[ \int_{\tilde{\mathcal{M}}} q^w_t(i) \, di - \int_{\tilde{\mathcal{P}}} q^p_t(i) \, di \right] \)
   \( 0 = \lambda_3 \left[ \int_{\tilde{\mathcal{C}}} q^r_t(i) \, di - \int_{\tilde{\mathcal{M}}} q^w_t(i) \, di \right] \)

3. \( \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \)

4. \( \int_{\mathcal{C} \cup \mathcal{M} \cup \mathcal{P}} x_t(i) \, di \leq \int_{\mathcal{C} \cup \mathcal{M} \cup \mathcal{P}} y_t(i) + \int_{\tilde{\mathcal{M}}} R(q^w_t(i) - q^r_t(i)) \, di + \int_{\tilde{\mathcal{M}}} Rq^w_t(i) \)
   \( \int_{\tilde{\mathcal{M}}} q^w_t(i) \, di \leq \int_{\tilde{\mathcal{P}}} q^p_t(i) \, di \)
   \( \int_{\tilde{\mathcal{C}}} q^r_t(i) \, di \leq \int_{\tilde{\mathcal{M}}} q^w_t(i) \, di \)
   \( q^r_t(i) \leq q^w_t(j) \) for each \((i, j)\) trading pair
   \( q^w_t(i) \leq q^p_t(j) \) for each \((i, j)\) trading pair
The first order conditions for \( x(i), y(i) \) immediately give \( \lambda_1 = 1 \).

\[
\frac{\partial L}{\partial x(i)} = - \frac{\partial L}{\partial y(i)} = - \int_{C \cup P \cup M} di + \lambda_1 \int_{C \cup P \cup M} di = 0
\]

The first order condition for \( q^p(i) \) reveals that \( \lambda > 0 \) or else \( q^p(i) = 0 \forall i \) which is surely not a maximum of the objective function.

\[
\frac{\partial L}{\partial q^p(i)} - \int_{\tilde{P}} c'(q^p(j))di + \lambda_2 \int_{\tilde{P}} di = 0
\]

Using the values for \( \lambda_1, \lambda_2 \), the first order conditions for \( q^w(i), q^r(i) \) are,

\[
\frac{\partial L}{\partial q^w(i)} = \int_{\tilde{M}_w \cap \tilde{M}_r} R di + \int_{\tilde{M}_w \cap \tilde{M}_r} R di \quad \frac{\partial L}{\partial q^p(i)} - \int_{\tilde{P}} c'(q^p(j))di \int_{\tilde{M}_w} di + \lambda_3 \int_{\tilde{M}_w} di = 0
\]

\[
\frac{\partial L}{\partial q^r(i)} = \int_{\tilde{C}} u'(q^r(i))di - \int_{\tilde{M}_w \cap \tilde{M}_r} R di \quad \lambda_3 \int_{\tilde{C}} di = 0
\]

We must consider two cases: (I) \( \lambda_3 = 0 \) and (II) \( \lambda_3 > 0 \). Finally, the first order condition for \( n \) is given by evaluating the measure of each subset according to matching functions \( \mu(n), \gamma(n) \),

\[
\frac{\partial L}{\partial n} = \mu'(n)u(q^r) - \gamma'(n)c(q^p) - k
\]

\[
- \lambda_1 \left[ R \gamma'(n)(q^w - \frac{\mu(n)}{n} q^r) - R \gamma(n)(\mu(n)/n)'q^r \right]
- \lambda_2 \left[ \gamma'(n)q^w - \gamma'(n)q^p \right]
- \lambda_3 \left[ \mu'(n)q^r - (\frac{\mu(n)}{n})'q^w \right]
\]

51
I’ll start with case I. If \( \lambda_3 = 0 \) then we have the following system,

\[
\int_{\tilde{C}} u'(q^r(i)) = \int_{\tilde{M}_w \cap \tilde{M}_r} Rdi \\
\int_{\tilde{M}_w \cap \tilde{M}_r} Rdi + \int_{\tilde{M}_w \cap \tilde{M}_{-r}} Rdi = \frac{\int_{\tilde{P}} c'(q^p(j))di}{\int_{\tilde{P}} di} \int_{\tilde{M}_w} di \\
\int_{\tilde{M}_w} q^w(i)di = \int_{\tilde{P}} q^p(i)di \\
\int_{C \cup M \cup P} x(i) - y(i)di = \int_{\tilde{M}_w \cap \tilde{M}_r} R(q^w(i) - q^r(i))di + \int_{\tilde{M}_w \cap \tilde{M}_{-r}} Rq^w(i) \\
\int_{\tilde{C}} q^r(i)di = \int_{\tilde{M}_w \cap \tilde{M}_r} q^w(i) \\
\int_{\tilde{M}_w} q^w(i)di = \int_{\tilde{P}} q^p(i)di \\
\int_{C \cup M \cup P} x(i) - y(i)di = \int_{\tilde{M}_w \cap \tilde{M}_r} R(q^w(i) - q^r(i))di + \int_{\tilde{M}_w \cap \tilde{M}_{-r}} Rq^w(i)
\]

Evaluating the measure of each subset according to matching functions \( \mu(n), \gamma(n) \) we have the following system for \( \lambda_3 = 0 \)

\[
u'(q^r) = \frac{\gamma(n)}{n} R \\
c'(q^p) = R \\
q^w = q^p \\
\int x(i) - y(i)di = R\gamma(n)[q^w - \mu(n)] \\
k = \mu'(n) \left[ u(q^r) + \frac{\gamma(n)}{n} Rq^r \right] - \gamma'(n) \left[ c'(q^w) + Rq^w - \mu(n) Rq^r \right] - \frac{\mu(n) \gamma(n)}{n} Rq^r
\]

Now for case II. When \( \lambda_3 > 0 \) we have the following system,

\[
\int_{\tilde{M}_w \cap \tilde{M}_r} Rdi + \int_{\tilde{M}_w \cap \tilde{M}_{-r}} Rdi - \frac{\int_{\tilde{P}} c'(q^p(j))di}{\int_{\tilde{P}} di} \int_{\tilde{M}_w} di + \left( \int_{\tilde{C}} u'(q^r(i))di - \int_{\tilde{M}_w \cap \tilde{M}_r} Rdi \right) \int_{\tilde{M}_w \cap \tilde{M}_r} di = 0 \\
\int_{\tilde{M}_w} q^w(i)di = \int_{\tilde{M}_w \cap \tilde{M}_r} q^w(i) \\
\int_{\tilde{M}_w} q^w(i)di = \int_{\tilde{P}} q^p(i)di \\
\int_{C \cup M \cup P} x(i) - y(i)di = \int_{\tilde{M}_w \cap \tilde{M}_r} R(q^w(i) - q^r(i))di + \int_{\tilde{M}_w \cap \tilde{M}_{-r}} Rq^w(i)
\]

Evaluating the measure of each subset according to matching functions \( \mu(n), \gamma(n) \) we have the
following system for $\lambda_3 > 0$

$$\frac{\mu(n)}{n}(u'(q') - \frac{\gamma(n)}{n}R) + R = c'(q^p)$$  \hspace{1cm} (47)

$$q' = \frac{\gamma(n)}{n}q^w$$  \hspace{1cm} (48)

$$q^w = q^p$$  \hspace{1cm} (49)

$$\int x(i) - y(i)di = R\gamma(n)[q^w - \frac{\mu(n)}{n}]$$  \hspace{1cm} (50)

$$k = \mu'(n)\left[u(q') + \frac{\gamma(n)}{n}Rq' + (u'(q') - \frac{\gamma(n)}{n}R)(q^w - q')\right]$$

$$- \gamma'(n)\left[c'(q^w) + Rq^w - \frac{\mu(n)}{n}Rq' - \frac{\mu(n)}{n}(u'(q') - \frac{\gamma(n)}{n}R)q^w\right]$$

$$- \frac{\mu(n)}{n}\gamma(n)\left[Rq' + (u'(q') - \frac{\gamma(n)}{n}R)q^w\right]$$  \hspace{1cm} (51)

### 9.2 Characterizing Equilibrium in Monetary Retail Market

First I show the solutions for consumers and middlemen and demonstrate analytically over which regions the solutions are interior v. boundary.

Begin with the interior solutions for consumers and middlemen given by (22),(23)

$$\mu(n)\theta_r\left[u'(q') - R\right] = i$$

$$(\mu(n)/n)(1 - \theta_r)(u'(q^w) - R) + R = c'(q^w)$$

For comparability, I rewrite the interior solutions (assuming $n \neq 0, \theta_r \neq \{0, 1\}$) as,

$$u'(q') - R = \frac{i((1 - \theta_r)u'(q') + \theta_rR)}{\mu(n)\theta_r}$$

$$u'(q^w) - R = \frac{c'(q^w) - R}{(\mu(n)/n)(1 - \theta_r)}$$
A priori the amount of inventory desired by middlemen can be less than or greater than the amount desired by consumers. That is, it will depend on the number of middlemen $n$ as well as bargaining power $\theta_r$ and the nominal interest rate $i$ whether middlemen or consumers are at an interior solution (and consequently which agent is at a boundary solution). To fix this idea, consider adjusting the number of middlemen $n$. Clearly $\partial A(q^r; n)/\partial n < 0$ and $\partial B(q^w; n)/\partial n > 0$ which implies that $\partial q^r/\partial n > 0$ and $\partial q^w/\partial n < 0$. Define the quantity traded as $q = \min\{q^w, q^r\}$ which takes into account the boundary conditions for each agent. For sufficiently low $n$, we have that $q = q^r$ which corresponds to an interior solution for consumers and a boundary solution for middlemen. For sufficiently high $n$, we have that $q = q^w$ which corresponds to an interior solution for middlemen and a boundary solution for consumers. There exists some threshold value $\bar{n}$ such that $q = q^r = q^w$.

The following expressions are derived from a simple application of the implicit function theorem,

$$\frac{\partial q^r}{\partial n} = -\frac{\mu'(n)(u'(q^r) - R)((1 - \theta_r)u'(q^r) + \theta_r R)}{Ru^0(q^r)} > 0$$

$$\frac{\partial q^w}{\partial n} = \frac{(\mu(n)/n)'(1 - \theta_r)(u'(q^w) - R)}{(\mu(n)/n)(1 - \theta_r)u''(q^w) - c''(q^w)} < 0$$

And the threshold $\bar{n}$ and corresponding $q = q^r = q^w$ is defined by the system,

$$u'(q) - R = A(q; n)$$
$$u'(q) - R = B(q; n)$$

In summary, for $n \in (0, \bar{n})$ we have $\partial q/\partial n \geq 0$ and for $n \in (\bar{n}, \infty)$ we have $\partial q/\partial n \leq 0$ with a kink at $\bar{n}$. This is depicted in Figure 6. I can derive $\bar{n}$ by first expressing (22) as,

$$u'(q^r) = \frac{\theta_r R(i + \mu(n))}{\mu(n)\theta_r - i(1 - \theta_r)}$$

for notational convenience define $f(q) = u'(q)$. I have now that,

$$q^r = f^{-1}\left(\frac{\theta_r R(i + \mu(n))}{\mu(n)\theta_r - i(1 - \theta_r)}\right)$$

54
Threshold \( \bar{n} \) is defined where \( q = q^r = q^w \). Using this and (23) I can write,

\[
(\mu(\bar{n})/\bar{n})(1 - \theta_r) \left( Ri \right) + R - c' \circ f^{-1} \left( \frac{\theta_r R(i + \mu(\bar{n}))}{\mu(\bar{n})\theta_r - i(1 - \theta_r)} \right) = 0
\]

After a tedious application of implicit differentiation, I show that

\[
\frac{\partial \bar{n}}{\partial \theta_r} < 0
\]

\[
(\mu(n)/n)(1 - \theta_r) \left( u'q^w - R \right) = \frac{(\gamma(n)/n)\theta_w + i\theta_w}{\gamma(n)/n - i(1 - \theta_w)} (c'(q^w) - R) - iR
\]

This form makes it especially comparable to the pure credit economy where inventory demand was given by (10). As a first point of comparison, not that the vertical intercept of RHS is identical between (10) and (35),

\[
\lim_{q^w \to 0} RHS = -R
\]

As a second point, we can easily verify that the Friedman rule obtains the inventory demand (10) \( RHS(i = 0) = c'(q^w) - R \). Now we need to evaluate the slope of the RHS to determine whether inventory demand will be higher or lower under a monetary wholesale market. The slope is given by,

\[
slope = \frac{(\gamma(n)/n)\theta_w + i\theta_w}{(\gamma(n)/n) - i(1 - \theta_w)}
\]

9.3 Inventory Demand under Monetary Wholesale

Inventory demand (35) can be rewritten as,

\[
(\mu(n)/n)(1 - \theta_r) \left( u'q^w - R \right) = \frac{(\gamma(n)/n)\theta_w + i\theta_w}{\gamma(n)/n - i(1 - \theta_w)} (c'(q^w) - R) - iR
\]

This form makes it especially comparable to the pure credit economy where inventory demand was given by (10). As a first point of comparison, not that the vertical intercept of RHS is identical between (10) and (35),

\[
\lim_{q^w \to 0} RHS = -R
\]

As a second point, we can easily verify that the Friedman rule obtains the inventory demand (10) \( RHS(i = 0) = c'(q^w) - R \). Now we need to evaluate the slope of the RHS to determine whether inventory demand will be higher or lower under a monetary wholesale market. The slope is given by,

\[
slope = \frac{(\gamma(n)/n)\theta_w + i\theta_w}{(\gamma(n)/n) - i(1 - \theta_w)}
\]
The following shows that this slope is greater than one,

\[ \theta_w ((\gamma(n)/n + i)) \leq \theta_w (\gamma(n)/n) - i(1 - \theta_w) \]
\[ \theta_w i \leq -i(1 - \theta_w) \]
\[ 0 \leq -i \]

where the ZLB condition guarantees the direction of the final inequality and proves \textit{slope} > 1. Therefore, the RHS of (35) lies strictly above the RHS of (10) so that there is less inventory demand for any given level of entry.