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

October 14, 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[1] 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.

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 Li 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/qiao-ziqi/middlemen, but here is a sample. In Rubinstein and Wolinsky (1987) middlemen are faster than producers at meeting customers; in Biglaiser (1993) and Li (1998) they have expertise in discerning quality; 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. Urias (2018) has an environment similar to ours, but with crucial differences in assumptions (producers are not allowed to trade directly with consumers) and applications (there is no credit, just cash). Also, Gu et al. (2025) like us study dynamics with middlemen, and cite related work, but that has nothing to do with credit. Another related paper, Hu et al. (2025), is discussed in Section 4.

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. We also note that no less an authority than Diamond (2023) argues that a critical role for intermediaries is the facilitation of credit, although he does not provide a model to analyze the idea rigorously.[3]

Our framework builds on recent work in the spirit of Rubinstein-Wolinsky by Gong and Wright (2024). That paper is not about credit, but the environment is natural for introducing 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, our framework replaces the usual three-sided market, where producers, consumers and middlemen all

[2] On imperfect credit, there are too many papers with exogenous debt limits, but on more-or-less endogenous limits some examples are 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 (with a different focus) Kiyotaki and Moore (1997). These do not have middlemen like those cited in fn. 1, although there is some work with banks, a kind of intermediary (see the survey by Gu et al. 2023).

[3] Diamond (2023) says "There are many conflicts of interest in finance, but the most important one involves getting borrowers to repay investors... How do we provide incentives to overcome these preexisting conflicts of interest? One way, monitoring, requires acquiring detailed information about the borrower's activities and requires decision making by investors... The other option to resolve the conflict is to write a contract that imposes a penalty on the borrower who does not repay." Again, he does not provide an explicit model to analyze this rigorously; we endeavor to do so.

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 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. 11), 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 nonstationary 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.[4] To be clear, the point we emphasize is *not* that intermediation *causes* instability; it is that limited commitment gives an endogenous role for intermediation and an endogenous source of dynamics, suggesting a more subtle but no less interesting link between intermediaries and volatility. An implication is that volatility in intermediated markets need not be eliminated simply by limiting intermediary activity.

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.[5]

[4] 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).

[5] It is well known that search models with increasing returns have multiplicity and belief-based dynamics (e.g. Diamond 1982; Diamond and Fudenberg 1989; Mortensen 1999). That is *not* relevant 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.

As more motivation, there is a growing interest in BNPL (buy now pay later) plans by academics and regulators.[6] 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 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.[7]

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 to actually produce the object being traded, and in some contexts it is better to interpret it as an endowment.

[6] See Han et al. (2024), Dong 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).

[7] In addition to work following Rubinstein and Wolinsky (1987), there are papers 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, but simply reallocate assets across investors via a frictionless interdealer market (Weill 2008 and Yang and Zeng 2021, e.g., are exceptions but they do not have endogenous dynamics; see Trejos and Wright 2016). Also those papers generally assume transferable utility (but see Martel et al. 2023).

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

In CM all agents have utility $U(x) - \ell$ with $U'(\cdot) > 0 > U''(\cdot)$ and discount the next CM by $\beta \in (0, 1)$. As usual, quasi-linear utility implies everyone starts next period with a clean slate (history independence), and we can, without loss of generality, use short-term debt cleared each CM. In RM, C gets $u(q)$ from 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 opt to go to RM seeking direct trade. There are two versions: Q can be indivisible or divisible (i.e., fixed or variable) at the production stage, but it is divisible in RM trade. Production costs $c(Q)$, in utils, with $c'(\cdot) > 0 \leq c''(\cdot)$, $c(0) = 0$ and $u'(0)/c'(0) = \infty$. Besides production costs, sellers may need to pay an entry cost κ to go to RM.

As in some related work, 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, but as a benchmark $A = 0$, and in any case $A \leq c'(0)$, so P will not produce just for scrap value. Also, P can only produce once per period, so trading in WM means skipping the RM, to capture 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, 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 we emphasize that having two two-sided markets is what allows the use of general meeting technologies, while related papers with three-sided markets only consider the 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 $\theta_{ji} = 1 - \theta_{ij}$. 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.[8]

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 credibly promise future payment – or sellers – e.g., their ability to collect debt, say by punishing renegers. In addition to p's and q's, endogenous choices are (allowing mixed strategies) τ, is the probability P and M trade in WM meetings, and ρ, the probability P goes to RM when not trading in WM.

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).

[8] Here is the intuition. First, the result says $q = Q$ on, but not necessarily off, the equilibrium path. If sellers in RM had a very large Q, they may sell $q < Q$, but then in equilibrium they do not bring such a large Q to RM. An analogy to monetary theory may be useful. In Lagos and Wright (2005), 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.

To get to the definition of equilibrium, first, 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

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

where ℓ is labor income since 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.[9]

Moving to WM, for P,

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

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 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. Notice in (3) that the term in brackets looks like immediate settlement due to the

[9] It also implies one-period debt is without loss of generality: agents are happy to settle up in CM. These result follows easily from quasi-linear CM utility, but also holds for any $\tilde{U}(x, 1-\ell)$ with $\tilde{U}_{11}\tilde{U}_{22} = \tilde{U}_{12}^2$ (Wong 2016). Moreover, since seminar participants asked about this, while here agents settle debt using income from ℓ in the CM it is equivalent to instead give them an endowment of x.

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.

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 $q_{ic} = Q_{ic}$ if there is no scrap value.

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,

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

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.[10] 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:

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

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:

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

[10] We mean static in the same sense that, e.g., vacancies are static in Pissarides (2000), although one may prefer to call them jump variables. There are genuine state variables below, the D's, but for now they are exogenous and constant over time so we need not keep track of them.

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 nonstationary equilibria, but that changes below when the D's are endogenous. So we use the SE label here, but when we say, e.g., there is a unique stationary equilibrium one can read that as meaning 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.[11]

[11] It is useful to identify factors leading to alternative Regimes, since empirically different markets have different degrees of intermediation. While many consumer goods are bought from middlemen, like grocery stores, there are still farmers' markets. Inputs are often bought from intermediaries, although high-end purveyors of coffee, chocolate and tea these days are buying direct from sources (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). We also mention middlemen chains – e.g., farmer to broker to distributor to retailer to consumer (see Wright and Wong 2014). While our formalization may be too stylized to capture every detail of these diverse markets, it provides guidance as to what factors may be relevant.

Using the algorithm we can construct the SE set. The Appendix proves:[12]

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.*

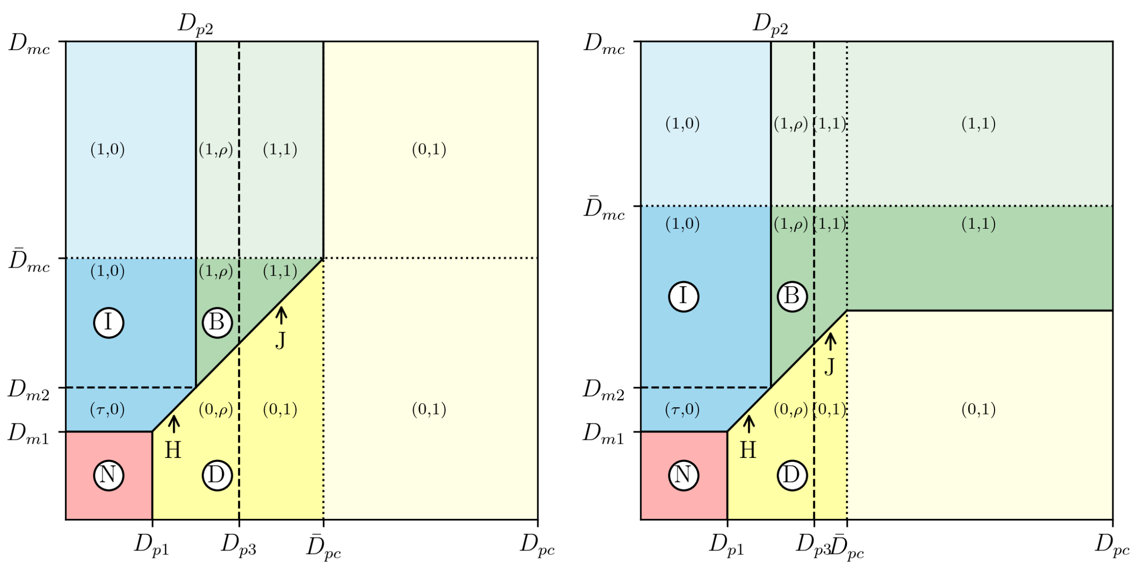

Figure 1: SE with Q indivisible and $D_{pm} = \infty$; $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

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

[12] Here we ignore outcomes with $\tau \in (0, 1)$ if they are nongeneric – e.g., if $D_{mc} = D_{pc}$ and $\theta_{mc} = \theta_{pc}$, then when P meets M in WM it is a matter of indifference who goes to RM, so any $\tau \in [0, 1]$ is a best response, but that is uninteresting and does not survive obvious refinements.

debt limits may or may not bind. So *we know exactly what happens for all parameters.*

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 we make it precise, is this: *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.

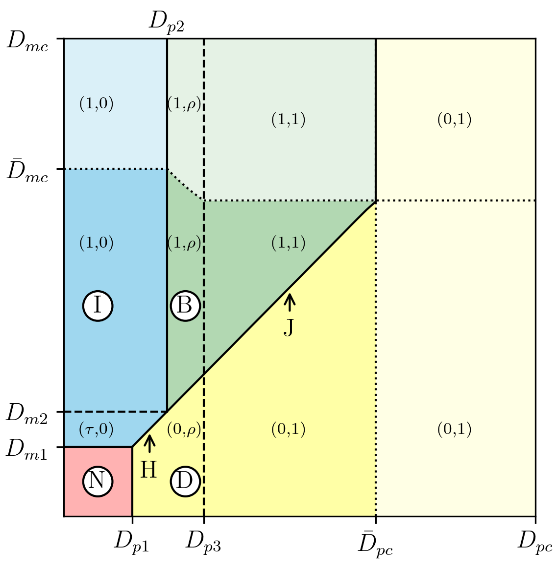

Figure 2: SE with Q divisible and $D_{pm} = \infty$; $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

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

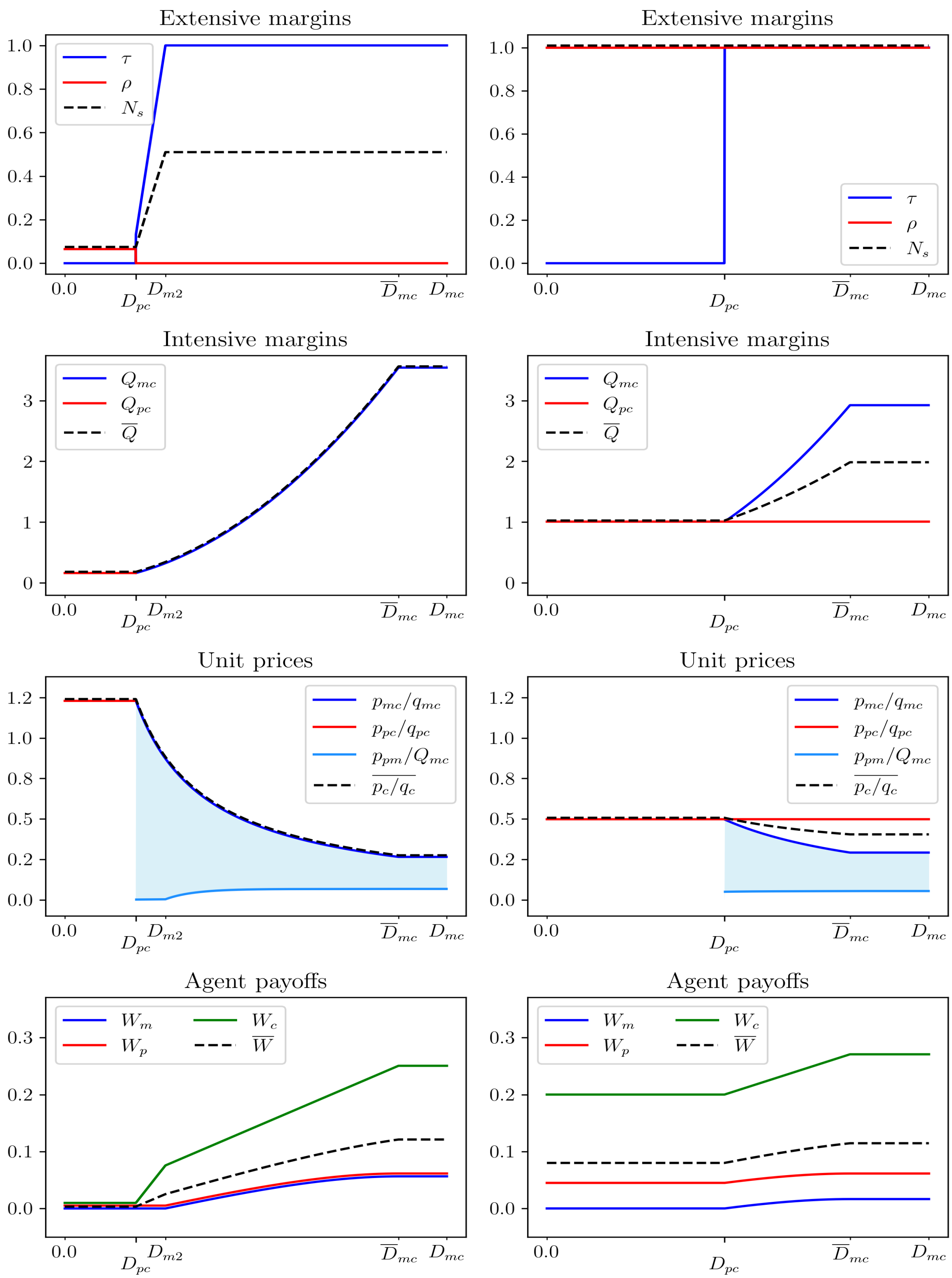

Figure 3: Effects of D_{mc} with Q divisible, $D_{pm} = \infty$ and D_{pc} low (left) or high (right).

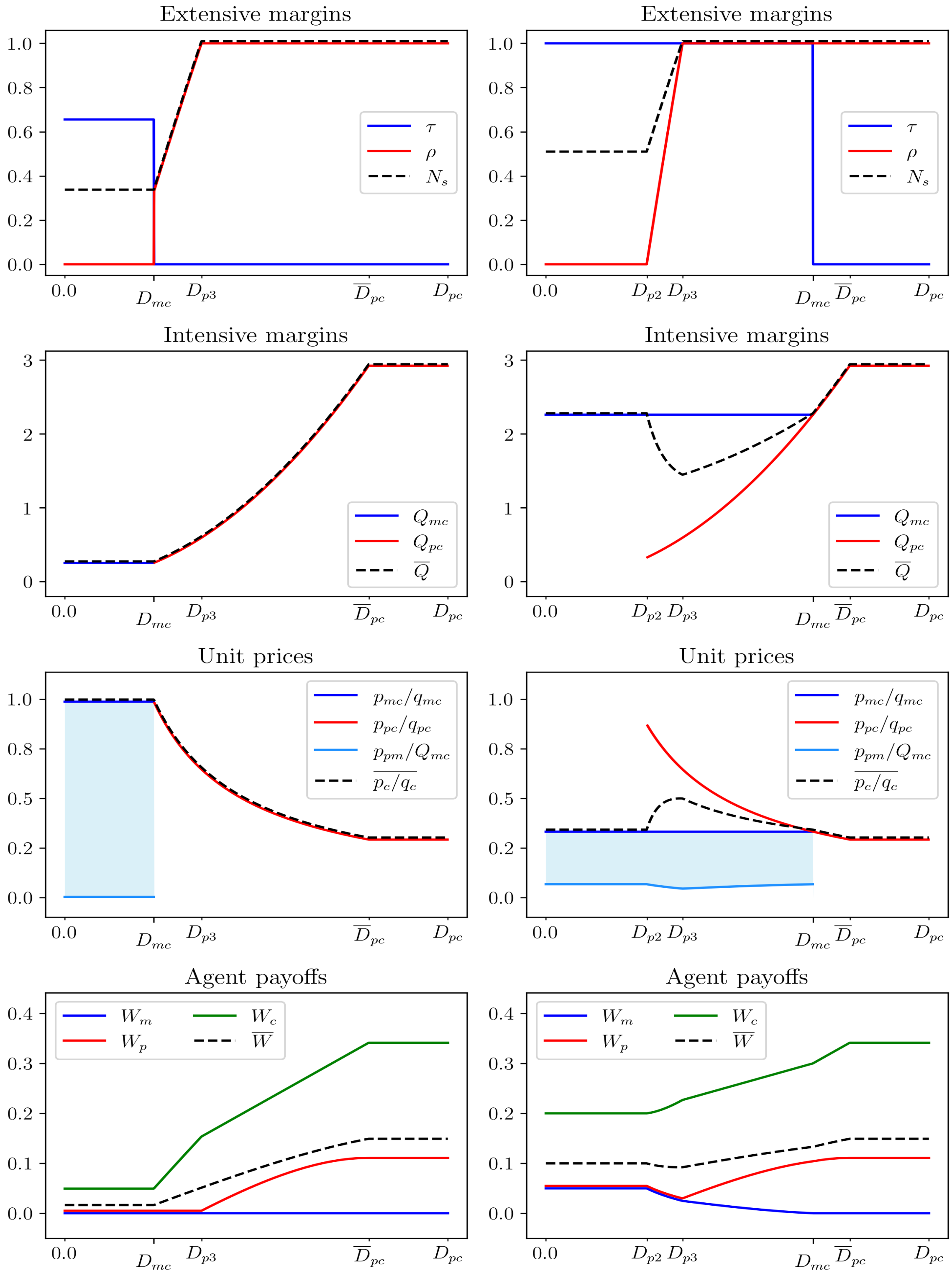

Figure 4: Effects of D_{pc} with Q divisible, $D_{pm} = \infty$ and D_{mc} low (left) or high (right).

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 less obvious one is *in intermediated markets average price and price dispersion can be nonmonotone in debt limits.*[13]

In Fig. 3 payoffs increase with D_{mc}, due to effects on both the extensive and intensive margins; in the right panel of Fig. 4, higher D_{pc} reduces payoffs for P and M over some range, but raises the payoff for C. In other examples, higher D's can also lower C's payoff (e.g., with Q indivisible raising D's means C pays more but does not get more). This can be summarized by saying *lower payment frictions can increase or decrease welfare.* While important, we do not dwell this because there are already many discussions of welfare in the middleman literature and it is known that the results depend on details (see Gong and Wright 2024 and references therein).

4 Extensions

To check robustness we now consider some extensions and alternative specifications, reporting in each case how the result change. However, although we keep this brief, readers more interested in endogenous debt limits and dynamics can skip to Section 5 without loss of continuity.

[13] As an aside, when we reduce search frictions by raising the constant in the meeting technology, average price and price dispersion can go up or can go down. This is relevant since some people are puzzled about average price and price dispersion not falling in the data with improvements in technology capturing higher search efficiency (see the discussion and references in Gong et al. 2025).

For the first extension, recall $D_{pm} = \infty$ is assumed above. Suppose $D_{pm} < \infty$, so M does not have unlimited credit or deep pockets in WM. Then τ is determined by (14), where both P and M need a positive surplus for trade, but if $S_{pm} + S_{mp} > 0$ a binding D_{pm} will not make $S_{mp} < 0$, and so we only need check $S_{pm} \geq 0$. One can check that Proposition 1 still holds: SE exists uniquely for indivisible or divisible Q. The details are slightly different, however: when $D_{pm} < c(Q)$, WM payments by M cannot cover P's production cost, so only Regimes N and D are possible; and when $D_{pm} > c(Q)$ M can cover P's production cost but maybe not the opportunity cost of skipping RM, so all Regimes are possible. Either way, imperfect WM credit is still tractable and does not change the key results.

Next suppose production occurs ex ante, before WM convenes. Let ϕ be the probability P produces, to be determined along with τ and ρ. The definition of SE changes just slightly (see the Appendix), but now there is a WM holdup problem 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 may produce but neither meet M in WM nor meet C in RM. Still SE exists uniquely. For indivisible Q, the results are shown in Fig. 5. Since each element of (ϕ, τ, ρ) can be 0, 1 or mixed, there are more cases, but the main economic insights are similar.

Another extension is to add scrap value $A > 0$. This increases the incentive for sellers to enter RM by reducing the downside risk of not meeting C. Also, when meeting C, it affects the bargaining outcome via sellers threat points. With Q divisible, $A > 0$ actually does not qualitatively change the results; with Q indivisible in WM and divisible in RM, it can matter, and in particular we cannot 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. Despite this detail, the main results are similar to Section 3.

Next, note that as in many related papers q is naturally interpreted as a consumption good and $u\,(q)$ as C's utility function. But we can also say q is an input and $u\,(q)$

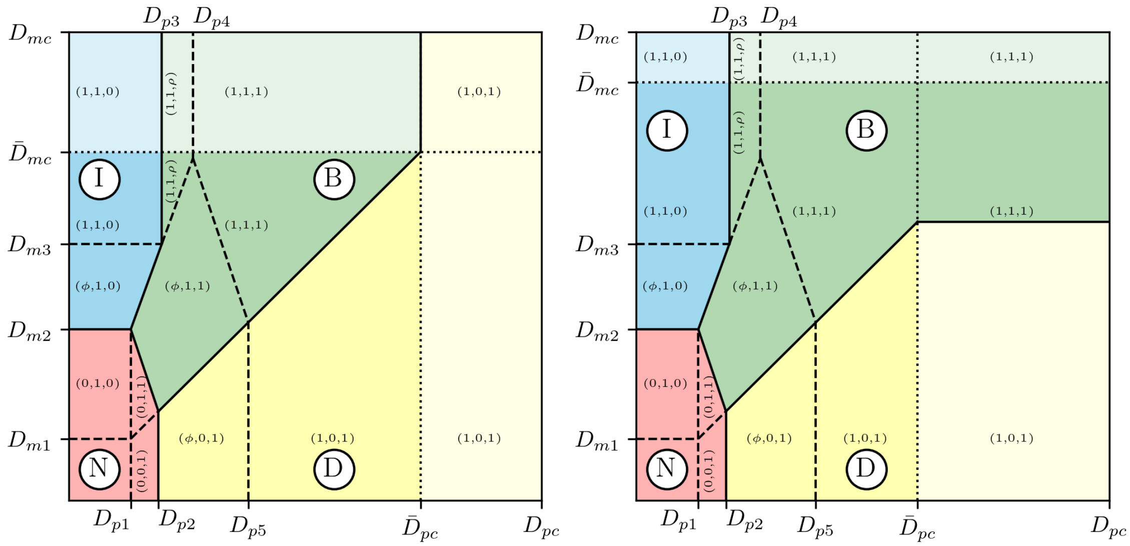

Figure 5: SE with ex ante prod. and $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

is a production function with output in CM numeraire, or q is an asset with $u(q)$ as C's return. These reinterpretations do not change the equations, but are interesting in light of recent work by Hu et al. (2025) on "supplier finance." When interpreting C as a producer or investor, rather than a consumer, the idea is that they need inputs or assets in RM and can get them directly from P or indirectly from M, with the outcome depending on parameters including debt limits as in the benchmark model. Hu et al. (2025) discuss why this is interesting, and while our model is different, it can be adapted to capture their idea.

As a final extension, in the spirit of Kiyotaki and Moore (1997), consider secured credit with repos (i.e., repossessions, not repurchase agreements). Suppose C wants a car that can potentially be obtained in two ways: buy it from P, where here we interpret P not as a producer but someone with a car for sale; or buy it from a dealer M. The idea is that if C fails to make required CM payments the seller can try to repo the car, where the probability type j's repo succeeds is χ_j. It is natural to let used-car dealers be relatively good at this, $\chi_m > \chi_p$.

Let us revert to $D_{pm} = \infty$ and $A = 0$. Also, assume 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 $u(Q_{ic}) - p_{ic} \geq (1-\chi_i)u(Q_{ic})$. In other words, the option to default generates an endogenous constraint $p_{ic} \leq D_{ic} \equiv \chi_i u(Q_{ic})$.

From (7) D_{ic} binds iff $\chi_i \leq \theta_{ic}$. With indivisible Q, this is the same as the baseline model with $D_{ic} = \chi_i u(Q)$. With divisible Q, $\chi_i > \theta_{ic}$ again implies equilibrium is the same, but $\chi_i \leq \theta_{ic}$ implies $p_{ic} = \chi_i u(Q_{ic})$ where Q_{ic} solves $\alpha_{sc}\chi_i u'(Q_{ic}) = c'(Q_{ic})$. So the results are similar to Section 3, although note that here reductions in credit frictions (higher χ) raise p and p/Q since χ is effectively seller bargaining power. While this extension does not affect the results too much, one may say it is nice since it endogenizes D. We go much further in that direction in Section 5.

5 Intermediation Dynamics

5.1 Endogenous Debt Limits

Suppose buyers can renege on payments, but risk being punished by taking away future credit, à la Kehoe-Levine (1993). That makes the model inherently dynamic. Since there can be two types of RM sellers, P and M, there are two types of future credit this mechanism can take away from C, and we discuss both below. To keep things manageable, let P and M have the same RM bargaining power and let Q be divisible (but see below).

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 $D_{pc,t}$, determined as follows. In the CM at t, C can renege on $p_{pc,t}$ owed to P at proportional cost $\lambda_p p_{pc,t}$, with $\lambda_p \in [0,1]$. Moreover, renegers are only caught and hence only punished with probability $\mu_p \in [0,1]$. Both λ and μ measure disincentives to misbehave, and while they are not crucial – we could set $\lambda = 0$ or $\mu = 1$ – they are included because they have interesting implications in related work.[14]

[14] In terms of λ, it captures resources used up by opportunistic behavior as in the "cash diversion" models of DeMarzo and Fishman (2007) or Biais et al. (2007). As for μ, in banking theory is is used by Gu et al. (2013b) and Huang (2015) to discuss who should be a banker and how many we should have;

When C is caught reneging on P, assume for now they lose future credit with P sellers but not M sellers. One rationale is that while taking away all future credit 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 C will do it again, since no further punishment is available. Later we consider taking away all future credit, effectively putting defaulters in autarky, perhaps simply by excluding them from RM.

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

$$V^C_{c,t}(0) - p_{pc,t} \geq -\lambda_p\, p_{pc,t} + \mu_p V^D_{c,t}(0) + (1-\mu_p) V^C_{c,t}(0), \tag{16}$$

where $V^D_{c,t}$ 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 (16) means

$$D_{pc,t} = R_p \left[V^C_{c,t}(0) - V^D_{c,t}(0) \right], \tag{17}$$

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

From (1) and (6), $V^C_{c,t}(0) - V^D_{c,t}(0) = \beta \left[V^C_{c,t+1}(0) - V^D_{c,t+1}(0) + \alpha_{cp,t+1} S_{cp,t+1} \right]$ where $\alpha_{cp,t+1}$ is C's probability of trading with P, and $S_{cp,t+1}$ is the corresponding surplus C loses from the punishment. Since both $\alpha_{cp,t}$ and $S_{cp,t}$ depend on $D_{pc,t}$ (17) reduces to the difference equation

$$D_{pc,t} = \Delta\left(D_{pc,t+1}\right) \equiv \beta D_{pc,t+1} + \beta R_p \alpha_{cp,t+1}\left(D_{pc,t+1}\right) S_{cp,t+1}\left(D_{pc,t+1}\right). \tag{18}$$

In words, $D_{pc,t}$ is the most a debtor would pay at t given the path of future debt limits described recursively in (18). 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 (18) we call dynamic equilibria, or DE, with time-varying debt limits.

---

while in monetary theory, Kocherlakota (1998) shows that fiat currency is never welfare enhancing 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.

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 obtain in equilibrium. Using the cutoffs from the above graphs we have:

$$\alpha_{cp,t}\left(D_{pc,t}\right) = \begin{cases} 0 & \text{if } D_{pc,t} \leq D_{p2} \\ m_R(N_{s,t}/N_c, 1)(1 - \alpha_{pm} N_p/N_{s,t}) & \text{if } D_{p2} < D_{pc,t} < D_{p3} \\ \bar{\alpha}_{cs}(1 - \alpha_{pm}) & \text{if } D_{p3} \leq D_{pc,t} < \bar{D}_{pc} \\ [\bar{\alpha}_{cs}(1 - \alpha_{pm}), \bar{\alpha}_{cs}] & \text{if } D_{pc,t} = \bar{D}_{pc} \\ \bar{\alpha}_{cs} & \text{if } D_{pc,t} > \bar{D}_{pc} \end{cases} \tag{19}$$

$$S_{cp,t}\left(D_{pc,t}\right) = \begin{cases} 0 & \text{if } D_{pc,t} \leq D_{p2} \\ (1/\theta_{pc} - 1) D_{pc,t} & \text{if } D_{p2} < D_{pc,t} < \bar{D}_{pc} \\ u(Q^*_{pc}) - \bar{D}_{pc} & \text{if } D_{pc,t} \geq \bar{D}_{pc} \end{cases} \tag{20}$$

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. 6.[15] 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 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. 6, with two panels drawn for different R_p, shows regions indicating the Regimes I, B or D (Regime N cannot happen here since $D_{mc} = D_{pm} = \infty$ and we assume κ is not too big). Both panels have three SS. In the left panel, with R_p low, the middle SS is $\bar{D}_{pc}$. In the right panel, 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 $\bar{D}_{pc}$.

[15] 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. 12 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: now we may need $\tau \in (0, 1)$ for equilibrium to exist.

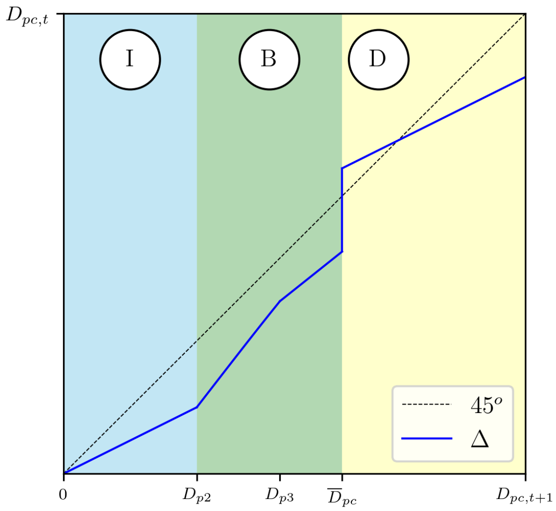

Figure 6: $D_{pc,t} = \Delta\left(D_{pc,t+1}\right)$ with $D_{pm} = D_{mc} = \infty$ and R_p low (left) or high (right).

Now consider DE, i.e., nonconstant solutions to the system. From Fig. 6 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 SS. All such 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 – in this setup, with Δ monotone, there are no cycles and a fortiori there are no cycle with recurrent Regime switching.

5.2 Stochastic Cycles

Monotone Δ rules out deterministic but not stochastic fluctuations, i.e., not 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 after CM closes. Saying s_t is a sunspot means it has no impact on fundamentals, but it may still affect endogenous variables.

It suffices here 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 equilibria, 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 $D_{pc} = \Delta(D_{pc})$ we reduce SSE to

$$D_{pc,1} = (1-\pi_1)\Delta(D_{pc,1}) + \pi_1\Delta(D_{pc,2}) \tag{21}$$

$$D_{pc,2} = \pi_2\Delta(D_{pc,1}) + (1-\pi_2)\Delta(D_{pc,2}). \tag{22}$$

Trivially, a SS solves these equations. Are there other solutions with, say, $D_{pc,2} > D_{pc,1}$?

Following textbook methods (Azariadis 1993), first solve (21)-(22) for the π's as functions of the D's:

$$\pi_1 = \frac{D_{pc,1} - \Delta(D_{pc,1})}{\Delta(D_{pc,2}) - \Delta(D_{pc,1})} \text{ and } \pi_2 = \frac{\Delta(D_{pc,2}) - D_{pc,2}}{\Delta(D_{pc,2}) - \Delta(D_{pc,1})}. \tag{23}$$

A pair $(D_{pc,1}, D_{pc,2})$ together with the probabilities in (23) constitutes a SSE as long as $\pi_1, \pi_2 \in (0,1)$. Whenever there are three SS, it is routine using Fig. 6 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.

Proposition 3: Assume Q is divisible, $D_{pm} = D_{mc} = \infty$ and 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 (23).

Similar outcomes appear in other credit models but the logic is different. Gu et al. (2013), e.g., take this approach: They have a deterministic system $D_t = \Delta(D_{t+1})$ with $D(0) = 0$, 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 have a SS where Δ crosses from below. An elementary result is that when $\Delta'(.) < -1$ at SS there exists a 2-cycle, i.e., $D_{pc,1} > 0$ and $D_{pc,2} > D_{pc,1}$ such that $D_{pc,1} = \Delta(D_{pc,1})$ and $D_{pc,2} = \Delta(D_{pc,1})$. Another standard result is that if a 2-cycle exists there also exist SSE (notice a 2-cycle is a limiting version of SSE with $\pi_1 = \pi_2 = 1$, then use 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 restrictions on price formation.[16]

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 a 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 are SSE fluctuating across $D_1 > 0$ and $D_2 \in (D_1, \hat{D})$; rigorously, that works iff Δ crosses the 45° from below.

Clearly SSE can involve recurrent Regime switches, across I and D, across B and D, or across I and B; or we could stay in Regime B. 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, we can show intermediaries attenuate fluctuations in a precise sense.

To verify this, 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. Further, M can be a buffer on the extensive margin, when SSE

[16] 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 for $\theta = 1$, nor for any θ under Kalai bargaining. This is not relevant here – we can use Kalai or Nash with any θ.

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. With no type M agents we can get SSE where RM shuts down when credit is tight. So on both margins, middlemen attenuate fluctuations – as Weill (2007) puts it, they are "leaning against the wind" – but this is not general, since as shown below in alternative formulations M might amplify fluctuations.

Before moving to other ideas, we can say what happens if the punishment is that defaulters lose all future credit, putting them in autarky (again, perhaps 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 hold with minor modification.

Next, consider making 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.[17] This means $V^D_{c,t} = (1-\beta)^{-1}[U(\ell) - \ell]$. Now there is always a SS with $D_{mc} = D_{pc} = 0$. What else is possible? The answer may seem complicated given we now have a bivariate system

$$D_{pc,t} = \Delta_p (D_{pc,t+1}, D_{mc,t+1}) \tag{24}$$

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

But notice (17) implies the two debt limits are proportional: $D_{mc,t}/D_{pc,t} = R \equiv R_m/R_p$. This is a very convenient result (that we did not anticipate) because it means we can first analyze a univariate system for $D_{mc,t}$ then get $D_{pc,t} = D_{mc,t}/R$.

If $R < 1$ then M is never active, so consider $R > 1$. The analog of (18) is

$$D_{mc,t} = \beta D_{mc,t+1} + \beta R_m (\alpha_{cp,t+1} S_{cp,t+1} + \alpha_{cm,t+1} S_{cm,t+1}) . \tag{26}$$

Notice C's expected surplus from trading with M and with P appear on the RHS, since now both are taken away by the punishment. The system is shown in Fig. 7, where

[17] Although this no longer gives the result in the preceding paragraph, that the lowest SS is $D > 0$, because that was for exogenous D_{mc}.

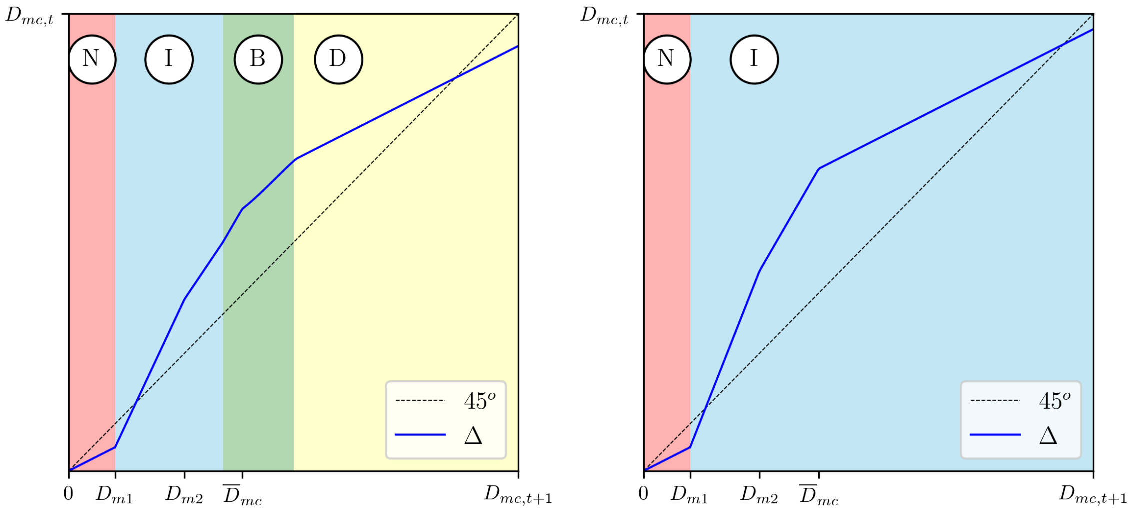

Figure 7: $D_{mc,t} = \Delta\left(D_{mc,t+1}\right)$ with $D_{pm} = \infty$; R low (left) or high (right).

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 – we suppose Weill (2007) would now say that M is "leaning with the wind." So whether intermediation attenuates or amplifies instability depends on details: theory does not unambiguously 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.[18]

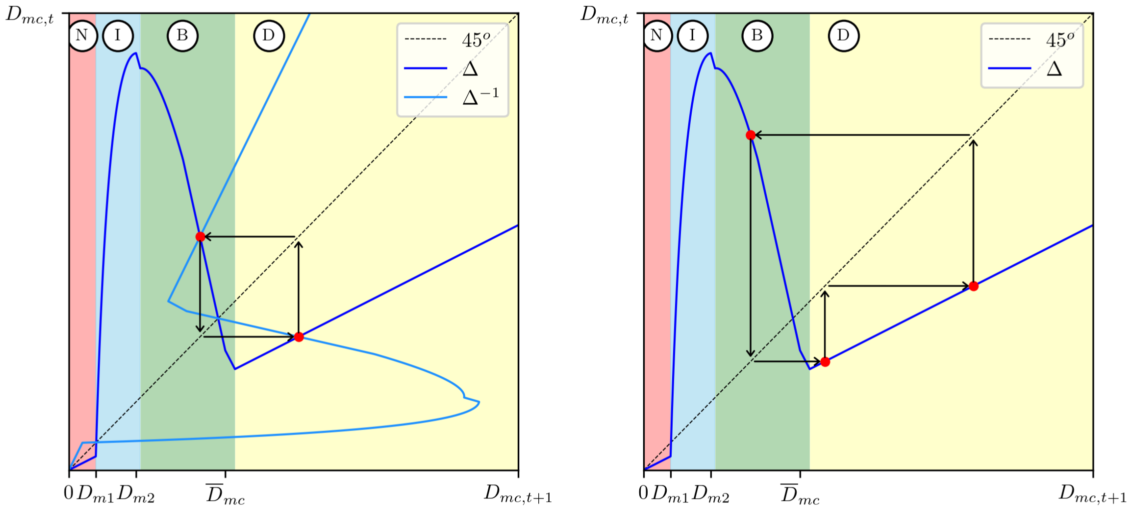

Figure 8: $D_{mc,t} = \Delta\left(D_{mc,t+1}\right)$ with $D_{pm} = \infty$; 2-cycle (left) and 3-cycle (right).

Having explained this, we point out that from Fig. 8 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).

Hence with indivisible Q we can get equilibria with deterministic or stochastic Regime switching. However, one may object to having Q indivisible. One reason has to do with the quantity/quality distinction – again, C may be interested in buying exactly 1 car, but that does not mean we should fix $Q = 1$ if cars can differ in quality

[18] Heuristically, here is how cycles may emerge when S_{ci} is decreasing in D_{ic}. It means low D_{ic} next period goes with high S_{ci} next period, and that makes C less inclined to default this period, and that makes this period's D_{ic} high. Hence there is a tendency for D_{ic} to oscillate.Now S_{ci} cannot be globally decreasing in D_{ic} but it can be for D_{ic} near the just-binding $\bar{D}_{ic}$. Also, having S_{ci} decreasing in is necessary but not sufficient for $\Delta' < 0$ since Δ also has an increasing linear term.

at the production stage. Another reason is technical: with indivisible goods agents generally may want to use lotteries for convexification. Suppose you want to sell your car but the buyer can pay at most a low p. What is there to bargain over? Propose to accept p and transfer the car with probability $\delta \in [0, 1]$. This effectively eliminates the indivisibility, although there can be ways to rule it out. We are agnostic and happy to consider models divisible or indivisible Q.

6 Conclusion

This paper studied environments with trade intermediated by middlemen and debt limits that were either exogenous or endogenous. It can be seen as putting payment frictions into models of intermediation or putting intermediaries into models of constrained credit. The approach applies to markets for goods, inputs or assets. It also works with indivisible or divisible objects being traded, and in the latter case there are interesting effects on both the intensive and extensive margin. With exogenous debt limits we characterized equilibrium in a benchmark specification and several variants, in each case proving existence and uniqueness, as well as showing how the pattern of exchange depends on parameters.

Endogenized debt limits arose from saying repayment must be incentive compatible. That generated multiple stationary equilibria plus dynamic equilibria with fluctuations in credit conditions, intermediation activity, etc., and these can display deterministic or stochastic Regime switching. Without arguing that real-world fluctuations are best explained as self-fulfilling prophecies, we think these results are relevant for the following reason: when simple model economies generate such outcomes, it lends credence to the view actual economies can, too. The emphasis was not that intermediation causes instability, but that limited commitment provides an endogenous role for intermediation and an endogenous source of dynamics, resulting in linkage between middlemen and volatility. Do middlemen attenuate or amplify fluctuations? We found the answer depends on details.

Our objective was not to explain one big empirical observation or prove one main theorem. Instead, the goal was to develop a flexible, tractable framework with intermediation and credit frictions that can be useful in future extensions and applications, and here we sketch a few ideas. One is to include occupational choice – who should be a middleman? Maybe those who are good at the things middlemen do, or maybe those who are bad at production, as discussed by Masters (2008), but in a much simpler model. Another idea is to further pursue welfare. When there are multiple equilibria, can we rank them and perhaps use policy to help select a good one? Also, since even the best equilibrium may be inefficient, how can we design corrective policies? Gong and Wright (2024) recently discussed this but in a simpler model.

On the technical front, it may be interesting to explore continuous-time versions of the framework since the implications for dynamics can be different in than discrete time, as discussed by Gu et al. (2025) in a middleman model, but one that is quite different than the framework here. A obvious extension would be to introduce money – maybe fiat currency, bank deposits, or e-money. While intermediation is one way to ameliorate trading frictions, money is another. Are they complements or substitutes? Our model has credit but no money; Urias (2018) is a related middleman model with money but no credit; in principle these could be merged.

Future work could further study the role of financial institutions as middlemen, as emphasized in Farboodi et al. (2023) in a different model. They highlight the core-periphery structure of financial networks, and their intermediaries have a dual motive for acquiring assets: they profit by using the instrument themselves (in our notation $\rho > 0$); and they facilitate further transactions (in our model they sell them to C). One can study how intermediation may lead to a core-periphery structure. This is related to the notion of middlemen chains, as discussed by Wright and Wong (2014), but without incorporating credit or many of the other ingredients here.

Another idea is to use version of the framework with more institutional detail to study particular markets for, say used cars, or certain financial assets. Another potentially fruitful path involves more work integrating middlemen models following Rubinstein and Wolinsky (1987), which focus on goods markets, and those 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 idiosyncratic shocks. An interesting feature of the former is the way dealers deal with inventories. Combining these could lead to new insights, with perfect or imperfect credit, but perhaps 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

$$\rho = \begin{cases} 0 & \text{if } \alpha_{sc} p_{pc} - c(Q_{pc}) - \kappa \leq 0 \\ [0,1] & \text{if } \alpha_{sc} p_{pc} - c(Q_{pc}) - \kappa = 0 \\ 1 & \text{if } \alpha_{sc} p_{pc} - c(Q_{pc}) - \kappa \geq 0 \end{cases} \tag{27}$$

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

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 (27)-(28) 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 (27)-(28). 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.

$$G(\tau, \rho) = \left[c(Q_{pc,\tau\rho}) + \kappa\right]/\alpha_{\tau\rho} \tag{29}$$

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

$$J(D_{pc}) = \left[\alpha_{01} D_{pc} - c(Q_{pc,01}) + c(Q_{mc,01})\right]/\alpha_{01} \tag{31}$$

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 Ex Ante Production: The analysis is slightly different. The CM value function for P is

$$\begin{aligned} V_p^C(\Omega) &= \max_{x,\ell,\phi} \left\{U(x) - \ell + \beta\left[\phi V_{p,+1}^W(Q) + (1-\phi)V_{p,+1}^W(0)\right]\right\} \qquad (32) \\ \text{st } x &= \Omega + \ell - \phi c(Q) \end{aligned}$$

where ϕ is the probability of producing. The CM problem for M or C is similar except they do not produce so $\phi = 0$. The WM value functions are

$$\begin{aligned} V_p^W(Q) &= V_p^C(0) + \alpha_{pm}\tau\left[V_p^C(p_{pm}) - V_p^C(0)\right] \qquad (33) \\ &\quad + (1-\alpha_{pm}\tau)\max_{\rho}\left\{\rho\left[V_p^R(Q) - V_p^C(0) - \kappa\right]\right\} \\ V_m^W &= V_m^C(0) + \alpha_{mp}\tau\left[V_m^R(Q) - V_m^C(0) - p_{pm} - \kappa\right] \qquad (34) \end{aligned}$$

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

$$\phi = \begin{cases} 0 & \text{if } \beta\left[V_{p,+1}^W(Q) - V_{p,+1}^W(0)\right] \leq c(Q) \\ [0,1] & \text{if } \beta\left[V_{p,+1}^W(Q) - V_{p,+1}^W(0)\right] = c(Q) \\ 1 & \text{if } \beta\left[V_{p,+1}^W(Q) - V_{p,+1}^W(0)\right] \geq c(Q) \end{cases} \qquad (35)$$

$$\rho = \begin{cases} 0 & \text{if } V_p^R(Q) - V_p^C(0) \leq \kappa \\ [0,1] & \text{if } V_p^R(Q) - V_p^C(0) = \kappa \\ 1 & \text{if } V_p^R(Q) - V_p^C(0) \geq \kappa \end{cases} \qquad (36)$$

SE is defined by: (V_i^n) for each type i in each market n satisfying (4)-(6) and (32)-(34); (p_{ij}, q_{ij}) satisfying (7)-(9); N_s satisfying (12); and (ϕ, τ, ρ) satisfying (15) and (35)-(36). The analysis follows the same procedure as the baseline model. ■

Parameters for 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 $D_{pc} \in \{0.2, 0.5\}$ 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 as follows. Fig. 6 has $R_p \in \{2, 4\}$, $(\omega_0, \omega_1) = (2, 0.2)$ and $\theta_{ic} = 0.5$. Fig. 7 has $R_p \in \{3, 1\}$, $R_m \in \{4, 5\}$, $(\omega_0, \omega_1) = (2, 0.1)$ and $\theta_{ic} = 0.5$. Fig. 8 has $R_p = 16$, $R_m = 17$, $(\omega_0, \omega_1) = (1, 0.5)$ and $\theta_{ic} = 0.98$.

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 (2023) "Nobel lecture: Financial intermediaries and financial crises," *JPE* 131, 2597-622.

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

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

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

[16] Y. Dong, J. Hu, Y. Huang, Y. Qiu and Y. Zhang (2025) "The use and disuse of fintech credit: When buy-now-pay-later meets credit reporting," mimeo.

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

[18] M. Farboodi, G. Jarosch and R. Shimer (2023) "The emergence of market structure," *RES* 90, 261-92.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[45] 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.

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

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

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

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

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

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

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

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

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

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

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

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

[58] 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.

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

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

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

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

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

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

[65] M. Yang and Y. Zeng (2021) "Coordination and fragility in liquidity provision," working paper available at SSRN 3265650.