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 30, 2016

Abstract

This paper studies agents' decisions to act as producers or middlemen, and hence endogenizes market composition, in a search-theoretic model of intermediation. We extend the standard framework to allow nonlinear utility, general bargaining, costs and returns. Also, we go beyond the usual steady state analysis by considering dynamics. The analysis remains tractable, delivering clean and sometimes surprising results. Intermediation can be essential, and equilibrium is efficient only under strict conditions. While the model with middlemen holding goods displays uniqueness, the version with middlemen holding assets has multiple steady states and interesting dynamics – suggesting there is something special about financial intermediation.

JEL Codes: G24, D83

Keywords: Middlemen, Intermediation, Search, Bargaining, Multiplicity

*We thank Douglas Gale, Chao Gu, David Laidler, George Selgin, Karl Shell, Alberto Trejos and Steve Williamson 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 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 seems to be room for additional work. This project revisits the classic search-based framework of Rubinstein and Wolinsky (1987), hereafter RW, extends it on several dimensions in terms of theory, and puts it to work in substantive applications.

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, as is efficient. Nosal et al. (2015) incorporate more general bargaining and costs, show that whether agents participate depends on various factors, and prove this may or may not be efficient. While this is interesting, the assumptions in the original and extended versions are still very special: utility is linear; the market is always in steady state; and the sets of producers and middlemen are fixed. We relax all of these, perhaps most importantly by letting individuals decide to act as *either* producers or middlemen. This "occupational choice" makes the market composition endogenous, which allows us to ask additional questions about efficiency, uniqueness vs multiplicity, and many other 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 which type i agents meet type j agents. This is legitimate because under mild conditions there exists a distribution of types, say $\mathbf{n} = \{n_i\}$, consistent with $\boldsymbol{\alpha}$, random

[1] Since it is easier to describe substantive issues after laying out modeling details, we defer an extended discussion for now (see, e.g., fn. 18). However, one feature to emphasize up front is that our model economy can only have more middlemen by having fewer producers – say, more MBA's and fewer engineers – which argubly captures an interesting real tradeoff.

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 $\boldsymbol{\alpha}$. We therefore 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 $\alpha_{pc} = \alpha_{mc}$. This means we must abandon RW's original idea that middlemen have a role iff $\alpha_{mc} > \alpha_{pc}$, but, fortunately, other factors in our environment can take over for $\boldsymbol{\alpha}$, including bargaining powers, storage costs or rates of return.

Despite these complications the framework remains tractable, delivering clean and sometimes surprising predictions. Consider a benchmark case where the objects being traded have positive storage costs, interpreted as a consumer goods in a retail market, and the storage costs generally differ between producers and middlemen. As one perhaps surprising result, increasing intermediaries' costs can lead to more of them. 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. Also, we show that equilibrium can have too few or too many middlemen, and efficiency obtains 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 go beyond the usual steady state analysis by studying dynamic equilibrium.

We then consider an application to financial intermediation. Suppose the objects being traded have negative storage costs, interpreted as assets bearing positive returns – e.g., houses, art, productive capital, etc. For instance, suppose

[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.

producers supply capital goods they can trade to end users who have more productive uses for them. Producers can also trade capital to middlemen who can invest it at a profit while waiting to pass the capital on to end users. This resembles financial intermediation.[3] Middlemen can have a higher or lower return than producers, but as long as it is positive, an interesting 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, active intermediation may or may not emerge in equilibrium.

Importantly, there are multiple equilibria. On the one hand, if middlemen decide to pass capital on to end users, lots of middlemen will be without capital. Since producers can trade with either end users or middlemen in search of capital, this raises their profit, and hence leads to more producers. With more producers, it is easier for middlemen to get capital, thus rationalizing their decision to trade it away, and making active intermediation an equilibrium. On the other hand, if middlemen decide to keep capital for their own use, in the long run most of them already have capital, so producers trade mainly with end users. This lowers their profit and leads to fewer producers. With fewer producers it is harder for middlemen to get capital, thus rationalizing their decision to not trade it away, and making no intermediation an equilibrium. This strategic effect implies there can be two pure-strategy equilibria, one with and one without active intermediation, as well as a mixed-strategy steady state equilibrium where middlemen randomize. There are also multiple dynamic equilibria where intermediation activity varies and can cycle over time.

An interesting aspect of this is result is that multiplicity is only possible when holding inventories entails positive returns – in the version with negative

[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 basic model to look like spot trade.

returns (storage costs) equilibrium is always unique. This is true even though we tried to get multiplicity with negative returns by going beyond linear steady-state analysis, based on the knowledge that in some related search-and-bargaining models multiplicity can arise iff we consider nonlinear utility or dynamics (Wright and Wong 2014; Trejos and Wright 2016). This does not mean that one can never generate multiplicity with negative returns, and we know lots of "tricks" to make that happen – e.g., going back to Diamond (1982), increasing returns in the meeting technology – but it shows that one does not need such devices when intermediation involves assets with positive yields.

To summarize, as a self-fulfilling prophecy, it can be a best response for middlemen to pass inventories on to end users, or to hoard them, if these inventories entail positive yields but not storage costs. This suggests intermediation in financial markets is special vis a vis retail markets, which we interpret as support for the long-standing notion, associated with names like Minsky, Kindleberger, and others, that financial activity is generally more susceptible to multiplicity, volatility or fragility. See Akerlof and Shiller (2009) for a broad discussion, and Reinhart and Rogoff (2009) for an empirical/historical account. As regards financial intermediation and banking in particular, Rolnick and Weber (1986) say "Historically, even some of the staunchest proponents of laissez-faire have viewed banking as inherently unstable and so requiring government intervention." One such proponent is Friedman (1960), who opposed regulation of most activities, yet advocated narrow banking as part of his "program for monetary stability." And a literature following Diamond and Dybvig (1983) is dedicated to studying multiplicity and instability in banking theory. All of this is evidence of a popular view that financial intermediation is special, consistent with our findings.[4]

[4] To be clear, based on their reading of history, Rolnick and Weber (1986) actually challenge the notion that financial intermediaries are inherently unstable; the point of the quotation is to establish that this is a conventional wisdom. Also, even if one accepts the premise that financial activity is subject to instability or volatility, the underlying reason is open to debate – e.g., Lacker (2014) argues it may be induced by policy and regulation.

The word hoard in the previous paragraph is apt because the results are 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 (see surveys by Nosal and Rocheteau 2011 and Lagos et al. 2016). Our result is somewhat different, however, more like showing that whether a buyer spends an asset can depend on what others are doing. Bad (low return) assets are always passed on by middlemen, while good (high return) assets are hoarded, in the spirit of Gresham's Law. If that is not surprising, it was less obvious to us that moderately good assets may be hoarded or may circulate depending on the equilibrium we select. Moreover, if we fix the sets of middlemen and producers, the strategic effects discussed above are inoperative, and equilibrium is again unique, showing how endogenous market composition is crucial. Finally, we emphasize our analysis does not concern instability in *credit arrangements* – on that, see Gu et al. (2013) and references therein – and instead features *intermediated asset markets*, but that makes it no less relevant.

The rest of the paper involves making our assumptions precise and verifying the results. Section 2 describes the environment. Sections 3 and 4 study equilibrium and efficiency in a stationary linear economy. Sections 5 and 6 consider nonlinear and nonstationary versions. Sections 7 and 8 analyze financial intermediation. Section 9 concludes. Technical results are relegated to an Appendix.[5]

[5] For more 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." The situation has improved since then, and in particular, work on intermediation with endogenous market composition 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), Watanabe (2010), and Geromichalos and Herrenbrueck (2016). Among other differences, we stay closer to RW by not giving middlemen better information or letting them hold larger inventories. For more on the literature see Wright and Wong (2014). In recent work, Farboodi et al. (2015,2016) also use search theory to study middlemen, but while complementary in style, the models and applications are quite different. In particular, while we do many things those papers do not, Farboodi et al. (2016) have a nice way to endogenize differential arrival rates that might be worth integrating into our framework in future work.

2 Environment

There is a continuum of infinitely-lived agents. Measure n_c of them are consumers, or end users, labeled C. The rest choose to be producers, middlemen or nonparticipants, labeled P, M or N, with measure n_p, n_m or n_n. Type P produce whenever they can and type M trade with them whenever they can (this is how we define P and M; those who do not want to act this way are called type N). Type C also trade whenever they can. All agents meet bilaterally in continuous time, with the Poisson rate at which any type i meets type j given by $\alpha_{ij} = \alpha n_j / \Sigma_h n_h$. Notice this displays constant returns: doubling n_j $\forall j$ doubles the number of meetings and leaves α_{ij} the same. To reduce notation, without loss of generality, set $n_c + n_p + n_m + n_n = 1$ and $\alpha = 1$, and write $\mathbf{n} = (n_c, n_p, n_m, n_n)$.

There are two tradeable objects, for now interpreted as consumption goods, x and y. Good x is indivisible. It is valued for consumption only by type C, who get utility u from it. Later we think of x as a productive input or asset that also yields a direct flow payoff. In any case, 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 we assume $\gamma_j > 0$; to reinterpret x as an asset we later assume $\gamma_j < 0$. While x can be produced by type P at cost κ, for most purposes we can set $\kappa = 0$ 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

[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.

$n_c n_m \mu$ (the measure of C meeting M with x). Thus, in steady state

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

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.

Nonconsumers can choose to be type P or M, and if they choose M they

[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.

start without x, for payoff V_0. Hence, the choice comes down to the following best-response considerations:

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

Obviously, $n_p, n_m > 0$ requires $V_p = V_0$. Given all this, 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 $n_m \mu$, etc.

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]

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

[8] Much of the Appendix involves manipulating quadratic equations that arise from random matching and endogenous inventories. It is worth re-emphasizing that this is not due to increasing returns: independent of market size, one meets other agents at fixed rates, but the outcome when one meets M depends on μ, and that depends on the type distribution $\mathbf{n}$.

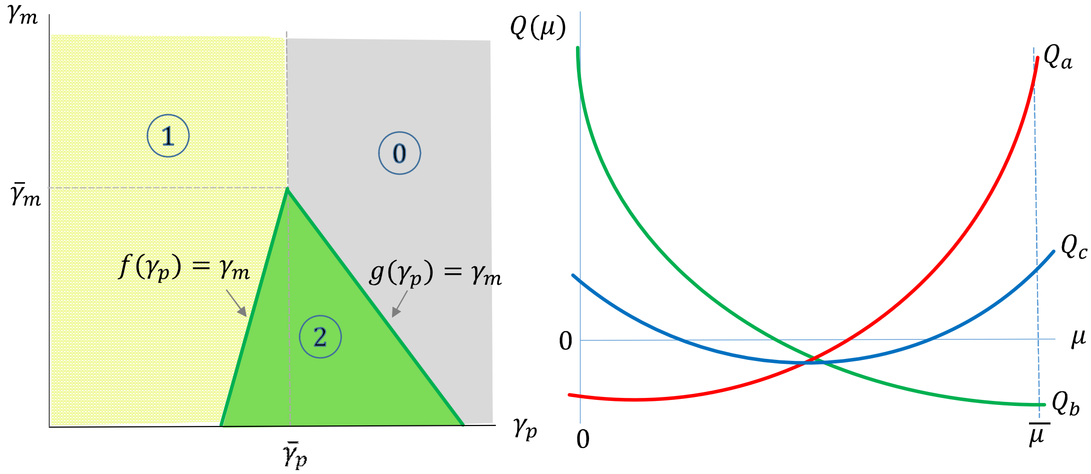

Figure 1: Equilibrium outcomes in (γ_m, γ_p) space

We seek $\mu \in (0, \bar{\mu})$ such that $Q(\mu) = 0$ and $V_0 \geq 0$. Now $V_0 \geq 0$ iff $\gamma_m \leq \bar{\gamma}_m$, which implies $\kappa_1 > 0$, and hence $Q(\mu)$ is convex. As shown by the curves Q_a, Q_b and 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 class 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 non-trivial, 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 out of existence, 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 as follows:

Proposition 1 *With $\gamma_j > 0$ equilibrium exists and is generically unique, as shown in Fig. 1. For some parameters intermediation is essential.*

Additional insights come from changing parameters in 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: how can we get more middlemen when γ_m is higher? This is answered in Section 4.

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,

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

Notice n_p on the RHS, because in addition to the direct effects of parameters,

[9]Here are some other effects. As regards r, we have this: $\gamma_p < \bar{\gamma}_p$ implies $\partial\mu/\partial r > 0$, $\partial n_p/\partial r > 0$ and $\partial n_m/\partial r < 0$; $\gamma_p > \bar{\gamma}_p$ implies the opposite. As regards a demand increase on the intensive margin, captured by higher u, we have this: if $\gamma_p > \bar{\gamma}_p$ 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.

there are 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, e.g., $\partial y_{mp}/\partial \gamma_p > 0$, $\partial y_{cm}/\partial \gamma_p < 0$ and $\partial s/\partial \gamma_p < 0$, where $s = y_{cm} - y_{mp}$ is the spread. One can also derive the effects on $\mathbf{y}$ and s of γ_m, θ_{ij}, etc. In general, the theory delivers predictions that are either unambiguous, or ambiguous due to indirect effects resulting from changes in market composition. Furthermore, all these results differ from settings where $\mathbf{n}$ is fixed exogenously, because then the indirect effects vanish. This is one reason to endogenize market composition. Another is to examine welfare.

4 Efficiency

We now discuss optimality by solving a planner's problem, focusing on the case where $r \to 0$: choose (μ^o, n_p^o, n_m^o) to maximize

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

where the first (second) term is the surplus from direct (indirect) trade.[10] On the one hand, consider $\gamma_m > n_c u$. Then intermediation is a bad idea – i.e., it necessarily contributes negatively to rW. In this case, $\gamma_p > n_c u$ implies $n_p^o = 0$ and the market shuts down, while $\gamma_p < n_c u$ implies $n_p^o = 1 - n_c$ and the market opens with direct trade only.

On the other hand, consider $\gamma_m < n_c u$, which means intermediation may or may not be a good idea. Substituting out 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] One can solve the problem with $r > 0$, then let $r \to 0$, but as usual it is much easier to take the limit first and the result is the same (we do this in a related model in Nosal et al. 2015). Note that $W \to \infty$ as $r \to 0$, but rW is well defined.

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

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

which is a quadratic that is decreasing in μ over the relevant range.

Hence, with $\gamma_m < n_c u$ there are three possibilities. First, the solution might be $\mu^o = 0$, 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, it might be $\mu^o = \bar{\mu}$, which corresponds to a class 1 outcome with $n_p^o = 1 - n_c$ and $n_m^o = 0$. 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 summarized as follows:

Lemma 6 *The efficient outcome has $n_p^o = n_m^o = 0$ iff $\gamma_p \geq n_c u$ and $\gamma_m \geq g^o(\gamma_p)$; it has $n_p^o = 1 - n_c$ and $n_m^o = 0$ iff $\gamma_b < n_c u$ and $\gamma_m \leq f^o(\gamma_p)$; and it has $n_p^o, n_m^o > 0$ 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. We say more about the comparison between the efficient and equilibrium outcomes below, after discussing the effects of parameters on the former when $\mu^o \in (0, \bar{\mu})$.

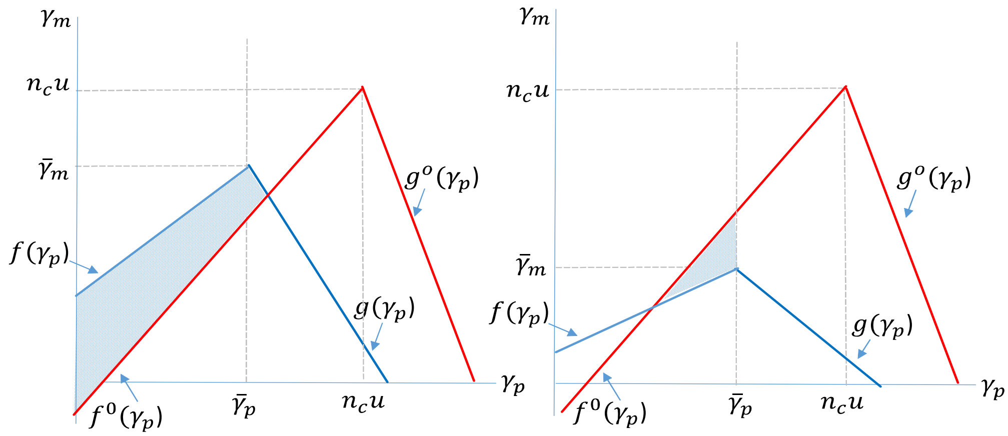

Figure 2: Comparing equilibrium and efficient outcomes

First, similar to the equilibrium results, we have

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

Also similar to the equilibrium results, we have

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

which should again be surprising. How can higher γ_m lead to more middlemen? To explain this, the following is useful:

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

This says that an increase in γ_m always reduces the stock of inventories held by middlemen, $n_m^o \mu^o$, but there are different ways to do so. One is to reduce n_m^o, which in steady state means higher μ^o; the other is to reduce μ^o, which means higher n_m^o. When $\gamma_p < n_c u$ it is optimal to use the extensive margin and reduce n_m^o; when $\gamma_p > n_c u$ it is optimal to use the intensive margin and reduce μ^o, which means higher n_m^o. 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.

Fig. 2 shows equilibrium can have too many or too few type M. In the shaded region in the left panel of , between $f(\gamma_p)$ and $f^o(\gamma_p)$, we have $n_m^o > 0 = n_m^o$ and equilibrium has too many. There is also a region where equilibrium has too few. The situation in the right panel is similar, except parameters are different. Also, even if the equilibrium and efficient outcomes are both class 2, $n_m = n_m^o$ only if bargaining powers are just right. Details are given in Proposition 2, but here is the idea. 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 an end user C. For θ_{pm}^o, there is also a holdup problem when P deals with M, but here other forces come into play. When someone chooses to be type P, he weighs 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. Summarizing all these results, we have this:

Proposition 2 *The efficient outcome exists and is generically unique. Let S_0^o, S_1^o and S_2^o be the sets of γ's where the efficient outcome is class 0, class 1 and class 2, resp. 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) $(\gamma_p, \gamma_m) \in S_2^o \Rightarrow$*

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

5 Concavity

Now suppose $U''(y) < 0$.[11] This is interesting for various reasons, but here is one big one. If $U(y) < y$ then the cost to the payer exceeds the value to the

[11] This extension is relevant for reasons dicussed below, and is a building block for the extension in Section 6, but without loss of continuity readers could skip to Section 7.

payee, which discourages intermediation because it requires two payments, M to P and C to M, rather than one, C to P. Nonlinearity can thus represent a transaction cost. Now, one can also capture this with, e.g., $U(y) = (1-t)y$, where t is a proportional cost, like an ad valorem tax, but there are other reasons to go beyond linear utility. In any case, we allow $U'(0) < 1$, so it may be that $U(y) < y \; \forall y > 0$, but for the record note that a version with $U'(0) > 1$ makes our environment similar to a standard model in monetary economics going back to Shi (1995) and Trejos and Wright (1995). Also, for tractability here we use the Kalai (1977) bargaining solution.[12]

In general the general dynamic programming equations are

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

However, for simplicity (and efficiency) consider $\theta_{pc} = \theta_{mc} = 1$, so $V_c = 0$ and $y_{cp} = y_{cm} = u$. Then letting $z = U(u)$, these equations reduce to

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

The definition of equilibrium is basically the same as the linear model. The solution method is also similar, although the algebra is more cumbersome – which is why $U(y) = y$ is our benchmark specification.

[12] The Kalai solution maximizes i's surplus in trade with j subject to i getting a share θ_{ij} of the total surplus. This is not the definition of Kalai bargaining, it is a result implied by his axioms, like maximizing the product of the surpluses is a result implied by Nash's axioms. If $U(y) = y$ here Nash and Kalai are the same; with $U'' < 0$ they are not, and Kalai has several advantages (see Aruoba et al. 2007 for a discussion in the context of related models).

Still, there are analogs to all the above results. The 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* $\gamma_p \leq F\left(\gamma_m\right)$*, 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 $V_p = V_0$, and y_{mp} is the bargaining solution. Now $V_p = V_0$ reduces to $\tilde{Q}\left(\mu, y_{mp}\right) = 0$, where $\tilde{Q}\left(\mu, y_{mp}\right) = \tilde{\kappa}_1 \mu^2 + \tilde{\kappa}_2 \mu + \tilde{\kappa}_3 = 0$ and

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

Here is the analog to Lemma 3:

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

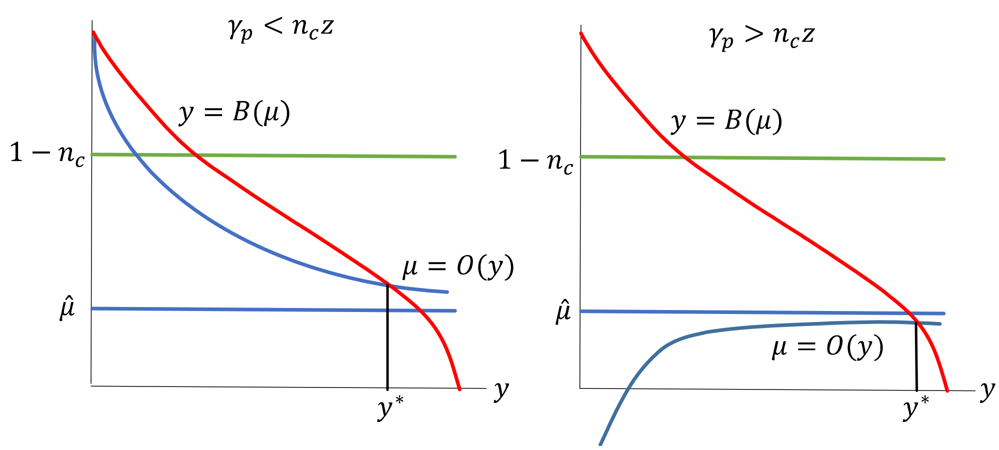

Figure 3: Equilibrium in (y_{mp}, μ) space

With a general $U(y)$ we cannot eliminate y_{mp} from the above conditions, so we work with two curves in (y_{mp}, μ) space representing bargaining and the choice to be type M or P. Setting $V_0 = V_p$ implies a quadratic that solves 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 $\mu = O(y_{mp})$, where O is 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, we get $y_{mp} = B(\mu)$, where B is for "bargaining." In fact, it can be solved for $\mu = B^{-1}(y_m)$ explicitly:

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

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

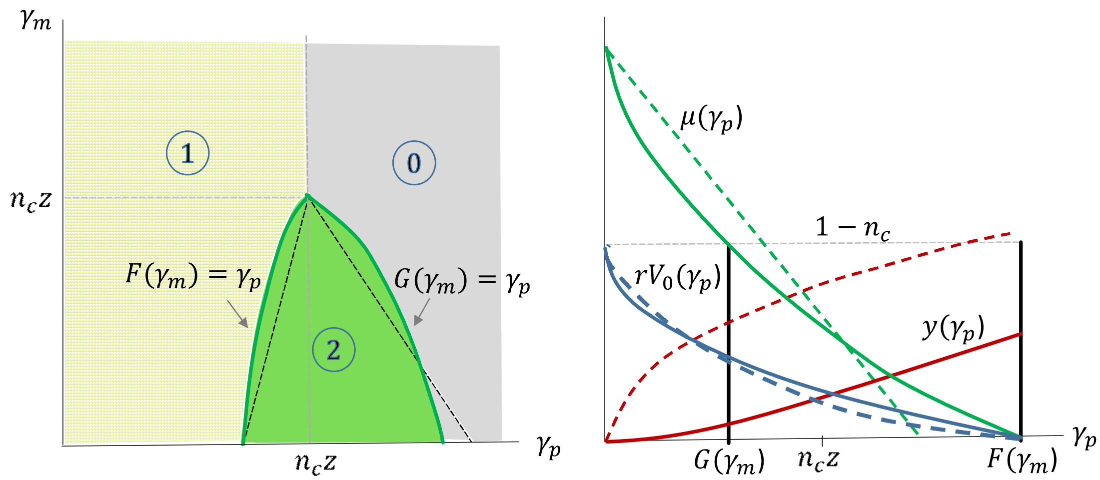

Figure 4: The nonlinear (solid) and linear (dashed) models

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* $(0, \infty) \times (0, \bar{\mu})$*.*

Proposition 3 *With* $\gamma_j > 0$ *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 dashed lines are for $\zeta = 1$.[13] In the right panel of Fig. 4, the solid curves are y_{mp}, μ and

[13]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 $U(y) = y^\zeta$ implies $U'(0) > 1$, and there is a $\hat{y} > 0$ such that in the relevant range $U(y) > y$ iff $y < \hat{y}$. Hence, nonlinearity tends to discourage intermediation when $y_{mp} > \hat{y}$ and encourage it when $y_{mp} < \hat{y}$. With $U'(0) < 1$, instead, nonlineality would unambiguously discourage intermediation.

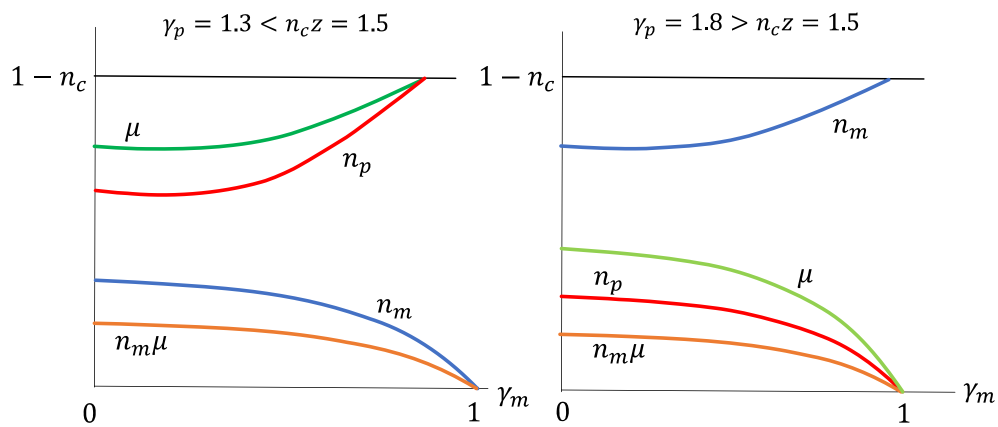

Figure 5: Effects of γ_m in the nonlinear model

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 in class 2 equilibrium.[14] As in Section 5, we allow concave U, but set $\theta_{pc} = \theta_{mc} = 1$ so that $y_{cp} = y_{cm} = u$, $V_c = 0$ and $z = U(u)$. At any point in time, a type P agent can 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 acquire the middleman technology. Also, we now work with the n's, rather than μ, and let n_1 be a state variable with law of motion

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

[14] Class 0 and 1 have no interesting dynamics, although we can potentially start with $n_m = 0$, then transit to $n_m > 0$, or vice versa.

In contrast to n_1, n_0 can jump to satisfy $V_0 = V_p$ at any point in time (like vacancies in Pissarides 2000). The other state variable is $\Delta = V_1 - V_0$, which represents "beliefs" about the value of holding, rather than searching for, x

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 to something manageable.

First, $V_p = V_0\ \forall t$ implies $\dot{V}_p = \dot{V}_0\ \forall t$, so from (22)-(23) $n_m \in (0, 1 - n_c)$ means

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

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

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

Then, substituting (25) and simplifying, we arrive at

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

Now (20) and (26) define a dynamical system in (n_1, Δ), with n_0 and y_{mp} implicit functions of the state.[15]

Definition 2 *Given an initial condition $\bar{n}_1$, an equilibrium is a nonnegative and bounded path for (n_1, Δ) solving (20) and (26).*

[15] Section 8 presents dynamics with $U(y) = y$, which is easier since we can solve for n_0 and y_{mp} explicitly; the goal here is to understand dynamic equilibria in the nonlinear model.

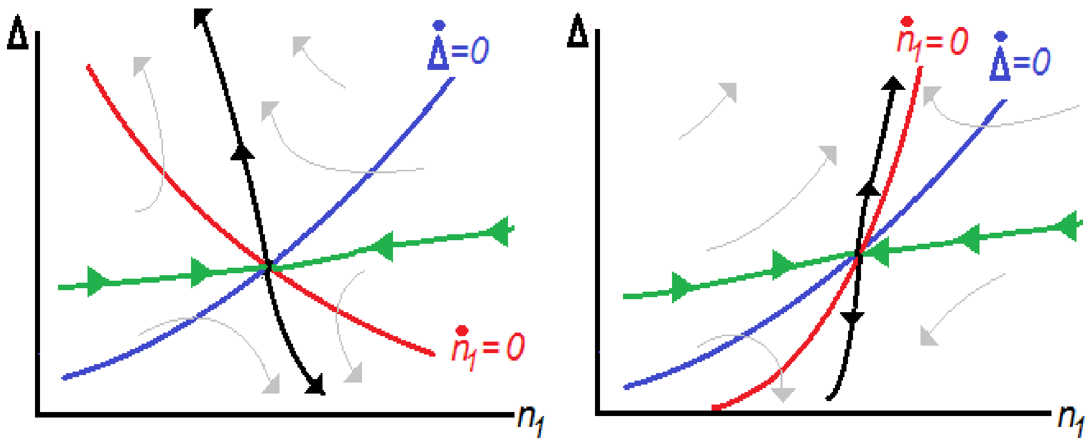

Figure 6: The phase plane with $\gamma_j > 0$

By Proposition 3 there exists a unique steady state, i.e., a unique intersection of the $\dot{n}_1 = 0$ and $\dot{\Delta} = 0$ curves. These curves have slopes

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

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, $\partial \dot{n}_1/\partial n_1 < 0$ and $\partial \dot{\Delta}/\partial n_1 < 0$. Hence the system looks like Fig. 6, which show the phase plane with $\dot{n}_1 = 0$ in red and $\dot{\Delta} = 0$ in blue. Whether $\dot{n}_1 = 0$ slopes up or down, the steady state is a saddle point, with the green and black curves showing the stable and unstable manifolds.

Proposition 4 *The unique class 2 steady state exhibits saddle path stability: given any initial condition $\bar{n}_1$, there is a unique $\bar{\Delta}$ such that (n_1, Δ) transits to the steady state, while any $\Delta \neq \bar{\Delta}$ implies an explosive path.*

We conclude that equilibrium, and not only steady state, is unique. This was not a foregone conclusion. In linear versions of a related model by Duffie et al. (2005), there is a unique equilibrium, but there can be multiplicity in the nonlinear dynamic version (Trejos and Wright 2016). Hence, saddle-path stability is not trivial in these kinds of models. An important aspect from the current perspective is this: the next part of the presentation establishes that when we consider asset market intermediation there is multiplicity. This contrasts with goods market intermediation, where we get uniqueness, even when we allow nonlinear utility and consider nonstationary equilibrium.

7 Intermediated Asset Markets

Above we assume a cost $\gamma_j > 0$ to storing x, as in retail establishments. Now consider $\gamma_j < 0$, so storing x is profitable. Perhaps, e.g., P is a painter, C is a collector, and M is a market-maker in fine art. Then the flow benefit of holding x is $\rho_j = -\gamma_j > 0$ if paintings generate positive utility, and $\rho_m > \rho_p$ corresponds to art dealers getting a higher payoff from holding a piece than the original artist, perhaps by charging admission to their galleries. What is relevant is $\rho_m > 0$; it is not qualitatively important whether $\rho_p > 0$ or $\rho_p < 0$. In either case, P sells his work to anyone, since he can always produce more. The interesting issue is this: if C enjoys x enough, M may sell it to him; but then again M may prefer to keep it for himself. This option is never relevant when $\gamma_j > 0$, since retailers obviously do not pay to store inventories unless they are planning to sell them.

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 has access to capital available for investment. An end user C may have the best ultimate use for the capital, but M can also generate a flow yield by putting it into temporary investments. Then $\rho_m > \rho_p$ means M has better investment opportunities than P, and ρ_m might even be high enough for M to keep x rather than passing it on to C. If M does pass x to C he is acting as a financial intermediary, by acquiring capital from the source agent and retrading it to those who need it. As another application, x can be a house providing utility as shelter, in which case M might either keep it as a residence or "flip" it to C.

These examples motivate consideration of $\gamma_j < 0$. To begin the analysis, write the dynamic programming equations as

$$rV_p = n_c\theta_{pc}u + n_m(1-\mu)\theta_{pm}(V_1 - V_0) - \gamma_p \tag{27}$$

$$rV_0 = n_p\theta_{mp}(V_1 - V_0) \tag{28}$$

$$rV_1 = n_c\theta_{mc}\tau(u + V_0 - V_1) - \gamma_m, \tag{29}$$

which are the same as (2)-(4), except for the appearance of τ, which is the probability that M trades x to C. This also affects the steady state, which is $\mu = n_p/\left(np + n_c\tau\right)$. There are now two best-response conditions: one for the decision to be type P or M, which is basically the same as above; plus one for the decision by M to keep or trade x when he meets C. He keeps it, $\tau = 0$, if $-\gamma_j/r$ exceeds the deviation payoff V_1^d, which is captured by a one-shot deviation – i.e., V_1^d is the payoff to trading x to C when an opportunity arises, then reverting to a hoarding strategy the next time M acquires x. Similarly, he trades it, $\tau = 1$, if the payoff exceeds the deviation payoff, $V_1^d = -\gamma_m/r$. And he can randomize, $\tau \in (0,1)$, if indifferent. Other than consideration of this new condition for τ, the definition of equilibrium is the same as in the baseline model.

To make the main point in a relatively simple way, let us set $U(y) = y$, and focus on steady state until Section 8. Also, when it facilitates the interpretation, let $\rho_j = -\gamma_j$. One can imagine 9 candidate equilibria, shown in Table 1. 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 would trade x to C. We call this a class 1^T equilibrium (T indicates M trades x). Similarly, we call $n_m = 0$ and $\tau = 0$ a class 1^K equilibrium, as off the equilibrium path M with x would not trade it to C (K indicates M keeps x). We would similarly call $n_m = 0$ and $\tau \in (0,1)$ a class 1^R equilibrium (R indicates M

randomizes), but ignore it because it does not exist for generic parameters. We also ignore all candidates with $n_m = 1 - n_c$.[16]

$n_m \backslash \tau$	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 $\rho_j = -\gamma_j > 0$.

Therefore, the remaining candidates are class 1^K and 1^T, plus $n_m \in (0, 1-n_c)$ and either $\tau = 1$, $\tau = 0$ or $\tau \in (0,1)$, which we call class 2^K, 2^T or 2^R (there are two types, P and M, and M either keeps x, trades it or randomizes). The following is proved in the Appendix:

Lemma 14 *Define*

$$\widehat{\gamma}_m \equiv -u[r + (1-n_c)\theta_{mp}] \tag{30}$$

$$\widehat{f}(\gamma_p) \equiv -(\overline{\gamma}_p - \gamma_p)\frac{r + (1-n_c)\theta_{mp}}{(1-n_c)\theta_{mp}} \tag{31}$$

$$k(\gamma_p) \equiv -ru - \overline{\gamma}_p + \gamma_p. \tag{32}$$

Also define $\widehat{k}(\gamma_p)$ to be the lower root of the quadratic given in the proof in the Appendix. Then as shown in Fig. 7, class 1^T equilibrium exists iff $\gamma_m \geq \max\{\widehat{\gamma}_m, f(\gamma_p)\}$; 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 $\widehat{k}(\gamma_p) < \gamma_m < k(\gamma_p)$; and class 2^T exists iff $\widehat{k}(\gamma_p) \leq \gamma_m \leq f(\gamma_p)$.

Fig. 7 illustrates Lemma 14 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 better for

[16] If $n_m = 1 - n_c$ there are no producers, so $\tau > 0$ (trading away x) leaves M with continuation value 0, which admits a profitable deviation where he becomes type P. We cannot rule out $n_m = 1 - n_c$ and $\tau = 0$ in this way, but ignore it because it is degenerate (production shuts down). Also notice there is nothing corresponding to a class 0 outcome in Table 1, since production obviously dominates nonparticipation when $\gamma_j < 0$.

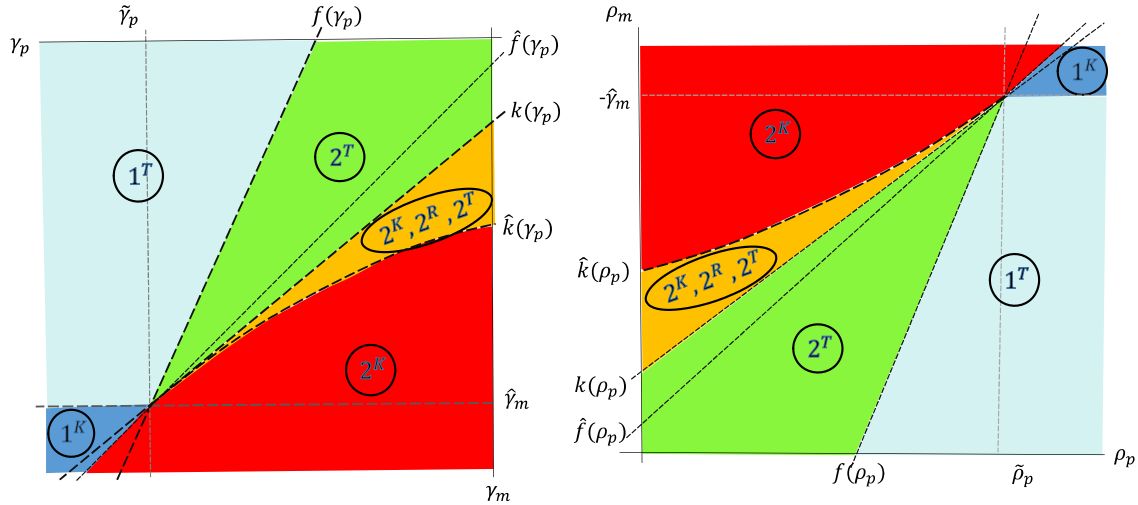

Figure 7: Outcomes in (γ_p, γ_m) and in (ρ_p, ρ_m) space

interpreting the results. Notice all of the equilibria in Table 1 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). 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_m there are multiple equilibria (orange region).

The economics is clear: If type M agents decide to trade x to end users, there will be more middlemen without x. Since type P can trade with either end users or middlemen in search of capital, this raises V_p, which raises n_p. That makes is easier for type M to acquire capital, rationalizing their decision to trade it away, and thus making active intermediation an equilibrium. If, however, type M agents instead decide to keep x for themselves, in steady state P trades only with end users, which lowers V_p and n_p. That makes it harder for type M agents to acquire capital, rationalizing their decision to keep it, and thus making no

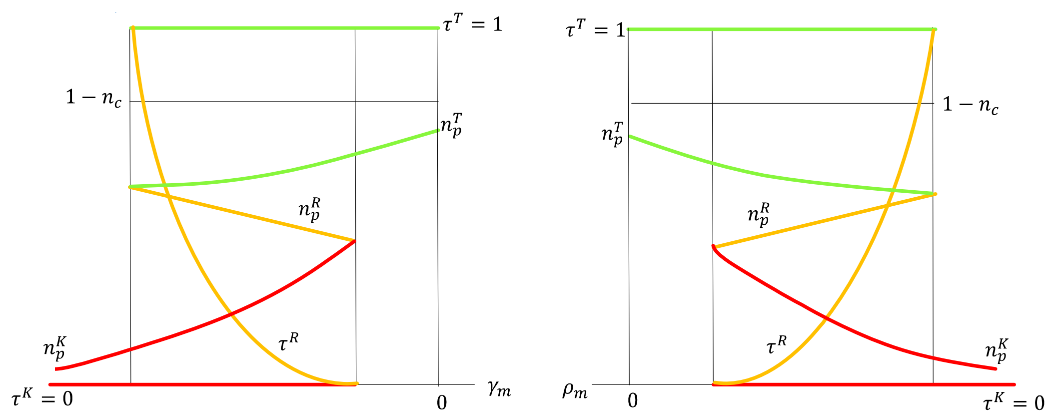

Figure 8: Equilibrium correspondence for τ and n_p

intermediation an equilibrium. This strategic effect implies $2T$ and $2K$ equilibria coexist in a region of parameter space with positive measure, and when they coexist, so does 2^R. Fig. 8 shows how n_p and τ vary within an equilibria as γ_m or ρ_m changes, and across equilibria for a fixed γ_m or ρ_m.

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 when $\gamma_j > 0$ – suggesting there is something special about intermediated asset markets, consistent with the discussion in the Introduction. Naturally, type M agents always trade away assets that give them low returns, always hoard those that give them high returns, and may or may not trade assets that give them moderately good returns, depending on the equilibrium selected. This is summarized as part (i) of the next Proposition. Part (ii) says that if $\mathbf{n}$ is fixed exogenously then equilibrium is once again unique, even with $\gamma_j < 0$. Intuitively, unless $\mathbf{n}$ can adjust, the strategic channel is severed and τ depends only on fundamentals.

Proposition 5 *(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 Intermediated Asset Market Dynamics

The last extension concerns dynamic equilibria with $\rho_j = -\gamma_j > 0$. As in Section 6, the focus is on class 2 equilibria, but now we need to be mindful of M's decision to keep or trade x. Also, we use $U(y) = y$, because this suffices to make the point. Similar dynamic systems have been studied in search models by, e.g., Diamond and Fudenberg (1989), Boldrin et al. (1993), Coles and Wright (1998) and Mortensen (1999). However, those models generally display interesting dynamics only under increasing returns in the matching or production technology, something we do not need. Still, a common set of mathematical methods is used in these papers, and readers are referred there for more detail; the presentation here focuses on some examples and general qualitative issues.

The law of motion for n_1 is similar to (20), except it includes τ,

$$\dot{n}_1 = n_0(1 - n_c - n_1 - n_0) - n_1 n_c \tau. \tag{33}$$

Bargaining with $U(y) = y$ can be solved explicitly, and implies

$$rV_p = n_c\theta_{pc}u + n_0\theta_{pm}\Delta - \gamma_p + \dot{V}_p \tag{34}$$

$$rV_0 = (1 - n_c - n_1 - n_0)\theta_{mp}\Delta + \dot{V}_0 \tag{35}$$

$$rV_1 = n_c\tau\theta_{mc}(u - \Delta) - \gamma_m + \dot{V}_1. \tag{36}$$

Also, with $U(y) = y$ the condition $V_p = V_0$ can be solved for

$$n_0 = \frac{\gamma_p - n_c\theta_{pc}u}{\Delta} + (1 - n_c - n_1)\theta_{mp}.$$

This gives n_0 as an explicit function of the state, $n_0(n_1, \Delta)$. The other decision rule $\tau = \tau(\Delta)$ is even easier: $\tau = 0$, $[0, 1]$ or 1 as $\Delta - u$ is positive, 0 or negative.

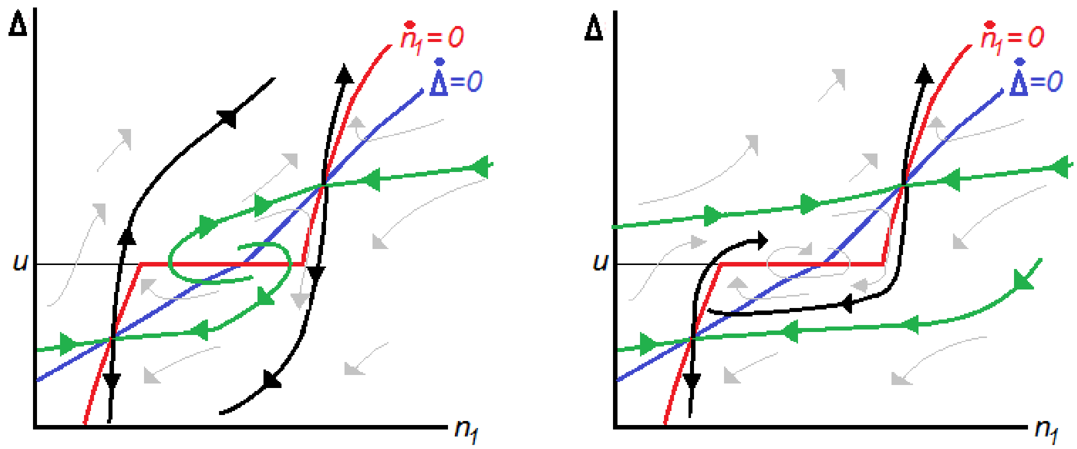

Figure 9: The phase plane with $\rho_j > 0$

As usual, subtracting (35)-(36), we get

$$\dot{\Delta} = r\Delta - n_c \tau \theta_{mc} (u - \Delta) + \gamma_m + (1 - n_c - n_1 - n_0)\theta_{mp}\Delta. \tag{37}$$

Inserting $\tau(\Delta)$ and $n_0(n_1, \Delta)$ into (33) and (37), we arrive at the system

$$\dot{n}_1 = n_0(n_1, \Delta)[1 - n_c - n_1 - n_0(n_1, \Delta)] - n_1 n_c \tau(\Delta) \tag{38}$$

$$\dot{\Delta} = r\Delta - n_c \tau(\Delta)\theta_{mc}(u - \Delta) + \gamma_m + [1 - n_c - n_1 - n_0(n_1, \Delta)]\theta_{mp}\Delta \tag{39}$$

Given an initial $\bar{n}_1$, any path (n_1, Δ) satisfying these conditions while remaining bounded and nonnegative constitutes an equilibrium. Again one can interpret Δ as "beliefs" about the value of holding an asset. As is standard in this kind of analysis, the initial value $\bar{\Delta}$ is not given by nature, and equilibrium "beliefs" can be anything as long as the path emanating from $(\bar{n}_1, \bar{\Delta})$ is bounded and nonnegative.

We know from Section 7 that for some parameters there are three steady states. Fig. 9 shows this situation, where the red curve is $\dot{n}_1 = 0$ and the blue curve is $\dot{\Delta} = 0$. Compared to Fig. 6, notice that the $\dot{n}_1 = 0$ curve has a flat spot, while the $\dot{\Delta} = 0$ curve has a kink, at $\Delta = u$, which is where τ switches from

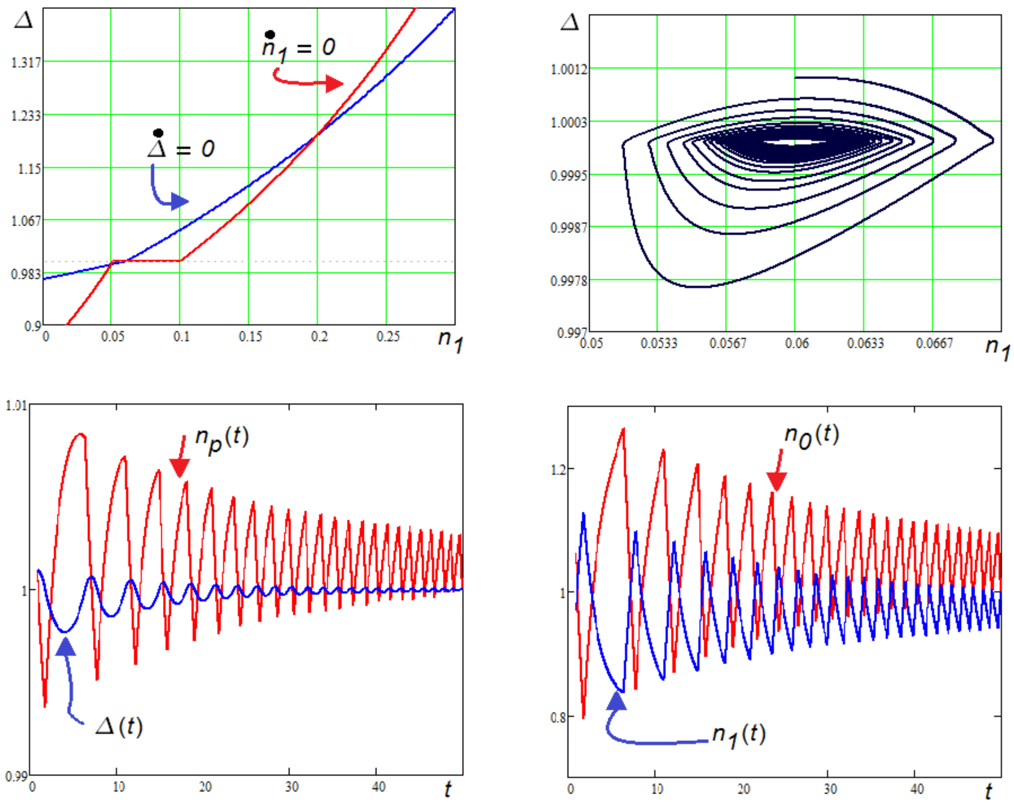

Figure 10: Dynamics in an example with $\rho_m > 0$

0 to 1. The lower steady state, call it (n_1^L, Δ^L), has $\Delta < u$ and $\tau = 1$, so the asset circulates; the higher one (n_1^H, Δ^H) has $\Delta > u$ and $\tau = 0$, so the asset is hoarded; and the one in between (n_1^M, Δ^M) has $\Delta = u$ and $\tau \in (0, 1)$. The results in Section 6 apply to (n_1^L, Δ^L) and (n_1^H, Δ^H), which are again saddle points, and the green and black curves are their stable and unstable manifolds.

We say more about the general case below, but consider first an example. Let $u = 1$, $r = 0.05$, $n_c = 0.3$, $\theta_{pm} = 0.5$, $\theta_{pc} = \theta_{mc} = 1$, $\rho_m = 0.36$ and $\rho = 0$. The upper left panel of Fig. 10 shows three steady states, approximately $n_1^L = 0.05$, $n_1^M = 0.07$ and $n_1^H = 0.25$. As is the case general, $\Delta^L < u$, $\Delta^M = u$ and $\Delta^H > u$, so M trades x to C in the low steady state, hoards x in the low steady, and randomizes in the middle. The upper right panel zooms in to show local dynamics, with approximate convergence after a hundred iterations to either to $\left(n_1^M, \Delta^M\right)$, or possibly to a limit cycle around it. While we cannot literally know this from the numerical output, in principle we could check the local stability of $\left(n_1^M, \Delta^M\right)$ directly, but that is slightly complicated because the $\dot{n}_1 = 0$ curve is nondifferentiable at $\Delta = u$. However, it is clear from other examples (e.g., reduce θ_{pc} from 1.0 to 0.1) that $\left(n_1^M, \Delta^M\right)$ can be a source: no matter how close we start to it, orbits rapidly move away.[17]

To emphasize the qualitative features of the system, notice that starting with any $\bar{n}_1$ in the neighborhood of n_1^M there is a continuum of equilibria indexed by an arbitrary choice of $\bar{\Delta}$ in some interval, because all paths starting at $\left(\bar{n}_1, \bar{\Delta}\right)$ remain bounded and nonnegative. In particular, after a shock to the system – e.g., an unexpected reduction in μ due to, say, bad weather – there are many equilibria that cycle around $\left(n_1^M, \Delta^M\right)$, as in Fig. 10. Hence, small changes in fundamentals can lead to very volatile reactions. Notice that while the fluctuations in Δ or n_p are not that big relative to their long-run average values, as shown in the lower left panel, the fluctuations in n_1 and n_0 are around 10% and 20% relative to their long-run average values (check the units on the vertical axis). While obviously this is not a calibration exercise, we mention that this outcome resembles the data in that percentage fluctuations in inventories are much bigger than output.

More about global dynamics in general can be discerned from Fig. 9. In the left panel, the stable manifolds of the two saddle points are trapped inside the unstable manifolds. This means the stable manifolds must wrap around the middle steady state, either emanating from it or from a limit cycle around it. In either case, again, starting with any $\bar{n}_1$ in the neighborhood of n_1^M, there is

[17] A "trick" that might make local analysis easier is to purify the mixed-strategy equilibrium by giving every type M a slightly different ρ_m to smooth out the $\dot{n}_1 = 0$ curve; the model in the text is the limiting version where these differences get small. Also, the papers mentioned above on dynamics in search models, especially Coles and Wright (1998) and Mortensen (1999), show their analog of $\left(n_1^M, \Delta^M\right)$ is a source for some parameters, and use this to prove there are stable limit cycles around it by application of a saddle-loop bifircation and the Poincare-Bendixson theorem. Going into these extended calculations and purifying the mixed-strategy equilibrium is beyond the scope of this project.

a continuum of equilibria indexed by $\bar{\Delta}$, and after a shock, there are a great many different equilibrium paths that fluctuate around $\left(n_1^M, \Delta^M\right)$, as well as equilibrium paths starting on one of the stable manifolds and asymtoptically approaching either $\left(n_1^L, \Delta^L\right)$ or $\left(n_1^H, \Delta^H\right)$. The economics is similar in the right panel, except now for any $\bar{n}_1$ in a very large range, depending on initial beliefs, the system can transit to $\left(n_1^L, \Delta^L\right)$ or $\left(n_1^H, \Delta^H\right)$, as well as oscillate on its way to $\left(n_1^M, \Delta^M\right)$ or to a cycle around it.

The general conclusion is that our intermediated asset market is not only subject to multiplicity, but dynamic indeterminacy, and excess volatility defined as fluctuations in endogenous variables while fundamentals are constant. Our market is also subject to fragility, where a small change in fundamentals can lead to a structural change in the equilibrium set. In the upper left panel of Fig. 10, e.g., a small increase in ρ_m shifts the $\dot{n}_1 = 0$ curve down, leaving one instead of three steady states, and hence potentially having a discrete impact on outcomes if we start in a steady states that disappears. Similar results have been discussed related models, as mentioned above, but they require increasing returns; we rely instead only on the self-referential nature of trading strategies in frictional markets. At the risk of repetition, a key assumption is that the object being traded bears a positive yield, not a storage cost, suggesting that intermediation in asset (as compared to consumer goods) markets may indeed be special. Another key assumption is that the composition of the market is endogenous. Absent both these features, equilibrium is unique and exhibits saddle-path stability.

9 Conclusion

This project began by asking "who wants to be a middleman?" – something we view as continuing the development of search-based theories of intermediation. We built on the standard framework introduced by Rubinstein and Wolinsky (1987), extended to incorporate more general bargaining, technology, and utility.

We went beyond steady states by analyzing dynamics. Perhaps most importantly, we let agents choose to act as either producers or middlemen, which we think is relevant for many issues.[18] 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, although we were able to reconcile that with intuition. Perhaps the most surprising finding was that equilibrium is unique when holding goods involved storage costs, while multiplicity, indeterminacy, volatility and fragility emerged when holding assets involved positive returns, as at least some people seem to think characterizes real-world asset markets, .

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), or 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 models in Burdett and Coles (1997) or Shimer and Smith (2000), where a person can only have $n \in \{0,1\}$ partners. But sometimes the simplicity of unit inventories is worth the loss in generality and realism, as demonstrated in search theory going back at least to Diamond (1982). Moreover, we think $x \in \{0,1\}$ is not critical to the economic insights. Suppose middlemen can hold, say, $x \in \{0,1,...\bar{x}\}$. Without working through the details, which might get messy, higher n_m and lower n_p should still entail

[18] 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 $\{0,1\}$. We say more on this below, but 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, even when the equilibrium is unique, with n_m, n_p and μ varying during transitions or after shocks. Weill (2007) and Duffie et al. (2007) study these issues in a related but different model, and discuss why such transitions are interesting. Of course, when there are multiple equilibria the situation is even more interesting.

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 their hoarding (trading) decisions. Hence, multiplicity in intermediated asset markets seems robust.

Still, future research should pursue such generalizations explicitly. With or without relaxing inventory restrictions, another idea is to add heterogeneity *across middlemen*, to generate interdealer trade, where goods or assets get passed from one intermediary to another before reaching end users. Or, one can add differences *across goods*, say in γ, to further develop the interpretation of Gresham's Law. There is also more one could do on dynamics. A natural idea is to try to construct sunspot equilibria, where intermediation activity varies stochastically over time as a self-fulfilling prophecy, even though fundamentals are deterministic, as another manifestation of excess volatility. Standard methods could be used for this construction, in principle, although it may 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 the study of intermediation, and in search theory 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] M. Boldrin, N. Kiyotaki and R. Wright (1993) "A Dynamic Equilibrium Model of Search, Production and Exchange," *JEDC* 17, 723-58.

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

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

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

[8] M. Coles and R. Wright (1998) "A Dynamic Model of Search, Bargaining, and Money," *JET* 78, 32-54.

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

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

[11] P. Diamond and D. Fudenberg (1989) "Rational Expectations Business Cycles in Search Equilibrium," *JPE* 97, 606-19.

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

[13] D. Duffie, N. Gârleanu and L. Pederson (2007) "Valuation in Over-the-Counter Markets," *RFS* 20, 1865-1900.

[14] M. Farboodi, G. Jarosch and G. Menzio (2015) "Tough Middlemen," mimeo.

[15] M. Farboodi, G. Jarosch and R. Shimer (2016) "Meeting Technologies in Decentralized Asset Markets," mimeo.

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

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

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

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

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

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

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

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

[24] 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.

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

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

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

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

[29] D. Mortensen (1999) "Equilibrium Unemployment Dynamics," *IER* 40, 889-914.

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

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

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

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

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

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

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

[37] S. Shi (1995) "Money and Prices: A Model of Search and Bargaining," *JET* 67, 467-96.

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

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

[40] A. Trejos and R. Wright (1995) "Search, Bargaining, Money, and Prices," *JPE* 103, 118-41.

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

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

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

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

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

[46] 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 $\gamma_p > \bar{\gamma}_p + (1 - n_c)\theta_{pm}(\bar{\gamma}_m - \gamma_m)/(r + n_c\theta_{mc})$, 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

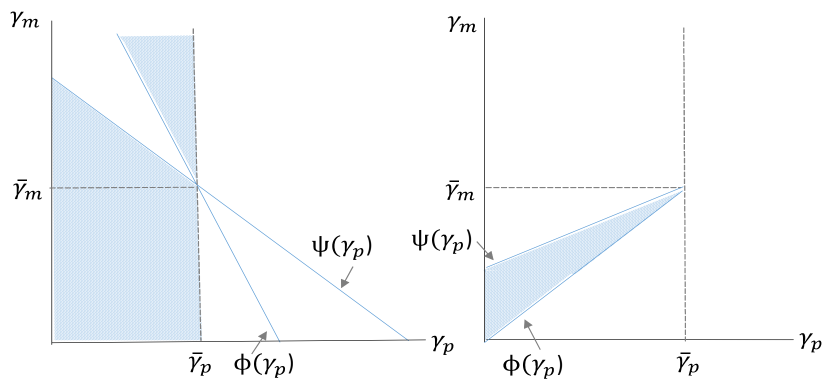

Figure 11: The functions $\phi(\gamma_p)$ and $\psi(\gamma_p)$

the left panel of Fig. 11, 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. 11

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. 12.

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

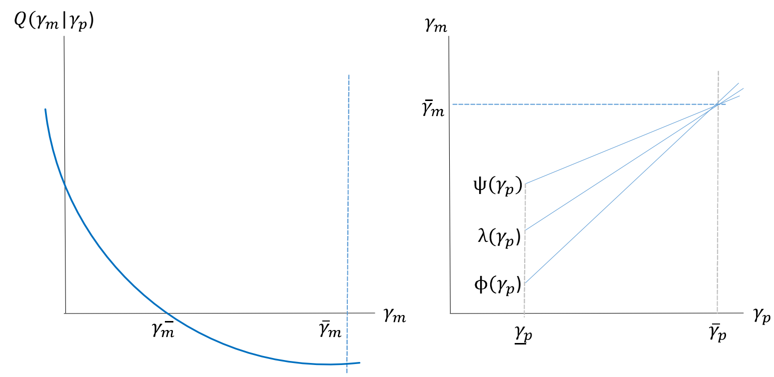

Figure 12: The functions $\hat{Q}(\gamma_m|\gamma_p)$ and $\lambda(\gamma_p)$

$\Lambda = \hat{\kappa}_2^2 - 4\hat{\kappa}_1\hat{\kappa}_3 > 0$. 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. 12 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. 12.

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 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,

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

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 2: 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

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

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

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

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

$$G'(\gamma_m) = \frac{-\theta_{pm}(1 - n_c)U'(y_0)}{(r + n_c)[\theta_{pm} + (1 - \theta_{pm})U'(y_0)]} < 0$$

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

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

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

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

Lemma 14: For preliminaries, first solve equations (27)-(29) for

$$rV_p = \frac{(r+\tau n_c\theta_{mc}+n_p\theta_{mp})(\overline{\gamma}_p-\gamma_p)+n_m(1-\mu)\theta_{pm}(\tau\overline{\gamma}_m-\gamma_m)}{r+\tau n_c\theta_{mc}+n_p\theta_{mp}} \quad (40)$$

$$rV_0 = \frac{n_p\theta_{mp}(\tau\overline{\gamma}_m-\gamma_m)}{r+\tau n_c\theta_{mc}+n_p\theta_{mp}} \quad (41)$$

$$rV_1 = \frac{(r+n_p\theta_{mp})(\tau\overline{\gamma}_m-\gamma_m)}{r+\tau n_c\theta_{mc}+n_p\theta_{mp}}, \quad (42)$$

and notice the steady state condition (**??**) implies

$$n_p = \frac{n_c\tau\mu}{1-\mu} \text{ and } n_m = \frac{(1-n_c)(1-\mu)-n_c\tau\mu}{1-\mu}. \quad (43)$$

We now consider each five candidate equilibria in Table 1.

Equilibrium 1^K: In a candidate equilibrium with $\tau = 0$ and $n_m = 0$, (40)-(42) reduce to

$$rV_p = \overline{\gamma}_p - \gamma_p \quad (44)$$
$$rV_0 = (1-n_c)\theta_{mp}(V_1 - V_0) \quad (45)$$
$$rV_1 = -\gamma_m. \quad (46)$$

The best response condition for $\tau = 0$ is $V_1^D \leq V_1$, where V_1^D is the value to setting $\tau = 1$, then reverting to the candidate equilibrium strategy, with payoff V_0 given by (45):

$$(r+n_c\theta_{mc})V_1^D = \overline{\gamma}_m - \gamma_m - \frac{n_c\theta_{mc}(1-n_c)\theta_{mp}}{[r+(1-n_c)\theta_{mp}]}\frac{\gamma_m}{r}.$$

Simplifying, $\tau = 0$ is a best response iff $\gamma_m \leq \widehat{\gamma}_m$ where $\widehat{\gamma}_m$ is defined in (30).

Similarly, the best response condition for $n_m = 0$, $V_p \geq V_0$, simplifies to $\gamma_m \geq \widehat{f}(\gamma_p)$ where $\widehat{f}\left(\gamma_p\right)$ is defined in (31). Hence, a class 1^K equilibrium exists iff $\widehat{f}(\gamma_p) \leq \gamma_m \leq \widehat{\gamma}_m$.

Equilibrium 1^T: Consider next the candidate $\tau = 1$ and $n_m = 0$. It is routine to check the best response for $\tau = 1$ is $\gamma_m \geq \widehat{\gamma}_m$. The best response for $n_m = 0$ is $V_p \geq V_0$, which simplifies to $\gamma_m \geq f(\gamma_p)$, where $f\left(\gamma_p\right)$ is defined in (7) in the text. Hence, a class 1^T equilibrium with $\tau = 1$ and $n_m = 0$ exists iff $\gamma_m \geq \max\{\widehat{\gamma}_m, f(\gamma_p)\}$.

Equilibrium 2^K**:** Consider next the candidate $\tau = 0$ and $n_m \in (0, 1 - n_c)$, where the dynamic programming equations are the same as (44)-(46). The best response condition for $\tau = 0$ is $V_1^D \leq V_1$, where V_1^D is the the value to setting $\tau = 1$, but reverting to the candidate strategy for continuation payoff $V_0 = V_p$. Algebra implies

$$V_1^D = \frac{r(\overline{\gamma}_m - \gamma_m) + n_c\theta_{mc}(\overline{\gamma}_p - \gamma_p)}{r(r + n_c\theta_{mc})}.$$

It is now easy to check $V_1^D \leq V_1$ iff $\gamma_m \leq k(\gamma_p)$ where $k(\gamma_p)$ is defined in (32).

The best response condition for $n_m \in (0, 1 - n_c)$, $V_p = V_0$, now implies

$$n_p = \frac{r(\overline{\gamma}_p - \gamma_p)}{\theta_{mp}(\gamma_m + \overline{\gamma}_p - \gamma_p)}.$$

We need to check $n_p \in (0, 1 - n_c)$. It turns out $n_p < 1 - n_c$ is the binding condition, and holds iff $\gamma_m \leq \widehat{f}(\gamma_p)$. So class 2^K equilibrium exists iff $\gamma_m \leq \min\{k(\gamma_p), \widehat{f}(\gamma_p)\}$.

Equilibrium 2^R: Consider next $\tau \in (0, 1)$ and $n_m \in (0, 1 - n_c)$. The best response condition for $\tau \in (0, 1)$ is $rV_1 = -\gamma_m$, which holds iff $u = V_1 - V_0$ iff $\gamma_m = -[r + (1 - n_c - n_m)\theta_{mp}]u$. Solving for n_m, we get

$$n_m = \frac{\gamma_m + [r + (1 - n_c)\theta_{mp}]u}{u\theta_{mp}}.$$

Clearly, $n_m \in (0, 1 - n_c)$ iff $-ru - (1 - n_c)\theta_{mp}u < \gamma_m < -ru$.

The best response condition for $n_m \in (0, 1 - n_c)$ now implies

$$\tau = \frac{(\gamma_m + ru)(\gamma_m + ru + \overline{\gamma}_p - \gamma_p)}{n_c u[\gamma_m + ru + \theta_{mp}(\overline{\gamma}_p - \gamma_p) + \theta_{pm}\theta_{mp}(1 - n_c)u]}.$$

We next obtain the set of parameters such that $\tau \in (0,1)$ and $n_m \in (0, 1-n_c)$. To see when $\tau > 0$, denote the denominator of τ by D. There are two possibilities, $D < 0$ and $D > 0$. The former can be shown to be inconsistent with $\tau > 0$, so we are left with $D > 0$, which holds iff

$$\gamma_m > \Psi\left(\gamma_p\right) \equiv -ru - \theta_{mp}(\overline{\gamma}_p - \gamma_p) - \theta_{pm}\theta_{mp}(1-n_c)u.$$

Given this, $\tau > 0$ iff $\gamma_m < k(\gamma_p)$. So $\tau > 0$ iff $\Psi\left(\gamma_p\right) < \gamma_m < k(\gamma_p)$.

Also, using $D > 0$, algebra implies $\tau < 1$ iff $Q_1(\gamma_m) < 0$ where $Q_1(\gamma_m) = a_1\gamma_m^2 + b_1\gamma_m + c_1 < 0$, with $a_1 = 1$,

$$\begin{aligned} b_1 &= \overline{\gamma}_p - \gamma_p + 2ru - n_c u \\ c_1 &= ru(\overline{\gamma}_p - \gamma_p + ru) - n_c u[ru + \theta_{mp}(\overline{\gamma}_p - \gamma_p) + \theta_{pm}\theta_{mp}(1-n_c)u]. \end{aligned}$$

We cannot sign b or c, but we can show $Q_1(\gamma_m^*) < 0$, where γ_m^* solves $Q_1'(\gamma_m^*) = 0$. Hence, there are three possibilities for $Q_1(\gamma_m) < 0$, all of which reduce to $\gamma_m > \widehat{k}(\gamma_p)$, where $\widehat{k}(\gamma_p)$ represents the lower root of $Q_1(\gamma_m) = 0$:

1. if $\gamma_m^* > 0$ then $Q_1(0) < 0$ and so $Q_1(\gamma_m) < 0$ iff $\gamma_m > \widehat{k}(\gamma_p)$;
2. if $\gamma_m^* < 0$ and $Q_1(0) < 0$ then $Q_1(\gamma_m) < 0$ also implies $\gamma_m > \widehat{k}(\gamma_p)$;
3. if $\gamma_m^* < 0$ and $Q_1(0) > 0$ then $Q_1(\gamma_m) < 0$ implies $\gamma_m^+ > \gamma_m > \widehat{k}(\gamma_p)$, where γ_m^+ is the upper root of $Q_1(\gamma_m) = 0$, but $\tau > 0$ implies $\gamma_m < k(\gamma_p)$, and we can show $k(\gamma_p) < \gamma_m^+$, so $\gamma_m^+ > \gamma_m$ is not binding.

In sum, all these possibilities imply, given the other conditions, that $\tau < 1$ iff $Q_1(\gamma_m) < 0$ iff $\gamma_m > \widehat{k}(\gamma_p)$. Now we claim that when this holds, the earlier condition $\gamma_m \geq \Psi\left(\gamma_p\right)$ is not binding. To see this, check that $\widehat{k}(\gamma_p)$ intersects Ψ at $(\widetilde{\gamma}_p, -ru-(1-n_c)\theta_{mp}u)$ and $(\overline{\gamma}_p+(1-n_c)\theta_{pm}u, -ru)$, and that $\widehat{k}(\gamma_p)$ is increasing and concave; hence $\widehat{k}(\gamma_p) > \Psi\left(\gamma_p\right)$, so the binding constraint is $\gamma_m \geq \widehat{k}(\gamma_p)$. Putting this together, class 2^R equilibrium exists iff $\widehat{k}(\gamma_p) \leq \gamma_m \leq k(\gamma_p)$.

Equilibrium 2^T: Consider $\tau = 1$ and $n_m \in (0, 1-n_c)$. We first solve $V_p = V_0$ for μ and check when $\mu \in (0, \overline{\mu})$, as that is equivalent to $n_m \in (0, 1-n_c)$, where $\overline{\mu} = 1 - n_c$ when $\tau = 1$. By (40)-(41), $V_p = V_0$ iff

$$(r + n_c\theta_{mc} + n_p\theta_{mp})(\overline{\gamma}_p - \gamma_p) + [n_m\theta_{pm}(1-\mu) - n_p\theta_{mp}](\overline{\gamma}_m - \gamma_m) = 0.$$

Using (43) to eliminate n_p and n_m, then simplifying, we get $Q_2(\mu) = a_2\mu^2 + b_2\mu + c_2 = 0$ where

$$\begin{aligned} a_2 &= \theta_{pm}(\overline{\gamma}_m - \gamma_m) \\ b_2 &= -(r + n_c\theta_{mc} - n_c\theta_{mp})(\overline{\gamma}_p - \gamma_p) - [n_c + 2(1-n_c)\theta_{pm}](\overline{\gamma}_m - \gamma_m) \\ c_2 &= (r + n_c\theta_{mc})(\overline{\gamma}_p - \gamma_p) + (1-n_c)\theta_{pm}(\overline{\gamma}_m - \gamma_m). \end{aligned}$$

We need to check when the solution $Q_2(\mu) = 0$ is in $(0, \overline{\mu})$. In principle $Q_2(\mu)$ can have one or two roots. Since $Q_2(0) > 0$ and $\mu > 0$, the one root result corresponds to the lower-root of $Q_2(\mu)$, call it μ^-. Because μ^- exists in either results, we start analyzing μ^-. First, $\mu^- > 0$ requires $\gamma_m > \overline{\gamma}_m + (r + n_c\theta_{mc})(\overline{\gamma}_p - \gamma_p)/[(1-n_c)\theta_{pm}]$, which is non-binding. Second, $\mu^- < 1 - n_c$ requires $\gamma_m < f(\gamma_p)$, which this is equivalent to $Q_2(\overline{\mu}) < 0$. This last result is indicative that the two-root result is impossible because we have $Q_2(0) > 0$ and $Q_2(\overline{\mu}) < 0$, only one root for $Q_2(\mu) = 0$ can occur.

Finally, we check the best response condition for $\tau = 1$, which reduces to $\gamma_m \geq -(r + n_p\theta_{mp})u$, or

$$\mu \leq \frac{\gamma_m + ru}{\gamma_m + ru - n_c\theta_{mp}u}.$$

Substituting out μ using μ^- and solving for γ_m surprisingly yields $Q_1(\gamma_m) = a_1\gamma_m^2 + b_1\gamma_m + c_1 \leq 0$, as in the class 2^R equilibrium. So the set of γ_m consistent with $Q_1(\gamma_m) \leq 0$ is $\gamma_m \geq \widehat{k}(\gamma_p)$. Hence, a class $2T$ equilibrium exists iff $\widehat{k}(\gamma_p) \leq \gamma_m \leq f(\gamma_p)$. ■

Proposition 5: 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

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

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. ■