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

May 11, 2016

Abstract

We study agents' decisions to be producers or middlemen in Rubinstein and Wolinsky's search model of intermediation, extended to allow general bargaining, cost and utility. This requires a different approach, but the analysis remains tractable, delivering clean and sometimes surprising results. We characterize equilibrium, show intermediation can be essential, and prove equilibrium is efficient iff bargaining satisfies versions of Hosios' conditions generalized to three-sided markets. We also go beyond the usual linear (transferable) utility specification to capture payment frictions, and beyond the usual steady state analysis by establishing saddle path stability. Applications, including one related to banking and liquidity, are discussed.

JEL Codes: G24, D83
Keywords: Middlemen, Intermediation, Search, Bargaining

*We thank Chao Gu and Alberto Trejos for input. We also thank the Chicago Fed and for providing an excellent research environment. Wright acknowledges support from the Ray Zemon Chair in Liquid Assets at the Wisconsin School of Business, and the hospitality of the the University of Chicago Economics Department during the Fall Quarter of 2015. The usual disclaimers apply.

1 Introduction

The various roles of middlemen, or intermediaries, have been studied by a number of authors (see fn. 3 below). However, given the importance of middlemen in real-world economic activity, from wholesale trade in producer or consumer goods, to financial intermediation, there would seem to be room for additional work. This project revisits the classic search-based framework introduced by Rubinstein and Wolinsky (1987), hereafter RW, extends it on a number of dimensions, and uses it to analyze a variety of issues, including the occupational choice of agents acting as producers or intermediaries.

In fact, our starting point is the version of RW in Nosal et al. (2015) that extends their specification to incorporate general bargaining powers and costs of production or storage. Still, the assumptions in those papers are quite special: utility is linear; the market is always in steady state; and the populations of producers and middlemen are fixed, although agents may or may not participate in the market. In the original formulation, all producers participate, middlemen participate iff they have a better search technology than producers, and equilibrium is necessarily efficient. In the generalized version, whether they participate depends on technologies, bargaining powers and other factors, and this may or may not be efficient. A novelty here is that an individual chooses to be *either* a producer or a middlemen, allowing us to study how this decision depends on the environment, to ask whether this is socially optimal, and to address additional substantive issues.[1]

Occupational choice complicates the analysis. Earlier studies of RW use a "trick" by taking $\boldsymbol{\alpha} = \{\alpha_{ij}\}$ as fixed, where α_{ij} is the rate at which type i agents meet type j agents. This is legitimate because there always exists a distribution

[1] Since it is easier to describe these substantive issues after laying out the model, we defer discussion for now (see fn. 14), but one feature to emphasize up front is that our economy can only have more middlemen by having fewer producers, capturing a very real tradeoff.

of types, say $\mathbf{n} = \{n_i\}$, consistent with $\boldsymbol{\alpha}$, random matching, and the identities implied by bilateral meetings, $n_i \alpha_{ij} = n_j \alpha_{ji}$. Hence, those papers 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 $\mathbf{n}$ can affect $\boldsymbol{\alpha}$. Therefore we determine endogenously $\alpha_{ij} = \alpha_i n_j$, where α_i is a baseline arrival rate for type i. However, then the relevant identities imply $\alpha_i = \alpha$ is the same for all i, and in particular $\alpha_{pc} = \alpha_{mc} = \alpha n_c$, so we must abandon the original RW idea that middlemen have a role iff $\alpha_{mc} > \alpha_{pc}$. Fortunately other factors here can take over for $\boldsymbol{\alpha}$, including costs and bargaining powers.

Despite these complications, the framework is tractable, delivering clean and sometimes surprising predictions – e.g., increasing the cost of intermediation can lead to more intermediaries. 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. In general, equilibrium can have too few or too many middlemen, and is efficient iff bargaining powers are just right.[2] We also go beyond the usual linear (transferable) utility specification by allowing strict concavity, which is relevant because, as discussed below, nonlinearity in payments interacts with intermediation. And we go beyond the usual steady state analysis by establishing saddle path stability of dynamic equilibrium. Finally, by reinterpreting some parameters, we introduce new applications, including financial intermediation, which has some novel implications.

Based on these results, we suggest the model is an advance over previous specifications for economists interested in intermediation, or in search theory more generally. The rest of the paper involves making the assumptions precise and verifying the results. Section 2 describes a benchmark environment. Sections

[2] This is related to standard results going back to Mortensen (1982) and Hosios (1990), but is also slightly different, because RW-style environments concern three-sided markets, with producers, consumers and middlemen.

3 and 4 study equilibrium and efficiency in the linear economy. Sections 5 and 6 consider nonlinear utility and dynamics. Section 7 introduces new applications. Section 8 concludes. The Appendix contains technical results.[3]

2 Environment

There is a continuum of infinitely-lived agents. Some are called consumers, and labeled C, with measure n_c. The rest choose to be producers, middlemen or nonparticipants, labeled P, M or N, with measure n_p, n_m or n_n, where $n_c + n_p + n_m + n_n = 1$. Let $\mathbf{n} = (n_c, n_p, n_m, n_n)$. Type P agents produce whenever they can and type M agents trade whenever they can (this is how we define P and M; those who do not want to act this way are type N). Type C agents trade and consume whenever they can. Agents meet bilaterally in continuous time according to a uniform random-matching process, with αn_j the Poisson rate at which anyone meets type j. Without loss of generality, normalize $\alpha = 1$.

There are two goods, x and y. Good x is indivisible, and is valued for consumption only by type C. They get utility u from consuming a unit of x. This good is storable, but only 1 unit at a time, at cost γ_p for type P and cost γ_m for type M. It is produced by type P at cost k, but for most purposes we can normalize $k = 0$ without loss of generality – as in Pissarides (2000), what matters is the total expected discounted cost, including entry, production and search, so we do not need them all. While M cannot produce x he can acquire it from P

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

to retrade it to C. 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 the original RW model, $U(y) = y$ means that transferable utility is used to pay for x, so there are no frictions in the payment process.

Type P agents always have 1 unit of x and type C agents always have 0, since the former produce and the latter consume right after trade, while type M agents can have 0 or 1 unit of x in inventory. Given that M accepts x from P, let μ denote the fraction of M with x. Then μ increases at rate $n_p n_m(1-\mu)$ (the measure of P meeting M without x) and decreases at rate $n_c n_m \mu$ (the measure of C meeting M with x). The steady state is therefore given by

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

We focus for now on steady states, and consider dynamics in Section 6.

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 $\theta_{ji} = 1 - \theta_{ij}$. As in previous analyses of RW, with transferable utility, this follows from various solution concepts, including Nash, Kalai and various strategic bargaining games (see Wright and Wong 2014 for more discussion). 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.[4] Similar expressions hold for the other surpluses, and allow us to eliminate $\mathbf{y} = (y_{cp}, y_{mp}, y_{cm})$ from the payoffs.

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, we get the dynamic

[4] 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.

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

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

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

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

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

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

3 Equilibrium

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 intermediation. The labels are chosen because class 0 implies no active

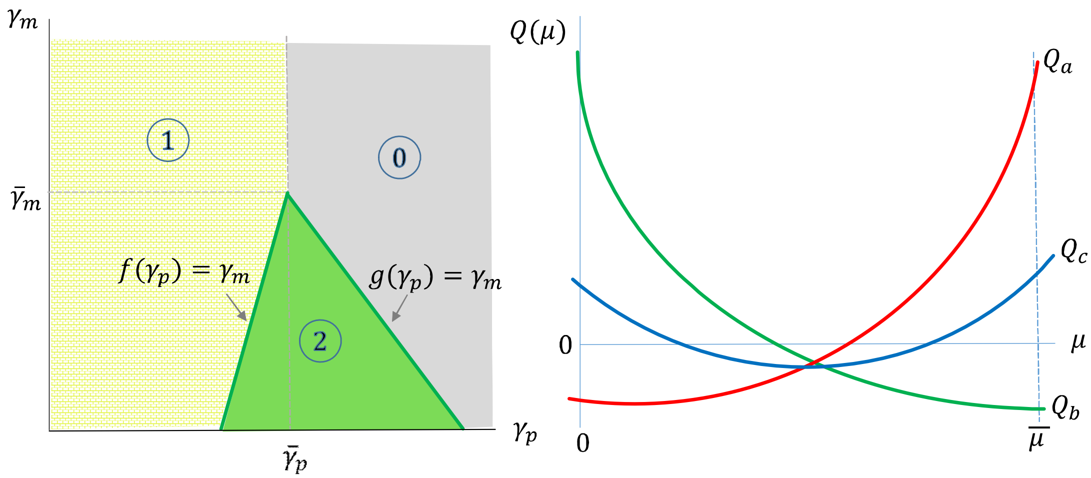

Figure 1: Equilibrium outcomes in (γ_m, γ_p) space

agents, class 1 has type P active but not type M, and class 2 has type P and M both active.[5]

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, $\gamma_p \geq \bar{\gamma}_p$ may violate subgame perfection, and in those cases we ignore those equilibria (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

[5] In principle there could be equilibria where $n_n > 0$, $n_p > 0$ and $n_m > 0$, but it is easy to check that this is possible only for a measure 0 set of parameters.

$\gamma_p \leq \bar{\gamma}_p$, 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.*

Now consider class 2 equilibrium, with $n_p, n_m > 0$ and $n_p + n_m = 1 - n_c$, where $V_p = V_0 \geq 0$. It is convenient to characterize the outcome in terms of μ, then use steady state conditions to recover $\mathbf{n}$. Clearly we need $\mu \in (0, \bar{\mu})$, where $\bar{\mu} = 1 - n_c$. Then routine algebra reduces $V_p = V_0$ to $Q(\mu) = 0$, where $Q(\mu)$ is obtained by replacing n_p and n_m with their values in terms of μ. The result is

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

a quadratic with coefficients[6]

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

We seek $\mu \in (0, \bar{\mu})$ such that $Q(\mu) = 0$ and $V_0 \geq 0$. Since $V_0 \geq 0$ iff $\gamma_m \leq \bar{\gamma}_m$, we restrict attention to $\kappa_1 > 0$, so $Q(\mu)$ is convex. Thus, as shown by the

[6] Much of the analysis in the Appendix deals with quadratic equations, and in one case, in the proof of Lemma 14, a cubic. This is unavoidable, and natural, given random matching and the inventory condition (1). In particular, the rate at which P can trade with M is $\alpha_{pm}(1-\mu) = n_m n_c/(1-n_m)$, which renders several equilibrium conditions quadratic. Of course it would be easier if $\mathbf{n}$ were fixed, as in previous work, but one of our main objectives is to make it endogenous. Also note that payoffs depend on $\mathbf{n}$ even though our matching technology has constant returns to scale: one meets potential counterparties at a constant rate but the outcome depends on whom one meets, and, for P or C, depends on μ.

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 assuming $f(0) < 0$, although $f(0) > 0$ obtains for other parameters.[7] Note that equilibrium is unique generically (i.e., except on a boundary between two regions). In any case, while it is no surprise that n_p and n_m depend on (γ_p, γ_m), the preceding analysis yields exact cutoffs for the different outcomes. 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. We summarize the above results as follows:

Proposition 1 *Equilibrium exists and is generically unique, as shown in Fig. 1.*

Intermediation can be *essential* in the technical sense used by monetary theorists: an institution like money is said to be essential if the set of outcomes

[7] This is relevant because it determines whether we get class 1 or 2 equilibrium at the origin. In particular, we can have $n_m > 0$ even when $\gamma_m > \gamma_p$, or $n_m = 0$ even when $\gamma_m < \gamma_p$, indicating there are forces at work other than just storage costs.

that can be supported as equilibria expands when money is introduced. Surveys by Nosal and Rocheteau (2011) and Lagos et al. (2015) discuss work on the essentiality of money, banking and related institutions. For both money and intermediation, the notion is nontrivial because, e.g., they are clearly *not* essential in the standard environment used in general equilibrium theory. In our environment, in the region where class 2 equilibrium exists with $\gamma_p > \bar{\gamma}_p$, economic activity depends on middlemen being active: if we were to exogenously eliminate type M, say by taxing them, the market would shut down. Thus, intermediation may be necessary for production and consumption to be viable. Even if they are viable without intermediation, welfare may be enhanced by having some type M agents, but it may also be diminished, as discussed in Section 4.

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

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

Based on these observations, it is immediate that

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

This accords well with intuition: when γ_p is higher, we get fewer producers. However, it is also immediate that

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

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

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

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

The reason that an increase in θ_{mc} works much 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.[8]

We now bring back the terms of trade, $\mathbf{y}$. In direct exchange, where C gets x from P, $y_{cp} = \theta_{pc} u$ is independent of the sunk storage costs, and increasing with P's bargaining power and C's valuation. In wholesale trade, where M gets x from P,

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

The endogenous n_p is left on the RHS to illustrate a point: there is a direct impact on y_{mp} from γ_m, but not from γ_p; plus there are indirect effects from both through $\mathbf{n}$. Similarly, in retail trade, where C gets x from M,

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

One can check $\partial y_{mp}/\partial \gamma_p > 0$, $\partial y_{cm}/\partial \gamma_p < 0$ and $\partial s/\partial \gamma_p < 0$, where $s = y_{cm} - y_{mp}$ is the spread. Less straightforwardly, $\gamma_p > \bar{\gamma}_p$ implies $\partial y_{mp}/\partial \gamma_m < 0$,

[8] For completeness we mention how the other parameters affect $\mathbf{n}$. The effect of r, like γ_m, depends on γ_p: $\gamma_p < \bar{\gamma}_p$ implies $\partial \mu/\partial r > 0$, $\partial n_p/\partial r > 0$ and $\partial n_m/\partial r < 0$, while $\gamma_p > \bar{\gamma}_p$ implies $\partial \mu/\partial r < 0$, $\partial n_p/\partial r < 0$ and $\partial n_m/\partial r > 0$. A demand increase on the intensive margin, captured by higher u, is less clear: Since what matters is γ_p/u and γ_m/u, raising u has the same impact as lowering both γ_p and γ_m. If $\gamma_p > \bar{\gamma}_p$ then higher u raises n_p and lowers n_m; if $\gamma_p < \bar{\gamma}_p$ then the effect can go either way. Similarly ambiguous is an increase in demand on the extensive margin, captured by higher n_c.

$\partial y_{cm}/\partial\gamma_m > 0$ and $\partial s/\partial\gamma_m > 0$, but $\gamma_p < \bar{\gamma}_p$ implies the effects are ambiguous. Some changes in bargaining powers are ambiguous, too, while others are not. In any case, the results would be different if $\mathbf{n}$ were exogenous, as then the indirect effects vanish. This is one reason to study occupational choice. Another is to examine the welfare implications.

4 Efficiency

We now solve a planner's problem, where for simplicity the focus is on $r \approx 0$. The problem is to choose (μ^o, n_p^o, n_m^o) to maximize:[9]

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

Consider first $\gamma_m > n_c u$, which means intermediation is not viable, because it contributes negatively to rW. Given $\gamma_m > n_c u$ we have: $\gamma_p > n_c u$ implies $n_p^o = 0$ and the market shuts down; and $\gamma_p < n_c u$ implies $n_p^o = 1 - n_c$ and the market opens with direct trade only.

Next consider $\gamma_m < n_c u$, which means intermediation is viable but may or may not be optimal. Eliminating n_p and n_m we reduce the planner's problem to

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

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 and decreasing in μ over the relevant range.

[9]The first (second) term is the net social surplus from direct (indirect) trade. Similar to the related analysis in Nosal et al. (2015), one can solve the dynamic problem with $r > 0$ and, as usual, the outcome is the same as maximizing rW when $r \to 0$. Also, as is standard $V_j \to \infty$ when $r \to 0$, but rV_j and rW are well defined. Of course, small r can be interpreted as saying search frictions are not overly severe.

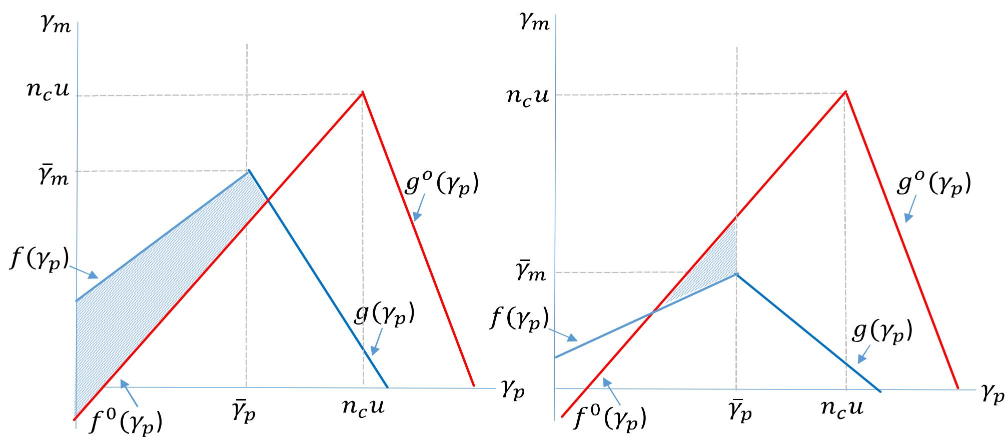

Figure 2: Comparing equilibrium and efficient outcomes

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

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

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

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

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

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

Fig. 2 shows for comparison f and g as well as f^o and g^o. Notice f^o and g^o intersect at $(n_c u, n_c u)$, and the region where $n_m^o > 0$ lies strictly below the 45^o line. Hence, for intermediation to be optimal, we need $\gamma_m < \gamma_p$. This is different from equilibrium, where $\gamma_m < \gamma_p$ is neither necessary nor sufficient for $n_m > 0$. It is also different from models with fixed $\mathbf{n}$. In such models, if γ_m is close to γ_p it is always a good idea 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 on his own. This is summarized as follows:

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

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

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

which is similar to the equilibrium result, and intuitively clear. Next,

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

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

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

Here is the economic explanation: If γ_m increases, the natural response is to reduce inventories held by M, given by $n_m \mu$, but there are different ways to do so. One is to reduce n_m, which in steady state means higher μ; the other is to reduce μ, which means higher n_m. When $\gamma_p < n_c u$ it is optimal to use the extensive margin and reduce n_m; when $\gamma_p > n_c u$ it is optimal to use the intensive margin and reduce μ, which means higher n_m. This explains the planner's choices. The idea is similar for equilibrium, but less transparent, as complications can make that different from the efficient outcome, as we now discuss.

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

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

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

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

5 Concavity

Now suppose $U''(y) < 0$, while continuing to assume $U(0) = 0$. It is interesting to go beyond linear (transferrable) utility, for various reasons, but here is a big one. Let $\hat{y} > 0$ solve $U(\hat{y}) = \hat{y}$. If an equilibrium payment is $y > \hat{y}$ then the transfer is such that the cost to the payer exceeds the value to the payee. Hence, $y > \hat{y}$ discourages, and symmetrically $y < \hat{y}$ encourages, intermediation. This is because indirect trade entails two payments, M to P and C to M, rather than one, C to P. The nonlinear specification can be interpreted as transaction costs in the settlement process. Note $y > \hat{y}$ is possible, since even with $U(y) < y$ the surplus can be positive due to the gains from trading x. Indeed, we do not impose $U'(0) > 1$, so it may be that $U(y) < y \ \forall y > 0$.

For tractability, with $U'' < 0$, we use Kalai's (1977) bargaining solution: when i trades with j, maximize i's surplus subject to i getting a share θ_{ij} of the total surplus.[10] Then we have

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

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

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

The solution method mimics that used above, although the algebra is more involved, which is why we use $U(y) = y$ as a benchmark model. The analog to

[10] This is not the definition of Kalai bargaining, it is a result about the outcome implied by his axioms, like maximizing the Nash product is a result about the outcome implied by his axioms. If $U(y) = y$, Nash and Kalai are the here; with $U'' < 0$, while we could use Nash, Kalai has some advantages (as Aruoba et al. 2007 argue in the context of a related model).

Lemma 1 (with the proof left as an exercise) 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 the occupational choice condition $V_p = V_0$, and y_{mp} is the bargaining solution. Here $V_p = V_0$ 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* $\tilde{Q}\left(0, y_0\right) > 0 > \tilde{Q}\left(\bar{\mu}, \bar{y}\right)$*.*

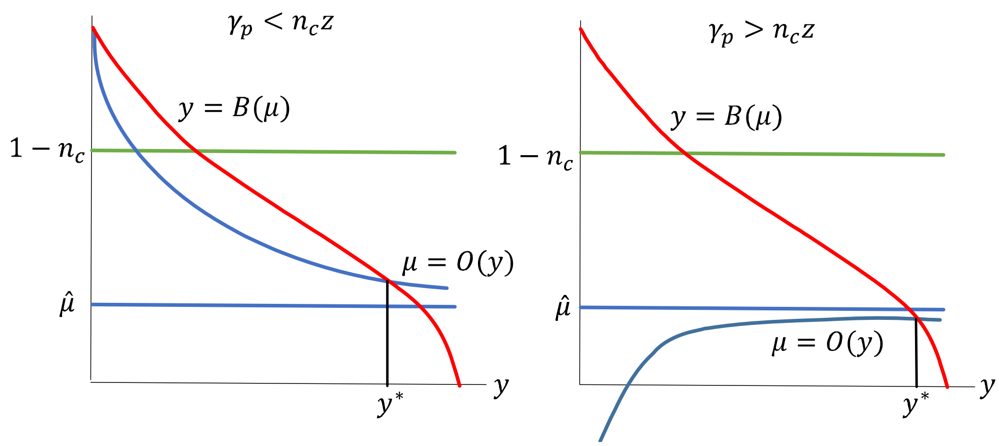

Figure 3: Equilibrium in (y_{mp}, μ) space

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

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

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

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

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

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

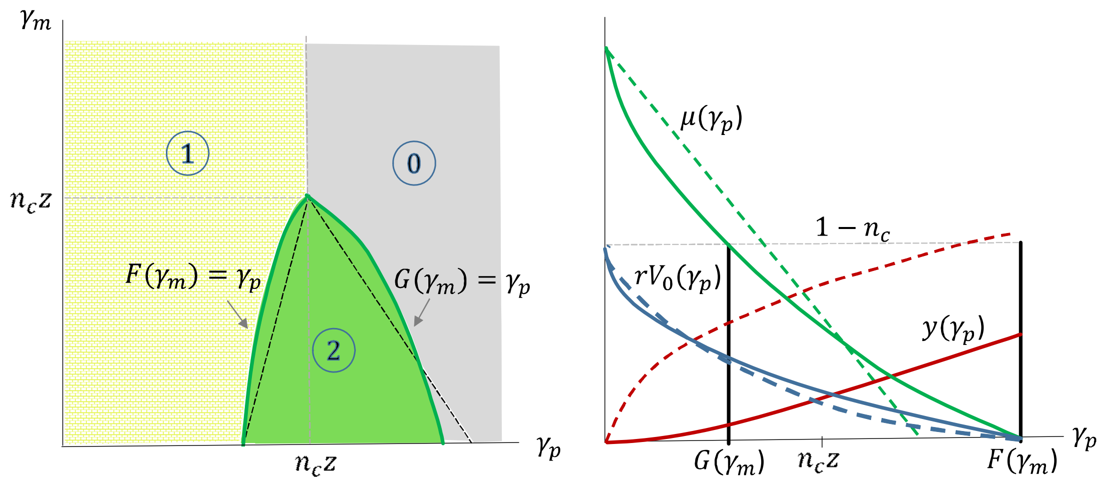

Figure 4: The nonlinear (solid) and linear (dashed) models

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

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

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

In the left panel of Fig. 4, the solid curves are F and G for $U(y) = y^{\zeta}$ with $\zeta = 0.3$, $\theta_{pm} = 1/2$, $n_c = 1/2$, $z = 2.2$ and $r = 0.01$. For comparison the dashed lines are for $\zeta = 1$. Notice the set in (γ_p, γ_m) space where $n_m > 0$ does not necessarily expand or contract with increased curvature in $U(\cdot)$, as evidenced

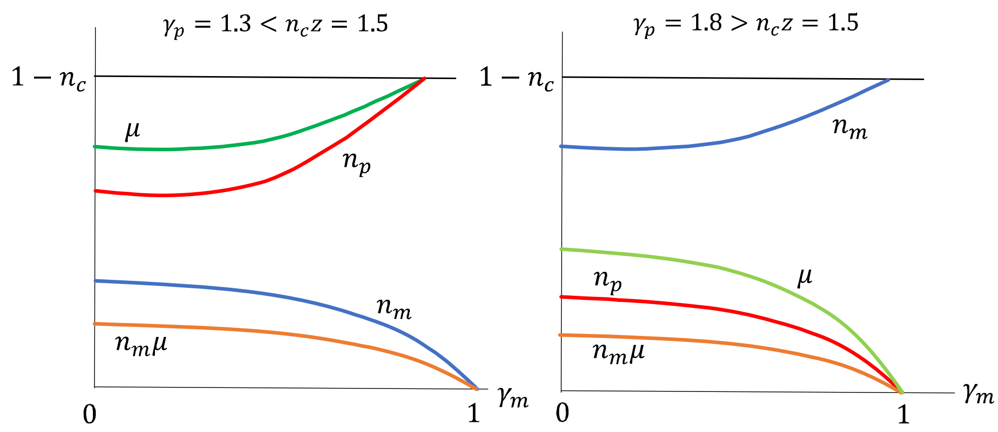

Figure 5: Effects of γ_m in the nonlinear model

by the dashed and solid curves crossing; this is because nonlinearity tends to discourage intermediation when $y_{mp} > \hat{y}$ and encourage it when $y_{mp} < \hat{y}$. In the right panel of Fig. 4, the solid curves are y_{mp}, μ and rV_0 as functions of γ_p for the nonlinear model, while the dashed curves are for the linear model. The impact of γ_m is shown in Fig. 5, where higher γ_m implies lower (higher) n_m in the left (right) panel. These are general results, as was the case in the linear specification, as 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 transitions in class 2 equilibrium given an initial μ.[11] As above we set $\theta_{pc} = \theta_{mc} = 1$, so that $y_{cp} = y_{cm} = u$, $V_c = 0$ and $z = U(u)$, and let $\Delta = V_1 - V_0$. Also, at any time we assume a type P can dispose of x and

[11] Class 0 and 1 equilibria have no interesting dynamics. However, the economy can potentially start with, e.g., $n_m = 0$, then transit to $n_m > 0$, or vice versa.

become type M, but agents cannot start as type M with their own output, say, because they must spend x to acquire the middleman technology. Here we work with the n's, rather than μ, since n_1 is a state variable, with law of motion

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

In contrast, n_0 can jump at any time to satisfy occupational choice, $V_0 = V_p$, just like vacancies jump in the well-known labor-market model of Pissarides (2000). The bargaining solution for y_{mp} is

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

and 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 reduce this dynamic system to something manageable. First notice that $V_p = V_0$ implies $\dot{V}_p = \dot{V}_0$, and then from (22)-(23) the occupational choice condition becomes

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

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

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

Substituting (25) and simplifying, we arrive at

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

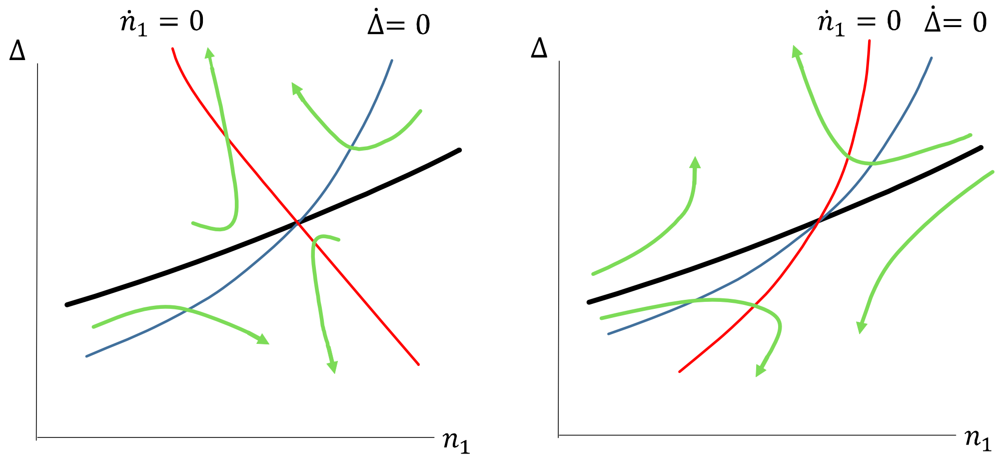

Figure 6: Saddle path stability

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

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

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

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

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

Therefore, given an initial condition for $\bar{n}_1$, there is a unique initial $\bar{\Delta}$ such that (n_1, Δ) transits to the steady state, and any $\Delta \neq \bar{\Delta}$ implies an explosive path that cannot be an equilibrium. So equilibrium, not only steady state, is unique – which was not a foregone conclusion.[12]

7 Negative γ's

The theory applies to many types of middlemen with comparative advantage in storage or bargaining. But storing inventories is not always costly. Suppose P is producing and M is dealing in fine art. Then the net benefit of holding x can be $\rho_j = -\gamma_j > 0$, given art generates positive utility. If $\rho_m > \rho_p$, e.g., an art dealer, perhaps by charging admission to his gallery, gets more from the piece than the artist. If an art consumer/collector C enjoys it even more, M may retrade it, or he may prefer to keep it – an option not relevant in the baseline model with $\gamma_j > 0$ (note that P never prefers to keep x, regardless of γ_p, since as soon as he sells it he can produce another). Given the γ's and θ's, we can again analyze when the market is open, $n_p > 0$, when there are dealers, $n_m > 0$, and how do they behave – i.e., do they keep x or retrade it?

Another application concerns financial intermediation. When P has access to capital available for investment in projects by type C, we can ask the same questions. In this case, $\rho_j = -\gamma_j \geq 0$ can represent the flow return to an agent putting capital into a temporary investment, presumably a liquid one, so that it is readily available to pass on to C if so desired. In this case $\rho_m > \rho_p$ means that M has better liquid investment opportunities than P, and ρ_m might even be high enough for M to keep x rather than pass it to C. Again, we also have to consider M's negotiating power. If M does pass x to C, he is a financial intermediary. As another example, x might be a house that provides a flow payoff as shelter. Then M might want to acquire x as a residence, or to retrade it to C, as in real

[12] While there is a unique equilibrium in the linear version of the related model of Duffie et al. (2005), e.g., there is multplicity in the nonlinear version (Trejos and Wright 2015). Hence, saddle path stability with $U'' < 0$ is not a triviality.

estate flipping. All of these applications make it interesting to consider $\gamma_j < 0$. Moreover, in terms of theory, $\gamma_j < 0$ generates some novel results.

The dynamic programming equations are the same, but we need a new endogenous variable τ, for the probability that M trades x to C. Also, V_1 is still M's payoff to holding x with the intention of trading it to C, but he actually prefers to keep it if $rV_1 < -\gamma_j$. It is now possible in principle to have $n_m = 1 - n_c$ and $n_p = 0$, but if so, then M must hold on to x (if he trades it he never gets x again since $n_p = 0$); in this case the no-deviation condition is $-\gamma_j \geq rV_p$ since the relevant deviation is to become type P. Hence we have this:

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

The other change is that the possibility of $\tau < 1$ makes the steady state condition

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

We study steady state equilibria in terms of τ and n_m, with $U(y) = y$ to ease the presentation. There are 9 candidates, shown in Table 1, none of which correspond to a class 0 outcome because production always dominates nonparticipation with $\gamma_j < 0$. If $n_m = 0$, in the first row of Table 1, $\tau = 1$ corresponds to a class 1 outcome in the baseline model, where there are no type M agents on the equilibrium path, but if there were, off the equilibrium path, they set $\tau = 1$. We call this a class 1^T equilibrium (T indicates M trades x). Similarly, if $n_m = 0$ we call $\tau = 0$ a class 1^K equilibrium, because if there were a type M with x he would not trade it to C (K indicates M keeps x). And if $n_m = 0$ we call $\tau \in (0, 1)$ a class 1^R equilibrium (R indicates M randomizes), but in fact this can be ruled out: M only chooses $\tau \in (0, 1)$ if he is indifferent, which might happen if $n_m \in (0, 1 - n_c)$ is set endogenously, but generically not if n_m is 0 or $1 - n_c$.

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

One can also rule out $n_m = 1 - n_c$ and either $\tau \in (0,1]$: if $n_m = 1 - n_c$ there are no producers, so trading away x leaves M with a continuation value 0, which implies a profitable deviation because he can become type P. We cannot rule out $n_m = 1 - n_c$ and $\tau = 0$, but we ignore it in what follows because it is a degenerate outcome with no production.[13] The remaining candidates are $n_m \in [0, 1-n_c]$ and $\tau = 1$, $\tau = 0$ or $\tau \in (0,1)$, called class 2, 2^K or 2^R (there are both type P and M, and M either trades x, keeps it or randomizes). The following is proved in the Appendix and illustrated in Fig. 7.

Lemma 14 *The Appendix defines $h(\cdot)$, $k(\cdot)$, $v(\cdot)$, $\widetilde{\gamma}_p$and γ_m^*. Class 1^T equilibrium exists iff $\gamma_m \geq h(0), f(\gamma_p)$. Class 1^K exists iff $f(\gamma_p) \leq \gamma_m \leq h(0)$. Class 2^K exists iff $\gamma_m \leq k(\gamma_p), f(\gamma_p)$. Class 2^R exists iff $v(\gamma_p) < \gamma_m < h(1-n_c)$. Class 2^T exists iff $\gamma_m^* \leq \gamma_m \leq f(\gamma_p)$.*

The left panel of Fig. 7, with $\widetilde{\gamma}_p > 0$, is simple: $\gamma_m > f\left(\gamma_p\right) \Rightarrow n_m = 0$; and $\gamma_m < f\left(\gamma_p\right) \Rightarrow n_p, n_m > 0$, but type M agents keep x. In the right panel, with $\widetilde{\gamma}_p < 0$, those outcomes are still possible, but so are class 2^T and 2^R, with active intermediation. Clearly we lose uniqueness. Is there something about financial intermediation that contributes to this? Yes. Heuristically, the multiplicity is due to a strategic complementarity. When M keeps x with probability $1-\tau > 0$, the

[13] This equilibrium, with type M agents simply sitting on x, can be shown to exist iff $\gamma_m \leq k(\gamma_p)$, where $k(\cdot)$ is defined below. When it exists, there coexists another equilibrium, so we do not need it to establish existence. However, we ignore it mainly to avoid cluttering the graphs; we do not claim it should be ignored based on stability considerations, even if one might ask, how can all the M's be holding x when there are no P's to produce it? The answer is that $n_p > 0$ along the transition path, with $n_p \to 0$ only as $t \to \infty$.

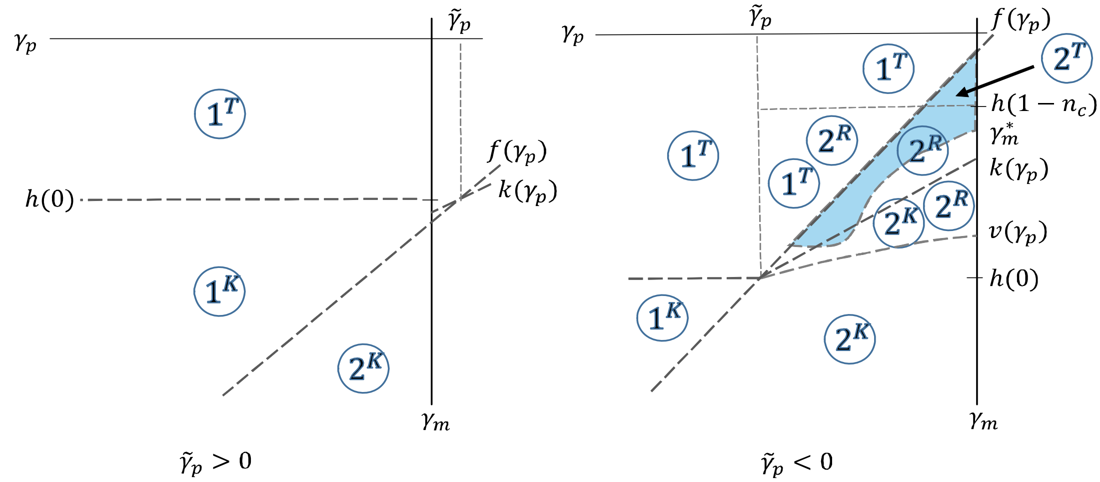

Figure 7: Equilibrium outcomes in negative (γ_p, γ_m) space

payoff to P can go down rather than up when n_m rises. Hence, having more M sitting on x can encourage even more to do so. Thus the market can be *illiquid*, in the sense that M will not give x to C, precisely because M expects others are doing so. That cannot happen with $\gamma_j > 0$. While multiplicity arising from liquidity considerations is well known in general (see the survey by Nosal and Rocheteau 2011 or Lagos et al. 2015), this version seems new.

It is also worth mentioning that this result is not especially related to occupational choice – the complementarity applies to τ, not n_m – we just happened to discover it in a model with endogenous n_m. Again, if more type M agents stop trading x to C, others are inclined to follow suit. This is reminiscent of monetary economics, where it can be the case that you should not accept an asset if others do not accept it. Here the idea is that you should hoard rather than trade an asset if others are hoarding it, somewhat related to Gresham's law (again see the above-mentioned surveys). We formalize the main results as follows.

Proposition 6 *With $\gamma_j < 0$, equilibrium exists but is not generically unique, as shown in Fig. 7.*

8 Conclusion

This project has continued the development of search-and-bargaining theories of intermediation. We built on the classic model of Rubinstein and Wolinsky (1987), extended to allow general bargaining powers and costs, but rather than fixing the numbers of producers and middlemen we let agents choose their types. This is natural for investigating many issues.[14] The theory delivered clean and sometimes surprising results – e.g., $\partial n_m/\partial \gamma_m > 0$ is possible, for reasons explained above. We established existence and generic uniqueness for the baseline model, although with $\gamma_j < 0$ an interesting multiplicity can emerge. We discussed how middlemen can be essential, and showed equilibrium is efficient iff bargaining powers are just right; otherwise there can be too much or too little intermediation. Extensions including strictly concave utility and dynamics were presented.

Many other extensions and applications should be possible. Clearly one would like to go beyond unit inventories, just like it was desirable to move beyond unit inventories in monetary models like Kiyotaki and Wright (1993). This has been accomplished in search-based theories of money and finance by several authors using a variety of techniques (again see Nosal and Rocheteau 2011 or Lagos et al. 2015). Something similar could work for middlemen, too, if one were willing to adopt similar assumptions. This is left for future work. Based on the results developed here, we think the framework should become a benchmark model in intermediation theory, and in search theory more generally.

[14] One issue is that the only way to get more intermediaries here is to have fewer producers, capturing a very real economic trade-off (e.g., more MBA's means fewer engineers). Also, our setup eliminates some effects in earlier models that are artifacts of simplifying assumptions, one of which concerns the restriction of M's inventory be 0 or 1. In other models, when M takes P's good, the latter can produce again, leading to more output. That is not relevant here, because if M does not take P's good, he can become a producer and make his own x. Hence, intermediation is useful not merely because it gets around the unit-inventory restriction. Other features of the model also allow one to consider additional issues, including $U'' < 0$, which captures the idea that payments are not necessary perfect (linear), and this naturally affects the incentives to engage in intermediation. Dynamics are also interesting, with n_m, n_p and μ varying along the saddle path over time, somewhat similar to Weill (2007).

References

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

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

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

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

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

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

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

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

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

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

[11] R. Lagos, G. Rocheteau and R. Wright (2015) "Liquidity: A New Monetarist Perspective," *JEL*, in press.

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

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

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

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

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

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

[18] A. Rubinstein and A. Wolinsky (1987) “Middlemen,” *QJE* 102, 581-594.

[19] A. Shevchenko (2004)“Middlemen,” *IER* 45, 1-24.

[20] E. Smith (2004) “Intermediated Search,” *Economica* 71, 619-636.

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

[22] C. Tse (2009) “The Spatial Origin of Commerce,” mimeo.

[23] M. Watanabe (2010) “A Model of Merchants,” *JET* 145, 1865-1889.

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

[25] R. Wright (1995) “Search, Evolution and Money,” *JEDC* 19, 181-206.

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

Appendix

Here we provide proofs for results that are not obvious.

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

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

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

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

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

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

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

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 the left panel of Fig. 8, 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

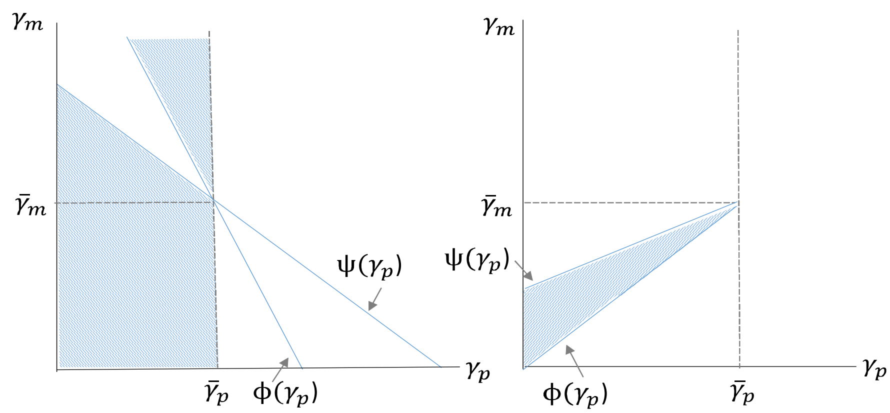

Figure 8: The functions $\phi(\gamma_p)$ and $\psi(\gamma_p)$

(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. 8

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. 9.

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

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

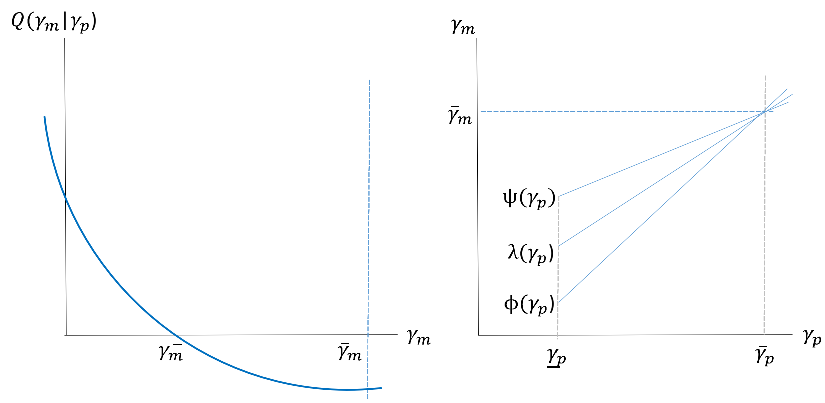

Figure 9: The functions $\hat{Q}(\gamma_m|\gamma_p)$ and $\lambda(\gamma_p)$

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. 9 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. 9.

Hence, (c) requires $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \varnothing$ and $\lambda'(\gamma_p) < \phi'(\gamma_p)$. But the latter inequality can be simplified to

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

when we ignore terms with r, which only strengthens the inequality. This inequality 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,

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

where $\mu = \tilde{\mu}$ is inserted after the derivatives are taken. This implies $\partial\mu/\partial\gamma_m > 0$ 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

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

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

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

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

If $\theta^o_{pm} = 1$ then $g(\gamma_p) = g^0(\gamma_p)$; so for $(\gamma_p, \gamma_m) \in S^o_0$, $n_j = n^o_j = 0$. If $\theta^o_{pm} = 0$ then $f(\gamma_p) = f^0(\gamma_p)$; so for $(\gamma_p, \gamma_m) \in S^o_1$, again $n_j = n^o_j$. If $\theta^o_{pm} \in (0,1)$ then $\gamma_m \le f(\gamma_p)$ implies $\gamma_m \le f^o(\gamma_p)$ and $\gamma_m \le g(\gamma_p)$ implies $\gamma_m \le g^o(\gamma_p)$. 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 ways for $\tilde{Q}(\mu, y_{mp}) = 0$ to have a solution in $(0, \bar{\mu})$: (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, which requires (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 $\partial\tilde{Q}(\mu^*, y_{mp})/\partial\mu = 0$. Notice

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

As in Lemma 3, it is easy to check that case (a) is not possible.

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

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

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

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

We need $\tilde{Q}(\mu^*, y_{mp}) < 0$. Although it is an abuse of notation, let $\tilde{Q}(\mu^*, y_{mp}) \equiv \tilde{Q}(\gamma_p) < 0$ where

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

For (c5) we seek the set of γ_p such that $\tilde{Q}(\gamma_p) < 0$. There are three possibilities: (c5.1) one root with $\tilde{Q}(K) < 0 < \tilde{Q}(F)$; (c5.2) one root with $\tilde{Q}(K) > 0 > \tilde{Q}(F)$; (c5.3) two roots, which requires $\tilde{Q}(K) < 0 < \tilde{Q}(F)$, $\tilde{Q}'(K) > 0 > \tilde{Q}'(F)$, and $\tilde{Q}(\gamma_p^*) < 0$, where $\tilde{Q}'(\gamma_p^*) = 0$. Given $\gamma_p = K$ and $n_c z - \gamma_p = U(\bar{y})n_c/\theta_{pm}$,

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

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

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

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

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

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

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

To establish the properties of F and G, derive

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

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

Lemma 12: In class 2 equilibrium we have

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

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

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

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

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

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

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

where

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

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

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

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

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

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

Lemma 14: We consider each class of equilibrium from Table 1 in turn. Consider first a class 1^K equilibrium, where $n_m = \tau = 0$ and

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

Now $rV_1 \le -\gamma_m$ iff $\gamma_m \le h(0)$, where $h(n_m) \equiv -u[r + (1-n_c-n_m)\theta_{mp}]$ and $V_p \ge V_0$ iff $\gamma_m \ge f(\gamma_p)$. Hence, the equilibrium exists under the stated conditions. Similarly, class 1^T equilibrium exists iff $\gamma_m \ge h(0)$ and $\gamma_m \ge f(\gamma_p)$.

Now consider $\tau = 0$ and $n_m \in (0, 1-n_c)$, a class 2^K equilibrium, which requires $rV_1 \le -\gamma_m$ and $V_p = V_0$. The latter solves for

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

One can check $n_m < 1-n_c$. Also, $n_m > 0$ requires either: (i) the denominator and numerator are both positive, which is true iff $\gamma_m < f(\gamma_p)$; or (ii) they are both negative, which is true iff $\gamma_m > \ell(\gamma_p) \equiv \overline{\gamma}_m - \overline{\gamma}_p + \gamma_p$. We also need $rV_1 \le -\gamma_m$, which is true iff

$$
\underline{Q}(\gamma_m) = \gamma_m^2 + [ru - (\overline{\gamma}_m - \overline{\gamma}_p + \gamma_p)]\gamma_m - \overline{\gamma}_m(ru + \overline{\gamma}_p - \gamma_p) \ge 0.
$$

Note $\underline{Q}(\gamma_m)$ is convex, $\underline{Q}(0) < 0$ and $\underline{Q}(\gamma_m^*) < 0$ where $\gamma_m^* < 0$ solves $\underline{Q}'(\gamma_m^*) = 0$. Thus, $\underline{Q}(\gamma_m) \ge 0$ iff $\gamma_m \le k(\gamma_p)$, where where $k(\gamma_p) \equiv -\left(ru + \bar{\gamma}_p\right) + \gamma_p$ is the lower root of $\underline{Q}(\gamma_m)$. Note it is linear and lies below $\ell(\gamma_p)$; hence, $\gamma_m > \ell(\gamma_p)$ cannot occur. In sum, the equilibrium exists iff $\gamma_m \le k(\gamma_p),\ f(\gamma_p)$.

Now consider $\tau = 1$ and $n_m \in (0, 1-n_c)$, a class 2^T equilibrium. This requires $rV_1 \ge -\gamma_m$ and $V_p = V_0$. The latter reduces to $\bar{Q}(n_m) = \bar{a}n_m^2 + \bar{b}n_m + \bar{c} = 0$, where

$$
\begin{aligned}
\bar{a} &= \theta_{mp}[\gamma_m - \ell(\gamma_p)] \\
\bar{b} &= -\theta_{mp}(2-n_c)[\gamma_m - \ell(\gamma_p)] - (r + n_c\theta_{mc})(\overline{\gamma}_p - \gamma_p) + n_c\theta_{pm}(\overline{\gamma}_m - \gamma_m) \\
\bar{c} &= \theta_{mp}(1-n_c)[\gamma_m - \ell(\gamma_p)] + (r + n_c\theta_{mc})(\overline{\gamma}_p - \gamma_p).
\end{aligned}
$$

As usual, the upper root and two root case are not possible, so consider the lower root n_m. Now $n_m > 0$ exists iff $\gamma_m < f\left(\gamma_p\right)$, and $n_m < 1 - n_c$ iff $\gamma_m < \ell\left(\gamma_p\right)$. Since $\ell\left(\gamma_p\right) > f\left(\gamma_p\right)$, the binding condition is $\gamma_m < f\left(\gamma_p\right)$.

Next, to check $rV_1 \geq -\gamma_m$, substitute the root of $\bar{Q}(n_m) = 0$ into $\gamma_m \geq h(n_m)$ to get a cubic $C(\gamma_m) = \gamma_m^3 + \widetilde{a}\gamma_m^2 + \widetilde{b}\gamma_m + \widetilde{c} \geq 0$, where

$$
\begin{aligned}
\widetilde{a} &= -u(n_c - 2r) - \ell\left(\gamma_p\right) \\
\widetilde{b} &= -u^2[r + \theta_{mp}(1 - n_c)][n_c(1 - \theta_{mp}) - r + \theta_{mp}] + u^2\theta_{mp}^2(1 - n_c) \\
&\quad + u\{n\theta_{pm}\overline{\gamma}_m - (\overline{\gamma}_p - \gamma_p)(r + n_c\theta_{mc}) + \theta_{mp}(2 - n_c)\ell\left(\gamma_p\right) \\
&\quad - 2[r + \theta_{mp}(1 - n_c)]\ell\left(\gamma_p\right)\} \\
\widetilde{c} &= -u^2\theta_{mp}^2(1 - n_c)\ell\left(\gamma_p\right) + u^2\theta_{mp}(\overline{\gamma}_p - \gamma_p)(r + n_c\theta_{mc}) \\
&\quad + u^2[r + \theta_{mp}(1 - n_c)][n\theta_{pm}\overline{\gamma}_m - (\overline{\gamma}_p - \gamma_p)(r + n_c\theta_{mc}) + \ell\left(\gamma_p\right)(\theta_{mp} - r)]
\end{aligned}
$$

To solve the cubic, we employ Cardano's method (see Jacobson 2009, p. 210). First, rewrite $C(\gamma_m) = 0$ using the transformation $z = \gamma_m + \widetilde{a}/3$ to get

$$
z^3 + \frac{3\widetilde{b} - \widetilde{a}^2}{3}z + \frac{2\widetilde{a}^3 - 9\widetilde{a}\widetilde{b} + 27\widetilde{c}}{27} = 0.
$$

Cardano's method implies there is one real root given by

$$
\begin{aligned}
z^* &= \sqrt[3]{\frac{\widetilde{a}\widetilde{b}}{6} - \frac{\widetilde{c}}{2} - \frac{\widetilde{a}^3}{27} + \sqrt{\left(\frac{\widetilde{a}\widetilde{b}}{6} - \frac{\widetilde{c}}{2} - \frac{\widetilde{a}^3}{27}\right)^2 + \left(\frac{\widetilde{b}}{3} - \frac{\widetilde{a}^2}{9}\right)^3}} \\
&\quad + \sqrt[3]{\frac{\widetilde{a}\widetilde{b}}{6} - \frac{\widetilde{c}}{2} - \frac{\widetilde{a}^3}{27} - \sqrt{\left(\frac{\widetilde{a}\widetilde{b}}{6} - \frac{\widetilde{c}}{2} - \frac{\widetilde{a}^3}{27}\right)^2 + \left(\frac{\widetilde{b}}{3} - \frac{\widetilde{a}^2}{9}\right)^3}}.
\end{aligned}
$$

One can show $\gamma_m^* = z^* - \widetilde{a}/3 < 0$, $C(0) > 0$, and $C'\left(\gamma_m\right) > 0$ for $0 > \gamma_m \geq \gamma_m^*$. Therefore, given a unique root $\gamma_m^* < 0$, it is clear that $C\left(\gamma_m\right) \geq 0$ for $\gamma_m \geq \gamma_m^*$; hence $rV_1 \leq -\gamma_m$ holds, iff $\gamma_m \geq \gamma_m^*$. To summarize, this equilibrium exists iff $\gamma_m^* \leq \gamma_m \leq f(\gamma_p)$.

Finally, consider $\tau \in (0, 1)$ and $n_m \in (0, 1 - n_c)$, a class 2^R equilibrium. This equilibrium requires $rV_1 = -\gamma_m$ and $V_p = V_0$. Now $rV_1 = -\gamma_m$ solves for

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

Note $0 < n_m < 1 - n_c$ requires $h(0) < \gamma_m < h(1 - n_c)$. Then $V_p = V_0$ solves for

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

Imposing $\gamma_m < h(1 - n_c) = -ru$, $\tau > 0$ requires $\gamma_m < k(\gamma_p)$ and $\gamma_m > k(\gamma_p)$ if $\gamma_p \geq \widetilde{\gamma}_p \equiv \overline{\gamma}_p - (1 - n_c)\theta_{mp}u$. When $\widetilde{\gamma}_p > 0$, $\gamma_m > k(\gamma_p)$ cannot occur.

The condition $\tau < 1$ requires $Q(\gamma_m) < 0$, where $Q(\gamma_m) = \gamma_m^2 + b\gamma_m + c$, with

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

One can show $Q(\gamma_m^*) < 0$ where γ_m^* is the solution to $Q'(\gamma_m^*) = 0$. There are two cases. (1) if $\gamma_m^* > 0$ then one can show $Q(0) < 0$, and $Q(\gamma_m) < 0 \Rightarrow \gamma_m \geq v(\gamma_p)$ where $v(\gamma_p)$ is the lower root of $Q(\gamma_m)$. (2) if $\gamma_m^* < 0$ then (2.1) if $Q(0) < 0$, $Q(\gamma_m) < 0 \Rightarrow \gamma_m \geq v(\gamma_p)$; (2.2) if $Q(0) > 0$, then $Q(\gamma_m) < 0 \Rightarrow u(\gamma_p) \geq \gamma_m \geq v(\gamma_p)$, where $u(\gamma_p)$ is the upper root of $Q(\gamma_m) = 0$. However, since $\gamma_m < k(\gamma_p)$ and $k(\gamma_p) < u(\gamma_p)$, $u(\gamma_p) \geq \gamma_m$ is not binding. So, $\tau < 1$ requires $\gamma_m \geq v(\gamma_p)$. To summarize, the conditions for $0 < \tau < 1$ are $\gamma_m < k(\gamma_p)$, $\gamma_m > k(\gamma_p)$ if $\gamma_p \geq \widetilde{\gamma}_p$, and $\gamma_m \geq v(\gamma_p)$. Note that $v(\gamma_p)$ and $k(\gamma_p)$ intersect at $\widetilde{\gamma}_p$, with $v'(\gamma_p) > 0$, which is less than $k'(\gamma_p)$. Therefore, the binding constraints for this equilibrium are $v(\gamma_p) < \gamma_m < h(1 - n_c)$ and $\gamma_p \geq \widetilde{\gamma}_p$. ■