Who Wants To Be A Middleman?*

Ed Nosal
FRB Chicago

Yuet-Yee Wong
Binghamton University

Randall Wright
University of Wisconsin-Madison,
FRB Chicago, FRB Minneapolis and NBER

August 4, 2016

Abstract

We study agents' decisions to act as producers or middlemen, and hence endogenize the composition of the market, in a Rubinstein-Wolinsky model of intermediation extended to allow more general bargaining, technology and preferences. While requiring a different approach, the analysis remains tractable, delivering clean and sometimes surprising results. We show intermediation can be essential, and prove equilibrium is efficient iff bargaining satisfies generalized Hosios conditions. Our analysis goes beyond the usual linear specification to capture payment frictions, and beyond steady state by considering dynamics. Finally, we prove that while the model with middlemen storing goods has a unique equilibrium, a version with middlemen holding assets displays multiplicity attributable to composition effects. This suggests there is something special about financial intermediation.

JEL Codes: G24, D83

Keywords: Middlemen, Intermediation, Search, Bargaining

*We thank Zack Bethune, Chao Gu, Alberto Trejos, Steve Williamson, Mark Gertler, Karl Shell, George Selgin, David Laidler and Douglas Gale for input. We also thank the Chicago Fed and for providing an excellent research environment. Wright acknowledges support from the Ray Zemon Chair in Liquid Assets at the Wisconsin School of Business, and the hospitality of the the University of Chicago Economics Department during the Fall Quarter of 2015. The usual disclaimers apply.

1 Introduction

The various roles of middlemen, or intermediaries, have been studied by a number of authors (see fn. 5 below). However, given the importance of middlemen in real-world economic activity, from wholesale and retail trade in producer or consumer goods, to financial intermediation, there would seem to be room for additional work. This project revisits the classic search-based framework of Rubinstein and Wolinsky (1987), hereafter RW, extends it on a number of dimensions, and uses it to analyze a variety of issues, including the choice of agents to become middlemen.

The original RW formulation has no cost of production or search, equal numbers of producers and consumers, a fixed number of potential middlemen, and symmetric bargaining. In equilibrium, all producers participate in the market, while middlemen participate iff they have a better search technology than producers and this is efficient. Nosal et al. (2015) extend the model to incorporate more general bargaining and costs, show that whether agents participate depends on various factors including bargaining power, and this may or may not be efficient. However, even in this version, like RW, the assumptions are very special: utility is linear; the market is in steady state; and the numbers of producers and middlemen are fixed. We relax all of these restrictions, perhaps most importantly by letting individuals decide to act as *either* producers or middlemen. This occupational choice makes the composition of market endogenous, allows us to study how it depends on the environment, ask additional questions about optimality, and address other substantive issues.[1]

Endogenous market composition complicates the analysis. Earlier studies of RW use a "trick" by taking $\boldsymbol{\alpha} = \{\alpha_{ij}\}$ as exogenous, where α_{ij} is the rate at

[1]Since it is easier to describe substantive issues after laying out more details, we defer an extended discussion of these for now (see, e.g., fn. 17). However, one feature to emphasize up front is that our model economy can only have more middlemen by having fewer producers – for the sake of argument, say, more MBA's and fewer engineers – which argubly captures an interesting real tradeoff.

which type i agents meet type j agents. This is legitimate because there always exists a distribution of types, say $\mathbf{n} = \{n_i\}$, consistent with $\boldsymbol{\alpha}$, random matching, and the identities implied by bilateral meetings, $n_i \alpha_{ij} = n_j \alpha_{ji}$. Hence, those papers can conveniently take $\boldsymbol{\alpha}$ as fixed when characterizing equilibrium – but that won't work when agents get to choose their types, since anything that affects $\mathbf{n}$ generally affects $\boldsymbol{\alpha}$. So we must determine endogenously $\alpha_{ij} = \alpha_i n_j$, where α_i is a baseline arrival rate for type i. However, now the relevant identities imply $\alpha_i = \alpha$ is the same for all i, and in particular $\alpha_{pc} = \alpha_{mc}$, so we must abandon RW's original idea that middlemen have a role iff $\alpha_{mc} > \alpha_{pc}$. Fortunately, other factors in our environment can take over for $\boldsymbol{\alpha}$, including bargaining powers, and type-specific storage costs or rates of return.

Despite the complications necessitated by endogenous occupational choice, the framework remains tractable, delivering clean and sometimes surprising predictions. Consider a baseline model where agents are interpreted as storing consumer goods at positive costs, like retailers, where these costs generally differ between producers and middlemen. As one perhaps surprising result, we show that increasing intermediaries' costs can lead to more of them, although we also explain why that is not unnatural. For this baseline model, we establish existence and uniqueness of equilibrium, and show how intermediation can be essential – e.g., the market may shut down if middlemen are prohibited, say, by taxing them out of existence. Also, we find that equilibrium can have either too few or too many middlemen, and is efficient iff bargaining powers are just right.[2] Additionally, we go beyond the usual linear (transferable) utility specification by allowing strict concavity, which is relevant because the nonlinearity can be interpreted in terms of frictions in the payment process that interact with intermediation. And we

[2]This is related to standard results going back to Mortensen (1982) and Hosios (1990); see Julien and Mangin (2016) for an updated discussion. However, our results are also slightly different, because RW-style environments involve three-sided markets, with producers, consumers and middlemen that all meet and bargain.

go beyond the usual steady state analysis by establishing saddle path stability of dynamic equilibrium.

We then go on to consider an application to financial intermediation. If the cost of storing goods is negative, one can reinterpret the goods as assets bearing positive returns. Producers, e.g., might supply capital goods that they can invest at a profit while waiting to trade them to someone else. They can trade them to end users that have a more productive use for capital, or they might trade them to middlemen. Middlemen that get capital goods from producers can also invest them at a profit while waiting to pass them on to end users, and it is reasonable to think of this as financial intermediation.[3] Middlemen can have a higher or lower return than producers, but as long as it is positive, an interesting new possibility arises: they might prefer to keep the capital for themselves when their return on investment is large compared to what they can get from end users.

Hence, financial intermediation may or may not emerge in equilibrium. Importantly, there are multiple equilibria in large regions of parameter space. On the one hand, if middlemen decide to pass capital on to end users, lots of them will be without capital in steady state. Since producers can trade with either end users or middlemen in search of capital, this raises their profit, and hence leads to more producers and fewer middlemen. With more producers, it is easier for middlemen to get capital, thus rationalizing their decision to trade it to end users and making active intermediation an equilibrium. On the other hand, if middlemen instead decide to keep capital for their own investments, in steady state, producers can only trade with end users. This lowers their profit and leads to fewer producers. With fewer producers it is harder for middlemen to get cap-

[3] A minor detail is that, from a narrow perspective, exchange in the model looks like spot trade, but it is easy to reconsider it as intertemporal where, e.g., an end user that receives capital, either directly from a supplier or indirectly from an intermediary, remits payment at some future date when his investment pans out. As long as contracts are enforceable, the results are identical. Also, it would not matter in our environment if these are debt or equity contracts. Hence, for simplicity we set up the base model to look like spot trade.

ital, thus rationalizing their decision to not trade it to end users and making no intermediation an equilibrium.

This strategic effect implies there can be two pure-strategy equilibria, one with and one without intermediation, as well as a mixed-strategy equilibrium where middlemen sometimes pass capital to end users and sometimes keep it. In the context of intermediation theory, this multiplicity seems interesting because it is only possible when holding assets entails positive returns – in the version with negative returns (storage costs) the equilibrium is always unique. This is true even though we tried hard to get multiplicity in that version by going beyond a linear steady-state setting, based on the knowledge that in some related search-and-bargaining models multiplicity can arise only if utility is nonlinear or if we consider dynamics (e.g., Wright and Wong 2014; Trejos and Wright 2016). This does not mean that one can never generate multiplicity with negative returns – we know lots of "tricks" to make that happen, e.g., going back to Diamond (1982), increasing returns – but it shows that one does not need such devises when intermediation involves assets with positive returns.

The finding that middlemen may pass assets on to end users, or hoard them, as a self-fulfilling prophecy, suggests there is something special about financial intermediation as compared to, say, retail.[4] The word hoard is apt because this is reminiscent of monetary economics, where it is standard to show that whether a seller accepts or rejects an asset in payment can depend on what others are doing

[4]While the model is stylized, we interpret the results as providing support for the long-standing notion, associated with names like Minsky and Kindleberger, among others, that financial intermediation and financial markets generally may be particularly susceptible to multiplicity or volatility. See Akerlof and Shiller (2009) for a broad discussion, and Reinhart and Rogoff (2009) for an empirical account. As regards intermediation, in particular, Rolnick and Weber (1986) say that "Historically, even some of the staunchest proponents of laissez-faire have viewed banking as inherently unstable and so requiring government intervention." One such "staunch proponent" proponent is Friedman (1960), who advocated narrow banking regulation in his program for monetary stability. A large literature following Diamond and Dybvig (1983) studies banking instability theoretically. Even if one accepts the view that financial intermediation is subject to instability or multiplicity, however, the reason is open to debate (e.g., Lacker 2014 argues it may be induced by policy and regulation).

(see surveys by Nosal and Rocheteau 2011 and Lagos et al. 2016). Our result is similar, but more like showing that whether a buyer spends (trades away) an asset can depend on what others are doing. Poor assets are always traded by middlemen, while excellent assets are hoarded, in the spirit of Gresham's Law. Whether or not that is surprising, it was less obvious that moderately good assets may be traded or hoarded depending on the equilibrium we select. Moreover, if we fix the sets of middlemen and producers, the strategic complementarity discussed above is inoperative and the equilibrium is unique, showing how endogenizing market composition is crucial for the demonstration. Finally, to be clear, it is worth emphasizing that our analysis does not concern the instability of *credit arrangements* – on that, e.g., see Gu et al. (2013) – but multiple equilibria in *intermediated asset markets*, which seems at least as relevant.

The rest of the paper involves making our assumptions precise and verifying the results. Section 2 describes the environment, while Sections 3 and 4 study equilibrium and efficiency in a stationary linear economy. Sections 5 and 6 consider nonlinear utility and dynamics. Section 7 takes up financial intermediation and Section 8 concludes. Technical results are relegated to an Appendix.[5]

2 Environment

There is a continuum of infinitely-lived agents. Measure n_c of them are end users, or consumers, labeled C. The rest choose to be producers, middlemen or

[5] As additional motivation, it is hard to improve on RW: "Despite the important role played by intermediation in most markets, it is largely ignored by the standard theoretical literature. This is because a study of intermediation requires a basic model that describes explicitly the trade frictions that give rise to the function of intermediation. But this is missing from the standard market models, where the actual process of trading is left unmodeled." The situation has improved since then, and related work on intermediation with occupation choice includes Bigalser (1993), Wright (1995), Li (1998), Camera (2001), Johri and Leach (2002), Shevchenko (2004), Smith (2004), Duffie et al. (2005), Masters (2007), Tse (2009), Lagos and Rocheteau (2009), Watanable (2010), and Geromichalos and Herrenbrueck (2016). Among many other differences, we stay closer to RW by not giving middlemen better infomation or letting them hold larger inventories. For more on the literature see Wright and Wong (2014).

nonparticipants, labeled P, M or N, with measure n_p, n_m or n_n, where $n_c + n_p + n_m + n_n = 1$ and $\mathbf{n} = (n_c, n_p, n_m, n_n)$. Type P produce whenever they can and type M trade with them whenever they can (this is how we define types P and M; those who do not want to act this way are called type N). Type C trade whenever they can, too. Agents meet bilaterally in continuous time according to a uniform random-matching process, with αn_j the Poisson rate at which anyone meets type j, and without loss of generality we set $\alpha = 1$.

There are two goods, x and y. Good x is indivisible, and valued for consumption by type C, who get utility u from consuming a unit. While this is a standard interpretation that we adopt for now, later we think of x as a productive input or asset that produces final goods. In any case, good x is storable, but only 1 unit at a time, at cost γ_p for type P and γ_m for type M. For now $\gamma_j > 0$; to reinterpret x as an asset simply set $\gamma_j < 0$. While x can be produced by type P at cost k, for most purposes we can set $k = 0$ without loss of generality.[6] While M cannot produce x he can acquire it from P to potentially retrade it to C. The other good y is divisible but nonstorable. All agents can produce y at constant marginal cost in terms of utility, normalized to 1, and can consume it for utility $U(y)$, where for now $U(y) = y$, but later we consider $U''(y) < 0$. As in RW, $U(y) = y$ means that transferable utility is used to pay for x.

Type P agents always have 1 unit of x and type C agents always have 0, as the former produce and the latter consume right after trade. Type M can have 0 or 1 unit of x in inventory. If μ is the fraction of M holding x, it increases at rate $n_p n_m (1 - \mu)$ (the measure of P meeting M without x) and decreases at rate $n_c n_m \mu$ (the measure of C meeting M with x). Hence, the steady state is

$$\mu = \frac{n_p}{n_p + n_c}. \tag{1}$$

[6] As in many models (e.g., Pissarides 2000), what matters is the total expected discounted cost, including entry, production and search, so we do not need them all.

We focus for now on steady states, and consider dynamics in Section 6. Also, in Section 7, M might not trade x to C, which requires amendment of (1).

Bargaining determines the terms of trade: agents i and j split the total surplus with θ_{ij} denoting the share, or bargaining power, of i and $\theta_{ji} = 1 - \theta_{ij}$. With transferable utility, this follows from various solution concepts, including Nash, Kalai or strategic bargaining games. The surplus of type C meeting type P is $u - y_{cp} = \theta_{cp}u$, because $y_{cp} = \theta_{pc}u$, given that for both C and P the continuation values and outside options cancel.[7] Similar expressions hold for the other y_{ij}, and we write $\mathbf{y} = (y_{cp}, y_{mp}, y_{cm})$.

Let V_p be P's payoff or value function. Let V_0 or V_1 be M's value function when he has 0 or 1 unit of x. Let V_c and $V_n = 0$ be C's and N's value functions, and $\mathbf{V} = (V_p, V_0, V_1, V_c, V_n)$. Eliminating the y's from the V's using the bargaining solution, we get the dynamic programming equations

$$rV_p = n_c\theta_{pc}u + n_m(1-\mu)\theta_{pm}(V_1 - V_0) - \gamma_p \tag{2}$$

$$rV_0 = n_p\theta_{mp}(V_1 - V_0) \tag{3}$$

$$rV_1 = n_c\theta_{mc}(u + V_0 - V_1) - \gamma_m \tag{4}$$

$$rV_c = n_p\theta_{cp}u + n_m\mu\theta_{cm}(u + V_0 - V_1). \tag{5}$$

In (2), the flow value rV_p is the rate at which P meets C times his share of the surplus, plus the rate at which he meets M without x times his share of that surplus, minus the flow storage cost γ_p. The other equations are similar.

Agents that choose to be type M start without x, for payoff V_0. Hence, occupational choice comes down to the following considerations:

$$n_p > 0 \Rightarrow V_p \geq \max\{V_0, 0\} \text{ and } n_m > 0 \Rightarrow V_0 \geq \max\{V_p, 0\} \tag{6}$$

[7] This is because our agents all stay in the market forever. In the original RW setup, P and C exit after trading, to be replaced by clones, while M stays forever. Nosal et al. (2015) nest these formulations by having agents stay after trading with a type-specific probability; having them stay with probability 1 reduces the algebra without affecting the results too much.

Obviously, $n_p, n_m > 0$ requires $V_p = V_0$. In any case, we have:

Definition 1 *A (steady state) equilibrium is a nonnegative list* $\langle \mu, \mathbf{V}, \mathbf{n} \rangle$ *such that* μ *satisfies (1),* $\mathbf{V}$ *satisfies (2)-(5) and* $\mathbf{n}$ *satisfies (6).*

From this we can compute the terms of trade $\mathbf{y}$, the spread $s = y_{cm} - y_{mp}$, the stock of inventories held by middlemen $n_m \mu$, and other interesting variables.

3 Equilibrium

Given $\gamma_j > 0$, there are three kinds of outcomes. A *class 0 equilibrium* is one where $n_p = n_m = 0$ and $n_n = 1 - n_c$, which means the market shuts down. A *class 1 equilibrium* is one where $n_p = 1 - n_c$ and $n_m = n_n = 0$, with production but no intermediation. A *class 2 equilibrium* is one where $n_p > 0$, $n_m > 0$ and $n_n = 0$, with production and intermediation. In case it is not obvious, the labels are chosen because class 0 implies no active agents, class 1 has one active type, P, and class 2 has two active types, P and M. Also, while in principle there can be equilibria where $n_n > 0$, $n_p > 0$ and $n_m > 0$, it is easy to show this is possible only for a measure 0 set of parameters.

To begin the analysis, consider a candidate class 0 equilibrium, with $n_p = n_m = 0$. This is an equilibrium iff $V_p \leq 0$ and $V_0 \leq 0$. When $n_m = 0$, $V_p \leq 0$ iff $\gamma_p \geq \bar{\gamma}_p \equiv n_c \theta_{pc} u$, and $V_0 = 0$ for all parameters. So class 0 equilibrium exists iff $\gamma_p \geq \bar{\gamma}_p$, and obviously there are not multiple class 0 equilibria. However, unless parameters satisfy the condition in Lemma 1, class 0 equilibrium violates subgame perfection, and hence is ignored (proofs of all results that are not clear from the discussion in the text are contained in the Appendix).

Lemma 1 *A (subgame perfect) class 0 equilibrium exists iff* $\gamma_p \geq \bar{\gamma}_p$ *and* $\gamma_m \geq g(\gamma_p)$*, where* g *is defined in (9) below. When it exists it is unique.*

Consider next a candidate class 1 equilibrium, with $n_p = 1 - n_c$ and $n_m = 0$. For this to be an equilibrium we need $V_p \geq 0$ and $V_p \geq V_0$, so that type P agents do not want to deviate and become type N or M. It is easy to check $V_p \geq 0$ iff $\gamma_p \leq \bar{\gamma}_p$, and $V_p \geq V_0$ iff

$$\gamma_m \geq f(\gamma_p) \equiv \bar{\gamma}_m - \frac{r + n_c\theta_{mc} + (1 - n_c)\,\theta_{mp}}{(1 - n_c)\,\theta_{mp}}(\bar{\gamma}_p - \gamma_p), \tag{7}$$

where $\bar{\gamma}_m \equiv n_c\theta_{mc}u$. Since (2)-(5) are linear, there cannot be multiple class 1 equilibria. This proves:

Lemma 2 *A class 1 equilibrium exists iff $\gamma_p \leq \bar{\gamma}_p$ and $\gamma_m \geq f(\gamma_p)$, where f is defined in (7). When it exists it is unique.*

Consider next class 2, with $n_p, n_m > 0$ and $V_p = V_0 \geq 0$. It is convenient to proceed using μ, and later use steady state to recover $\mathbf{n}$. Then we need $\mu \in (0, \bar{\mu})$, where $\bar{\mu} = 1 - n_c$. Now routine algebra reduces $V_p = V_0$ to $Q(\mu) = 0$, where

$$Q(\mu) = \kappa_1\mu^2 + \kappa_2\mu + \kappa_3 \tag{8}$$

is obtained by replacing n_p and n_m with their values in terms of μ, and the coefficients are[8]

$$\begin{aligned}
\kappa_1 &= \theta_{pm}(\bar{\gamma}_m - \gamma_m) \\
\kappa_2 &= -[2(1 - n_c)\theta_{pm} + n_c](\bar{\gamma}_m - \gamma_m) - (r + n_c\theta_{mc} - n_c\theta_{mp})(\bar{\gamma}_p - \gamma_p) \\
\kappa_3 &= (1 - n_c)\theta_{pm}(\bar{\gamma}_m - \gamma_m) + (r + n_c\theta_{mc})(\bar{\gamma}_p - \gamma_p).
\end{aligned}$$

We seek $\mu \in (0, \bar{\mu})$ such that $Q(\mu) = 0$ and $V_0 \geq 0$. Since $V_0 \geq 0$ iff $\gamma_m \leq \bar{\gamma}_m$, this requires $\kappa_1 > 0$, and hence $Q(\mu)$ convex. As shown by the curves Q_a, Q_b and

[8] Much of the Appendix involves manipulating quadratic equations that naturally arise from random matching and endogenous inventories. It is worth emphasizing that this is *not* due to a matching technology with increasing returns: independent of market size, one meets other agents at a fixed rate, and in a meeting the counterparty is type j with probability n_j; howerever, the outcome when one meets M depends on μ, and μ depends on the type distribution $\mathbf{n}$.

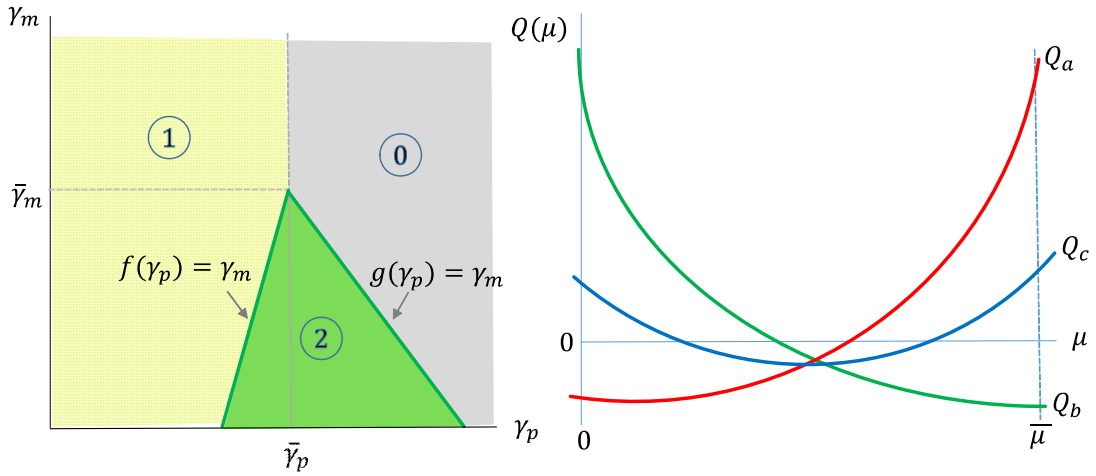


Figure 1: Equilibrium outcomes in (γ_m, γ_p) space

Q_c in the right panel of Fig. 1, there are three ways $Q(\mu)$ can have a solution in $(0, \bar{\mu})$: (a) one root with $Q(0) < 0 < Q(\bar{\mu})$; (b) one root with $Q(0) > 0 > Q(\bar{\mu})$; or (c) two roots. The Appendix rules out cases (a) and (c):

Lemma 3 *A class 2 equilibrium exists iff* $Q(0) > 0 > Q(\bar{\mu})$.

To see when the conditions in Lemma 3 hold, note that $Q(\bar{\mu}) < 0$ iff $\gamma_m < f(\gamma_p)$ where f is defined above, while $Q(0) > 0$ iff $\gamma_m < g(\gamma_p)$ where

$$g(\gamma_p) \equiv \bar{\gamma}_m + \frac{r + n_c\theta_{mc}}{(1 - n_c)\theta_{pm}}(\bar{\gamma}_p - \gamma_p). \tag{9}$$

Also, since there is exactly one $\mu \in (0, \bar{\mu})$ with $Q(\mu) = 0$, and again (2)-(5) are linear, there cannot be multiple class 2 equilibria.

Lemma 4 *A class 2 equilibrium exists iff* $\gamma_m < f(\gamma_p)$ *and* $\gamma_m < g(\gamma_p)$. *When it exists it is unique.*

The outcome is illustrated in the left panel of Fig. 1, drawn with $f(0) < 0$, although $f(0) > 0$ is also possible, so we can get class 1 or 2 equilibrium at the origin. Clearly equilibrium is unique for generic parameter values. The

equilibrating force is this: When n_p increases, P is less likely to meet M, and when he does it is more likely that M already has x. For both reasons higher n_p lowers the incentive to become type P.

Intermediation can be essential in the sense used by monetary theorists: an institution like money is said to be essential if the set of outcomes that can be supported as equilibria expands when money is introduced. Nosal and Rocheteau (2011) and Lagos et al. (2016) discuss work on the essentiality of money, banking and related institutions. For both money and intermediation the concept is nontrivial, since they are clearly *not* essential in the standard environment of general equilibrium theory. Here, in the region where class 2 equilibrium exists with $\gamma_p > \bar{\gamma}_p$, economic activity depends on middlemen: if we were to exogenously eliminate type M, say, by taxing them, the market would shut down. Thus, intermediation may be necessary for production and consumption to be viable. Even if they are viable without intermediation, welfare can be enhanced by middlemen, but can also be diminished, as discussed in Section 4. In any case, we summarize the above results for the benchmark model as follows:

Proposition 1 *Equilibrium exists and is generically unique, as shown in Fig. 1. For some parameters intermediation is essential.*

Additional insights come from changing parameters in a class 2 equilibrium, where μ solves $Q(\mu) = 0$. First, notice anything that shifts $Q(\mu)$ up (down) causes μ to increase (decrease). The Appendix proves the following:

Lemma 5 *An increase in γ_p shifts $Q(\mu)$ down; an increase in γ_m shifts $Q(\mu)$ down if $\gamma_p < \bar{\gamma}_p$ and up if $\gamma_p > \bar{\gamma}_p$.*

Based on these observations, it is immediate that

$$\frac{\partial \mu}{\partial \gamma_p} < 0, \frac{\partial n_p}{\partial \gamma_p} < 0 \text{ and } \frac{\partial n_m}{\partial \gamma_p} > 0.$$

This accords well with intuition: when γ_p is higher, we get fewer producers and more middlemen, and therefore the latter hold x with a lower probability. Less intuitively, we have

$$\begin{aligned}\gamma_p &< \bar{\gamma}_p \Rightarrow \frac{\partial \mu}{\partial \gamma_m} > 0, \frac{\partial n_p}{\partial \gamma_m} > 0 \text{ and } \frac{\partial n_m}{\partial \gamma_m} < 0 \\ \gamma_p &> \bar{\gamma}_p \Rightarrow \frac{\partial \mu}{\partial \gamma_m} < 0, \frac{\partial n_p}{\partial \gamma_m} < 0 \text{ and } \frac{\partial n_m}{\partial \gamma_m} > 0.\end{aligned}$$

The case $\gamma_p > \bar{\gamma}_p$ should be surprising: why are there more middlemen when γ_m is higher? This is answered in Section 4 in the context of efficiency.

In terms of bargaining power, one can check that an increase in θ_{pc} or θ_{pm} shifts $Q(\mu)$ up, raising μ and n_p while lowering n_m, as again accords with intuition. However, just like γ_m, an increase in θ_{mc} can shift $Q(\mu)$ up or down depending on the sign of $\gamma_p - \bar{\gamma}_p$, and therefore

$$\begin{aligned}\gamma_p &< \bar{\gamma}_p \Rightarrow \frac{\partial \mu}{\partial \theta_{mc}} < 0, \frac{\partial n_p}{\partial \theta_{mc}} < 0 \text{ and } \frac{\partial n_m}{\partial \theta_{mc}} > 0 \\ \gamma_p &> \bar{\gamma}_p \Rightarrow \frac{\partial \mu}{\partial \theta_{mc}} > 0, \frac{\partial n_p}{\partial \theta_{mc}} > 0 \text{ and } \frac{\partial n_m}{\partial \theta_{mc}} < 0.\end{aligned}$$

The reason an increase in θ_{mc} works like a decrease in γ_m is that both make intermediation more profitable, with γ_m operating during the search process and θ_{mc} operating during the bargaining process.[9]

We now bring back the terms of trade, $\mathbf{y}$. In direct trade between C and P, $y_{cp} = \theta_{pc} u$ is independent of the sunk cost γ_p, and increases with θ_{pc} or u. In wholesale trade where M gets x from P,

$$y_{mp} = \theta_{pm}(V_1 - V_0) = \frac{\theta_{pm}(n_c \theta_{mc} u - \gamma_m)}{r + n_c \theta_{mc} + n_p \theta_{mp}}.$$

[9] For completness, we mention the impact on $\mathbf{n}$ of r, like γ_p or θ_{mc}, depends on γ_p: $\gamma_p < \bar{\gamma}_p$ implies $\partial\mu/\partial r > 0$, $\partial n_p/\partial r > 0$ and $\partial n_m/\partial r < 0$, while $\gamma_p > \bar{\gamma}_p$ implies the opposite. A demand increase on the intensive margin, captured by higher u, is less clear: Since what matters is γ_p/u and γ_m/u, raising u has the same impact as lowering both γ_p and γ_m. If $\gamma_p > \bar{\gamma}_p$, e.g., then $\partial n_p/\partial u > 0$ and $\partial n_m/\partial u < 0$, but if $\gamma_p < \bar{\gamma}_p$ then the effects can go either way. Similarly ambiguous is an increase in demand on the extensive margin, captured by higher n_c.

Notice n_p on the RHS. Generally, there are direct effects of parameters, plus indirect effects through $\mathbf{n}$. Similarly, in retail trade where C gets x from M,

$$y_{cm} = \theta_{mc}(u + V_0 - V_1) = \theta_{mc}u - \frac{\theta_{mc}(n_c\theta_{mc}u - \gamma_m)}{r + n_c\theta_{mc} + n_p\theta_{mp}}.$$

One can check $\partial y_{mp}/\partial\gamma_p > 0$, $\partial y_{cm}/\partial\gamma_p < 0$ and $\partial s/\partial\gamma_p < 0$, where $s = y_{cm} - y_{mp}$ is the spread. Also, $\gamma_p > \bar{\gamma}_p$ implies $\partial y_{mp}/\partial\gamma_m < 0$, $\partial y_{cm}/\partial\gamma_m > 0$ and $\partial s/\partial\gamma_m > 0$, but $\gamma_p < \bar{\gamma}_p$ implies the effects are ambiguous, as are the effects of bargaining powers. All these results are different when $\mathbf{n}$ is exogenous, since the indirect effects vanish. That is one reason to study occupational choice. Another is to examine the welfare implications.

4 Efficiency

We now solve a planner's problem, where for simplicity the focus is on $r \approx 0$. The problem is to choose $\left(\mu^o, n_p^o, n_m^o\right)$ to maximize:[10]

$$rW = n_p(n_c u - \gamma_p) + \mu n_m(n_c u - \gamma_m).$$

Consider first $\gamma_m > n_c u$, which means intermediation is a bad idea, because it contributes negatively to rW. Given $\gamma_m > n_c u$ we have: $\gamma_p > n_c u$ implies $n_p^o = 0$ and the market shuts down; and $\gamma_p < n_c u$ implies $n_p^o = 1 - n_c$ and the market opens with direct trade only.

Next consider $\gamma_m < n_c u$, which means intermediation may or may not be a good idea. Eliminating n_p and n_m we reduce the problem to

$$\max_{\mu\in[0,\bar{\mu}]} \left\{ \mu n_c u - n_c \frac{\mu}{1-\mu}\gamma_p - \mu\frac{1-n_c-\mu}{1-\mu}\gamma_m \right\}.$$

[10] The first (second) term is the net social surplus from direct (indirect) trade. Similar to a related analysis in Nosal et al. (2015), one can solve the much harder dynamic problem with $r > 0$ and, as usual, the outcome is the same as maximizing rW when $r \to 0$. As is standard, small r can be interpreted as saying search frictions are not too severe.

After simplification, the derivative of the objective function is proportional to

$$Q^o(\mu) = (1-\mu)^2(n_c u - \gamma_m) + n_c(\gamma_m - \gamma_p), \tag{10}$$

a quadratic that decreases with μ over the relevant range.

Hence, with $\gamma_m < n_c u$ there are three possibilities. First, $\mu^o = 0$, which corresponds to a class 0 outcome with $n_p^o = 0$. It is easy to check this occurs iff

$$\gamma_m > g^o(\gamma_p) \equiv \frac{n_c(u - \gamma_p)}{1 - n_c}. \tag{11}$$

Second, $\mu^o = \bar{\mu}$, which corresponds to a class 1 outcome with $n_p^o = 1 - n_c$ and $n_m = 0$. This occurs iff

$$\gamma_m > f^o(\gamma_p) \equiv \frac{-n_c^2 u + \gamma_p}{1 - n_c}. \tag{12}$$

Finally, if $\gamma_m < f^o(\gamma_p), g^o(\gamma_p)$ there is a unique $\mu^o \in (0, \bar{\mu})$ solving $Q^o(\mu^o) = 0$, a class 2 outcome. This is all summarized as follows:

Lemma 6 *The efficient outcome has $n_p^o = n_m^o = 0$ iff $\gamma_p \geq n_c u$ and $\gamma_m \geq g^o(\gamma_p)$; $n_p^o = 1 - n_c$ and $n_m^o = 0$ iff $\gamma_b < n_c u$ and $\gamma_m \leq f^o(\gamma_p)$; and $n_p^o, n_m^o > 0$ iff $\gamma_m \leq g^o(\gamma_p)$ and $\gamma_m \leq f^o(\gamma_p)$.*

Fig. 2 shows f^o and g^o, as well as the analogs from the equilibrium analysis, f and g. Notice the region where $n_m^o > 0$ lies strictly below the 45^o line, so for intermediation to be optimal we need $\gamma_m < \gamma_p$, By comparison, $\gamma_m < \gamma_p$ is neither necessary nor sufficient for $n_m > 0$ to be an equilibrium. Also note that the results are different from models with fixed $\mathbf{n}$. In such models, if γ_m is close to γ_p it is always efficient for P to trade x to M, so P can produce another unit and put more x on the market. The economics is different here because M can turn into P and produce output on his own. This is summarized as follows:

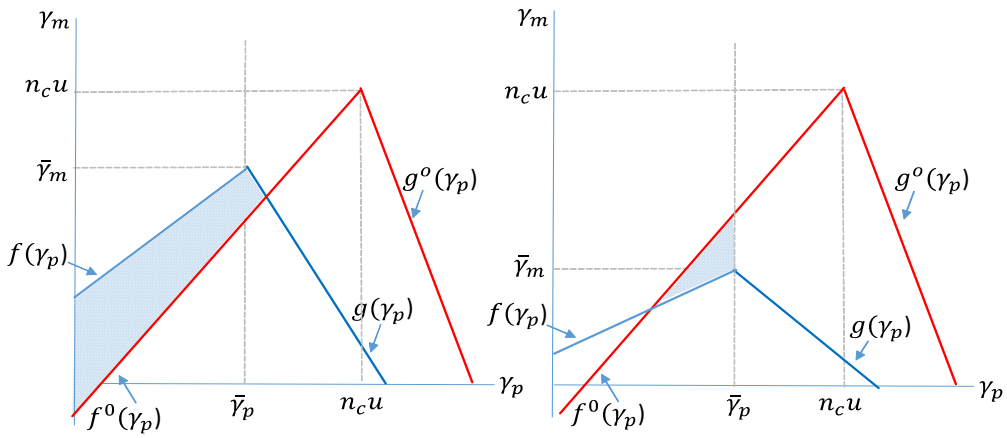


Figure 2: Comparing equilibrium and efficient outcomes

Proposition 2 *The efficient outcome exists and is generically unique, as shown in Fig. 2.*

Before further comparing the efficient and equilibrium outcomes, consider the effects of parameters on the planner's solution when $\mu^o \in (0, \bar{\mu})$. First,

$$\frac{\partial \mu^o}{\partial \gamma_p} < 0, \frac{\partial n_p^o}{\partial \gamma_p} < 0 \text{ and } \frac{\partial n_m^o}{\partial \gamma_p} > 0,$$

which is similar to the equilibrium results. Next,

$$\begin{aligned} \gamma_p &< n_c u \Rightarrow \frac{\partial \mu^o}{\partial \gamma_m} > 0, \frac{\partial n_p^o}{\partial \gamma_m} > 0 \text{ and } \frac{\partial n_m^o}{\partial \gamma_m} < 0 \\ \gamma_p &> n_c u \Rightarrow \frac{\partial \mu^o}{\partial \gamma_m} < 0, \frac{\partial n_p^o}{\partial \gamma_m} < 0 \text{ and } \frac{\partial n_m^o}{\partial \gamma_m} > 0, \end{aligned}$$

which is similar to the equilibrium results and again surprising. To explain why higher γ_m can lead to more middlemen, the following is useful:

Lemma 7 *For all parameters,* $\partial(n_m^o \mu^o)/\partial \gamma_m < 0$.

Here is the explanation: If γ_m increases, the natural response is to reduce inventories held by M, given by $n_m \mu$, but there are different ways to do so. One

is to reduce n_m, which in steady state means higher μ; the other is to reduce μ, which means higher n_m. When $\gamma_p < n_c u$ it is optimal to use the extensive margin and reduce n_m; when $\gamma_p > n_c u$ it is optimal to use the intensive margin and reduce μ, which means higher n_m. This explains the planner's choices. The idea is similar for equilibrium, but less transparent, as complications can make that different from the efficient outcome, as we now discuss.

First note that equilibrium can involve too many type P and too few type M, or vice versa. In the shaded region in the left panel of Fig. 2, between $f(\gamma_p)$ and $f^o(\gamma_p)$, we have $n_m > 0 = n_m^o$ and equilibrium has too many middlemen. There is also a region where equilibrium has too few. The situation in the right panel, drawn for different parameters, is similar. Also, even if the equilibrium and efficient outcomes are both class 2, we only get $n_m = n_m^o$ if bargaining powers are just right. To see this, define S_0^o, S_1^o and S_2^o as the sets of γ's where the efficient outcome is class 0, class 1 and class 2, respectively. Then we have:

Proposition 3 *Equilibrium is efficient iff* $\theta_{pc}^o = \theta_{mc}^o = 1$ *and: (i)* $(\gamma_p, \gamma_m) \in S_0^o \Rightarrow \theta_{pm}^o = 1$*; (ii)* $(\gamma_p, \gamma_m) \in S_1^o \Rightarrow \theta_{pm}^o = 0$*; and (iii)* $(\gamma_p, \gamma_m) \in S_2^o \Rightarrow$

$$\theta_{pm}^o = \frac{(1-\mu^o)(1-n_c-\mu^o)}{(1-\mu^o)(1-n_c-\mu^o) + \mu^o n_c[1-(n_c u-\gamma_p)/(n_c u-\gamma_m)]} \in (0,1).$$

Heuristically, $\theta_{pc}^o = 1$ and $\theta_{mc}^o = 1$ avoid holdup problems associated with the costs γ_p and γ_m, which are sunk when P and M deal with the end user C. For θ_{pm}^o, there is also a holdup problem when P deals with M, but in this case other forces come into play. When someone chooses to be type P, he considers his own benefit and cost, but neglects the fact that at the margin he makes it harder for other P's to meet M's and easier for M's to meet P's. In addition, having more P's increases μ, and that makes it harder for a type P agent to trade when he does meet M's. Balancing these considerations delivers θ_{pm}^o.

5 Concavity

Now suppose $U''(y) < 0$, with $U(0) = 0$ and $U(\hat{y}) = \hat{y}$ for some $\hat{y} > 0$.[11] It is interesting to go beyond transferrable utility, for various reasons, but here's a big one. If $y > \hat{y}$, the cost to the payer exceeds the value to the payee, which discourages intermediation, since indirect trade requires two payments, M to P and C to M, rather than one, C to P. Nonlinearity thus can represent a transaction cost on payments. Note $y > \hat{y}$ is possible, as even with $U(y) < y$ the surplus can be positive due to the gain from trading x. Indeed, we do not impose $U'(0) > 1$, so it may be that $U(y) < y \ \forall y > 0$.

For tractability here we use Kalai's (1977) proportional bargaining solution.[12] While we specialize below, the general dynamic programming equations are

$$
\begin{aligned}
rV_p &= n_c\theta_{pc}\left[U(y_{cp}) - y_{cp} + u\right] + n_m(1-\mu)\theta_{pm}\left[U(y_{mp}) - y_{mp} + V_1 - V_0\right] - \gamma_p \\
rV_0 &= n_p\theta_{mp}\left[U(y_{mp}) - y_{mp} + V_1 - V_0\right] \\
rV_1 &= n_c\theta_{mc}\left[U(y_{cm}) - y_{cm} + u + V_0 - V_1\right] - \gamma_m \\
rV_c &= n_p\theta_{cp}\left[U(y_{cp}) - y_{cp} + u\right] + n_m\mu\theta_{cm}\left[U(y_{cm}) - y_{cm} + u + V_0 - V_1\right].
\end{aligned}
$$

Now for simplicity, and efficiency, consider $\theta_{pc} = \theta_{mc} = 1$, so that $V_c = 0$ and $y_{cp} = y_{cm} = u$. Then letting $z = U(u)$, we simplify the above equations to

$$
\begin{aligned}
rV_p &= n_c z + (1 - n_c - \mu)\theta_{pm} U(y_{mp}) - \gamma_p && (13) \\
rV_0 &= \frac{n_c\mu(1-\theta_{pm})}{(1-\mu)\theta_{pm}} U(y_{mp}) && (14) \\
rV_1 &= n_c(z + V_0 - V_1) - \gamma_m. && (15)
\end{aligned}
$$

[11] Although we think this extension is important for reasons dicussed below, in can be skipped with no loss in continuity.

[12] The Kalai solution says that when i trades with j we maximize i's surplus subject to him getting a share θ_{ij} of the total surplus. This is not the definition of Kalai bargaining, it is a result about the outcome implied by his axioms, like maximizing the product of the surpluses is a result implied by Nash's axioms. If $U(y) = y$ in our setting Nash and Kalai are the same; with $U'' < 0$ they are not. While we could use Nash, Kalai has several advantages, including simplicity, as argued by Aruoba et al. (2007) in the context of related models.

The solution method is similar to the linear model, although the algebra is more cumbersome – which, by the way, explains why we take $U(y) = y$ as our benchmark. The obvious analog to Lemma 1 is:

Lemma 8 *A (subgame perfect) class 0 equilibrium exists iff* $\gamma_p \geq n_c z$ *and* $\gamma_p \geq G(\gamma_m)$*, where*

$$G(\gamma_m) \equiv nz + U(y_0)(1-n), \tag{16}$$

and y_0 *is given by the bargaining solution for* y_{mp} *with* $\mu = 0$*.*

Notice $\gamma_p \geq G(\gamma_m)$ replaces $\gamma_m \geq g\left(\gamma_p\right)$ from the linear specification, since now we can solve for γ_p but not γ_m. Also, to be clear, $G(\gamma_m)$ depends on γ_m because y_{mp} in (16) solves the bargaining problem given γ_m. Similarly, the analog to Lemma 2 is:

Lemma 9 *A class 1 equilibrium exists iff* $\gamma_p \leq n_c z$ *and* $\gamma_p \leq F(\gamma_m)$*, where*

$$F(\gamma_m) \equiv n_c z - U(\bar{y})(1-n_c)\frac{(1-\theta_{pm})}{\theta_{pm}}, \tag{17}$$

and $\bar{y}$ *is the bargaining solution for* y_{mp} *when* $\mu = \bar{\mu}$*.*

Here $\gamma_p \leq F(\gamma_m)$ replaces $\gamma_m \geq f\left(\gamma_p\right)$, and F depends on γ_m, similar to the discussion of $G(\gamma_m)$.

For class 2 equilibrium, it must be that $n_c z > \gamma_m$, μ satisfies the occupational choice condition $V_p = V_0$, and y_{mp} is the bargaining solution. Now $V_p = V_0$ reduces to $\tilde{Q}(\mu, y_{mp}) = 0$, where $\tilde{Q}(\mu, y_{mp}) = \tilde{\kappa}_1 \mu^2 + \tilde{\kappa}_2 \mu + \tilde{\kappa}_3 = 0$ and

$$\begin{aligned}
\tilde{\kappa}_1 &= \theta_{pm} U(y_{mp}) \\
\tilde{\kappa}_2 &= -\left[2\theta_{pm}(1-n_c) + n_c\right] U(y_{mp}) - \theta_{pm}\left(n_c z - \gamma_p\right) \\
\tilde{\kappa}_3 &= \theta_{pm}\left(n_c z - \gamma_p\right) + \theta_{pm}(1-n_c) U(y_{mp}).
\end{aligned}$$

Here is the analog to Lemma 3:

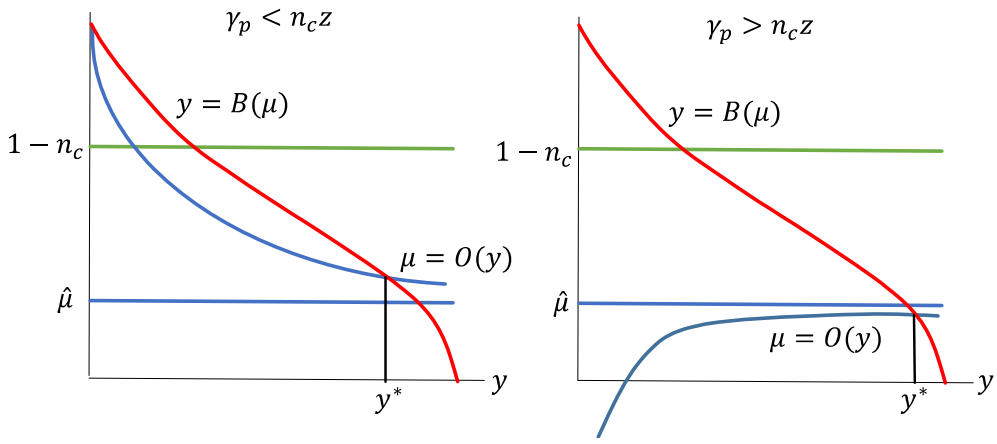


Figure 3: Equilibrium in (y_{mp}, μ) space

Lemma 10 *A class 2 equilibrium exists iff* $\tilde{Q}(0, y_0) > 0 > \tilde{Q}(\bar{\mu}, \bar{y})$.

With a general $U(y)$ we cannot eliminate y_{mp} from the equilibrium conditions, so we work with two curves in (y_{mp}, μ) space representing bargaining and occupational choice. Setting $V_0 = V_p$ implies a quadratic we can solve for

$$\mu = \frac{[2\theta_{pm}(1 - n_c) + n_c] U(y_{mp}) + \theta_{pm}(n_c z - \gamma_p) - \sqrt{\tilde{D}}}{2\theta_{pm} U(y_{mp})} \tag{18}$$

where $\tilde{D}$ is the discriminant. This defines a function $\mu = O(y_{mp})$, for occupational choice. One can check $\partial O / \partial y_{mp} \simeq -(n_c z - \gamma_p)$, where $a \simeq b$ means a and b have the same sign. As shown in Fig. 3, this traces a curve in (y_{mp}, μ) space that slopes up or down, depending on the sign of $n_c z - \gamma_p$, but in any case $\lim_{y_{mp} \to \infty} O(y_{mp}) = \hat{\mu} \in (0, \bar{\mu})$.

Next, using (14)-(15) to solve for $V_1 - V_0$ and eliminating it from the Kalai solution, $U(y_{mp}) = \theta_{pm}[U(y_{mp}) + V_1 - V_0]$, we get $y_{mp} = B(\mu)$, for bargaining. In fact, it can be solved for $\mu = B^{-1}(y_m)$ explicitly:

$$\mu = \frac{\theta_{pm}(n_c z - \gamma_m) - \Upsilon}{\theta_{pm}(n_c z - \gamma_m) - \Upsilon + n_c(1 - \theta_{pm}) U(y_{mp})} \tag{19}$$

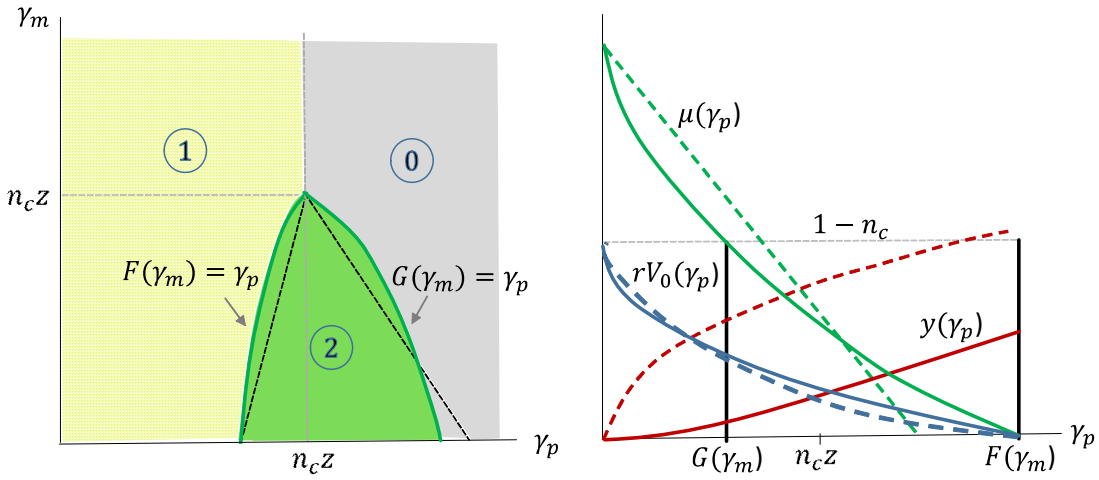


Figure 4: The nonlinear (solid) and linear (dashed) models

where $\Upsilon \equiv (r + n_c)\left[\theta_{pm} y_{mp} + (1 - \theta_{pm})\, U\left(y_{mp}\right)\right]$. This traces a downward-sloping curve, as shown in Fig. 3. The Appendix proves the following results:

Lemma 11 *The curves $y_{mp} = B(\mu)$ and $\mu = O(y_{mp})$ shown in Fig. 3 intersect at $(y_{mp}, \mu) \in (0, \infty) \times (0, \bar{\mu})$, and hence a class 2 equilibrium exists iff $F(\gamma_m) < \gamma_p < G(\gamma_m)$, where F and G are defined in (16) and (17). Moreover, in (γ_p, γ_m) space, F is increasing and concave, G is decreasing and concave, and $F(n_c z) = G(n_c z) = n_c z$.*

Lemma 12 *The curves $y_{mp} = B(\mu)$ and $\mu = O(y_{mp})$ in Fig. 3 cannot intersect more than once in $(0, \infty) \times (0, \bar{\mu})$.*

Combining these results, and once again ignoring equilibrium when it is not subgame perfect, we have this:

Proposition 4 *Equilibrium exists and is generically unique in the nonlinear model, as shown in Fig. 4.*

In the left panel of Fig. 4, the solid curves are F and G for $U(y) = y^{\zeta}$ with $\zeta = 0.3$, $\theta_{pm} = 1/2$, $n_c = 1/2$, $z = 2.2$ and $r = 0.01$. For comparison the

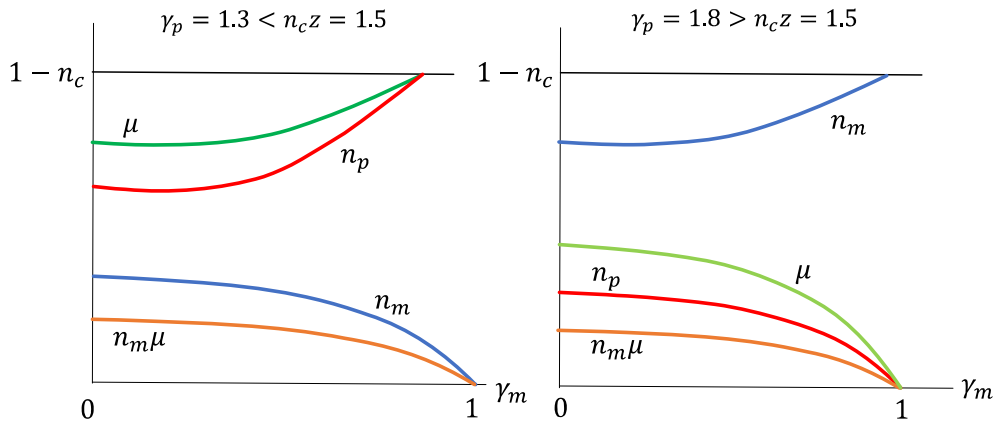


Figure 5: Effects of γ_m in the nonlinear model

dashed lines are for $\zeta = 1$. Notice the set in (γ_p, γ_m) space where $n_m > 0$ does not necessarily expand or contract with increased curvature in $U(\cdot)$, as the dashed and solid curves cross; this is because nonlinearity tends to discourage intermediation when $y_{mp} > \hat{y}$ and encourage it when $y_{mp} < \hat{y}$. In the right panel of Fig. 4, the solid curves are y_{mp}, μ and rV_0 as functions of γ_p for the nonlinear model, while the dashed curves are for the linear model. Fig. 5 shows where higher γ_m can lower or raise n_m, as in the linear specification, which can be proved using the following easily-verified result:

Lemma 13 *An increase in γ_p shifts the O curve down and does not affect the B curve in Fig. 3, while an increase in γ_m shifts the B curve down and does not affect the O curve.*

6 Dynamics

The next extension concerns dynamic transitions given an initial μ.[13] We focus on class 2 equilibrium, since class 0 and 1 have no interesting dynamics (although

[13] As with Section 5, we think this extension is useful, but it can be skipped without loss of continuity.

we can potentially start with, e.g., $n_m = 0$, then transit to $n_m > 0$, or vice versa). As above, set $\theta_{pc} = \theta_{mc} = 1$, so that $y_{cp} = y_{cm} = u$, $V_c = 0$ and $z = U(u)$, and let $\Delta = V_1 - V_0$. Also, at any point in time time we let a type P agent dispose of x and become type M, but he cannot start as type M with his own output, say because he must use x to adopt the middleman technology. Also, we now work with the n's, rather than μ, since n_1 is a state variable, with law of motion

$$\dot{n}_1 = n_0(1 - n_c - n_1 - n_0) - n_1 n_c. \tag{20}$$

In contrast, n_0 can jump at any time to satisfy $V_0 = V_p$, just like vacancies can jump in Pissarides' (2000) labor-market model.

The bargaining solution for y_{mp} is

$$U(y_{mp}) = \theta_{pm}[U(y_{mp}) - y_{mp} + \Delta]. \tag{21}$$

The analogs to (13)-(15), without imposing steady state, are

$$\begin{aligned} rV_p &= n_c z + n_0 U(y_{mp}) - \gamma_p + \dot{V}_p & (22)\\ rV_0 &= (1 - n_c - n_1 - n_0)\frac{1 - \theta_{pm}}{\theta_{pm}} U(y_{mp}) + \dot{V}_0 & (23)\\ rV_1 &= n_c(z - \Delta) - \gamma_m + \dot{V}_1. & (24) \end{aligned}$$

We now show how to reduce this dynamic system to something manageable.

First notice that $V_p = V_0$ implies $\dot{V}_p = \dot{V}_0$, and then from (22)-(23) the occupational choice condition becomes

$$n_c z + n_0 U(y_{mp}) - \gamma_p - (1 - n_c - n_0 - n_1)\frac{1 - \theta_{pm}}{\theta_{pm}} U(y_{mp}) = 0. \tag{25}$$

Next, subtracting (23)-(24), we get

$$r\Delta = n_c(z - \Delta) - \gamma_m + \dot{\Delta} - (1 - n_c - n_0 - n_1)\frac{1 - \theta_{pm}}{\theta_{pm}} U(y_{mp}) = 0.$$

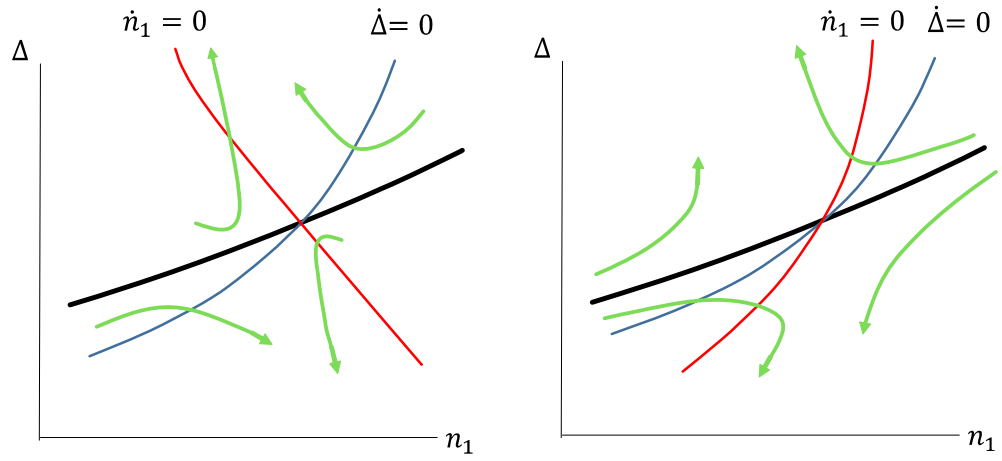


Figure 6: Saddle path stability

Substituting (25) and simplifying, we arrive at

$$\dot{\Delta} = (r + n_c)\Delta + \gamma_m - \gamma_p + n_0 U(y_{mp}). \tag{26}$$

Then (20) and (26) define a two-dimensional system in (n_1, Δ) space, where n_0 and y_{mp} are functions of (n_1, Δ) given by the free entry and bargaining conditions.

In Section 5 it was verified there exists a unique steady state, which is the intersection of the curves along which $\dot{n}_1 = 0$ and $\dot{\Delta} = 0$. These curves have slopes after simplification given by

$$\begin{aligned}
\frac{\partial \Delta}{\partial n_1}|_{\dot{n}_1=0} &= \frac{[(n_0 + n_c) + (1 - n_c - n_1 - 2n_0)(1 - \theta_{pm})]\,[(1 - \theta_{pm})U' + \theta_{pm}]}{(1 - n_c - n_1 - 2n_0)\frac{U'}{U}[(1 - \theta_{pm})(1 - n_c - n_1) - n_0]\theta_{pm}} \\
\frac{\partial \Delta}{\partial n_1}|_{\dot{\Delta}_1=0} &= \frac{U(1 - \theta_{pm})\,[(1 - \theta_{pm})U' + \theta_{pm}]}{(r + n_c)\,[(1 - \theta_{pm})U' + \theta_{pm}] + U'(1 - \theta_{pm})(1 - n_c - n_1)\theta_{pm}}.
\end{aligned}$$

The slope of the $\dot{\Delta} = 0$ curve is strictly positive. The slope of the $\dot{n}_1 = 0$ curve can be positive or negative, but if it is positive one can check it is steeper than the $\dot{\Delta} = 0$ curve. Also note that $\partial \dot{n}_1 / \partial n_1 < 0$ and $\partial \dot{\Delta} / \partial n_1 < 0$. Hence, the system looks like Fig. 6. Whether the $\dot{n}_1 = 0$ curve slopes down (left panel) or up (right panel), the steady state exhibits saddle-path stability.

Proposition 5 *The class 2 steady state is a saddle point.*

Therefore, given an initial condition for $\bar{n}_1$, there is a unique initial $\bar{\Delta}$ such that (n_1, Δ) transits to the steady state, and any $\Delta \neq \bar{\Delta}$ implies an explosive path that cannot be an equilibrium. So equilibrium, not only steady state, is unique. This was not ex ante obvious – e.g., by analogy, there is a unique equilibrium in the linear version of a related model by Duffie et al. (2005), but there can be multiplicity in the nonlinear version (Trejos and Wright 2016). Hence, establishing saddle-path stability in these kinds of models is useful and not trivial.

7 Intermediated Asset Markets

The theory developed above applies to many types of intermediation where there is a cost of storing inventories, as in a typical retail establishment. Now, to consider a different application, consider changing the benchmark specification so that $\gamma_j < 0$ – i.e., storing inventories is profitable rather than costly – as surely applies in some interesting situations. Suppose, e.g., that P is producing and M is dealing in fine art. Then the net benefit of holding x is $\rho_j = -\gamma_j > 0$ if fine art generates positive utility. In such an environment, $\rho_m > \rho_p$ corresponds to an art dealer getting more from the piece than the original artist, perhaps by charging admission to his gallery. If an end user, here an art consumer/collector C, enjoys it enough M may sell it to him; then again, he may prefer to keep it. This an option is never relevant when $\gamma_j > 0$ – retailers obviously will not pay to store inventories unless they are planning to sell them.[14]

More generally, x can be any asset that generates positive returns, and this leads us to an interpretation in terms of financial intermediation. Suppose P produces or otherwise gets access to capital available for investment. An end

[14]The way the model is designed, type P agents always prefer to trade x rather than keep it, since they can always produce more. Of course we can change this by making different assumptions, but the current specification puts the focus squarely on M's decisions.

user C may have the best ultimate use for the capital, but P and M can also generate flow yields by putting it into temporary investments – presumably liquid ones if they are planning on trading the capital to someone else. Then $\rho_m > \rho_p$ means that M has better liquid investment opportunities than P, and ρ_m might even be high enough for M to keep x, rather than passing it on to C, depending on various considerations including bargaining power. If M does pass x to C he is acting as a financial intermediary, by acquiring capital from one agent and retrading it to another.[15]

These applications make it interesting to consider $\gamma_j < 0$, and in terms of theory $\gamma_j < 0$ turns out to generate novel results, without otherwise changing the environment. The dynamic programming equations are in fact the same, although we need a new endogenous variable τ, for the probability that M trades x to C. Now V_1 is M's payoff to holding x with the intention of retrading it, but he actually prefers to keep it if $rV_1 < -\gamma_j$. Generally, the best response conditions are:

$$n_m = \begin{cases} 0 & \text{if } V_p > V_0 \\ [0, 1-n_c] & \text{if } V_p = V_0 \\ 1-n_c & \text{if } V_p < -\gamma_m/r \end{cases} \quad \text{and } \tau = \begin{cases} 0 & \text{if } rV_1 < -\gamma_j \\ [0,1] & \text{if } rV_1 = -\gamma_j \\ 1 & \text{if } rV_1 > -\gamma_j \end{cases}$$

The only other change is that now τ affects the steady state,

$$\mu = \frac{1-n_c-n_m}{1-n_c-n_m+n_c\tau}.$$

Consider steady state equilibria, with $U(y) = y$, which allows us to make the key points in a relatively simple way.[16] One can imagine 9 candidate equilibria,

[15] Other interpretations and applications are easy to describe – e.g., x might be a house that provides a flow payoff as shelter, in which case M might want to acquire x as a residence, or might to retrade (flip) it to type C.

[16] One of these points is that there can be multple equilibria, which is obviously true with concavity or out of steady state if we can show it with linerity in steady state. Indeed, a reason to analyze concavity and dynamics above is that it shows uniquness in the benchmark model is somewhat robust, making the multplicity below all the more interesting.

shown in Table 1, which does not include anything corresponding to a class 0 outcome, because production always dominates nonparticipation with $\gamma_j < 0$. If $n_m = 0$, in the first row of Table 1, $\tau = 1$ corresponds to a class 1 outcome with no type M agents on the equilibrium path, but off the equilibrium path they trade x to C. We call this a class 1^T equilibrium (T indicates M trades x). Similarly, if $n_m = 0$ we call $\tau = 0$ a class 1^K equilibrium, because off the equilibrium path M with x would not trade it to C (K indicates M keeps x). And if $n_m = 0$ we call $\tau \in (0,1)$ a class 1^R equilibrium (R indicates M randomizes), but such an equilibrium can be ruled out for generic parameters.

$n_m \backslash \tau$	0	$[0,1]$	1
0	1^K	$\times$	1^T
$[0, 1-n_c]$	2^K	2^R	2^T
$1-n_c$	$\times$	$\times$	$\times$

Table 1: Candidate equilibria with $\rho_j = -\gamma_j > 0$.

We can also rule out $n_m = 1-n_c$ and either $\tau \in (0,1)$ or $\tau = 1$: if $n_m = 1-n_c$ there are no producers, so trading away x leaves M with continuation value 0, which implies a profitable deviation, since M can become P. We cannot rule out $n_m = 1 - n_c$ and $\tau = 0$, but ignore it in what follows because it is degenerate (no production). In addition to class and 1^K and 1^T, therefore, the remaining candidates are $n_m \in [0, 1-n_c]$ and either $\tau = 1$, $\tau = 0$ or $\tau \in (0,1)$, called class 2^K, 2^T or 2^R (there are both type P and M, and M either trades x, keeps it or randomizes). The following is proved in the Appendix and illustrated in Fig. 7.

Lemma 14 *The Appendix defines $\widehat{\gamma}_m$, $\widehat{f}(\cdot)$, $k(\cdot)$, and $\widehat{k}(\cdot)$, shown in Fig. 7. Class 1^T equilibrium exists iff $\gamma_m \geq \widehat{\gamma}_m, f(\gamma_p)$. Class 1^K exists iff $\widehat{f}(\gamma_p) \leq \gamma_m \leq \widehat{\gamma}_m$. Class 2^K exists iff $\gamma_m \leq \min\{k(\gamma_p), \widehat{f}(\gamma_p)\}$. Class 2^R exists iff $\widehat{k}(\gamma_p) < \gamma_m < k(\gamma_p)$. Class 2^T exists iff $\widehat{k}(\gamma_p) \leq \gamma_m \leq f(\gamma_p)$.*

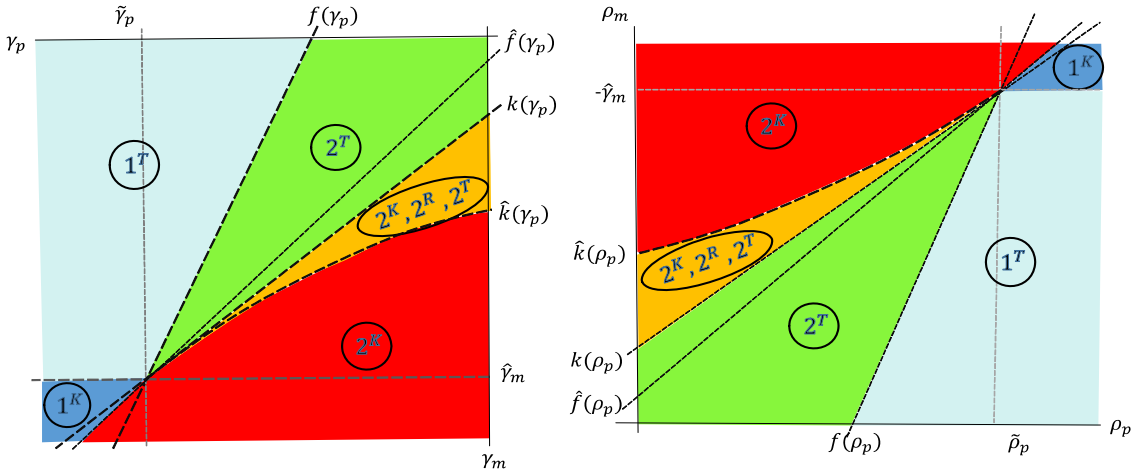


Figure 7: Outcomes in (γ_p, γ_m) and in (ρ_p, ρ_m) space

Fig. 7 shows the results in two ways: in the negative quadrant of (γ_p, γ_m) space; and equivalently, in the positive quadrant of (ρ_p, ρ_m) space, which may be easier for interpreting the economics. Notice that all of the above-mentioned candidate equilibria exist for some parameters. Naturally, if ρ_p is high relative to ρ_m, no one wants to be a middleman, but if, off the equilibrium path, there were a middleman with x he would trade it to C at low values of ρ_m (light blue region), and keep it at high values of ρ_m (dark blue region). But for $\rho_p < \tilde{\rho}_p$, someone always wants to be a middleman, $n_m \in (0, 1 - n_c)$. In this situation there are three possibilities: if ρ_m is small type M agents always trade x to C; if ρ_m is big they never trade x to C; and if ρ_m is neither too big nor too small they randomize. Importantly, for intermediate values of (ρ_p, ρ_m) there are multiple equilibria (orange region).

The economics is clear: If type M agents decide to trade assets to end users, at any point in time there are more middlemen without assets. Since type P can trade with either end users or middlemen in search of capital, this raises V_p, which raises n_p and lowers n_m. Higher n_p makes is easier for type M to acquire capital, rationalizing the decision of middlemen to trade it to end users, and

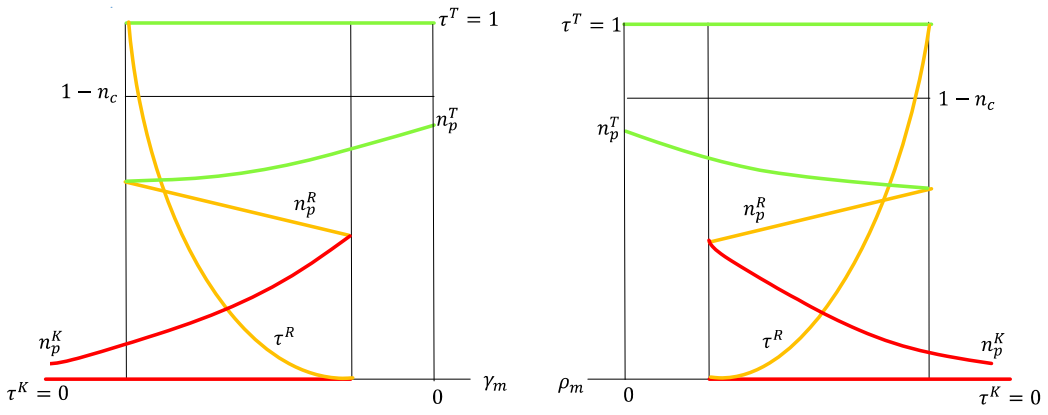


Figure 8: Equilibrium correspondence for τ and n_p

making active intermediation an equilibrium. However, if type M agents instead decide to keep capital for their own uses, type P can only trade with end users, which lowers V_p and n_p. Lower n_p makes it harder for type M agents to get capital, rationalizing their decision to keep it, and making no intermediation an equilibrium. This strategic effect implies class $2T$ and class $2K$ equilibria coexist in a region of parameter space with positive measure; and as usual, when the two pure-strategy equilibria coexist, so does the mixed-strategy equilibrium 2^R.

Hence, whether assets circulate or get hoarded can be a self-fulfilling prophecy. We emphasize that this is only possible when holding assets generates positive returns – equilibrium is unique with storage costs, $\gamma_j > 0$. This suggests there is something special about intermediated asset markets, as compared to, say, retail good markets, consistent with the discussion in fn. 4. Intuitively, type M agents always trade away assets that give them very low returns, always hoard assets that give them very high returns, and may or may not trade assets that give them moderately good returns, depending on the equilibrium selected. As Fig. 6 show, n_p and $n_m = 1 - n_c - n_p$ vary within an equilibria as γ_m changes, and across equilibria as τ changes for a fixed γ_m. This is important: part (ii) of the next result says that if $\mathbf{n}$ is fixed exogenously then equilibrium is unique, even with

$\gamma_j < 0$. Intuitively, unless $\mathbf{n}$ adjusts in response, the strategic chanel discussed above is severed, and hence τ depends on only on fundamentals. Theoretically, this is one more reason to endogenize market composition.

Proposition 6 *(i) With $\gamma_j < 0$ equilibrium exists. As shown in Fig. 7, $\forall \gamma_p < \tilde{\gamma}_p$ where $\tilde{\gamma}_p < 0$, and γ_m neither too high nor too low (equivalently, $\forall \rho_p < \tilde{\rho}_p$ where $\tilde{\rho}_p > 0$, and ρ_m neither too high nor too low) class 2^K, 2^T and 2^R equilibria coexist. Otherwise equilibrium is unique. (ii) If n_m and n_p are fixed exogenously uniqueness re-emerges: $\rho_m = -\gamma_m > (r + n_p \theta_{mp})\, u \Rightarrow \tau = 0$ and $\rho_m < (r + n_p \theta_{mp})\, u \Rightarrow \tau = 1$.*

8 Conclusion

This project has continued the development of search-and-bargaining theories of intermediation. We built on a standard framework extended to allow general bargaining, technology (search or storage costs), and preferences, and most importantly to allow agents to choose to act as either producers or middlemen. This is relevant for many subtantive issues.[17] We established existence, and discussed essentiality as well as efficiency. The theory delivered clean and sometimes surprising predictions – e.g., $\partial n_m / \partial \gamma_m > 0$ for some parameters. Perhaps the most surprising finding was that equilibrium was unique when holding goods involved storage costs, while multiplicity emerged when holding assets involved positive returns, suggesting that there is something special about asset markets.

[17] We reiterate that in our model economy the only way to get more intermediaries is to have fewer producers, arguably capturing a realistic trade-off. We also mention that our setup is less dependent than earlier models on certain simplifying assumptions, including the restriction of inventories to $\{0, 1\}$ (as discussed more in the next paragraph in the text). In particular, in those earlier models, when M takes x from P, the latter can produce again, leading to more output; here, in contrast, if M does not take x from P he can become a producer and make his own, so intermediation does more than try to get around the unit-inventory restriction. Other features of the model let us consider additional issues, including $U'' < 0$, which captures payment frictions that affect the incentive to intermediate. Dynamics are also interesting, with n_m, n_p and μ varying during transitions, e.g., after any shock. Weill (2007) studies this kind of transition in a related but different model, and discusses why it is relevant.

There are many potential extensions and applications. Of course it is desirable to go beyond unit inventories, $x \in \{0, 1\}$, just like it is in search-based models of money by Kiyotaki and Wright (1993), banking by Cavalcanti and Wallace (1999), and over-the-counter markets by Duffie et al. (2005). It is also desirable to generalize the labor model in Pissarides' (2000), where a firm can only employ $n \in \{0, 1\}$ workers, and the marriage model in Burdett and Coles (1997) or Shimer and Smith (2000), where one can only have $n \in \{0, 1\}$ partners. But sometimes the simplicity of unit inventories is worth the loss in generality, going back in search theory at least to Diamond (1982). Moreover, suppose middlemen can hold, say, $x \in \{0, 1, ...\bar{x}\}$. Without working through the details, higher n_m and lower n_p should still entail higher V_p, so we could still endogenize $\mathbf{n}$. Also, if M hoards (trades) capital, V_p and hence n_p should fall (rise), making it harder (easier) for M to acquire capital and rationalizing the hoarding (trading) decision. Hence, multiplicity in intermediated asset markets seems robust.

Still, future research should pursue such generalizations explicitly. With or without relaxing the inventory restriction, another idea is to introduce heterogeneity in technology, bargaining power, or other variables *within* the set of middlemen. This should lead to interdealer trade, where goods or assets get passed from one middlemen to another before ultimately being consumed. There is also more work to be done on dynamics. We suspect one can construct cyclic or sunspot equilibria, where middlemen's actions vary deterministically or stochastically over time, as a self-fulfilling prophecy. That would deliver volatility in addition to multiplicity in the market. Standard methods can be used for this construction, in principle, although it would be complicated by the endogenous inventory distribution. In any case, based on the results developed so far, we think the framework should become a benchmark model in search-based theories of intermediation, and in search theory more generally.

References

[1] G. Akerlof and R. Shiller (2009) *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press.

[2] S. Aruoba, G. Rocheteau and C. Waller (2007) "Bargaining and the Value of Money," *JME* 54, 2636-55.

[3] G. Biglaiser (1993) "Middlemen as Experts," *RAND* 24, 212-223.

[4] K. Burdett and M. Coles (1997) "Marriage and Class," *QJE* 112, 141-68.

[5] G. Camera (2001) "Search, Dealers, and the Terms of Trade," *RED* 4, 680-694.

[6] R. Cavalcanti and N. Wallace (1999) "A Model of Private Banknote Issue," *RED* 2, 104-36.

[7] D. Diamond and P. Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity," *JPE* 91, 401-419.

[8] P. Diamond (1982) "Aggregate Demand Management in Search Equilibrium," *JPE* 90, 881-94.

[9] D. Duffie, N. Gârleanu and L. Pederson (2005) "Over-the-Counter Markets," *Econometrica* 73, 1815-47.

[10] M. Friedman (1960) *A Program for Monetary Stability*, Fordham University Press, New York.

[11] A. Geromichalos and L. Herrenbrueck (2016) "Monetary Policy, Asset Prices, and Liquidity in Over-the-Counter Markets," *JMCB* 48, 35-79.

[12] C. Gu, F. Mattesini, C. Monnet and R. Wright (2013) "Endogenous Credit Cycles," *JPE* 121, 940-65.

[13] A. Hosios (1990) "On the Efficiency of Matching and Related Models of Search and Unemployment," *RES* 57, 279-298.

[14] N. Jacobson (2009), *Basic Algebra 1*, 2nd ed. Dover.

[15] A. Johri and J. Leach (2002) "Middlemen and the Allocation of Heterogeneous Goods," *IER* 43, 347-361.

[16] B. Julien and S. Mangin (2016) "Efficiency in Search and Matching Models: A Generalized Hosios Condition," mimeo.

[17] E. Kalai (1977) "Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons," *Econometrica* 45, 1623-30.

[18] Y. Li (1998) "Middlemen and Private Information," *JME* 42, 131-159.

[19] J. Lacker (2014) "Economics After the Crisis: Models, Markets, and Implications for Policy," https://www.richmondfed.org/press_room/speeches/president_jeff_lacker/2014/lacker_speech_20140221.

[20] R. Lagos and G. Rocheteau (2009) "Liquidity in Asset Markets with Search Frictions," *Econometrica* 77, 403-26.

[21] R. Lagos, G. Rocheteau and R. Wright (2016) "Liquidity: A New Monetarist Perspective," *JEL*, in press.

[22] A. Masters (2007) "Middlemen in Search Equilibrium," *IER* 48, 343-62.

[23] D. Mortensen (1982) "Property Rights and Efficiency of Mating, Racing, and Related Games," *AER* 72, 968-79.

[24] E. Nosal and G. Rocheteau (2011) *Money, Payments, and Liquidity*. MIT Press.

[25] E. Nosal, Y.-Y. Wong and R. Wright (2015) "More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets," *JMCB* 47, 7-37.

[26] C. Pissarides (2000) *Equilibrium Unemployment Theory*. MIT Press.

[27] C. Reinhart and K. Rogoff (2009) *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press.

[28] A. Rolnick and W. Weber (1986) "Inherent Instability in Banking: The Free Banking Experience," *Cato Journal* 5, 877-890.

[29] A. Rubinstein and A. Wolinsky (1987) "Middlemen," *QJE* 102, 581-594.

[30] A. Shevchenko (2004)"Middlemen," *IER* 45, 1-24.

[31] R. Shimer and L. Smith (2000) "Assortative Matching and Search," *Econometrica*, 68, 343-69.

[32] E. Smith (2004) "Intermediated Search," *Economica* 71, 619-636.

[33] A. Trejos and R. Wright (2015) "Search-Based Models of Money and Finance: An Integrated Approach," *JET*, in press.

[34] C. Tse (2009) "The Spatial Origin of Commerce," mimeo.

[35] M. Watanabe (2010) "A Model of Merchants," *JET* 145, 1865-1889.

[36] P.-O. Weill (2007) "Leaning against the Wind," *RES* 74, 1329-14.

[37] R. Wright (1995) "Search, Evolution and Money," *JEDC* 19, 181-206.

[38] R. Wright and Y.-Y. Wong (2014) "Buyers, Sellers and Middlemen: Variations on Search-Theoretic Themes," *IER* 55, 375-398.

Appendix

Here we provide proofs for results that are not obvious.

Lemma 1: Class 0 and class 2 equilibria coexist in the region where $\gamma_p \geq \bar{\gamma}_p$ and $\gamma_m < g(\gamma_p)$, but we claim the former is not subgame perfect. Notice $\gamma_m < \bar{\gamma}_m$ in this region, and consider a class 0 candidate equilibrium. Suppose a nonparticipant deviates and produces. When he meets another nonparticipant, which happens with positive probability, that agent has a strict incentive to accept his good and act like type M because $\gamma_m < \bar{\gamma}_m$ (i.e., it is not credible to think he would reject it). This constitutes a profitable deviation. ■

Lemma 3: There are three ways for a convex $Q(\mu) = 0$ to have solutions in $(0, \bar{\mu})$: (a) one root with $Q(0) < 0 < Q(\bar{\mu})$; (b) one root with $Q(0) > 0 > Q(\bar{\mu})$; (c) two-roots, which requires (c1) $Q(\bar{\mu}) > 0$, (c2) $Q(0) > 0$, (c3) $Q'(\bar{\mu}) > 0$, (c4) $Q'(0) < 0$, and (c5) $Q(\mu^*) < 0$, where $Q'(\mu^*) = 0$. Notice that

$$\begin{aligned} Q(0) &= (1-n_c)\theta_{pm}(\bar{\gamma}_m - \gamma_m) + (r + n_c\theta_{mc})(\bar{\gamma}_p - \gamma_p) \\ Q(\bar{\mu}) &= n_c[r + n_c\theta_{mc} + (1-n_c)\theta_{mp}](\bar{\gamma}_p - \gamma_p) - n_c(1-n_c)\theta_{mp}(\bar{\gamma}_m - \gamma_m). \end{aligned}$$

In case (a), it is easy to see $Q(0) < 0$ iff $\gamma_p > (1-n_c)\theta_{pm}(\bar{\gamma}_m - \gamma_m)/(r + n_c\theta_{mc}) + \bar{\gamma}_p$, and $Q(\bar{\mu}) > 0$ iff $\gamma_p < \bar{\gamma}_p - (1-n_c)\theta_{mp}(\bar{\gamma}_m - \gamma_m)/[r + n_c\theta_{mc} + (1-n_c)\theta_{mp}]$. As these conditions are contradictory, case (a) cannot occur.

Turning to case (c), (c1) $\Rightarrow \gamma_p < \bar{\gamma}_p$ while (c2) $\Rightarrow \kappa_3 > 0 \Rightarrow \gamma_m < g(\gamma_p)$, which is redundant given (c1) and that equilibrium requires that $\gamma_m \leq \bar{\gamma}_m$. Also, (c3) and (c4) $\Rightarrow$

$$\begin{aligned} \gamma_m &> \phi(\gamma_p) \equiv \bar{\gamma}_m + \frac{r + n_c\theta_{mc} - n_c\theta_{mp}}{n_c}(\bar{\gamma}_p - \gamma_p) \\ \gamma_m &< \psi(\gamma_p) \equiv \bar{\gamma}_m + \frac{r + n_c\theta_{mc} - n_c\theta_{mp}}{2(1-n_c)\theta_{pm} + n_c}(\bar{\gamma}_p - \gamma_p). \end{aligned}$$

Finally, (c5) is equivalent to $D > 0$, where D is the discriminant of $Q(\mu)$.

We now show $r + n_c\theta_{mc} - n_c\theta_{mp} < 0$ is necessary for (c3) and (c4). Suppose that $r + n_c\theta_{mc} - n_c\theta_{mp} > 0$. This implies $\phi'(\gamma_p) < 0$ and $\psi'(\gamma_p) < 0$, and both of the lines $\gamma_m = \phi(\gamma_p)$ and $\gamma_m = \psi(\gamma_p)$ go through $(\bar{\gamma}_p, \bar{\gamma}_m)$. Since equilibrium requires $\gamma_m \leq \bar{\gamma}_m$ and $\gamma_p \leq \bar{\gamma}_p$, condition (c3) is violated, i.e., as illustrated in

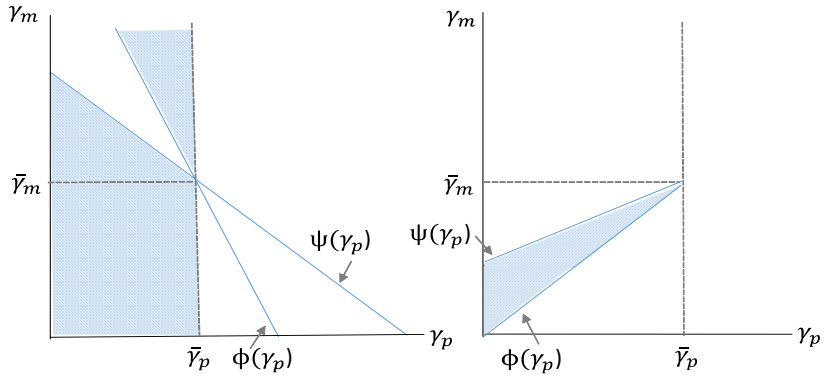


Figure 9: The functions $\phi(\gamma_p)$ and $\psi(\gamma_p)$

the left panel of Fig. 9, the intersection of conditions (c3) and (c4) is the empty set when $\gamma_p \leq \bar{\gamma}_p$. Suppose now that $r + n_c\theta_{mc} - n_c\theta_{mp} < 0$. It is easy to show (c3) and (c4) are satisfied. The parameter set consistent with the conditions c(1), c(3) and c(4) is given by $\mathcal{S}_1 \equiv \{(\gamma_p, \gamma_m)|0 < \gamma_p \leq \bar{\gamma}_p, \phi(\gamma_p) < \gamma_m < \psi\left(\gamma_p\right)\}$, shown in the right panel of Fig. 9

Similarly, let $\mathcal{S}_2$ be the set consistent with (c5). To characterize $\mathcal{S}_2$, the discriminant of $Q(\mu)$, D, can itself be written as a quadratic in γ_m given γ_p, $\hat{Q}(\gamma_m|\gamma_p) = \hat{\kappa}_1\gamma_m^2 + \hat{\kappa}_2\gamma_m + \hat{\kappa}_3$, where

$$\begin{aligned}
\hat{\kappa}_1 &= {n_c}^2 + 4n_c(1-n_c)\theta_{pm}\theta_{mp} \\
\hat{\kappa}_2 &= -2\bar{\gamma}_m[{n_c}^2 + 4n_c(1-n_c)\theta_{pm}\theta_{mp}] \\
&\quad -2n_c(\bar{\gamma}_p - \gamma_p)[(r + n_c\theta_{mc} - n_c\theta_{mp})(1-2\theta_{pm}) - 2\theta_{mp}\theta_{pm}] \\
\hat{\kappa}_3 &= \bar{\gamma}_m^2[{n_c}^2 + 4n_c(1-n_c)\theta_{pm}\theta_{mp}] + (\bar{\gamma}_p - \gamma_p)^2(r + n_c\theta_{mc} - n_c\theta_{mp}) \\
&\quad +2n_c\bar{\gamma}_m(\bar{\gamma}_p - \gamma_p)[(r + n_c\theta_{mc} - n_c\theta_{mp})(1-2\theta_{pm}) - 2\theta_{mp}\theta_{pm}].
\end{aligned}$$

Since $\hat{\kappa}_1 > 0$, $\hat{Q}$ is strictly convex. Also, it is straightforward to show that $\hat{Q}(\bar{\gamma}_m|\gamma_p) < 0 \ \forall \gamma_p \in [0, \bar{\gamma}_p)$. Thus, since $\hat{Q}$ is strictly convex and $\hat{Q}(\bar{\gamma}_m|\gamma_p) < 0$, $\mathcal{S}_2 \neq \varnothing \Rightarrow \hat{Q}(0|\gamma_p) > 0 \Rightarrow \hat{\kappa}_3 > 0$, as shown in the left panel of Fig. 10.

It can be shown that $\hat{Q}(\gamma_m|\bar{\gamma}_p) > 0 \ \forall \gamma_m \in [0, \bar{\gamma}_m)$ and $\hat{Q}(\bar{\gamma}_m|\bar{\gamma}_p) = 0$. Since $\hat{Q}$ is continuous in (γ_m, γ_p), $\hat{Q}(\gamma_m|\gamma_p) > 0$ for some $\gamma_m < \bar{\gamma}_m$ if $\bar{\gamma}_p - \gamma_p$ is small. The admissible set of γ_p for which $\hat{Q}(\gamma_m|\gamma_p) > 0$ is pinned down by the lower root of $\hat{Q}(\gamma_m|\gamma_p) = 0$ being positive, i.e., $\gamma_m^-(\gamma_p) = (-\hat{\kappa}_2 - \sqrt{\Lambda})/2\hat{\kappa}_1 > 0$, where

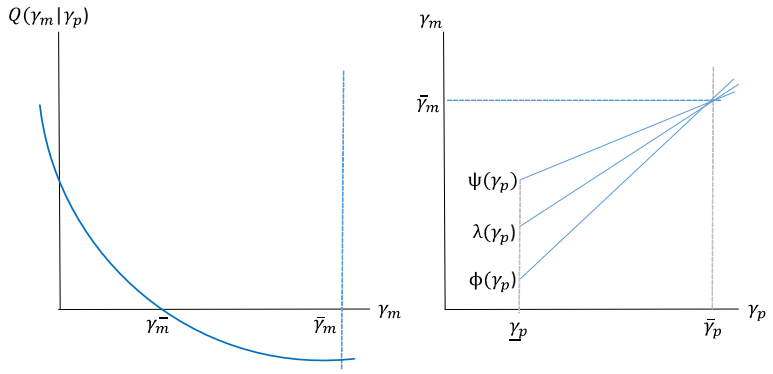


Figure 10: The functions $\hat{Q}(\gamma_m|\gamma_p)$ and $\lambda(\gamma_p)$

$\Lambda = \hat{\kappa}_2^2 - 4\hat{\kappa}_1\hat{\kappa}_3 > 0$. One can show $\gamma_m^-(\gamma_p) > 0 \Rightarrow \hat{\kappa}_2 > 0 \Rightarrow \gamma_p > \underline{\gamma}_p$ with

$$\underline{\gamma}_p \equiv \bar{\gamma}_p + \bar{\gamma}_m[n_c + 4(1-n_c)\theta_{pm}\theta_{mp}][(r + n_c\theta_{mc} - n_c\theta_{mp})(1-2\theta_{pm}) - 2\theta_{mp}\theta_{pm}].$$

Hence, for a given γ_p, the set of γ_m such that $\hat{Q}(\gamma_m|\gamma_p) > 0$ is $[0, \gamma_m^-(\gamma_p))$. Therefore, $\mathcal{S}_2 = \{(\gamma_p, \gamma_m)|\underline{\gamma}_p < \gamma_p < \bar{\gamma}_p, 0 < \gamma_m < \gamma_m^-(\gamma_p)\}$. Suppose for a given γ_p there exists $\gamma_m^-(\gamma_p) > 0$ such that $\hat{Q}(\gamma_m^-(\gamma_p)) = 0$. We express the lower root as $\gamma_m^-(\gamma_p) = \lambda(\gamma_p)$, where

$$\lambda(\gamma_p) \equiv \bar{\gamma}_m + (\bar{\gamma}_p - \gamma_p)\frac{n_c[(r + n_c\theta_{mc} - n_c\theta_{mp})(1-2\theta_{pm}) - 2\theta_{mp}\theta_{pm}] - \sqrt{\Lambda}}{{n_c}^2 + 4n_c(1-n_c)\theta_{pm}\theta_{mp}}.$$

One can show $\lambda'(\gamma_p) > 0$. The right panel of Fig. 10 depicts $\gamma_m = \phi(\gamma_p)$, $\gamma_m = \psi(\gamma_p)$ and $\gamma_m = \lambda(\gamma_p)$. Since $\hat{Q} \equiv D > 0 \Rightarrow \gamma_m < \lambda(\gamma_h)$, a necessary condition for case (c) is $\lambda'(\gamma_p) < \phi'(\gamma_p)$, as in the right panel of Fig. 10.

Hence, (c) requires $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \varnothing$ and $\lambda'(\gamma_p) < \phi'(\gamma_p)$. This inequality implies

$$\begin{aligned}(\theta_{mc} - \theta_{mp})[n_c + 4(1-n_c)\theta_{pm}\theta_{mp}] < {} & [n_c(\theta_{mc} - \theta_{mp})(1-2\theta_{pm}) - 2\theta_{mp}\theta_{pm}] \\ & - \{[n_c(\theta_{mc} - \theta_{mp})(1-2\theta_{pm}) - 2\theta_{mp}\theta_{pm}]^2 \\ & - (\theta_{mc} - \theta_{mp})[n_c + 4(1-n_c)\theta_{pm}\theta_{mp}]\}^{1/2},\end{aligned}$$

ignoring terms with r that only strengthen the inequality. This implies

$$-1 + n_c(\theta_{mc} - \theta_{mp}) > 4n_c\theta_{mc}\theta_{pm} + 4(1-n_c)\theta_{pm}\theta_{mp}(\theta_{mc} + \theta_{pm}).$$

But the LHS is negative and the RHS positive – a contradiction. ■

Lemma 5: It is straightforward to derive $\partial Q(\mu)/\partial\gamma_p < 0$, so consider the effect of γ_m. In the particular case of $\gamma_p = \bar{\gamma}_p$, the relevant root is

$$\mu = \frac{2(1-n_c)\theta_{pm} + n_c - [4(1-n_c)\theta_{pm}n_c(1-\theta_{pm}n_c) + n_c^2]^{1/2}}{2\theta_{pm}} \equiv \tilde{\mu}.$$

Hence, $\partial\mu/\partial\gamma_m = 0$ when $\gamma_p = \bar{\gamma}_p$. More generally,

$$\frac{\partial Q(\mu)}{\partial\gamma_m} = n_c\mu + \theta_{pm}[2(1-n_c)\mu - (1-n_c) - \mu^2],$$

which vanishes when $\mu = \tilde{\mu}$ or $\gamma_p = \bar{\gamma}_p$. Moreover,

$$\frac{\partial}{\partial\mu}\frac{\partial Q(\mu)}{\partial\gamma_m}\bigg|_{\mu=\tilde{\mu}} = n_c + \theta_{pm}2(1-n_c) - 2\theta_{pm}\mu > 0.$$

Hence, $\partial\mu/\partial\gamma_m > 0$ if $\gamma_p < \bar{\gamma}_p$ and $\partial\mu/\partial\gamma_m < 0$ if $\gamma_p > \bar{\gamma}_p$. ■

Lemma 7: As $\partial(n_m\mu)/\partial\gamma_m = \partial(n_m\mu)/\partial\mu \times \partial\mu/\partial\gamma_m$ we need to sign $\partial(n_m\mu)/\partial\mu$. Notice $n_m\mu = \mu - n_c\mu/(1-\mu)$, which implies

$$\frac{\partial(n_m\mu)}{\partial\mu} \simeq (1-\mu)^2 - n_c \simeq \gamma_p - n_c u,$$

using (10) to eliminate $(1-\mu)^2$, and $a \simeq b$ means a and b have the same sign. When $\gamma_p < n_c u$, $\partial\mu/\partial\gamma_m > 0$ and $\partial(n_m\mu)/\partial\mu < 0$, so $\partial(n_m\mu)/\partial\gamma_m < 0$; when $\gamma_p > n_c u$, $\partial\mu/\partial\gamma_m < 0$ and $\partial(n_m\mu)/\partial\mu > 0$, so again $\partial(n_m\mu)/\partial\gamma_m < 0$. ■

Proposition 3: The efficient and equilibrium outcomes only correspond in general if $\theta_{mc} = \theta_{pc} = 1$, as that needed for $\bar{\gamma}_p = \bar{\gamma}_m = n_c u$. Given $\theta_{mc} = \theta_{pc} = 1$,

$$\begin{aligned} f(\gamma_p) &= \frac{-n_c^2 u + [1 - \theta_{pm}(1-n_c)]\gamma_p}{(1-n_c)(1-\theta_{pm})} \\ g(\gamma_p) &= \frac{n_c u[n_c + \theta_{pm}(1-n_c)] - n_c\gamma_p}{(1-n_c)\theta_{pm}}. \end{aligned}$$

If $\theta_{pm}^o = 1$ then $g(\gamma_p) = g^0(\gamma_p)$; so for $(\gamma_p, \gamma_m) \in S_0^o$, $n_j = n_j^o = 0$. If $\theta_{pm}^o = 0$ then $f(\gamma_p) = f^0(\gamma_p)$; so for $(\gamma_p, \gamma_m) \in S_1^o$, again $n_j = n_j^o$. If $\theta_{pm}^o \in (0,1)$ then $\gamma_m \leq f(\gamma_p)$ implies $\gamma_m \leq f^o(\gamma_p)$ and $\gamma_m \leq g(\gamma_p)$ implies $\gamma_m \leq g^o(\gamma_p)$. If we set

$\theta_{mc} = \theta_{pc} = 1$ and equate the roots of (8) and (10), so that $\mu = \mu^o$, we get θ^o_{pm}. To check $\theta^o_{pm} \in (0,1)$, note the numerator is positive since $\mu^o < 1 - n_c$, and the denominator is even bigger since $n^o_m > 0$ requires $\gamma_m < \gamma_p$. ■

Lemma 10: There are again three cases for $\tilde{Q}(\mu, y_{mp}) = 0$: (a) one root with $\tilde{Q}(0, y_0) < 0 < \tilde{Q}(\bar{\mu}, \bar{y})$; (b) one root with $\tilde{Q}(0, y_0) > 0 > \tilde{Q}(\bar{\mu}, \bar{y})$; and (c) two roots, requiring (c1) $\tilde{Q}(0, y_0) > 0$, (c2) $\tilde{Q}(\bar{\mu}, \bar{y}) > 0$, (c3) $\partial\tilde{Q}(\mu, y_{mp}|\mu = \bar{\mu})/\partial\mu > 0$, (c4) $\partial\tilde{Q}(\mu, y_{mp}|\mu = 0)/\partial\mu < 0$, and (c5) $\tilde{Q}(\mu^*, y_{mp}) < 0$, where $\partial\tilde{Q}(\mu^*, y_{mp})/\partial\mu = 0$. As in Lemma 3, case (a) is impossible. Now notice

$$\begin{aligned}
\tilde{Q}(0, y_0) &= \theta_{pm}[n_c z - \gamma_p + U(y_0)(1 - n_c)] \\
\tilde{Q}(\bar{\mu}, \bar{y}) &= n_c[\theta_{pm}(n_c z - \gamma_p) - U(\bar{y})(1 - n_c)(1 - \theta_{pm})].
\end{aligned}$$

In case (c), (c1) implies $\gamma_p < G(\gamma_m)$ and (c2) implies $\gamma_p < F(\gamma_m)$. From (c3) and (c4), we have

$$\frac{\partial\tilde{Q}(\mu, y_{mp})}{\partial\mu} = 2\theta_{pm}\mu U(y_{mp}) - (n_c z - \gamma_p)\theta_{pm} - U(y_{mp})[n_c + 2(1 - n_c)\theta_{pm}].$$

We need this positive at $\mu = \bar{\mu}$, which means $\gamma_p > K \equiv n_c z - U(\bar{y})n_c/\theta_{pm}$, and at $\mu = 0$, which means $\gamma_p < n_c z + U(y_0)[n_c + 2(1 - n_c)\theta_{pm}]/\theta_{pm}$. Given (c2), (c1) and (c4) are not binding. Also, (c2) and (c3) imply γ_p is between K and F, which holds iff $\theta_{pm} < (1 - 2n_c) / (1 - n_c)$. Assume this is true and consider (c5). To get μ^*, solve $\partial Q/\partial\mu = 0$ to get

$$\begin{aligned}
\tilde{Q}(\mu^*, y_{mp}) \simeq\ & -(n_c z - \gamma_p)[(n_c z - \gamma_p)\theta_{pm} + 2n_c U(y_{mp})(1 - 2\theta_{pm})] \\
& -U(y_{mp})^2 n_c[1 + 4\theta_{pm}(1 - \theta_{pm})(1 - n_c)].
\end{aligned}$$

We need $\tilde{Q}(\mu^*, y_{mp}) < 0$. With a small abuse of notation, let $\tilde{Q}(\mu^*, y_{mp}) \equiv \tilde{Q}(\gamma_p) < 0$ where

$$\begin{aligned}
\tilde{Q}(\gamma_p) = & -\theta_{pm}\gamma_p^2 + 2n_c[U(y_{mp})(1 - 2\theta_{pm}) + \theta_{pm} z]\gamma_p - n_c^2 z^2 \theta_{pm} \\
& -n_c^2 z U(y_{mp})(1 - 2\theta_{pm}) - n_c U(y_{mp})^2[1 + 4\theta_{pm}(1 - \theta_{pm})(1 - n_c)].
\end{aligned}$$

For (c5) we seek the set of γ_p such that $\tilde{Q}(\gamma_p) < 0$. There are three possibilities: (c5.1) one root with $\tilde{Q}(K) < 0 < \tilde{Q}(F)$; (c5.2) one root with $\tilde{Q}(K) > 0 >$

$\tilde{Q}(F)$; (c5.3) two roots, which requires $\tilde{Q}(K) < 0 < \tilde{Q}(F)$, $\tilde{Q}'(K) > 0 > \tilde{Q}'(F)$, and $\tilde{Q}(\gamma_p^*) < 0$, where $\tilde{Q}'(\gamma_p^*) = 0$. Given $\gamma_p = K$ and $n_c z - \gamma_p = U(\bar{y}) n_c / \theta_{pm}$,

$$\tilde{Q}(K) = -U(\bar{y})^2 \frac{n_c^2}{\theta_{pm}} [1 + 2(1 - 2\theta_{pm})] - U(\bar{y})^2 n_c [1 + 4\theta_{pm}(1 - \theta_{pm})(1 - n_c)] < 0.$$

Given $\gamma_p = F(\gamma_m)$ and $n_c z - \gamma_p = U(\bar{y})(1 - \theta_{pm})(1 - n_c)/\theta_{pm}$,

$$\tilde{Q}(F) \simeq -U(\bar{y})^2 \{(1 - \theta_{pm})(1 - n_c)[1 + n_c - \theta_{pm}(1 + 3n_c) + 4n_c \theta_{pm}^2] + n_c \theta_{pm}\} < 0,$$

for $(1 - 2n_c)/(1 - n_c) > \theta_{pm} > 0$. This rules out (c5.1) and (c5.2). To check (c5.3), consider

$$\tilde{Q}'(\gamma_p) = -2\theta_{pm}\gamma_p + 2n_c[U(y_{mp})(1 - 2\theta_{pm}) + \theta_{pm} z].$$

Now $\tilde{Q}'(\gamma_p) > 0$ at $\gamma_p = K$, and $\tilde{Q}'(\gamma_p) > 0$ at $\gamma_p = F(\gamma_m)$. As $\tilde{Q}'(F) > 0$ violates (c5.3), there is no γ_p^* between K and F such that $\tilde{Q}(\gamma_p^*) < 0$. ■

Lemma 11: We need B and O in Fig. 3 cross at $(y_{mp}, \mu) \in (0, \infty) \times (0, \bar{\mu})$, plus $\gamma_m < n_c z$. For $\mu \in (0, \bar{\mu})$, we check $\tilde{Q}(0, y_0) > 0 > \tilde{Q}(\bar{\mu}, \bar{y})$, where $y_0 = B(0)$ and $\bar{y} = B(\bar{\mu})$. Now $\tilde{Q}(0, y_0) > 0$ iff $\gamma_p < G(\gamma_m)$. At $\gamma_m = n_c z$, bargaining implies $y_0 = 0$, and $\gamma_p < G(\gamma_m)$ becomes $\gamma_p < n_c z$. As we lower γ_m, y_0 rises, and we need $\gamma_p < G(\gamma_m)$. In (γ_p, γ_m) space $G(\gamma_m)$ traces a curve that downward sloping and concave (see below), and $\tilde{Q}(0, y_0) > 0$ to the left of $\gamma_p = G(\gamma_m)$. The $\tilde{Q}(\bar{\mu}, \bar{y}) < 0$ iff $\gamma_p > F(\gamma_m)$. At $\gamma_m = n_c z$, bargaining implies $\bar{y} = 0$, and $\gamma_p > F(\gamma_m)$ becomes $\gamma_p > n_c z$. As we lower γ_m, $\bar{y}$ rises, and we need $\gamma_p > F(\gamma_m)$. In (γ_p, γ_m) space, $F(\gamma_m)$ traces a curve that is upward sloping and concave. Hence $\exists \mu \in (0, \bar{\mu})$ solving $\tilde{Q}(\mu, y_{mp}) = 0$ iff $F(\gamma_m) < \gamma_p < G(\gamma_m)$. To check $y_{mp} > 0$, note from Fig. 4 that it lies to the right of $\bar{y}$, and $\bar{y} \geq 0$ as long as $n_c z \geq \gamma_m$. To check $V_p = V_0 \geq 0$, note by construction $V_0 \geq 0$ if $\mu \geq 0$.

To establish the properties of F and G, derive

$$\begin{aligned} G'(\gamma_m) &= \frac{-\theta_{pm}(1 - n_c)U'(y_0)}{(r + n_c)[\theta_{pm} + (1 - \theta_{pm})U'(y_0)]} < 0 \\ G''(\gamma_m) &\simeq \frac{-\theta_{pm}^2(1 - n_c)U''(y_0)\, y_0'(\gamma_m)}{(r + n_c)[\theta_{pm} + (1 - \theta_{pm})U'(y_0)]^2} < 0. \end{aligned}$$

Thus $G(\cdot)$ is decreasing and concave in (γ_m, γ_p) space or (γ_p, γ_m) space . Similarly, $F'(\gamma_m) > 0$ and $F''(\gamma_m) > 0$. Thus $F(\cdot)$ is increasing and convex in (γ_m, γ_p) space, or increasing and concave in (γ_p, γ_m) space. ■

Lemma 12: In class 2 equilibrium we have

$$\mu = \frac{A - (r + n_c)(1 - \theta_{pm}) U(y_{mp})}{A - r(1 - \theta_{pm}) U(y_{mp})} \equiv \mu^*,$$

with $A = \theta_{pm}(n_c z - \gamma_m) - (r + n_c)\theta_{pm} y_{mp} > 0$, from the bargaining solution. Note $\mu > 0 \Rightarrow A > (r + n_c)(1 - \theta_{pm}) U(y_{mp})$, and $\mu < \bar{\mu} \Rightarrow A < (1 + r)(1 - \theta_{pm}) U(y_{mp})$. Then

$$\begin{aligned} \frac{\partial O(y_{mp})}{\partial y_{mp}} &= -\frac{U'\theta_{pm}(n_c z - \gamma_p)}{U\sqrt{\tilde{D}}}(1 - \mu) \\ \frac{\partial B^{-1}(y_{mp})}{\partial y_{mp}} &= -\frac{n_c(1 - \theta_{pm})}{[A - r(1 - \theta_{pm}) U]^2}[AU' + \theta_{pm}(r + n_c) U] \end{aligned}$$

If $n_c z < \gamma_p$ the equilibrium is obviously unique. If $n_c z > \gamma_p$, we claim $\partial O/\partial y_{mp} > \partial B^{-1}/\partial y_{mp}$ when they cross. To verify this, insert $\mu = \mu^*$ to get

$$\frac{\partial O(y_{mp})}{\partial y_{mp}} = -\frac{U'\theta_{pm}(n_c z - \gamma_p)}{\sqrt{\tilde{D}}} \frac{n(1 - \theta_{pm})}{A - r(1 - \theta_{pm}) U},$$

where $\tilde{D}$ is the discriminant of $\tilde{Q}(\mu, y_{mp})$. Using (18) to replace $\sqrt{\tilde{D}}$ and inserting $\mu = \mu^*$, we get

$$\frac{\partial O(y_{mp})}{\partial y_{mp}} = -\frac{U'\theta_{pm}(n_c z - \gamma_p) n_c(1 - \theta_{pm})}{[A - r(1 - \theta_{pm}) U]\Omega},$$

where

$$\Omega \equiv [2\theta_{pm}(1 - n_c) + n_c] U + \theta_{pm}(n_c z - \gamma_p) - \frac{2\theta_{pm} U[A - (r + n_c)(1 - \theta_{pm}) U]}{A - r(1 - \theta_{pm}) U}.$$

In equilibrium, $A = \theta(n_c z - \gamma_m) - (r + n_c)\theta y^*$ and $U = U(y^*)$ solves

$$\begin{aligned} & \theta U[A - (r + n)(1 - \theta) U]^2 + \theta[A - r(1 - \theta) U]^2[nz - \gamma_p + (1 - n)U] \\ = \; & [A - (r + n)(1 - \theta) U][A - r(1 - \theta) U]\{[2\theta(1 - n) + n] U + \theta(nz - \gamma_p)\} \end{aligned}$$

Routine algebra implies $\partial O(y)/\partial y - \partial B^{-1}(y)/\partial y$ is proportional to

$$U\theta\left(nz-\gamma_p\right)\left[A-r\left(1-\theta\right)U\right]\left[U'r\left(1-\theta\right)+\left(r+n\right)\theta\right]$$
$$+\left[AU'+\theta\left(r+n\right)U\right]U\{n[A-r\left(1-\theta\right)U]+2\theta n[\left(1+r\right)\left(1-\theta\right)U-A]\}$$

Since $\left(1+r\right)\left(1-\theta\right)U > A > r\left(1-\theta\right)U$, in equilibrium, this is positive, thus establishing the desired result. ■

Lemma 14: To be added. ■

Proposition 6: Part (i) of the result follows directly from Lemma 14. For part (ii), consider the model with $\gamma_j < 0$, and fix the numbers n_p and n_m. Everthing else is the same, and in particular

$$\begin{aligned} rV_p &= n_c\theta_{pc}u + n_m\theta_{pm}\left(1-\mu\right)\left(V_1-V_0\right)-\gamma_m \\ rV_1 &= n_c\theta_{mc}\tau(u+V_0-V_1)-\gamma_m \\ rV_0 &= n_c\theta_{mp}\left(V_1-V_0\right). \end{aligned}$$

Consider a candidate equilibrium with $\tau = 0$. Then $\mu = 1$ and the above equations imply

$$V_0 = \frac{-n_c\theta_{mp}\gamma_m}{r\left(r+n_c\theta_{np}\right)} \text{ and } V_1 = \frac{\gamma_m}{r}.$$

A deviation by M to $\tau = 1$ implies $rV_1^d = n_c\theta_{mc}\left(u+V_0-V_1^d\right)-\gamma_m$. After inserting V_0 we get

$$V_1^d = \frac{r\left(r+n_c\theta_{mp}\right)\left(n_c\theta_{mc}u-\gamma_m\right)-n_c^2\theta_{mc}\theta_{mp}\gamma_m}{r\left(r+n_c\theta_{mp}\right)\left(r+n_c\theta_{mc}\right)}.$$

The deviation is not profitable, and hence $\tau = 0$ is an equilibrium, iff $V_1^d \leq V_1$. This reduces to $-\gamma_m \geq \left(r+n_c\theta_{mp}\right)u$.

Now consider a candidate equilibrium with $\tau = 1$. Then we solve in the usual way for

$$V_1 = \frac{\left(r+n_c\theta_{mp}\right)\left(n_c\theta_{mc}u-\gamma_m\right)}{r\left(r+n_c\theta_{mc}+n_c\theta_{mp}\right)}.$$

A deviation to $\tau = 0$ implies $V_1^d = -\gamma_m/r$. This is not profitable, and hence $\tau = 1$ is an equilibrium, iff $V_1^d \leq V_1$. This reduces to $-\gamma_m \leq \left(r+n_c\theta_{mp}\right)u$. Equilibrium is generically unique. ■