Intermediation and Imperfect Credit*

Grace Xun Gong
School of Economics and Academy of Financial Research, Zhejiang University

Ziqi Qiao
University of Wisconsin - Madison

Yuet-Yee Wong
Binghamton University

Randall Wright
Zhejiang University and Wisconsin School of Business

August 5, 2025

Abstract

We study environments with intermediation (trade via middlemen) and credit that is constrained (either exogenously, or endogenously due to limited commitment). Existing models of middlemen assume they have advantages in search, bargaining, information, etc. Here they have advantages using credit, because they are better at credibly promising payments or enforcing others' payments. With exogenous debt limits there is a unique equilibrium transaction pattern – direct trade, indirect trade, or both – depending on parameters. With endogenous limits, there are multiple equilibria, including ones where credit conditions fluctuate as self-fulfilling prophecies. Depending on details, intermediaries may attenuate or amplify credit cycles.

JEL codes: D51, D61, D83
Keywords: Middlemen, Intermediation, Search, Debt Limits, Credit Cycles

*We thank Benoit Julien, Chao Gu and Yu Zhu for input. We also thank participants in several seminars or workshops. Gong acknowledges support from the Humanities and Social Sciences Project of the Ministry of Education of China (No. 24YJC790052). Wright acknowledges support from the Raymond Zemon Chair in Liquid Assets and the Kenneth Burdett Professorship in Search Theory and Applications at UW. The usual disclaimers apply.

Many middlemen provide credit... This financial support helps businesses sustain operations and enhances market liquidity. Marketing for Managers, The Critical Role of Middlemen in Enhancing Market Dynamics.

Intermediaries who are able to make credible commitments bring advantages over contracts between buyers and sellers that are subject to renegotiation. Market Microstructure: Intermediaries and the Theory of the Firm.

1 Introduction

This paper studies environments with intermediation, by which we mean trade via middlemen, and credit that may be constrained, either exogenously, or endogenously due to limited commitment. As background, Rubinstein and Wolinsky (1987) emphasized that intermediated trade is a feature of many if not most real-world markets, yet there is no role for this in classical equilibrium theory, and proposed a framework based on search and bargaining. Following their work, a literature emerged where middlemen arise endogenously when certain agents have various advantages over others: they may be faster at locating trading partners; they may have lower search or storage costs; they may be able to hold larger or more diverse inventories; they may possess superior information; or they may be good at bargaining.[1]

We introduce a new dimension along which middlemen may have advantages: they are better at using credit, by more reliably promising future payment, or enforcing payment by others. This is realistic in many contexts. Consider trying to sell your car, which is a leading example because it rings true, and because there is nice empirical work on intermediation in automobile markets (see Murry et al. 2025 and references therein). If potential buyers do not have sufficient liquidity for immediate settlement,

[1] As there are many related papers, we posted an online bibliography at https://github.com/qiaoziqi/middlemen, but here is a sample. In Rubinstein and Wolinsky (1987) middlemen are faster than producers at contacting customers; in Biglaiser (1993) and Li (1998) they have superior information; in Masters (2008) and Nosal et al. (2015) they have higher bargaining power or lower costs; in Shevchenko (2004) and Watanabe (2010) they hold bigger or better inventories. We also mention Urias (2018), who has an environment similar to ours, but with crucial differences in assumptions (producers are not allowed to trade directly with consumers) and applications (credit is not analyzed). Also see Hu et al. (2025), although again there are differences in assumptions and applications (they are interested in supply chains where intermediaries provide working capital). Gu et al. (2025), like this paper, study dynamics with middlemen and cite related work, but it has nothing to do with credit.

deferred payment is an option, but you might worry about them reneging. An alternative is to sell your car to a dealer, where default can be less of a concern, either because they have liquidity for immediate settlement (deep pockets) or they are less likely to renege on deferred settlement (due to reputational considerations). Of course, when dealers sell, their customers may have commitment issues, but it is no stretch to think used-car dealers are better than you at collecting debt.[2]

Our contribution can be seen as either putting debt limits into theories of intermediated trade, or introducing middlemen into theories of imperfect credit. The result is greater than the sum of its parts: there are interesting interactions between credit and middlemen that would be missed if one only examined each in isolation.

In terms of modeling, we build on recent work in the spirit of Rubinstein-Wolinsky by Gong and Wright (2024). That paper is not about credit, but the environment is a natural one in which to introduce deferred settlement. In particular, as in Lagos and Wright (2005), it has both centralized and decentralized trade, and features an asynchronicity of expenditures and receipts: sometimes agents want to buy in the decentralized market while their incomes accrue in the centralized market. It is also flexible since, as is known from other applications, it easily incorporates various frictions and alternative ways to determine the terms of trade. Also, different from most related work, this framework replaces the usual three-sided market, where producers, consumers and middlemen all interact, with two two-sided markets, one with wholesale trade between producers and middlemen, and one with retail trade between buyers and sellers, which is arguably more realistic and definitely more tractable.

The objects being traded can be goods, inputs or assets, which is relevant to the extent that in reality there is intermediated trade in all three. Another key feature is that these objects can be either indivisible or divisible, different from much related

[2] On imperfect credit, there are too many papers with exogenous debt limits, but on more-or-less endogenous limits relevant examples include Kehoe and Levine (1993), Alvarez and Jermann (2000), Gu et al. (2013a,b), Gu et al. (2016), Azariadis and Kass (2007,2013), Lorenzoni (2008), Hellwig and Lorenzoni (2009), Sanches and Williamson (2010), Carapella and Williamson (2015) and Kiyotaki and Moore (1997). These do not have middlemen like the papers mentioned in fn. 1, although there is some work with banks, a kind of intermediary (see the survey by Gu et al. 2023).

work. This lets us analyze the intensive margin – the size of trades – not just the extensive margin – the number of trades. Also note that divisibility can be interpreted in terms of quantity or quality, and the latter is interesting because even if buyers want just one unit when shopping for, say, a car, obviously cars and most other things can vary in quality. Moreover, divisibility is important for modeling debt limits, technically, and substantively: it is not merely that buyers may not have enough credit to get something, it may be that they can but have to settle for a lower quantity/quality.

First we characterize equilibrium with exogenous debt limits to capture the above-mentioned forces, middlemen may be good at getting credit from producers or extending/enforcing credit to consumers. We show equilibrium exists and is unique. Depending on parameters – debt limits, bargaining power, etc. – we pin down the endogenous pattern of exchange, either no trade, only direct trade from producers to consumers, only indirect trade from producers to middlemen, then from middlemen to consumers, or both direct and indirect trade. This is useful because in reality some markets have mostly direct trade, others have mostly indirect trade, and still others are in between (see fn. 13), and our results elucidate fundamental factors determining which of these trade patterns might emerge.

Then we endogenize debt limits by saying repayment must be incentive compatible given the punishment for agents who renege is the loss of future credit. Note that while we often use the word renege, suggesting that the creditor gets nothing from the debtor, this may be too strong of an interpretation. Below we discuss versions of the model where a creditor gets a fraction of what is owed, or, equivalently for our purposes, gets what is owed with some probability. One can think of this as having agents "write down the debt," which is common in reality, and which is how we interpret "renegotiation" in the second epigraph (due to Spulber 1999). Hence commitment here can involve a credible promise to not renegotiate, i.e. to not partially default.

In any case, endogenizing debt limits generates multiple equilibria, including non-stationary equilibria with fluctuations in credit conditions, intermediation activity, prices, etc. In some versions these fluctuations are deterministic, in others they are necessarily stochastic. Also, they can involve regime switching with recurrent changes in the trade pattern (no trade, direct trade from producers to consumers, etc.). Also, they are self-fulfilling prophecies, not due to fundamental factors, consistent with the long-standing notion that intermediation is excessively volatile or unstable.[3] To be clear, the point we emphasize is not that intermediation *causes* instability, it is that limited commitment provides an endogenous role for intermediation and an endogenous source of dynamics, suggesting a more subtle but no less interesting link between intermediaries and volatility.

Characterizing these dynamics is one of our main goals. In one version of the model, we prove there are stochastic (sunspot) equilibria but not deterministic cycles; in another version both are possible. In either case we can ask if middlemen tend to attenuate or amplify fluctuations. The answer is that it can go either way on both the extensive and intensive margins: over the cycle, when there are fewer producers in the market, middlemen may enter with a higher or lower probability; and when producers bring lower quantity/quality to the market middlemen may bring higher or lower quantity/quality.[4]

As more motivation, there is a growing interest in BNPL (buy now pay later) plans by academics and regulators.[5] Of course, credit issued to consumers by retailers is not new, with an early example being the Charga-plate system popular in the 1930's-50's (Frankel 2024). Even earlier, in the 1900s cards were launched by department stores

[3]Myerson (2012) suggests such cycles are interesting, but his model is unrelated. More generally, it is an old idea that financial intermediaries are susceptible to instability, including banks (Diamond and Dybvig 1982) and other financial institutions (again see the Gu et al. 2023 survey).

[4]Further on dynamics, it is well known search models with increasing returns can have multiplicity and belief-based dynamics (e.g. Diamond 1982; Diamond and Fudenberg 1989; Mortensen 1999). That is not what is going on here: we have constant returns. It is also understood that monetary search models generate multiplicity and belief-based dynamics since what you accept in payment can depend on what others accept (see surveys by Lagos et al. 2017 and Rocheteau and Nosal 2017). That is not what is going on here, although there is a similarity: payment frictions take center stage.

[5]See, e.g., Han et al. (2024) and Stavins (2024). According to the latter, BNPL is "a short-term, interest-free credit option for retail purchases that is becoming increasingly popular, and evidence indicates that its use is significantly higher among financially vulnerable consumers... BNPL can thus provide short-term credit to consumers who lack alternative sources of credit." In terms of size, BNPL usage grew from \$50 billion in 2019 to \$370 billion in 2023 (Mojon et al. 2023).

and oil companies that, different from many cards today, could only be used at specific vendors (Tretina and Little 2004). With the payment landscape evolving rapidly, these issues seem worth study, but that is just one application – the bigger idea is to study how credit and intermediated trade interact in a general setting.

In what follows, Section 2 describes the environment. Sections 3 and 4 analyze stationary equilibria with exogenous debt limits in a baseline model and extensions. Section 5 endogenizes debt limits and discusses dynamics. Section 6 concludes.[6]

2 Environment

A continuum of agents live forever in discrete time. There are three types: consumers C, producers P, and middlemen M, with population measures N_c, N_p and N_m. In each period, three markets convene sequentially. First there is a wholesale market WM, where type P agents may trade an object to type M, with Q denoting quantity or quality. Type C cannot participate in WM – an assumption about spatial/temporal separation – but there is a retail market RM where they may buy $q \leq Q$ from sellers that are either M agents that have traded in WM or P agents that go to RM in search of direct trade with C. In fact, we do not need P agents to literally produce the object being traded, and in some contexts it is better to interpret them as endowed with it.

Both WM and RM are decentralized, with bilateral random matching, bargaining, and payment frictions. After they close, there is a frictionless centralized market CM, where all agents sell labor ℓ, buy a numeraire good x and settle debts. The idea is that agents use credit for RM and WM purchases, to be honored in the next CM. This is different from related models where spot payments are made in transferable utility.

[6] In addition to work following Rubinstein and Wolinsky (1987), there are papers using a different search model, focusing on dealers in OTC asset markets, following Duffie et al. (2005); see Hugonnier et al. (2025) for a survey. These differ in various ways – e.g., in those models dealers typically hold no inventories, they simply reallocate assets across investors via a frictionless interdealer market. An exception is Weill (2008), who has inventories and studies dynamics, but only transitions to steady state, not endogenous fluctuations, which as Trejos and Wright (2016) show cannot occur in such a model. Also those papers generally assume transferable utility, but it would be good to add payment frictions (see Martel et al. 2023 for a step in that direction, although they not mention imperfect commitment or credit, they simply impose an exogenous payment ceiling $P \leq \bar{P}$).

Transferable utility is equivalent to a special case of our setup, with perfect credit, but we are interested in imperfect credit. That is microfounded below, with endogenous debt limits, but it is useful to first study exogenous limits since they are interesting in their own right and provide a stepping stone to later analysis.

In CM all agents have discount factor $\beta \in (0,1)$ for next period, and instantaneous utility $U(x) - \ell$ with $U'(\cdot) > 0 > U''(\cdot)$. As usual, quasi-linear utility implies everyone starts next period with a clean slate: history independence. It also lets us restrict attention, without loss of generality, to short-term debt cleared in every CM. In RM, C gets $u(q)$ from acquiring q, with $u'(\cdot) > 0 > u''(\cdot)$ and $u(0) = 0$. Type P produce Q (when it is produced and not an endowment) in WM after they meet type M and decide to trade with them, or decide to go to RM seeking direct trade. There are two versions: Q can be indivisible (i.e., fixed at the production stage), or divisible (a choice at the production stage). In either version, it is divisible in RM trade. Production costs $c(Q)$ in units of numeraire, or labor, since we assume that 1 unit of ℓ produces q unit of x, with $c'(\cdot) > 0 \leq c''(\cdot)$, $c(0) = 0$ and $u'(0)/c'(0) = \infty$. Besides production cost, P and M need to pay an entry cost κ whenever going to RM.[7]

Later we say that any $Q - q$ left over after RM trade can be carried into CM where it turns into $A(Q - q)$ units of numeraire, called scrap value. While for some results below $A > 0$ is interesting, for now $A = 0$. In any case, assume $A \leq c'(0)$, meaning P does not want to produce only for scrap value. Also assume P can only produce once per period, so trading in WM means sitting out the RM (we analyzed a version where P can produce twice, for both WM and RM trade, but ruling that out simplifies the analysis without dramatically affecting the results and captures the idea that for P RM trade is an opportunity cost of WM trade).

Meetings are determined by a CRS technology. The measure of WM meetings is $m_W(N_m, N_p)$ where N_m and N_p are the fixed measures of WM buyers and sellers,

[7]While q can be a good and $u(q)$ a utility function, one can also interpret q as an input and $u(q)$ a production function with output in units of CM numeraire x, or q can be an asset with $u(q)$ the buyers' return and $c(q)$ the sellers' return in the same units. In general, when P does not produce Q, but it is an endowment, $c(q)$ is P's opportunity cost of selling it.

M and P. Similarly, the measure of RM meetings is $m_R(N_c, N_s)$ where the measure of RM buyers N_C is fixed but the measure of RM sellers N_s is endogenous. In either case, the buyer-seller ratio, called market tightness, determines the meeting probabilities. This is all standard, but it is worth emphasizing that having two two-sided markets is what allows the use of general meeting technologies, while papers with three-sided markets typically use only a special case where α_{ij} is proportional to $N_j/\Sigma_h N_h$.

Terms of trade are determined by generalized Nash bargaining, with θ_{ij} the bargaining power of i when trading with j, and $\theta_{ji} = 1 - \theta_{ij}$.[8] For WM trade, P produces Q and sells $Q_{pm} = Q$ to M (this is obvious when Q is indivisible; when it is divisible, it is assumed P cannot produce Q and sell $Q_{pm} < Q$ to M, then take $Q - Q_{pm}$ to RM to sell to C, as that is like producing twice, which is ruled out). However, when it is divisible P can produce one Q_{pm} for trade with M and a different Q_{pc} for going to RM in search of direct trade. Then $q_{mc} \leq Q_{pm}$ is what M sells to C and $q_{pc} \leq Q_{pc}$ is what P sells to C in RM. By way of preview, in equilibrium $q_{ic} = Q_{ic}$, even if there is scrap value $A > 0$, but that is a result, not an assumption.[9]

Associated with Q_{pm}, q_{pc} and q_{mc} are payments p_{pm}, p_{pc} and p_{mc} due in the next CM. These are constrained by $p_{ij} \leq D_{ij}$, which can reflect properties of buyers – e.g., their ability to commit or credibly promise future payment – as well as sellers – e.g., their ability to collect debt, perhaps by punishing those who renege. In addition to the p's and q's, we have these endogenous choices: Allowing for mixed strategies, τ is the probability P and M trade in WM meetings, and ρ is the probability P goes to RM after not trading in WM.

[8] We also tried Kalai's alternative to Nash, since it is known that the bargaining solution matters when there are payment constraints (e.g., Aruoba et al. 2007). Most, if not all, results below are similar with Kalai bargaining. One can also use alternatives like price posting with random or directed search, as well as Walrasian price taking, which might make more sense when meetings are not bilateral but involve large numbers as in, e.g., the labor-search model of Lucas and Prescott (1974) or the money-search model of Rocheteau and Wright (2005).

[9] Still, we can give the intuition here. First, the result says $q = Q$ on, but not necessarily off, the equilibrium path. If sellers were to find themselves in RM with a large Q, they might sell $q < Q$, but then in equilibrium they would not bring such a large Q to RM. An analogy to monetary theory is useful. In Lagos and Wright (2005), e.g., on the equilibrium path buyers spend all their money when they meet sellers; off the equilibrium path, if they had a lot of money they might not spend it all, but then they do not bring so much money.

3 Baseline Results

We start with the situation where the WM debt limit does not bind, as can be guaranteed by $D_{pm} = \infty$. This is a leading case because it captures the idea that M has excellent credit, or equivalently deep pockets, in dealing with P, and lets us focus on imperfect credit in RM (but see Section 4.1). To define equilibrium we exploit a general feature of search-and-bargaining models: equilibrium is a list of value functions satisfying dynamic programming equations, terms of trade satisfying bargaining solutions, state variables satisfying steady state conditions (more generally, laws of motion), and decisions satisfying best response conditions. We discuss each in turn.

Denote the value functions for type i in WM, RM and CM by V_i^W, V_i^R and V_i^C. The CM problem for type i is

$$V_i^C(\Omega) = \max_{x,\ell} \left\{ U(x) - \ell + \beta V_{i,+1}^W \right\} \text{ st } x = \Omega + \ell \quad (1)$$

where ℓ is labor income when 1 unit of ℓ produces 1 unit of x, and Ω is wealth. For P, Ω includes accounts receivable from either WM or RM; for C, it includes accounts payable from RM; and for M, it includes accounts receivable from RM minus accounts payable from WM. As is standard in models like this, given an interior solution for ℓ, it is immediate that V_i^C is linear with slope 1.[10]

Moving to WM, for P,

$$\begin{aligned} V_p^W &= V_p^C(0) + \alpha_{pm}\tau \left[p_{pm} - c(Q_{pm}) \right] \\ &\quad + (1 - \alpha_{pm}\tau)\rho \max_{Q_{pc}} \left[V_p^R(Q_{pc}) - V_p^C(0) - c(Q_{pc}) - \kappa \right]. \end{aligned} \quad (2)$$

The first term is the value of not producing and going to CM with $\Omega = 0$. The second is the probability of trading in WM times the surplus from continuing with accounts receivable p_{pm} minus cost $c(Q_{pm})$, using the result that $V_p^C(p_{pm}) - V_p^C(0) = p_{pm}$ by the linearity of V_{pt}^C. The third is the probability of not trading in WM, and with probability

[10] It is also immediate that x is independent of Ω, and that one-period credit is without loss of generality as agents are happy to settle debts in CM. That is what we mean by history independence, and it all follows easily from quasi-linear CM utility (as in Lagos and Wright 2005), but the results also hold for any $\tilde{U}(x, 1 - \ell)$ such that $\tilde{U}_{11}\tilde{U}_{22} = \tilde{U}_{12}^2$ (as in Wong 2016).

ρ producing and going to RM, which entails surplus $V_p^R(Q_{pc}) - V_p^C(0) - c(Q_{pc}) - \kappa$. For the choice of $Q_{mc}, Q_{pc} \in \mathcal{Q}$, where $\mathcal{Q}$ is the production set, we have two options: indivisibility, $\mathcal{Q} = \{Q\}$ for some constant Q, as assumed in most related papers; and divisibility, $\mathcal{Q} = [0, \infty)$, which generates more insights with additional work.

For M in WM,

$$V_m^W = V_m^C(0) + \alpha_{mp}\tau \left[V_m^R(Q_{pm}) - V_m^C(0) - p_{pm} - \kappa\right] \tag{3}$$

using the linearity of V_m^C and the fact that M only trades in WM if they then go to RM.[11] Moving to RM, for P

$$V_p^R(Q_{pc}) = V_p^C(AQ_{pc}) + \alpha_{sc}(p_{pc} - Aq_{pc}), \tag{4}$$

where the second term comes from $V_p^C(p_{pc} + AQ_{pc} - Aq_{pc}) - V_p^C(AQ_{pc}) = p_{pc} - Aq_{pc}$. Similarly, for M in RM

$$V_m^R(Q_{pm}) = V_m^C(AQ_{pm}) + \alpha_{sc}(p_{mc} - Aq_{mc}). \tag{5}$$

For both P and M these are conditional on being in RM, after paying the entry cost κ. For C in RM

$$V_c^R = V_c^C(0) + \alpha_{cm}[u(q_{mc}) - p_{mc}] + \alpha_{cp}[u(q_{pc}) - p_{pc}] \tag{6}$$

where $q_{ic} \leq Q_{ic}$, in general, but clearly $q_{ic} = Q_{ic}$ holds if, e.g., there is no scrap value, $A = 0$.

Terms of trade when i sells to j come from generalized Nash bargaining,

$$(p_{ij}, q_{ij}) = \arg\max_{(p,q)} S_{ij}(p,q)^{\theta_{ij}} S_{ji}(p,q)^{\theta_{ji}} \tag{7}$$

where the S's are surpluses defined as follows: When P sells to C,

$$S_{pc} = p_{pc} - Aq_{pc} \text{ and } S_{cp} = u(q_{pc}) - p_{pc}. \tag{8}$$

[11] Notice in (3) that the term in brackets looks like immediate settlement due to the appearance of $-p_{pm}$, but it actually is the anticipation of deferred settlement in the next CM; in this sense one can say perfect credit looks like deep pockets.

When M sells to C,

$$S_{mc} = p_{mc} - Aq_{mc} \text{ and } S_{cm} = u(q_{mc}) - p_{mc}. \tag{9}$$

And when P sells to M

$$S_{pm} = p_{pm} - c(Q_{pm}) - \rho \left[V_p^R (Q_{pc}) - V_p^C (0) - c(Q_{pc}) - \kappa \right] \tag{10}$$

$$S_{mp} = V_m^R (Q_{pm}) - V_m^C (0) - \kappa - p_{pm}. \tag{11}$$

There are constraints in (7): $p_{ij} \leq D_{ij}$; and $q_{ic} = Q_{ic}$, for now, but more generally $q_{ic} \leq Q_{ic}$. Note that Q_{pm} is determined bilaterally between P and M, while Q_{pc} is unilaterally chosen by P. Also note that there are holdup problems in RM, since when P meets C the production cost is sunk, when M meets C the WM debt is sunk, and for both P and M the RM entry costs κ is sunk.

Next we determine participation in WM and RM and hence the meeting probabilities. In WM, α_{pm} and α_{mp} come from $m_W (N_m, N_p)$ with N_m and N_p fixed. In RM, while the measure of buyers is fixed at N_c, the measure of sellers includes M that trade in WM plus P that do not trade in WM and go to RM,

$$N_s = N_p \left[\alpha_{pm} \tau + (1 - \alpha_{pm} \tau) \rho \right]. \tag{12}$$

We call (12) a steady state condition, but it is basically static.[12] Also, in RM the meeting technology treats all sellers the same in generating α_{cs} and $\alpha_{pc} = \alpha_{mc} = \alpha_{sc}$, but the outcome of a meeting depends on the seller type, P or M, determined by $\alpha_{cp} = \alpha_{cs} (1 - \alpha_{pm} \tau) \rho N_p / N_s$ and $\alpha_{cm} = \alpha_{cs} \alpha_{pm} \tau N_p / N_s$.

Now consider strategy profile (τ, ρ), where τ is the probability P and M trade in WM, and ρ the probability P goes to RM after not trading in WM. For ρ we have:

$$\rho = \begin{cases} 0 & \text{if } V_p^R(Q_{pc}) - V_p^C(0) \leq c(Q_{pc}) + \kappa \\ [0,1] & \text{if } V_p^R(Q_{pc}) - V_p^C(0) = c(Q_{pc}) + \kappa \\ 1 & \text{if } V_p^R(Q_{pc}) - V_p^C(0) \geq c(Q_{pc}) + \kappa \end{cases} \tag{13}$$

[12] Static in the same sense that, e.g., vacancies are static in Pissarides (2000). There are state variables in the model, the D's, but for now they are exogenous and constant over time so we do not bother to keep track of them here. That will change below.

For τ, if there is transferable utility P wants to trade with M iff M wants to trade with P iff $S_{pm} + S_{mp} > 0$. Since we do not have transferable utility, in general, WM trade requires the proverbial double coincidence of wants:

$$\tau = \begin{cases} 0 & \text{if } S_{mp} < 0 \text{ or } S_{pm} < 0 \\ [0,1] & \text{if } S_{mp}, S_{pm} \geq 0 \text{ and } S_{mp}S_{pm} = 0 \\ 1 & \text{if } S_{mp} > 0 \text{ and } S_{pm} > 0 \end{cases} \tag{14}$$

However, for now $D_{pm} = \infty$, so $p_{pm} \leq D_{pm}$ is slack and hence:

$$\tau = \begin{cases} 0 & \text{if } S_{pm} + S_{mp} < 0 \\ [0,1] & \text{if } S_{pm} + S_{mp} = 0 \\ 1 & \text{if } S_{pm} + S_{mp} > 0 \end{cases} \tag{15}$$

A *stationary equilibrium*, or SE, is defined by: (V_i^n) for each type i in each market n satisfying (1)-(6); (p_{ij}, q_{ij}) satisfying (7)-(11); N_s satisfying (12); (Q_{ic}) satisfying (2) and (7); and (τ, ρ) satisfying (13) and (15). In fact, the *stationary* qualification is not necessary, because when the D's are exogenous there are no other equilibria, something that changes when the D's are endogenous. So we use the SE label, but when we say below that, e.g., there is a unique stationary equilibrium it is trivial to generalize that by saying there is a unique equilibrium.

For indivisible Q, here is an algorithm for characterizing the SE set: (1) Pick a candidate strategy profile by specifying whether each element (τ, ρ) is 0, 1, or mixed. (2) Given (τ, ρ), determine N_s and hence the α's. (3) Then solve the dynamic programming equations for the V's taking p's as given (easy, since the system is linear). (4) Then use the bargaining solution to get the p's. (5) Given all that, check the best response conditions, since these variables just constructed constitute a SE iff those conditions are satisfied. (6) Repeat until exhausting possible strategy profiles. For divisible Q, the procedure is similar but we solve Q_{pc} from (2) and Q_{mc} from (7).

Since τ and ρ can be 0, 1 or in $(0, 1)$ there are $3^2 = 9$ candidate profiles which we classify into 4 different Regimes. Regime N, for *no trade*, has $\tau = \rho = 0$. Regime D, for *direct trade*, has $\tau = 0$ and $\rho > 0$, so P never trades in WM and goes to RM with positive probability. Regime I, for *indirect trade*, has $\tau > 0$ and $\rho = 0$, so P trades with M in WM with positive probability, and M goes to RM while P does not. Regime B, for *both*, has $\tau > 0$ and $\rho > 0$, so both P and M go to RM with positive probability. Within a Regime we also distinguish between pure- and mixed-strategy outcomes for τ and ρ, and between cases where debt limits bind and are slack.[13]

Using the algorithm we can construct the SE set. The Appendix proves:[14]

Proposition 1: When $D_{pm} = \infty$ SE exists and is unique. Its dependence on parameters is shown in (D_{pc}, D_{mc}) space by Fig. 1 for $\mathcal{Q} = \{Q\}$ and Fig. 2 for $\mathcal{Q} = [0, \infty)$, where $\bar{D}_{pc}$ and $\bar{D}_{mc}$ are values at which the debt limits just bind, and the other boundaries of the regions are defined in the Appendix.

Note that the graphs are not drawn free hand but come from numerical specifications, with parameters listed in the Appendix. Also notice each graph has two panels, with $\theta_{mc} = \theta_{pc}$ on the left and $\theta_{mc} > \theta_{pc}$ on the right, to show the impact of M and P having different bargaining power against C, in addition to different ability to collect debt from C. The first result to highlight is that this partitioning parameter space into regions each containing exactly one Regime that constitutes SE implies *we have existence and uniqueness.*

In the graphs, solid borders separate regions supporting different Regimes; the dashed borders show separate pure and mixed strategies; and the dotted borders separate binding and slack debt limits. For Regime B, e.g., we have $\rho = 1$ or $\rho \in (0, 1)$ and debt limits may or may not bind. So *we know exactly what happens for all parameters.*

[13]While this is a theory paper, it is empirically useful to identify factors leading to different Regimes, since as mentioned different markets have different degrees of intermediation. Obviously many consumer goods are bought from middlemen, like grocery stores, but there are still farmers' markets. Inputs are often bought from intermediaries, but apparently high-end purveyors of coffee, chocolate and tea these days are buying direct from sources for several reasons (see Charles 2024). In the used car market, 2/3 of sales go through dealers (Li et al. 2025). In asset markets, trade for fed funds is about 40% intermediated, NASDAQ is closer to 100%, and many OTC markets including corporate debt, munis, and emerging-market debt, are in between (Lagos and Rocheteau 2006). Also, there are often middlemen chains – e.g., farmer to broker to distributor to retailer to consumer (see Wright and Wong 2014). While formalizations like the one here are too stylized to capture every detail of these diverse markets, they provide guidance as to what factors may be relevant.

[14]A technical detail is that we ignore outcomes with $\tau \in (0, 1)$ if they are nongeneric – e.g., if $D_{mc} = D_{pc}$ and $\theta_{mc} = \theta_{pc}$, when P meets M in WM it is a matter of indifference who goes to RM, so any τ is a best response, but that is uninteresting, and would not survive obvious refinements like imposing an ε transaction cost to trade and letting $\varepsilon \to 0$. Also, in the graphs we show the cutoffs $\bar{D}_{ic}$, where debt limits just bind, in the range where $\tau = \rho = 1$ but in general this depends on parameters (this does not affect the substance of the results much).

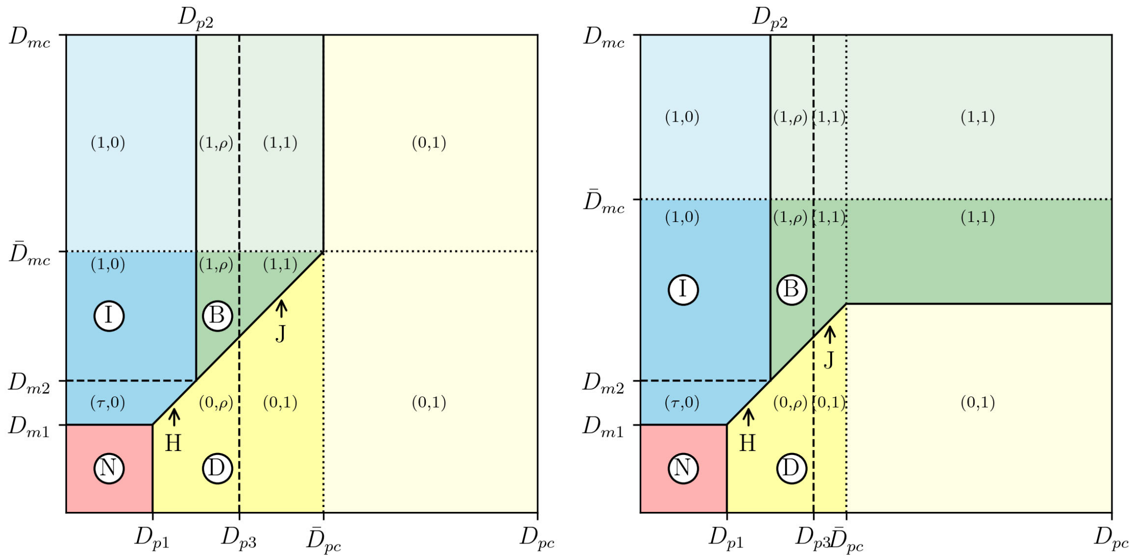

Figure 1: SE with Q indivisible and $D_{pm} = \infty$; $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

Consider the left panel in Fig. 1. When both D's are low (red region) Regime N obtains since payments are too constrained to justify RM entry by P or M. Outside that region various outcomes can occur. Regime D obtains if $D_{pc} > \bar{D}_{pc}$ (light yellow region) since P is unconstrained in RM and hence there are no gains from WM trade, or if $D_{pc} < \bar{D}_{pc}$ and we are below the 45° line (dark yellow region) since while P is constrained M is more constrained. Now suppose $D_{pc} < \bar{D}_{pc}$ and we are above the 45° line. When D_{pc} is low, Regime I obtains, as the D's justify RM entry by M but not P (dark blue region where M is constrained and light blue where M is not). When D_{pc} is somewhat higher, Regime B obtains, since D_{pc} makes RM profitable for P but $D_{mc} > D_{pc}$ makes it more profitable for M (dark green region where M is constrained and light green where M is not). The general conclusion, which may not be surprising, but at least we make it precise, is that *M is active when D_{pc} is low and D_{mc} high.*

For another perspective, fix D_{mc} and increase D_{pc}, moving horizontally through the graph. In this case Regimes can switch once or twice. When D_{mc} is very low we transit from Regime N to D as D_{pc} increases. When D_{mc} is somewhat higher we transit from Regime I to B and then from B to D. A general observation is that *relaxing credit frictions can lead to disintermediation.*

Now fix D_{pc} and increase D_{mc}, moving vertically through the graph. When D_{pc} is low we switch from Regime N to I, a case where RM is open with M but not without M (e.g., not if intermediation is shut down by regulation or taxation). For higher D_{pc} we switch from Regime D to I, so M is not crucial for RM to open, but could still improve welfare (see below). For even higher D_{pc} we switch from D to B, and for $D_{pc} \geq \bar{D}_{pc}$ we are always in Regime D. A switch from D to B shows *middlemen can be active solely due to their ability to enforce debt.*

The above discussion maintains $\theta_{mc} = \theta_{pc}$. If $\theta_{mc} > \theta_{pc}$, as in the right panel of Fig. 1, there can be WM trade when $D_{mc} < D_{pc}$, which shows *middlemen can be active solely due to bargaining power*, as is already known in the literature. In particular, the upper right has Regime D in the left panel and B in the right. Also notice in the left panel $\bar{D}_{mc} = \bar{D}_{pc}$ while in the right $\bar{D}_{mc} > \bar{D}_{pc}$, since greater bargaining power means M can fetch a higher price, so it takes bigger $\bar{D}$ for the constraint to slacken.

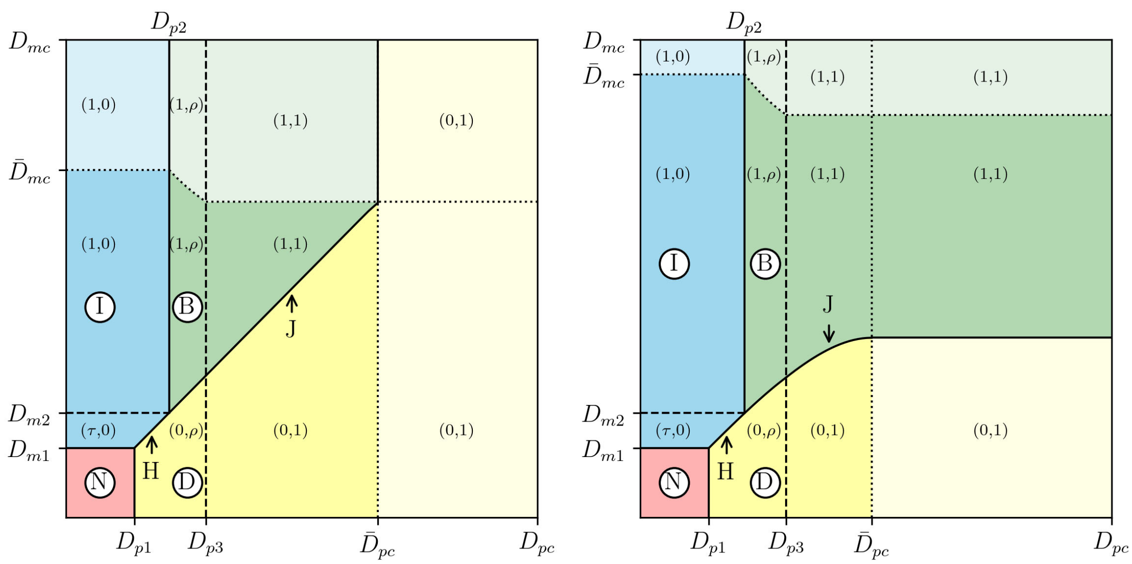

Figure 2: SE with Q divisible and $D_{pm} = \infty$; $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

The above discussion maintains indivisible Q. When it is divisible, Q is a choice in WM, but still $q = Q$ in RM. See Fig. 2, which is similar to Fig. 1 since, to facilitate comparison, they use the same parameters and the exogenous Q in the indivisible case is in the middle of its range in the divisible case. Making Q divisible affects the graphs in several ways, one being that the boundaries can now be nonlinear while they were linear in Fig. 1. In terms of substance, notice region N in now smaller than in Fig. 1, because *divisibility encourages entry by letting sellers better cater to market conditions.*

Fig. 3 varies D_{mc} and shows the effects on several variables with D_{pc} fixed at either a low (left panel) or high (right panel) value, where the curves are only drawn over relevant ranges – e.g., if M does not enter RM then Q_{mc} is not shown. Notice in the top row M becomes active exactly when $D_{mc} > D_{pc}$, as we already know, given $\theta_{pc} = \theta_{mc}$. In the left panel, when M starts going to RM P stops going, resulting in a switch from Regime D to I and an increase in N_s. In the right panel, when M starts going P does not stop, resulting in a switch from Regime D to B while N_s remains the same because M crowds out P one-for one. Hence, depending on parameters, *entry by middlemen may increase the number of sellers or simply crowd out producers.*

Rows two and three in Fig. 3 show quantity/quality Q and the unit price p/Q. In the left panel, since only one type of seller enters RM there is only one Q and one p/Q. In the right, when both M and P go to RM there is dispersion in Q and in p/Q. While here p_{pc}/Q_{pc} and p_{mc}/Q_{mc} monotonically decrease with D_{mc}, due to the concavity of $u(\cdot)$, Fig. 4 fixes D_{mc} and varies D_{pc}, and there average unit price in the right panel is nonmonotone due to composition effects (the mix of P and M in RM). Price dispersion is also nonmonotone: it is 0 for $D_{pc} < D_{p2}$ and $D_{pc} > D_{mc}$ and positive for $D_{pc} \in (D_{p2}, D_{mc})$. One implication is *intermediation can generate dispersion in price and quantity/quality.* A perhaps less obvious one is *in intermediated markets average price and price dispersion can be nonmonotone in debt limits.*[15]

Also notice in Fig. 3 that payoffs can increase with D_{mc}, due to effects on both the extensive and intensive margins, while in the right panel of Fig. 4, higher D_{pc} reduces payoffs of P and M, as well as the average payoff, over some range, even though it

[15] As an aside, when we reduce search frictions fall by raising the constant in the meeting technology, average price and price dispersion can both go up or down. This is relevant since some people are puzzled about average price and dispersion not falling in the data with improvements in information technology that seemingly capture higher search efficiency (see the discussion with references in Gong et al. 2025). In fact, it is easy to get price dispersion nonmonotone in frictions – models like Burdett and Judd (1983) already show that – while it is a little harder to get the average price nonmontone.

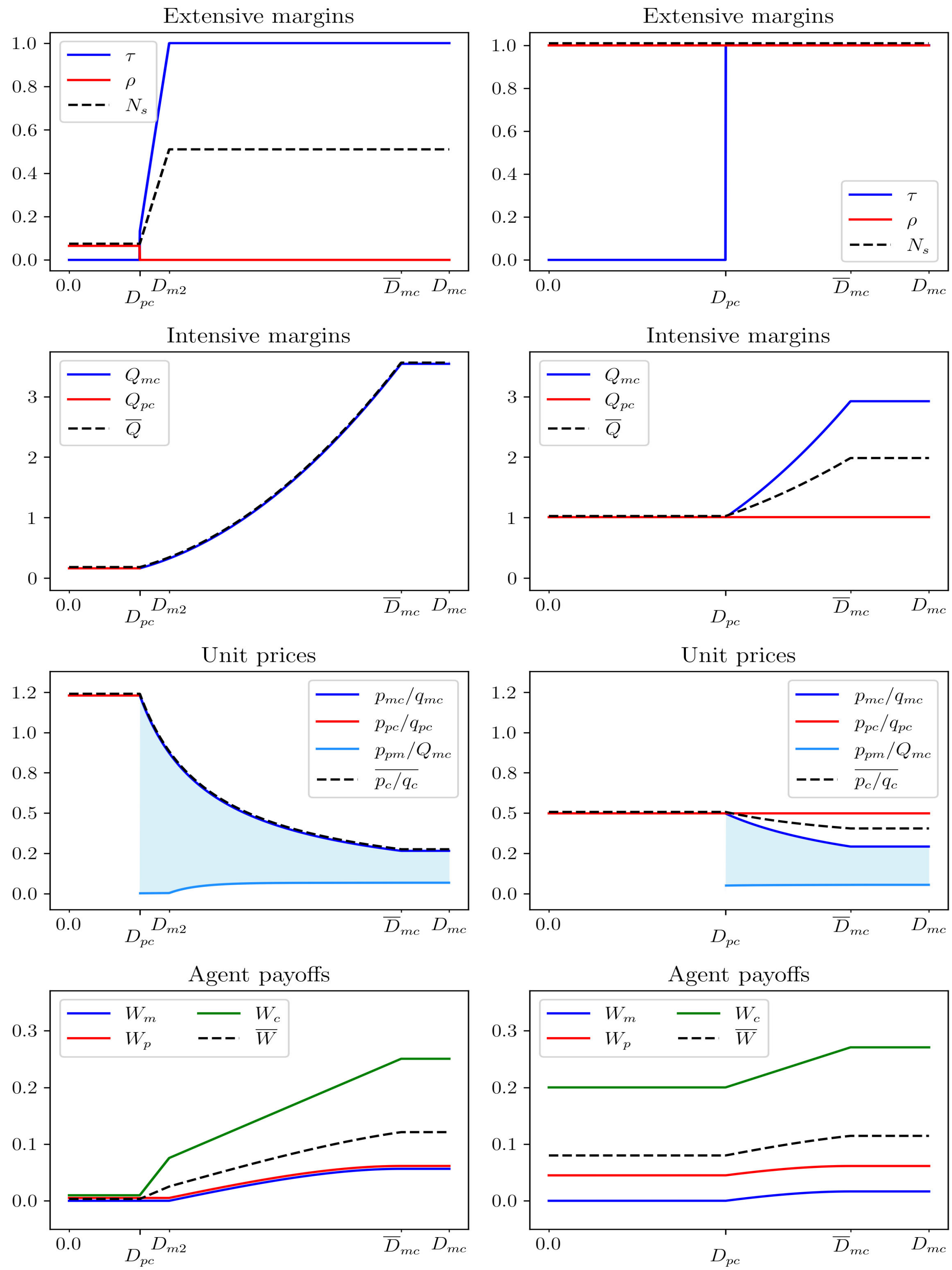

Figure 3: Effects of D_{mc} with Q divisible, $D_{pm} = \infty$ and D_{pc} low (left) or high (right).

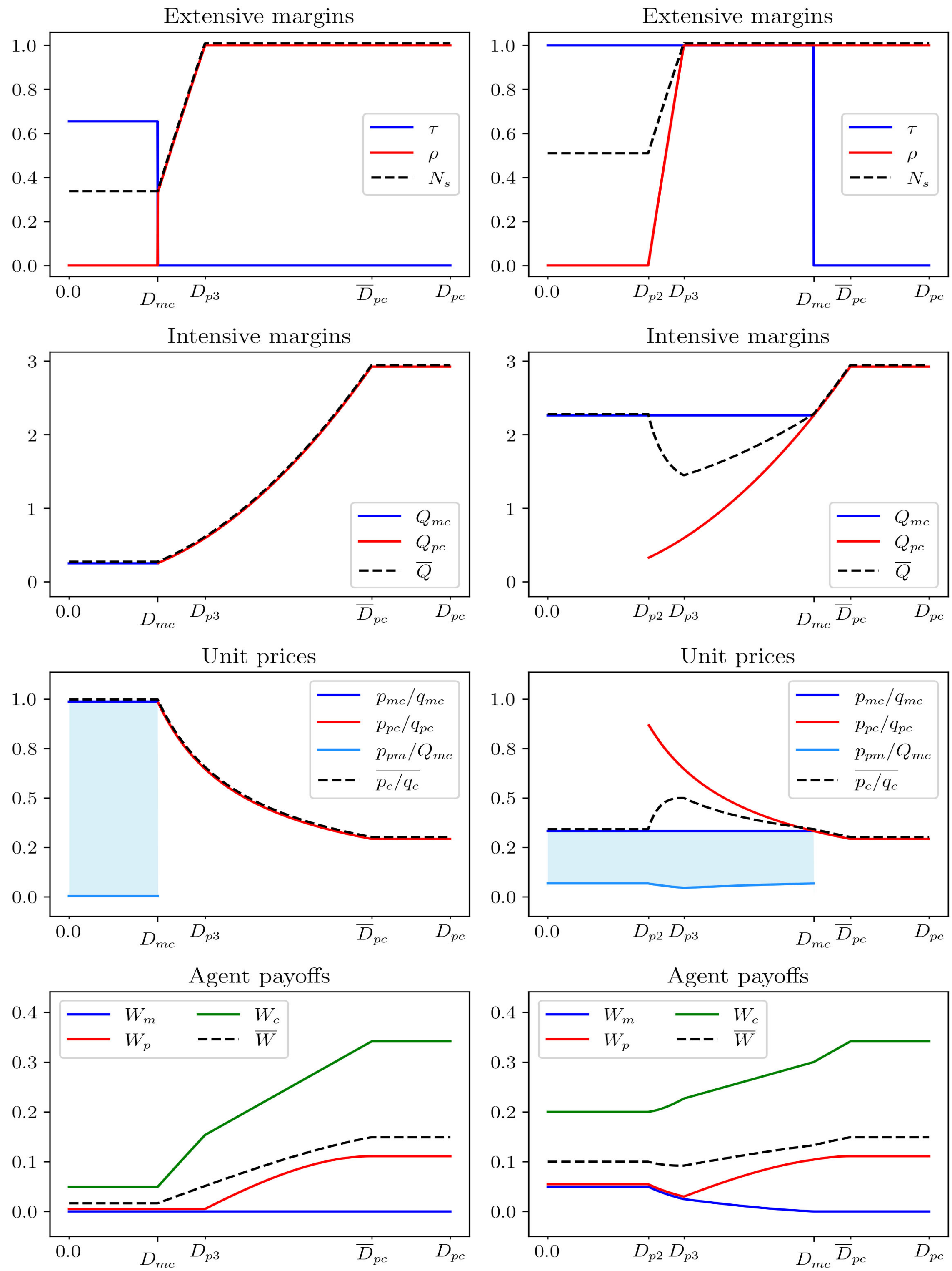

Figure 4: Effects of D_{pc} with Q divisible, $D_{pm} = \infty$ and D_{mc} low (left) or high (right).

raises the payoff for C. In other examples, higher D's can also lower C's payoff – the simplest case having Q indivisible, so raising D's means C pays more but does not get more. This is not novel – it is well known that buyers can be worse off when liquidity constraints are relaxed (more on this below). It is also not novel that P and M can be worse off at higher D's due to endogenous entry – as in many models, entry can be too high or too low. In summary, *lower payment frictions can increase or decrease welfare*, but we do not dwell on this because there are already many discussions in the literature.[16]

4 Extensions

4.1 Imperfect Wholesale Credit

Above $D_{pm} = \infty$ was assumed. Suppose now $D_{pm} < \infty$, so that M does not have unlimited credit, or deep pockets, in WM. Then τ is determined by (14), showing both P and M need a positive surplus for WM trade. However, if $S_{pm} + S_{mp} > 0$ then a binding D_{pm} will not make $S_{mp} < 0$, so we only need to check $S_{pm} \geq 0$.

It is not hard to verify that the conclusions of Proposition 1 continue to hold with $D_{pm} < \infty$: SE still exists uniquely. This is true for indivisible and divisible Q, but to save space we only show the partition for the former, in Fig. 5. The left panel has $D_{pm} < c(Q)$, so WM payments by M cannot even cover P's production cost, and hence only Regimes N and D are possible. The right panel has $D_{pm} > c(Q)$, so M can cover P's production cost and may or may not be able to cover P's opportunity cost of forgoing RM, so all Regimes are possible.

We could show the effects of D_{pm}, the way the effects of D_{mc} and D_{pc} are shown in Figs. 3 and 4, but instead simply mention one result: M's payoff can fall when D_{pm} increases. Intuitively, they can end up paying higher p_{pm} while not getting much, if

[16] See, e.g., Gong and Wright (2024), which is not about credit, but does have middlemen with both extensive and intensive margins. Their finding is that efficiency on the former margin requires bargaining powers satisfy the Hosios condition, while efficiency on the latter margin requires bargaining powers satisfy the Mortensen rule. Since it is generally hard to satisfy both simultaneously, tax-subsidy schemes have a role; we do not elaborate more to avoid diverting attention from the main messages.

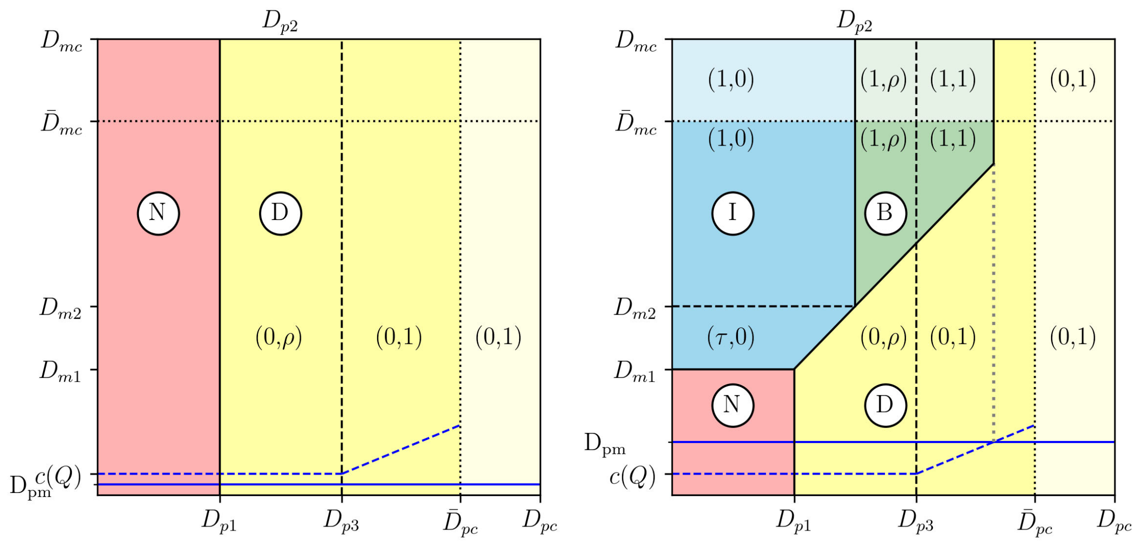

Figure 5: SE with Q indivisible and $D_{pm} < c(Q)$ (left) or $D_{pm} > c(Q)$ (right).

anything, more in terms of Q_{pm}, just like C's payoff can go down when D_{ic} increases in Section 3. Anyway, the conclusion here is that imperfect WM credit is interesting, and still tractable, although less so than when $D_{pm} = \infty$, so we still like that as a benchmark model.

4.2 Ex Ante Production

Next suppose production occurs before WM. Let ϕ be the probability P produces, to be determined along with τ and ρ. The definition of SE is similar to the baseline model (see the Appendix). However, now there is a WM holdup problem, in addition to the RM holdup problems mentioned above, since $c(Q)$ is sunk when P meets M. Also, there is a new source of wasted output: in addition to sellers failing to meet C in RM, now P might produce but neither meet M in WM nor go to RM.

One can check SE still exists uniquely. Reverting to $D_{pm} = \infty$ and restricting attention to indivisible Q, we present the results in Fig. 6. Since each element of (ϕ, τ, ρ) can be 0, 1 or mixed, there are more cases, although some are ruled out easily, e.g., $\phi = 1$ and $\tau = \rho = 0$, where P produces but never trades. Still, we must distinguish between, e.g., $\phi = 0$ with $\tau = 0$ and $\phi = 0$ with $\tau = 1$, even though both entail Regime N, because as usual to see whether $\phi = 0$ is a best response we must say what happens off the equilibrium path if P were to have Q and meet M in WM.

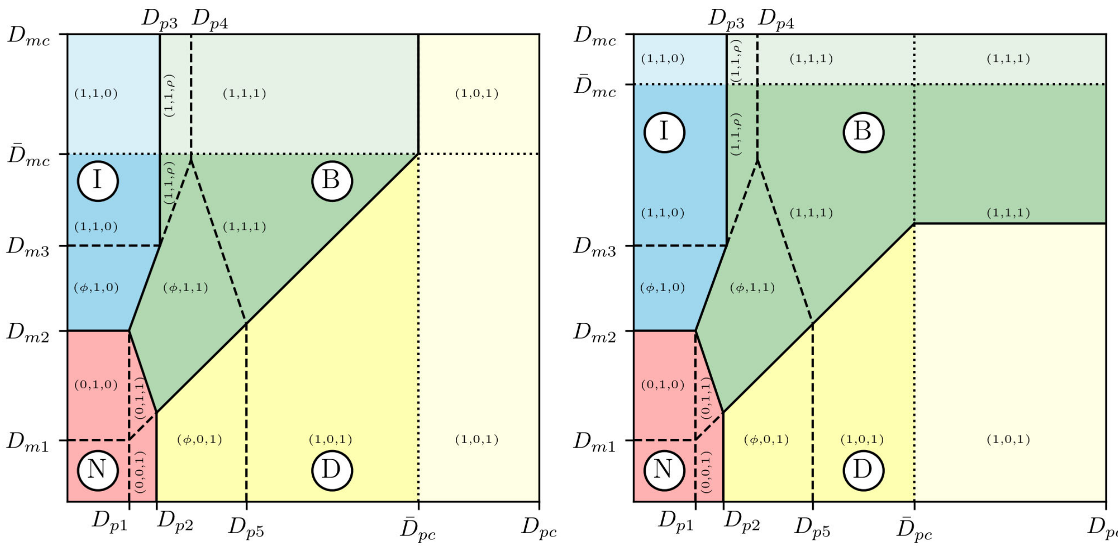

Figure 6: SE with ex ante prod. and $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

A difference here is that as D_{mc} increases, the Regime switches from N or D to B, then to I, making Regime B transitional. This contrasts with Section 3, where once we enter Regime B by increasing D_{mc} we never leave (technically this is because, over the relevant range, with ex ante production P agents mix using ϕ, while in Section 3 they mix using ρ). Also, now C's payoff can fall with D_{mc} for two reasons: there is a composition effect, since when M enters RM they crowd out P, and the former charge more; and again higher D_{mc} can result in C paying more without getting much more. In summary, ex ante production is interesting, but given the ex post version is more tractable it remains the preferred benchmark.

4.3 Scrap Value

Now consider scrap value $A > 0$. In principle, this could increase the incentive for sellers to enter RM by reducing the downside risk of not meeting C. Also, after meeting C, it could affect the bargaining outcome through sellers' threat points. Moreover, depending on whether Q is divisible, $A > 0$ can affect the intensive margin differently.

To see this, first, on the one hand when Q is divisible there is no benefit to carrying $Q_{ic} > q_{ic}$, where q_{ic} comes from RM bargaining, given the maintained assumption $c'(0) > A$. Therefore, $q_{ic} = Q_{ic}$ in equilibrium where $D_{ic} = \theta_{ic}u(Q_{ic}) + \theta_{ci}AQ_{ic}$. Hence Q_{ic} increases with D_{ic}, and so does sellers' surplus S_{ic}, just like the baseline model with $A = 0$. This means $A > 0$ does not qualitatively change equilibrium, and the Regime characterization is the same as before.

On the other hand, when Q is indivisible in WM but divisible in RM, with $A > 0$ we can no longer guarantee $q_{ic} = Q_{ic}$. Instead, q_{ic} is the minimum of three possibilities: the indivisible Q sellers bring to RM; the efficient q^* that solves $A = u'(q^*)$; or the generalized Nash bargaining solution that solves $D_{ic} = v_i(q_{ic})$ with

$$v_i(q) = \frac{\theta_{ci}u'(q)Aq + \theta_{ic}Au(q)}{\theta_{ci}u'(q) + \theta_{ic}A}. \tag{16}$$

This is the standard formula for generalized Nash bargaining with liquidity constraints (e.g., Lagos and Wright 2005), which reduces to $v_i(q) = \theta_{ic}u(q) + \theta_{ci}Aq$, as in the case of divisible Q, when the constraint is slack.

In any case, with $A > 0$, SE again exists uniquely, and characterization of Regimes in parameter space is similar to Section 3. Again, this extension is not too different from $A = 0$, so that remains the baseline model.

4.4 Repos

Next consider a type of secured credit and repos (i.e., repossessions, not repurchase agreements), which is one way to get D_{pc} and D_{mc} exogenous.[17] As a concrete example, suppose C wants a car that can potentially be obtained in two ways. One is to buy it from P, although it is better for this discussion to interpret P not as a producer, but simply someone with a car for sale. The other is to buy it from a used-car dealer M. The idea here is that if C fails to make a required payment in the CM, the seller can try to repo the car, where the probability that type j sellers succeed is χ_j for

[17] This setup is in the spirit of Kiyotaki and Moore (1997). Section 5 pursues a different approach, which delivers much more, but this one has the virtue of simplicity.

$j \in \{p, m\}$. It is eminently reasonable to say that used-car dealers are relatively good at this, $\chi_m > \chi_p$.

Suppose $D_{pm} = \infty$ and $A = 0$. Also, assume now that the Q_{ic} purchased in RM is not consumed until the end of the period. Thus, if the CM payment is made, C enjoys $u(Q_{ic})$; otherwise C gets $u(Q_{ic})$ with probability $1 - \chi_i$ and 0 with probability χ_i. This makes the incentive constraint for repayment $u(Q_{ic}) - p_{ic} \geq (1 - \chi_i)u(Q_{ic})$. Again there is no default on the equilibrium path, but the option to default generates an endogenous constraint $p_{ic} \leq D_{ic} \equiv \chi_i u(Q_{ic})$. Now from (7) D_{ic} binds iff $\chi_i \leq \theta_{ic}$. Note that we assume the seller does not get any value from the repo (one can assume otherwise but that raises some issues). In the car example, e.g., one can say the vehicle depreciates (has some scrap value that is small). Hence, the binding constraint on DM trade is still the incentive condition for buyers' repayment. This means the repo is a punishment for reneging buyers, not a recovery for sellers.

With indivisible Q, SE is basically the same as the baseline model – just set $D_{ic} = \chi_i u(Q)$. With divisible Q, $\chi_i > \theta_{ic}$ implies equilibrium is again the same as the baseline model, while $\chi_i \leq \theta_{ic}$ implies $p_{ic} = \chi_i u(Q_{ic})$ where Q_{ic} solves $\alpha_{sc}\chi_i u'(Q_{ic}) = c'(Q_{ic})$. In any case, again SE exists uniquely, and its dependence on parameters looks basically the same as the graphs in Section 3 with χ replacing D on the axes. However, there are some differences in implications. For one, reductions in credit frictions now mean higher χ, and in the range where debt limits bind, that raises p as well as p/Q, intuitively, since χ is effectively sellers' bargaining power as well as the constraint.

Like the other extensions, this is interesting, but does not obviously dominate the benchmark. One can say it at least takes a step toward endogenizing D, but we plan to go further in that direction in Section 5. First, we mention a reinterpretation. Instead of χ being the probability a seller recovers what is owed, it can be the fraction of what is owed – i.e., partial default can replace probabilistic repo.[18] Still we can maintain

[18] A detail is that if the seller takes back a fraction χ of q the buyers payoff should be $u[(1-\chi)q]$, not $(1-\chi)\, u(q)$. So perhaps it is better to say the seller can seize a fraction of the promised payment p rather than q.

the assumption that whatever gets recovered gives 0 value to the seller, so again repos affect the incentive of buyers to repay, not the incentive of sellers to grant credit in the first place, but it may be interesting to proceed differently in future work.

5 Intermediation Cycles

5.1 Endogenous Debt Limits

The idea is that now buyers can renege on promised payment, but risk being punished by taking away their future credit, as in Kehoe and Levine (1993), which makes the model inherently dynamic. Since there can be two types of RM sellers, P and M, there are two types of future credit the mechanism could take away from C, and we discuss both below. To keep the presentation manageable, assume P and M have the same RM bargaining power and $A = 0$. Also, Q is divisible for now, while the indivisible case is discussed in Section 5.3.

First suppose M faces no credit frictions in either WM or RM, $D_{pm,t} = D_{mc,t} = \infty$, but credit between C and P is limited by $D_{pc,t}$, which is determined as follows. In the CM at t, C can renege on $p_{pc,t}$ owed to P at a proportional cost, $\lambda_p \, p_{pc,t}$, with $\lambda_p \in [0, 1]$. This captures resources used up by opportunistic behavior, similar to "cash diversion" models (e.g., DeMarzo and Fishman 2007; Biais et al. 2007). Moreover, opportunistic agents are only caught and hence can only be punished with probability $\mu_p \in [0, 1]$, which plays a role similar to the random repo probability χ in Section 4.4.[19]

When C is caught reneging on P, assume for now that they lose future credit with P sellers but not M sellers. One rationale is that while taking away all future credit

[19]Both λ and μ measure disincentives to renege. While they are not critical (we could set $\lambda = 0$ or $\mu = 1$) they are included because they have interesting implications in related work using mechanism design. In banking theory, e.g., they are used by Gu et al. (2013b) and Huang (2015) to discuss the choice of who should be a banker as well as the optimal number of bankers. In monetary theory, e.g., Kocherlakota (1998) shows that fiat currency is never welfare enhancing (it is dominated by credit) if $\mu = 1$, but can be if $\mu < 1$. One interpretation of $\mu < 1$ is that debt payments are randomly monitored by the mechanism, like the IRS randomly audits tax payments. Another is that payees know with certainty if a default occurs, but can only communicate this to the mechanism randomly. Finally, on this topic, here C's cost of default λp is a deadweight loss, but as discussed in Section 4.4 it could go to the seller, so that it looks like partial default.

is harsher, it might not be viable: given M and C have gains from RM trade and $D_{mc,t} = \infty$, meaning M can enforce payment by C, having P deny credit to C might not be self-enforcing. This contrasts with having P deny C credit after a default, since if C did it before they will do it again given there is no further punishment available. Still, we later consider taking away all future credit, effectively putting defaulters in autarky, perhaps simply by excluding them from RM.

Under the current assumptions the incentive condition for C to pay debt $p_{pc,t}$ is

$$V_{c,t}^{C}(0) - p_{pc,t} \geq -\lambda_p \, p_{pc,t} + \mu_p V_{c,t}^{D}(0) + (1-\mu_p) V_{c,t}^{C}(0), \tag{17}$$

where $V_{c,t}^{D}$ is C's CM deviation payoff, which is the continuation value from future trade with M but not P. An endogenous debt limit satisfies $p_{pc,t} \leq D_{pc,t}$ with equality, which by (17) means

$$D_{pc,t} = R_p \left[V_{c,t}^{C}(0) - V_{c,t}^{D}(0) \right]. \tag{18}$$

where $R_p \equiv \mu_p/(1-\lambda_p)$ captures the combined disincentive to misbehave.

From (1) and (6), $V_{c,t}^{C}(0) - V_{c,t}^{D}(0) = \beta \left[V_{c,t+1}^{C}(0) - V_{c,t+1}^{D}(0) + \alpha_{cp,t+1} S_{cp,t+1} \right]$ where $\alpha_{cp,t+1}$ is C's probability of trading with P, and $S_{cp,t+1}$ is the corresponding surplus C would lose from the punishment. Since both $\alpha_{cp,t}$ and $S_{cp,t}$ depend on $D_{pc,t}$ (18) reduces to the difference equation

$$D_{pc,t} = \Delta \left(D_{pc,t+1} \right) \equiv \beta D_{pc,t+1} + \beta R_p \alpha_{cp,t+1} \left(D_{pc,t+1} \right) S_{cp,t+1} \left(D_{pc,t+1} \right). \tag{19}$$

In words, $D_{pc,t}$ is the most a debtor would pay at t given the path of future debt limits, described recursively in (19). A steady state, or SS, of this system, $D_{pc} = \Delta \left(D_{pc} \right)$, constitutes a SE with an endogenous D_{pc}. Other (nonnegative and bounded) solutions to (19) are dynamic equilibria, or DE, with time-varying debt limits.

To be explicit, $\alpha_{cp,t}$ depends on τ_t and ρ_t, which depend on $D_{pc,t}$ via (15) and (13), while $S_{cp,t}$ depends on $D_{pc,t}$ via (2) and (7). There are various possibilities given the (τ, ρ) that can be consistent with equilibrium. Using the cutoffs from the graphs shown above, we have:

$$\alpha_{cp,t}\left(D_{pc,t}\right)=\begin{cases}0 & \text{if } D_{pc,t}\leq D_{p2}\\ m_R(N_{s,t}/N_c,1)(1-\alpha_{pm}N_p/N_{s,t}) & \text{if } D_{p2}<D_{pc,t}<D_{p3}\\ \bar{\alpha}_{cs}(1-\alpha_{pm}) & \text{if } D_{p3}\leq D_{pc,t}<\bar{D}_{pc}\\ [\bar{\alpha}_{cs}(1-\alpha_{pm}),\bar{\alpha}_{cs}] & \text{if } D_{pc,t}=\bar{D}_{pc}\\ \bar{\alpha}_{cs} & \text{if } D_{pc,t}>\bar{D}_{pc}\end{cases} \tag{20}$$

$$S_{cp,t}\left(D_{pc,t}\right)=\begin{cases}0 & \text{if } D_{pc,t}\leq D_{p2}\\ (1/\theta_{pc}-1)D_{pc,t} & \text{if } D_{p2}<D_{pc,t}<\bar{D}_{pc}\\ u(Q^*_{pc})-\bar{D}_{pc} & \text{if } D_{pc,t}\geq\bar{D}_{pc}\end{cases} \tag{21}$$

where $m_R(\cdot)$ is the RM meeting technology, $\bar{\alpha}_{cs}=m_R(N_p/N_c,1)$ is C's maximum RM trading probability, Q^*_{pc} solves $\bar{\alpha}_{cs}\theta_{pc}u'(Q)=c'(Q)$, and $N_{s,t}$ satisfies P's entry condition $m_R(1,N_c/N_{s,t})D_{pc,t}=c(D_{pc,t}/\theta_{pc})+\kappa$.

The system $D_{pc,t}=\Delta\left(D_{pc,t+1}\right)$ is shown in Fig. 7.[20] Notice $\Delta(\cdot)$ is convex on the interval $[D_{p2},D_{p3}]$, and otherwise piecewise linear. Also, in the linear segment on the interval $(D_{p3},\bar{D}_{pc})$, $\Delta'>1$ if $\Delta(D_{p3})>D_{p3}$ and $\Delta'<1$ if $\Delta(D_{p3})<D_{p3}$. These observations immediately imply the following:

Proposition 2: With $D_{pm}=D_{mc}=\infty$, $D_{pc}=0$ is always a SS. It is unique if $(1-\beta)\bar{D}_{pc}>\beta R_p\bar{\alpha}_{cs}\left[u(Q^*_{pc})-\bar{D}_{pc}\right]$; otherwise, generically there is also a SS with $D_{pc}>\bar{D}_{pc}$ and one with $D_{pc}\leq\bar{D}_{pc}$. If $1-\beta<\beta R_p\bar{\alpha}_{cs}(1-\alpha_{pm})(1/\theta_{pc}-1)$ the middle SS has $D_{pc}<\bar{D}_{pc}$; otherwise, $D_{pc}=\bar{D}_{pc}$.

Fig. 7, with two panels drawn for different R_p, shows regions indicating the Regimes I, B or D (Regime N cannot happen here because $D_{mc}=D_{pm}=\infty$ and κ is assumed sufficiently small so at least M is always willing to go to RM). Both panels have three SS. In the left panel, with R_p low, the middle SS is $\bar{D}_{pc}$. In the right, with R_p higher, the incentive to default is reduced, which raises the high SS and lowers the middle SS so it is to the left of $\bar{D}_{pc}$.

Now consider DE, i.e., nonconstant solutions to the system. From Fig. 7 it is clear that a $D_{pc,t}$ path is a DE iff it is not a SS and starts anywhere between 0 and the high

[20]Notice $\Delta(\cdot)$ is set-valued at $D_{pc,t+1}=\bar{D}_{pc}$ since any $\tau\in[0,1]$ can be supported at this point. We argued in fn. 14 that, when debt limits are exogenous, if M and P are indifferent to trade it makes sense to set $\tau=0$. The situation is different here: we may need $\tau\in(0,1)$ to have equilibrium.

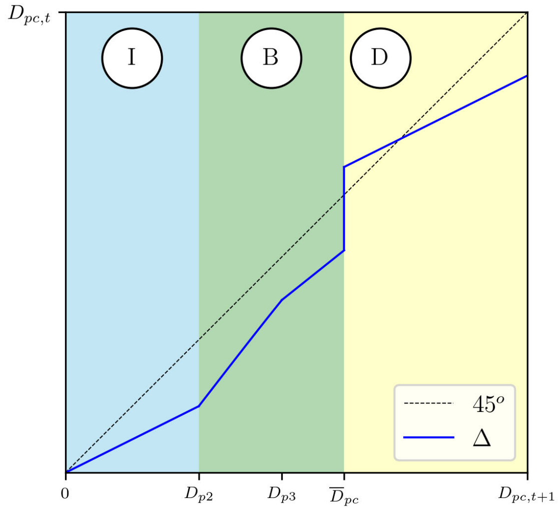

Figure 7: $D_{pc,t} = \Delta\left(D_{pc,t+1}\right)$ with $D_{pm} = D_{mc} = \infty$ and R_p low (left) or high (right).

SS. Clearly all DE paths converge to the middle SS. These paths can entail Regime switching, from either D to B or I to B, but there is at most one switch – when Δ is monotone increasing there are no cycles with recurrent Regime switching (yet).

5.2 Stochastic Cycles

While monotone increasing Δ rules out deterministic fluctuations it does not rule out stochastic fluctuations – i.e., sunspot equilibria. To pursue this, we keep the environment the same but generalize the equilibrium concept by introducing a random process for a sunspot variable s_t, with realizations of s_{t+1} observed by all at t after CM closes. Saying s_t is a sunspot means that it has no impact on fundamentals, but it might affect endogenous variables. For our purposes, it suffices to use a 2-state process with time-invariant transition probabilities $\pi_j = \text{prob}(s_{t+1} = s_{-j} | s_t = s_j)$ and focus on stationary sunspot equilibrium, or SSE.

One can rewrite the equilibrium conditions for the V's, (p, q), etc. as depending on s. The CM problem for C, e.g., when $s = s_1$ is

$$V^C_{i,1}(\Omega) = \max_{x,\ell} \left\{ U(x) - \ell + \beta \left[(1 - \pi_1) V^W_{i,1} + \pi_1 V^W_{i,2} \right] \right\} \text{ st } x = \Omega + \ell.$$

As the subscripts indicate, endogenous variables now depend on the state $s \in \{1, 2\}$ but not the date. Rewriting the other conditions in this way, after some algebra, analogous to reducing SS to $D_{pc} = \Delta\left(D_{pc}\right)$ we can reduce SSE to

$$D_{pc,1} = (1 - \pi_1)\Delta(D_{pc,1}) + \pi_1 \Delta(D_{pc,2}) \tag{22}$$

$$D_{pc,2} = \pi_2 \Delta(D_{pc,1}) + (1 - \pi_2)\Delta(D_{pc,2}). \tag{23}$$

Trivially, a SS solves these equations. We seek other solutions, say, without loss of generality, those with $D_{pc,2} > D_{pc,1}$.

Following textbook methods (Azariadis 1993), we first solve (22)-(23) for the π's former as functions of the D's:

$$\pi_1 = \frac{D_{pc,1} - \Delta(D_{pc,1})}{\Delta(D_{pc,2}) - \Delta(D_{pc,1})} \text{ and } \pi_2 = \frac{\Delta(D_{pc,2}) - D_{pc,2}}{\Delta(D_{pc,2}) - \Delta(D_{pc,1})}. \tag{24}$$

A pair $(D_{pc,1}, D_{pc,2})$ together with the probabilities in (24) constitutes a SSE as long as $\pi_1, \pi_2 \in (0,1)$. Whenever there are three SS, using Fig. 7 it is routine to show $\pi_1, \pi_2 \in (0,1)$ iff $D_{pc,1}$ is between 0 and the middle SS while $D_{pc,2}$ is between the middle SS and the high SS. Hence, there are many equilibria where D_{pc} fluctuates randomly as a self-fulfilling prophecy. This proves:

Proposition 3: Assume Q is divisible, and $D_{pm} = D_{mc} = \infty$ are exogenous while D_{pc} is endogenous. When there are three SS, there exist SSE with D_{pc} fluctuating across any $D_{pc,1}$ between 0 and the middle SS and any $D_{pc,2}$ between the middle and high SS, with transition probabilities given by (24).

Some similar outcomes appear in earlier credit papers, but the logic is very different. Gu et al. (2013), e.g., take this approach: They have a deterministic system $D_t = \Delta\left(D_{t+1}\right)$ with $D\left(0\right) = 0$, like we do, but look at a SS where Δ crosses the 45^o from above, unlike what we do. That is because their model does not, under reasonable conditions, have a SS where Δ crosses from below. An elementary result is that when $\Delta'\left(.\right) < -1$ at SS there exists a 2-cycle, i.e., there exists $D_{pc,1} > 0$ and $D_{pc,2} > D_{pc,1}$ such that $D_{pc,1} = \Delta\left(D_{pc,1}\right)$ and $D_{pc,2} = \Delta\left(D_{pc,1}\right)$. Another standard result is that if a 2-cycle exists there also exist SSE (notice that a 2-cycle is a limiting version of SSE with $\pi_1 = \pi_2 = 1$, then appeal to continuity).

That approach is irrelevant here because $\Delta' > 0$. But whenever we have three SS Δ crosses the 45^o line from below at the middle one, which implies the existence of SSE without cycles. Moreover, in a sense our SSE is more robust: it does not require anything in particular about how the terms of trade are determined, while $\Delta' < 0$ requires some restrictions on price formation.[21]

What is the intuition? First, it is no surprise that there can be multiple SS with endogenous debt limits: if agents believe $D_t = 0 \ \forall t > t_0$ then there is no punishment from reneging at t_0, so they will renege on any debt $\varepsilon > 0$, and that makes $D = 0$ a SS; but if they believe $D_t = \hat{D} > 0 \ \forall t > t_0$, then taking away future credit can dissuade them from reneging at t_0, so there can be a SS at some $\hat{D} > 0$. Now add sunspots. If $D_{t_0} = 0$ but agents believe D will increase at some stochastic $t_1 > t_0$, they may not risk punishment by reneging on small $\varepsilon > 0$, so there is a tendency for D to move up from $D = 0$; and if $D_{t_0} = \hat{D} > 0$ but agents believe D will decrease at some stochastic $t_2 > t_0$, there is a tendency for D to move down from the $D = \hat{D}$. Heuristically, this suggests there can be SSE fluctuating across $D_1 > 0$ and $D_2 \in (D_1, \hat{D})$, although technically that only works if Δ crosses the $45°$ from below.

It is clear that SSE can involve recurrent Regime switches, across I and D, across B and D, or across I and B: or we can always stay in Regime B. Now, to be clear, these fluctuations are not due to the presence of M *per se*, but to the self-referential nature of endogenous debt limits. Yet these fluctuations have implications for intermediation. In this specification, e.g., we can show intermediaries attenuate fluctuations in a precise sense, as we now discuss.

First, in the SSE described above one can check that when $D_{pc,1}$ is low the constraint binds, and when $D_{pc,1}$ is high it may or may not bind. This implies $Q_{pc,1} < Q_{pc,2}$. Also, one can check $Q_{mc,1} > Q_{mc,2}$. Hence, when P brings lower Q_{pc} to RM, due to tighter credit conditions, M brings higher Q_{mc}, acting as a buffer on the intensive margin.

[21] Gu et al. (2013) get $\Delta'(D) < -1$ at SS by having the surplus of borrowers higher when D is lower. In the bargaining version of their setup, they can get that using Nash bargaining when C's power is $\theta < 1$; it does not work with $\theta = 1$, nor with any θ when they use Kalai bargaining. These restrictions are not relevant here – it is no problem using Kalai or Nash with any θ.

Further, M can be a buffer on the extensive margin, when SSE has switching between Regime I and B. If $N_m > 0$ then RM is always open, with only M participating when credit is tight, and both P and M participating when credit is loose. If instead $N_m = 0$ we can get SSE where RM shuts down when credit is tight. So on both margins, middlemen can attenuate fluctuations.[22] However, this is not general: as shown below, in alternative formulations M may amplify fluctuations.

Before moving to other ideas, let us say what happens if the punishment is that defaulters lose all future credit, putting them in autarky (again, perhaps simply excluding them from RM). The outcome is similar except the lowest SS has $D > 0$ not $D = 0$. The reason is that punishment means losing trade with M, not just with P, so C would honor a small current debt to P even if future debt limits with P were 0. Otherwise, the results on SSE hold with minor modification.

The next scenario makes D_{mc} and D_{pc} both endogenous, still with $D_{pm} = \infty$. For this, it is easiest to use the punishment in the preceding paragraph, autarky, which means $V^D_{c,t} = (1-\beta)^{-1}[U(\ell) - \ell]$, although note that this no longer gives the result in the preceding paragraph, that the lowest SS is $D > 0$, because that was for exogenous D_{mc}. Here there is always a SS with $D_{mc} = D_{pc} = 0$. What else is possible? In principle the answer is more complicated with the bivariate system

$$D_{pc,t} = \Delta_p(D_{pc,t+1}, D_{mc,t+1}) \tag{25}$$

$$D_{mc,t} = \Delta_m(D_{pc,t+1}, D_{mc,t+1}). \tag{26}$$

But conveniently, (18) implies the two debt limits are proportional: $D_{mc,t}/D_{pc,t} = R \equiv R_m/R_p$. Thus we can analyze a univariate system for $D_{mc,t}$ then set $D_{pc,t} = D_{mc,t}/R$.

If $R < 1$ then M is never active, so consider $R > 1$. The analog of (19) is

$$D_{mc,t} = \beta D_{mc,t+1} + \beta R_m(\alpha_{cp,t+1}S_{cp,t+1} + \alpha_{cm,t+1}S_{cm,t+1}). \tag{27}$$

Notice C's expected surplus from trading with M and with P appear on the RHS, since

[22] This is what Weill (2007) calls middlemen "leaning against the wind," which may be bad language, since it commonly refers to government stabilization policy. Usage aside, although his model is different from ours the spirit is similar.

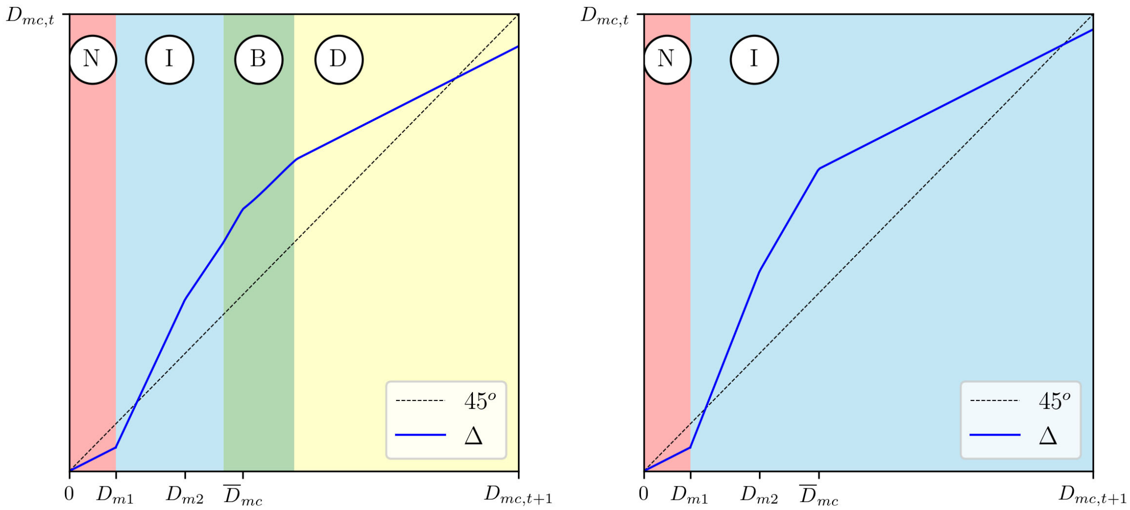

Figure 8: $D_{mc,t} = \Delta\left(D_{mc,t+1}\right)$ with $D_{pm} = \infty$; R low (left) or high (right).

now both are taken away by the punishment. The system is shown in Fig. 8, where the left panel has R close to 1, so M's advantage over P is small, and the right has R bigger. Still, the results on SS are similar to Proposition 2, and the results on SSE are similar to Proposition 3 except now D_{pc} and D_{mc} both fluctuate.

But some implications change. For one, now Regime N is possible. For another, in SSE M's activity is positively correlated with P's: when credit is tight, it is tight for M and P, so both Q's are low. Instead of acting as a buffer, here intermediation exacerbates cycles. So whether intermediation attenuates or amplifies instability depends on details – theory does not settle that debate.

5.3 Deterministic Cycles

In Section 5.2 deterministic cycles cannot exist. For completeness, we now show they can exist with indivisible Q. For this, assume $D_{pm} = \infty$ while both D_{pc} and D_{mc} are endogenous, and again let the punishment for default be autarky. One can check the results on SS are similar to Proposition 2. Moreover, the D's are still proportional, so we can again analyze a univariate system D_{mc}.

The difference from Section 5.2 is that now, with indivisible Q, it is possible to have $\Delta' < 0$, because S_{ci} can decrease with D_{ic}. This happens because higher D_{ic} means C

pays more but does not get more. That does not happen if Q is divisible because then P produces less when D is lower. It is the choice of Q made at the production stage that implies Δ is increasing in Section 5.2.[23]

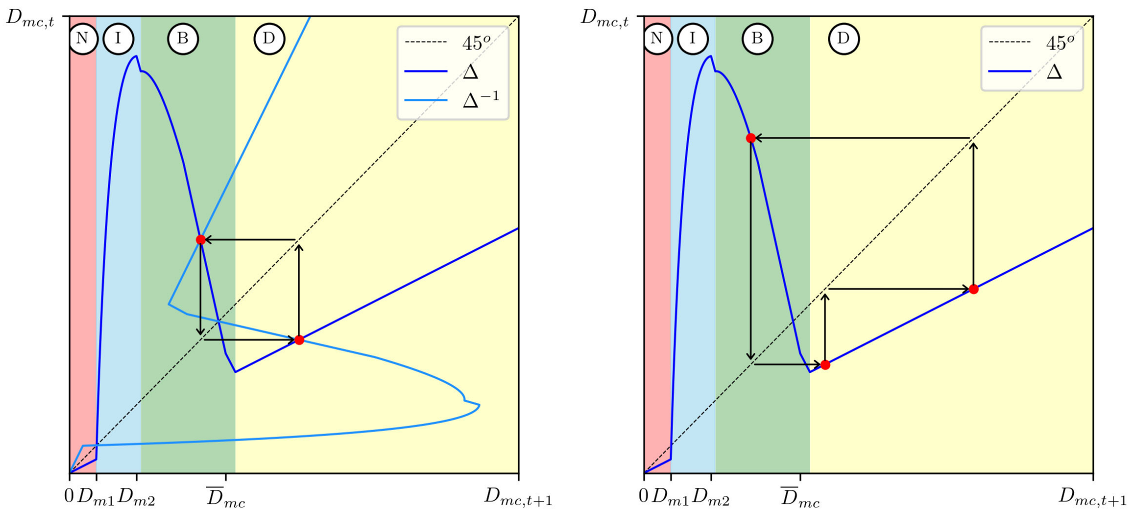

Figure 9: $D_{mc,t} = \Delta\left(D_{mc,t+1}\right)$ with $D_{pm} = \infty$; 2-cycle (left) and 3-cycle (right).

Having explained this, we point out that from Fig. 9 the SSE constructed above still exist with indivisible Q, happening around the middle SS where Δ crosses the 45^o line from below. What is new is that we can find cases with $\Delta' < -1$ at the high SS. Then there are 2-cycles as shown in the left panel, and hence there are SSE. Indeed, we can find parameters giving 3-cycles, fixed points of the third iterate $\Delta^3(\cdot)$, as shown in the right panel. The existence of 3-cycles implies the existence of n-cycles for any integer n, plus chaotic dynamics, by the Sharkovskii and Li-Yorke theorems (again see Azariadis 1993 for a textbook treatment).

Therefore, with indivisible Q the model can generate many dynamic equilibria with deterministic or stochastic Regime switching. However, we still like the version where

[23] In case it is not obvious, we can explain why cycles might emerge when S_{ci} is decreasing in D_{ic}. If D_{ic} is low next period and S_{ci} is decreasing, then S_{ci} is high next period, and that makes C is less inclined to default this period, so this period's endogenous D_{ic} is high. And vice versa if D_{ic} is high next period. Therefore there is a tendency for D_{ic} to oscillate. Note that S_{ci} cannot be globally decreasing in D_{ic}, of course, since $D_{ic} = 0$ implies $S_{ci} = 0$, but it can be decreasing if D_{ic} is close to the just-binding $\bar{D}_{ic}$. Also note that having S_{ci} decreasing in D_{ic} is necessary but not sufficient for $\Delta'(D_{ic}) < 0$ because $\Delta(D_{ic})$ also has an increasing linear term.

Q is divisible, since indivisibility is subject to possible objections. One has to do with the quantity/quality distinction. As mentioned, even if C is only interested in buying 1 car, that does not mean we should fix $Q = 1$ if cars can differ in quality at the production stage. There are also technical issues in models with indivisible goods, as special cases of models with nonconvexities. In particular, agents may have an incentive to trade using lotteries to convexify the feasible set.

By way of example, suppose you want to sell your car, which is fixed in quantity and quality. If a buyer can pay at most p, what is there to bargain over? You can propose to accept p and transfer the car to the buyer with probability δ. Since δ can be anything in $[0, 1]$ the indivisibility effectively vanishes. In any case, one can use the framework with or without indivisibilities.

6 Conclusion

This paper studied environments with trade intermediated by middlemen, and debt limits that were either exogenous or endogenous. The contribution can be seen as putting payment frictions into models of intermediated trade and/or putting intermediaries into models of constrained credit, leading to interesting interactions between two literatures. The theory applied to markets for goods, inputs or assets. We also considered both indivisible and divisible quantity/quality, so we could analyze both intensive and extensive margin effects. With exogenous debt limits we characterized equilibrium in a benchmark specification and several variants, and in each case proved existence and uniqueness. We also discussed how parameters (debt limits, bargaining powers, etc.) determine the pattern of exchange – no trade, only direct trade, only indirect trade, or both direct and indirect trade – which is relevant because in reality exchange patterns vary across markets and across time.

Then we endogenized debt limits by saying repayment must be incentive compatible given punishment for default was the loss of future credit. That generated multiple stationary equilibria, as well as equilibria with fluctuations in credit conditions, intermediation activity, prices, etc. These equilibria can display Regime switching, with recurrent spells where some types of trade diminish or even freeze for a while. These fluctuations, which can be deterministic or stochastic, were self-fulfilling prophecies. The point is not that real-world cycles are best explained as driven exclusively by beliefs; it is that when simple models generate such outcomes, it lends credence to the idea that these may be relevant in actual economies. Our emphasis was not that intermediation causes instability; it was that limited commitment provides an endogenous role for intermediation and an endogenous source of dynamics, resulting in interesting links between middlemen and volatility. We found middlemen can attenuate or amplify fluctuations, depended on details, on both the extensive and intensive margins.

The objective of the project was not to explain one major empirical observation or prove one big theorem. It was to develop a flexible, tractable framework for the analysis of intermediated trade and credit frictions. Future work may put the framework to use studying applications geared to particular kinds of markets, or perhaps designing tax/regulatory schemes to promote desirable outcomes. Potentially fruitful modeling may involve more work integrating middlemen models following Rubinstein and Wolinsky (1987), which generally focus on goods markets, with those in finance following Duffie et al. (2005), which focus on asset markets. An interesting feature of the latter is the way suppliers and demanders are determined by fundamental idiosyncratic shocks. An interesting feature of the former is the way dealers deal with inventories. Combining these might lead to new insights, with perfect or imperfect credit, but we think especially with the latter.

Appendix

Proof of Proposition 1: We derive the borders in Fig. 1 and 2. First, from the RM bargaining problem $p_{ic} = \min\{\theta_{ic}u(Q_{ic}), D_{ic}\}$ for $i \in \{P, M\}$. When $\mathcal{Q} = \{Q\}$, $Q_{ic} = Q$. When $\mathcal{Q} = [0, \infty)$, $Q_{ic} = \min\{u^{-1}(D_{ic}/\theta_{ic}), Q^*_{ic}\}$ where Q^*_{ic} satisfies $\alpha_{sc}\theta_{ic}u'(Q^*_{ic}) = c'(Q^*_{ic})$. Now rewrite the best response conditions as

$$\rho = \begin{cases} 0 & \text{if } \alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa \leq 0 \\ [0,1] & \text{if } \alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa = 0 \\ 1 & \text{if } \alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa \geq 0 \end{cases} \tag{28}$$

$$\tau = \begin{cases} 0 & \text{if } \alpha_{sc}p_{mc} - c(Q_{mc}) - \kappa \leq \rho\left[\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa\right] \\ [0,1] & \text{if } \alpha_{sc}p_{mc} - c(Q_{mc}) - \kappa = \rho\left[\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa\right] \\ 1 & \text{if } \alpha_{sc}p_{mc} - c(Q_{mc}) - \kappa \geq \rho\left[\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa\right] \end{cases} \tag{29}$$

Recall $\alpha_{sc} = m_R(1, N_c/N_s)$ and $N_s = N_p\left[\alpha_{pm}\tau + (1 - \alpha_{pm}\tau)\rho\right]$ where N_p, N_c and α_{pm} are fixed. Hence (28)-(29) are two equations in (τ, ρ), and the solution is an SE. We then derive the borders for Regimes by substituting the corresponding (τ, ρ) into (28)-(29). Let $\alpha_{\tau\rho} \equiv m_R\left(1, N_c/N_p[\alpha_{pm}\tau + (1 - \alpha_{pm}\tau)\rho]\right)$ and let $Q_{ic,\tau\rho}$ be the Q that type i sellers take to RM when the best responses are (τ, ρ). Define the following functions.

$$G(\tau, \rho) = \left[c(Q_{pc,\tau\rho}) + \kappa\right]/\alpha_{\tau\rho} \tag{30}$$

$$H(\tau, \rho) = \left[c(Q_{mc,\tau\rho}) + \kappa\right]/\alpha_{\tau\rho} \tag{31}$$

$$J(D_{pc}) = \left[\alpha_{01}D_{pc} - c(Q_{pc,01}) + c(Q_{mc,01})\right]/\alpha_{01} \tag{32}$$

The cutoffs where the debt limits just bind are $\bar{D}_{pc} = \theta_{pc}u\left(Q_{pc,\tau 1}\right)$ and $\bar{D}_{mc} = \theta_{pc}u\left(Q_{mc,1\rho}\right)$. Other cutoffs shown in the graphs are

$$D_{p1} = G(0,0),\ D_{p2} = G(1,0),\ D_{p3} = G(0,1),\ D_{m1} = H(0,0),\ D_{m2} = H(1,0).$$

We now go through the different Regimes:

Regime N: $\tau = \rho = 0$ implies $\alpha_{sc}p_{mc} - c(Q_{pm}) - \kappa \leq 0$ and $\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa \leq 0$. Hence, Regime N is a SE when $D_{pc} \leq D_{p1}$ and $D_{mc} \leq D_{m1}$.

Regime D: $\tau = 0$ while there are two cases for ρ. In the first case $\rho \in (0,1)$, $\alpha_{sc}p_{mc} - c(Q_{pm}) - \kappa \leq 0 = \alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa$. This is a SE when $D_{p1} < D_{pc} < D_{p3}$ and

$D_{mc} \leq H(0,\rho)$. In the second case $\rho = 1$, $\alpha_{sc}p_{mc} - c(Q_{pm}) - \kappa < \alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa$. This is a SE when $D_{pc} \geq D_{p3}$ and $D_{mc} \leq J(D_{pc})$.

Regime I: $\rho = 0$ implies $\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa \leq 0$. There are two cases for τ. In the case $\tau \in (0,1)$, $\alpha_{sc}p_{mc} - c(Q_{pm}) - \kappa = 0$. This is a SE when $D_{pc} \leq G(\tau, 0)$ and $D_{m1} < D_{mc} < D_{m2}$. In the case $\tau = 1$, $\alpha_{sc}p_{mc} - c(Q_{pm}) - \kappa > 0$. This is a SE when $D_{pc} \leq D_{p2}$ and $D_{mc} \geq D_{m2}$.

Regime B: $\tau = 1$ implies $\alpha_{sc}p_{mc} - c(Q_{pm}) - \kappa \geq \rho\left[\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa\right]$. There are two cases for ρ. In the case $\rho \in (0,1)$, $\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa = 0$. This is a SE when $D_{p2} < D_{pc} < D_{p3}$ and $D_{mc} > H(1,\rho)$. In the case of $\rho = 1$, $\alpha_{sc}p_{pc} - c(Q_{pc}) - \kappa > 0$. This is a SE when $D_{pc} \geq D_{p3}$ and $D_{mc} > J(D_{pc})$.

This partitions parameter space as shown in Fig. 1 and 2. Therefore, for all parameters there is one and only one SE. ■

Details for Section 4.2: The model is slightly different with ex ante production. The CM value function for P is

$$\begin{aligned} V_p^C(\Omega) &= \max_{x,\ell,\phi}\left\{U(x) - \ell + \beta\left[\phi V_{p,+1}^W(Q) + (1-\phi)V_{p,+1}^W(0)\right]\right\} \\ \text{st } x &= \Omega + \ell - \phi c(Q) \end{aligned} \tag{33}$$

where ϕ is the probability of producing. The CM problem for M or C is similar except that there is no production choice, so simply set $\phi = 0$.

The WM value functions are

$$\begin{aligned} V_p^W(Q) &= V_p^C(0) + \alpha_{pm}\tau\left[V_p^C(p_{pm}) - V_p^C(0)\right] \\ &\quad + (1-\alpha_{pm}\tau)\max_{\rho}\left\{\rho\left[V_p^R(Q) - V_p^C(0) - \kappa\right]\right\} \end{aligned} \tag{34}$$

$$V_m^W = V_m^C(0) + \alpha_{mp}\tau\left[V_m^R(Q) - V_m^C(0) - p_{pm} - \kappa\right] \tag{35}$$

The RM value functions, bargaining problems, steady state and best response conditions for τ are same as in the baseline model. The best response conditions for ϕ and ρ are

$$\phi = \begin{cases} 0 & \text{if } \beta\left[V_{p,+1}^W(Q) - V_{p,+1}^W(0)\right] \leq c(Q) \\ [0,1] & \text{if } \beta\left[V_{p,+1}^W(Q) - V_{p,+1}^W(0)\right] = c(Q) \\ 1 & \text{if } \beta\left[V_{p,+1}^W(Q) - V_{p,+1}^W(0)\right] \geq c(Q) \end{cases} \tag{36}$$

$$\rho = \begin{cases} 0 & \text{if } V_p^R(Q) - V_p^C(0) \leq \kappa \\ [0,1] & \text{if } V_p^R(Q) - V_p^C(0) = \kappa \\ 1 & \text{if } V_p^R(Q) - V_p^C(0) \geq \kappa \end{cases} \tag{37}$$

SE is defined by: (V_i^n) for each type i in each market n satisfying (4)-(6) and (33)-(35); (p_{ij}, q_{ij}) satisfying (7)-(9); N_s satisfying (12); and (ϕ, τ, ρ) satisfying (15) and (36)-(37). The analysis follows the same procedure as the baseline model. ■

Parameters for the figures: The matching function is $m_k(N_i, N_j) = a_k N_i N_j/(N_i + N_j)$ with $a_W = 1$ for WM and $a_R = 0.8$ for RM. The utility function is $u(q) = \omega_0 q^{\omega_1}$. The cost function is $c(Q) = \psi_0 Q^{\psi_1}$ with $(\psi_0, \psi_1) = (0.01, 2)$, $\kappa = 0.15$. Bargaining power in WM is $\theta_{pm} = 0.5$ and indivisible $Q = 2$.

For the Figs. in Section 3 and 4, $(\omega_0, \omega_1) = (1, 0.5)$ and $N_p = N_m = N_c = 1$. When RM bargaining powers equal, $\theta_{pc} = \theta_{mc} = 0.5$; otherwise, $(\theta_{pc}, \theta_{mc}) = (0.4, 0.6)$. The figure-specific parameters are as follows. Fig. 3 has $D_{pc} \in \{0.2, 0.5\}$. Fig. 4 has $D_{mc} \in \{0.25, 0.75\}$. Fig. 5 has $D_{pm} \in \{0.02, 0.1\}$. Fig. 6 has $D_{pc} \in \{0.2, 0.5\}$ and $D_{mc} = 0.75$. For the Figs. in Section 5, $(N_p, N_m, N_c) = (3, 3, 1)$ and $\beta = 0.5$. The other parameters are shown in Table 1.

Table 1: Other Fig. Parameters

Figure	R_p	R_m	(ω_0, ω_1)	θ_{ic}
7	{2, 4}	-	(2, 0.2)	0.5
8	{3, 1}	$\{4, 5\}$	(2, 0.1)	0.5
9	16	17	(1, 0.5)	0.98

References

[1] F. Alvarez and U. Jermann (2000) "Efficiency, equilibrium, and asset pricing with risk of default," *Econometrica* 68, 775-97.

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

[3] C. Azariadis (1993) *Intertemporal Macroeconomics*.

[4] C. Azariadis and L. Kass (2007) "Is dynamic general equilibrium a theory of everything?" *Econ Theory* 32, 13-41.

[5] C. Azariadis and L. Kass (2013) "Endogenous credit limits with small default costs," *JET* 148, 806-24.

[6] B. Biais, T. Mariotti, G. Plantin and J. Rochet (2007) "Dynamic security design: Convergence to continuous time and asset pricing implications," *RES* 74 , 345-90.

[7] G. Biglaiser (1993) "Middlemen as experts," *RAND* 24, 212-23.

[8] K. Burdett and K. Judd (1983) "Equilibrium price dispersion," *Econometrica* 51, 955-69.

[9] F. Carapella and S. Williamson (2015) "Credit markets, limited commitment, and government debt," *RES* 82, 963-90.

[10] S. Charles (2024) "Why roasters and producers cutting out the middleman," *Coffee Intelligence Newsletter*.

[11] P. DeMarzo and M. Fishman (2007) "Agency and optimal investment dynamics," *RFS* 20, 151-88.

[12] D. Diamond and P. Dybvig (1983) "Bank runs, deposit insurance, and liquidity," *JPE* 91, 401-19.

[13] P. Diamond (1982) "Aggregate demand management in search equilibrium," *JPE* 90, 881-94.

[14] P. Diamond and D. Fudenberg (1989) "Rational expectations business cycles in search equilibrium," *JPE* 97, 606-19.

[15] D. Duffie, N. Garleanu and L. Pederson (2005) "Over-the-counter markets," *Econometrica* 73, 1815-47.

[16] R. Frankel (2024) "History of credit cards: When were credit cards invented," https://www.forbes.com/advisor/credit-cards/history-of-credit-cards/.

[17] X. Gong and R. Wright (2024) "Middlemen in search equilibrium with intensive and extensive margins," *IER* 65, 1657-79.

[18] X. Gong, Z. Qiao and R. Wright (2025) "Middlemen redux," mimeo.

[19] C. Gu, L. Wang and R. Wright (2025) "Middlemen, inventories and economic dynamics," mimeo.

[20] C. Gu, F. Mattesini, C. Monnet and R. Wright (2013a) "Endogenous credit cycles," *JPE* 121, 803-1005.

[21] C. Gu, F. Mattesini, C. Monnet and R. Wright (2013b) "Banking: A new monetarist approach," *RES* 80, 636-662.

[22] C. Gu, F. Mattesini and R. Wright (2016) "Money and credit redux," *Econometrica* 84, 1-32.

[23] C. Gu, C. Monnet, E. Nosal and R. Wright (2023) "Diamond-Dybvig and beyond: On the instability of banking," *EER* 154.

[24] H. Han, B. Hu and M. Watanabe (2024) "From cash to buy-now-pay-later: Impacts of platform-provided credit on market efficiency," mimeo.

[25] C. Hellwig and G. Lorenzoni (2009) "Bubbles and self-enforcing debt," *Econometrica* 77, 1137-64.

[26] B. Hu, M. Watanabe and J. Zhang (2025) "A model of supplier finance," mimeo.

[27] A. Huang (2015) "Smaller or larger banks?" mimeo.

[28] J. Hugonnier, B. Lester and P. Weill (2025) *The Economics of Over-the-Counter Markets: A Toolkit for the Analysis of Decentralized Exchange.*

[29] T. Kehoe and D. Levine (1993) "Debt-constrained asset markets," *RES* 60, 865-88.

[30] N. Kiyotaki and J. Moore (1997) "Credit cycles," *JPE* 105, 211-48.

[31] N. Kocherlakota (1998) "Money is memory" *JET* 81, 863-904.

[32] R. Lagos and G. Rocheteau (2006) "Search in asset markets," *AER* 97, 198-202.

[33] R. Lagos, G. Rocheteau and R. Wright (2017) "Liquidity: A new monetarist perspective, " *JEL* 55, 371-440.

[34] R. Lagos and R. Wright (2005) "A unified framework for monetary theory and policy analysis," *JPE* 113, 463-84.

[35] F. Li , C. Murry, C. Tian, and Y. Zhou (2025), "Intermediaries in decentralized markets: Evidence from used-car transactions," mimeo.

[36] Y. Li (1998) "Middlemen and private information," *JME* 42, 131-59.

[37] G. Lorenzoni (2008) "Inefficient credit booms," *RES* 75, 809-833.

[38] R. Lucas and E. Prescott (1974) "Equilibrium search and unemployment" *JET* 7, 188-209.

[39] Marketing for Managers (2024) "The critical role of middlemen in enhancing market dynamics."

[40] J. Martel, E. Van Wesep and B. Waters (2023) "Constraints in decentralized markets," mimeo.

[41] A. Masters (2008) "Unpleasant middlemen," *JEBO* 68, 73-86.

[42] B. Mojon, H. Qiu, A. Schrimpf, H. Shin, N. Tarashev and K. Todorov (2023) "Buy now, pay later: A cross-country analysis." *BIS Quarterly Rev* 61.

[43] D. Mortensen (1999) "Equilibrium unemployment dynamics," *IER* 40, 889-914.

[44] R. Myerson (2012) "A model of moral-hazard credit cycles," *JPE* 120, 847-878.

[45] E. Nosal, Y. Wong and R. Wright (2015) "More on middlemen: Equilibrium entry and efficiency in markets with intermediation," *JMCB* 47, 7-37.

[46] C. Pissarides (2000) *Equilibrium Unemployment Theory.*

[47] G. Rocheteau and E. Nosal (2017) *Money, Payments, and Liquidity.*

[48] G. Rocheteau and R. Wright (2005) "Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium," *Econometrica* 73, 175-202.

[49] A. Rubinstein and A. Wolinsky (1987) "Middlemen," *QJE* 102, 581-94.

[50] D. Sanches and S. Williamson (2010) "Money and credit with limited commitment and theft," *JET* 145, 1525-49.

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

[52] D. Spulber (1999) *Market Microstructure: Intermediaries and the Theory of the Firm.*

[53] J. Stavins (2024) "Buy now, pay later: Who uses it and why," FRB Boston *Current Policy Perspectives* 24-3.

[54] A. Trejos and R. Wright (2016) "Search-based models of money and finance: An integrated approach,"*JET* 164, 10-31.

[55] K. Tretina and K. Little (2004) "When were credit cards invented?" https://finance.yahoo.com/personal-finance/credit-cards/article/when-were-credit-cards-invented-202145703.html.

[56] M. Urias (2018) "An integrated theory of intermediation and payments," mimeo.

[57] M. Watanabe (2010) "A model of merchants," *JET* 145, 1865-89.

[58] P. Weill (2007) "Leaning against the wind," *RES* 74, 1329-54.

[59] P. Weill (2008) "Liquidity premia in dynamic bargaining markets," *JET* 140, 66-96.

[60] T. Wong (2016) "A tractable monetary model under general preferences," *RES* 83, 402-20.

[61] R. Wright and Y. Wong (2014) "Buyers, sellers and middlemen: Variations on search-theoretic themes," *IER* 55, 375-97.