Middlemen and Imperfect Credit*

Grace Xun Gong

School of Economics and Academy of Financial Research, Zhejiang University

Ziqi Qiao

University of Bern

Yuet-Yee Wong

Binghamton University

Randall Wright

Zhejiang University and Wisconsin School of Business

August 29, 2026

Abstract

This paper studies environments with intermediation (trade via middlemen) and credit that is constrained, either exogenously or endogenously due to limited commitment. Existing models give middlemen advantages in search, bargaining, information, etc. We give them an advantage in using credit: they are better at credibly promising future payments or enforcing others' payments. Given debt limits, there is a unique equilibrium that determines the transaction pattern – direct trade, indirect trade, both, or no trade, depending on parameters. With endogenous debt limits, there are multiple equilibria, including some where credit conditions, intermediation, prices and quantities fluctuate, deterministically or stochastically, as self-fulfilling prophecies.

JEL codes: D51, D61, D83

Keywords: Middlemen, Intermediation, Search, Debt Limits, Credit Cycles

*We thank Benoit Julien, Chao Gu, Yu Zhu, and participants in several seminars for their input. The Associate Editor and two Referees also made very useful comments. Gong acknowledges support from the Natural Science Foundation of China (No. 72503222) and 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.

1 Introduction

The goal of this paper is to develop an integrated model of intermediated trade and imperfect credit. There are several ways to motivate the endeavor.

First, in terms of microeconomics, Rubinstein and Wolinsky (1987) break new ground by applying search theory to study intermediation, meaning indirect trade via middlemen. They argue as follows:

> Despite the important role played by intermediation in most markets, it is largely ignored by the standard theoretical literature. This is because a study of intermediation requires a basic model that describes explicitly the trade frictions that give rise to the function of intermediation. But this is missing from the standard market models, where the actual process of trading is left unmodeled.

The Rubinstein-Wolinsky model is simple and delivers one clear result: middlemen are active when they have a comparative advantage in search, meaning they can contact consumers faster than producers contact consumers. Following in their footsteps, a literature has emerged showing how middlemen can be active if they have other advantages, including a better storage or inventory technology, superior information or higher bargaining power.[1] We introduce a new possibility: they can be better at getting or giving credit, by more reliably promising future payments or by enforcing payments from others.

This is realistic. 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 consumers do not have sufficient liquidity for immediate settlement, 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 ready funds for immediate

[1] As there are many related papers, we posted an online bibliography at https://github.com/qiao-ziqi/middlemen. See Gong et al. (2026) for a survey, but here is a sample: in Biglaiser (1993) and Li (1998) middlemen have expertise in discerning quality; in Masters (2008) and Nosal et al. (2015, 2019) they have higher bargaining power or lower costs; in Shevchenko (2004) and Watanabe (2010) they can hold bigger inventories. See also Shi (1996), Hu et al. (2025) and Gu et al. (2025).

settlement or they are less likely to renege on deferred settlement due to reputational considerations. Of course, when dealers sell cars, their customers may have commitment issues, too, but it is no stretch to think that used-car dealers are better than you at collecting debts and/or repossessing vehicles.

A second motivation relates more to macroeconomics, where it is generally accepted that credit market imperfections are crucial. As Claessens and Kose (2017) say in their survey:

> Imperfect credit – referring to the inability of individuals or firms to borrow unlimited amounts due to information asymmetries and default risks – is central to macroeconomics. It causes economic shocks to amplify, impacts resource allocation, and is a primary channel through which central banks manage the economy.

There are many papers pursuing this line, including a large literature following Kiyotaki and Moore (1997).[2] However, it is important to recognize that resilient economies have ways to deal with credit frictions. One is the use of money and there is much research on that (see the surveys by Lagos et al. 2017 and Rocheteau and Nosal 2017). But another option is intermediated trade where middlemen can ameliorate credit frictions, and we know of no prior research studying this the way we do.

A third motivation relates more to finance, where it is often argued that intermediation and volatility/fragility are intimately related. This idea has been pushed over the years by many authors.[3] In particular, there is a large literature concerned with instability in banking following Diamond and Dybvig (1983), but other financial markets or institutions may also be volatile – as Ferguson (2008) says, "*all financial institutions* are at the mercy of our innate inclination to veer from euphoria to despondency" (emphasis added).

[2] Many papers impose exogenous debt limits. Models with endogenous limits include Kehoe and Levine (1993), Alvarez and Jermann (2000), Gu et al. (2013a,b), Gu et al. (2016), Azariadis and Kaas (2007, 2013), Lorenzoni (2008), Hellwig and Lorenzoni (2009), Sanches and Williamson (2010), Myerson (2012) and Carapella and Williamson (2015), but none of these have middlemen. Li (1998), Urias (2018) and Lotz and Zhang (2026) have models with middlemen and money, but no credit.

[3] These include Keynes (1936), Kindleberger (1978) and Minsky (1992), to name a few classics. More recently, Diamond (2023) provides a general discussion, while Reinhart and Rogoff (2009) provide a detailed historical account.

Moreover, as Shiller (1988) puts it:

> Financial market prices, prices of stocks, bonds, foreign exchange, and other investment assets, have shown striking changes in volatility through time. For each of these kinds of assets there are years when prices show enormous unpredictable movements from day to day or month to month, and there are years of stable, uneventful markets.

Do model economies with intermediation and credit frictions have equilibria where sometimes (but not always) there arise endogenous fluctuations even when fundamentals are constant? The answer is yes, but causation is delicate. The root source of the phenomena is limited commitment, which can engender volatility and can deliver a role for intermediated trade. Hence, we can say volatility and intermediation go together, but both can be ultimately traced back to commitment frictions.

Given all this motivation, we develop a model extending recent work by Gong and Wright (2024) in the Rubinstein-Wolinsky framework. That paper has nothing to do with credit, but the environment is natural for introducing it. The reason is that, following Lagos and Wright (2005), it has both centralized and decentralized trade and an asynchronicity of expenditures and receipts: sometimes someone wants something in the decentralized market, while their incomes accrue in the centralized market, generating a role for deferred payment.

Also, different from much related work, the model replaces the usual three-sided market, where producers, consumers and middlemen all interact, with two two-sided markets, a wholesale and a retail market, which is arguably more realistic and certainly more tractable. Moreover, the objects being traded can be goods, inputs or assets, which is relevant since in reality there is intermediated trade in all three. In addition, these objects can be indivisible or divisible, again different from related papers.[4]

[4]The original Rubinstein-Wolinsky middleman model and typical extensions only consider indivisible goods. This has much precedent in search-based models of markets for assets (e.g., Duffie et al. 2005), money (e.g., Kiyotaki and Wright 1993), labor (e.g., Pissarides 2000), etc. While these models are useful and interesting, it is surely worth exploring versions with divisibility. In particular, this lets us analyze the intensive margin (size of trades) not just the extensive margin (number of trades). It also turns out that, as shown below, divisibility vs indivisibility matters for the existence and nature of endogenous dynamics.

The first results concern outcomes with exogenous debt limits. We prove existence and uniqueness of equilibrium. Moreover, parameters pin down the endogenous pattern of exchange – no trade, direct trade from producers to consumers, indirect trade via middlemen, or both direct and indirect trade. This is useful because in reality some markets have mostly direct trade, while others have mostly indirect trade, and still others are in between (see fn. 10 for examples and discussion). Our results elucidate factors determining which pattern should emerge.

Then we endogenize credit limits: agents can renege, but risk exclusion from credit in the future (Kehoe and Levine 1993). Now there can be multiple equilibria, including nonstationary outcomes with fluctuations in credit conditions, intermediation activity, etc. These self-fulfilling prophecies can be deterministic or stochastic, and can involve regime switching with recurrent changes in trade patterns, consistent with the above remarks by Shiller (1988). One key point is this: it is not that middlemen cause instability, although that is possible, but limited commitment provides an endogenous role for intermediation and is a source of endogenous dynamics, suggesting a more subtle but no less interesting relationship between intermediation and volatility.

It is worth mentioning that our basic premise is consistent with practitioners' view: "Many middlemen provide credit... This financial support helps businesses sustain operations and enhances market liquidity" (Marketing for Managers 2024). Some theorists agree, like Spulber (1999), who in his study of market microstructure says "Intermediaries who are able to make credible commitments bring advantages over contracts between buyers and sellers." This is what we want to capture using dynamic search theory with constrained credit.

Another relevant point is that there is growing interest by academics and regulators in BNPL (buy-now-pay-later) plans, where usage has grown from \$50 billion in 2019 to \$370 billion in 2023 (Mojon et al. 2023). See also Han et al. (2024), Dong et al. (2025), and Stavins (2024), who says 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 fact, credit issued to consumers by retailers is not new, with an interesting example being the Charga-plate system in the 1930s-50s (Frankel 2024). Even earlier, in the early 1900s cards were launched by department stores and oil companies; unlike most credit cards today, these could only be used at specific vendors (Tretina and Little 2025). Theory can enhance our understanding of this, and more generally our understanding of the rapidly-evolving payment landscape. All of this goes back to what we might call the Rubinstein-Wolinsky dictum: modeling intermediated trade requires explicitly describing frictions absent from classical equilibrium theory.[5]

In what follows, Section 2 describes the environment. Section 3 analyzes equilibria with exogenous debt limits. Section 4 endogenizes these limits. Section 5 concludes.[6]

2 Environment

A continuum of agents lives 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. Each period three markets convene sequentially. First is a wholesale market WM, where type P agents may trade quantity Q_{pm} of some object to M. Type C cannot participate in WM, interpretable as an assumption about spatial/temporal separation. But after WM closes, there opens a retail market RM where C may buy $q_{ic} \leq Q_{ic}$ from type $i = m, p$, meaning either from type M that traded in WM or type P seeking direct trade. (Throughout the presentation, Q's denote quantities produced by P while q's denote quantities consumed by C). Both WM and RM are decentralized, with bilateral

[5]This is meant to mimic the dictum of Wallace (1996): "Money should not be a primitive in monetary theory – in the same way that firm should not be a primitive in industrial organization theory or bond a primitive in finance theory." Or, as Kareken and Wallace (1980) put it "Progress can be made in monetary theory and policy analysis only by modeling monetary arrangements explicitly."

[6]In addition to research on middlemen following Rubinstein-Wolinsky, there are search models of dealers in asset markets following Duffie et al. (2005); see Hugonnier et al. (2025) for a survey. Those models differ in various ways, e.g. their dealers hold no inventory, but simply reallocate assets across agents using frictionless interdealer markets (with a few exceptions, like Weill 2007 or Yang and Zeng 2021). Also those papers generally have transferable utility and hence no credit frictions.

random matching, bargaining and payment frictions, although one can turn off any of these to get closer to a Walrasian market.

After WM and RM close, there is a frictionless centralized market CM, where all agents sell labor ℓ, buy a numeraire consumption good x, and settle debt. The idea is that agents use credit for RM and WM purchases, to be honored in the next CM, which 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, or deep pockets, as discussed below.

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 we can focus on short-term debt without loss of generality, where all obligations are cleared each CM. In RM, C gets $u(q_{ic})$ from q_{ic} with $u'(\cdot) > 0 > u''(\cdot)$ and $u(0) = 0$. Type P can produce Q_{pm} at cost $c(Q_{pm})$ in WM when they meet type M and decide to trade, or can produce Q_{pc} at cost $c(Q_{pc})$ if they opt to go to RM seeking direct trade. We assume Q is divisible with $c'(\cdot) > 0$, $c''(\cdot) \geq 0$, $c(0) = 0$ and $u'(0)/c'(0) = \infty$.

A sometimes-useful reinterpretation (e.g., for used cars) is that P agents are endowed with Q and have an opportunity cost of giving it up, instead of a cost of producing it. Also, Appendix A discusses a version where Q is indivisible at the production stage. In any case, as in many search models there is a cost κ that sellers, be they type P or M, must pay to participate in RM. We emphasize that q can be a good and $u(q)$ a utility function, or q can be an input and $u(q)$ a production function, or q can be an asset with $u(q)$ the buyers' return and $c(q)$ the sellers' return. Also, P can produce only once per period, so trading in WM means skipping the RM.

Meetings in WM and RM are governed 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. 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. The buyer-seller ratio, or market tightness, determines α_{ij}, the probabilities of i meeting j

for $i, j \in \{p, m, c\}$. This is all standard but we emphasize that having two two-sided markets lets us use general meeting technologies, while papers with three-sided markets use a special case where α_{ij} is proportional to $N_j / \sum_h N_h$.

Terms of trade are determined by 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 $Q_{pm} \leq Q$ to M, but in equilibrium $Q_{pm} = Q$ (P sells M all the output). This is obvious when Q is indivisible; when it is divisible, we assume P cannot produce Q and sell $Q_{pm} \leq Q$ to M, then take $Q - Q_{pm}$ to RM. However, P can produce Q_{pm} for trade with M and a different Q_{pc} for direct trade if P wants to go to RM. Then $q_{mc} \leq Q_{pm}$ is what M sells to C and $q_{pc} \leq Q_{pc}$ is what P sells to C, but in equilibrium $q_{ic} = Q_{ic}$ since unsold Q has no scrap value (see Appendix A).

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. In addition to p's and q's, endogenous choices are, allowing mixed strategies, τ which is the probability that P and M trade in WM meetings, and ρ which is the probability that P goes to RM after not trading in WM. (We do not need the probability M goes to RM after trading in WM since it makes no sense to trade in WM and not go to RM.)

This is the baseline model, but Appendix A discusses extensions and alternative specifications, including indivisible goods, ex ante production (before WM meetings), scrap value for unsold output, and secured credit. Also, the benchmark studied below has $D_{pm} = \infty$, which lets us focus on D_{pc} and D_{mc} and captures the idea that M has either excellent credit or deep pockets in WM, but Appendix A solves the model with $D_{pm} < \infty$. In each case, details can matter, but the main message is similar.

3 Basic Results

We start by studying outcomes when the D_{ij}'s are constant. One can say they are exogenous, but they can generally be supported as endogenous debt limits with the right combinations of fundamentals, as discussed in Section 4.[7] As mentioned, as a benchmark here $D_{pm} = \infty$ (but see Appendix A). Then the first step is to present the value functions for type $i = C, P, M$ in WM, RM and CM, denoted V_i^W, V_i^R and V_i^C.

In the CM, for any type i we have

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

where ℓ is labor income, since we assume 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, V_i^C is linear with slope 1.[8]

Moving to WM, for P we have

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

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}, using the result that $V_p^C(p_{pm}) - V_p^C(0) = p_{pm}$ by the linearity of V_p^C. The third is the probability of not trading in WM, and with probability ρ producing for the RM, which entails surplus $V_p^R(Q_{pc}) - V_p^C(0) - c(Q_{pc}) - \kappa$. Again, Q_{pm} and

[7] To clarify, given the D's, we can back out deeper parameters supporting them as endogenous debt limits. Then many results – e.g., existence – still hold and apply to steady states of the dynamic system analyzed below. For other results – e.g., comparative statics – we must say the D's are genuinely exogenous since endogenous D's generally change with parameters (see Section 4). That said, many papers have exogenous debt limits and their authors should find the results interesting even without endogenous D's.

[8] This follows from quasi-linear CM utility – simply insert ℓ from the constraint into the objective function and it is obvious – but also holds for any $\tilde{U}(x, 1-\ell)$ with $\tilde{U}_{11}\tilde{U}_{22} = \tilde{U}_{12}^2$ (Wong 2016).

Q_{pc} can be any nonnegative quantity here, but we also mention what happens if Q is indivisible, with details in Appendix A.

For M in WM,

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

again using the linearity of V_m^C. In (3), the term in brackets looks like immediate settlement due to the appearance of $-p_{pm}$, but is actually the anticipation of settlement deferred to the next CM; in this sense perfect credit looks just like deep pockets.

Moving to RM, for P we have

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

where Q_{pc} is inventory P brings to market and A is the scrap value of whatever is unsold, which is 0 here but Appendix A discusses $A > 0$. For M we have

$$V_m^R(Q_{mc}) = V_m^C(0) + \alpha_{sc}p_{mc} \tag{5}$$

and for C

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

Terms of trade when i sells to j solve the generalized Nash bargaining problem

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

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,

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

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 $q_{ic} \leq Q_{ic}$, although the latter holds at equality. Notice Q_{pm} is determined bilaterally between P and M, while Q_{pc} is unilaterally chosen by P. Also notice 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 the entry cost κ is sunk.

Now consider the meeting probabilities. In WM, α_{pm} and α_{mp} come from the meeting technology $m_W(N_m, N_p)$ 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 trades in WM plus P that does not trade in WM and goes to RM. Hence,

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

As in related papers we call (12) a steady-state condition, even though it is basically static. More importantly, the RM meeting technology treats all sellers the same in terms of arrival rates, but the outcome of an arrival 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:

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

For τ, given $D_{pm} = \infty$, we have:[9]

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

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)

[9] More generally, if $D_{pm} < \infty$, as discussed in Appendix A, WM trade requires the proverbial double coincidence of wants:

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

and (7); and (τ, ρ) satisfying (13) and (15). Actually, the stationary qualification in the definition is unnecessary here since when D's are fixed there are no nonstationary equilibria, but that will change in Section 4.

We can characterize the SE set by checking each candidate strategy profile (τ, ρ) to see, after computing the other implied endogenous variables, which ones satisfy the best response conditions. Since τ and ρ can be 0, 1 or in $(0, 1)$ there are 9 candidate profiles that 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 with M and goes to RM with positive probability. Regime I, for *indirect trade*, has $\tau > 0$ and $\rho = 0$, so P and M trade with positive probability, and M goes to RM while P does not. Regime B, for *both*, has $\tau > 0$ and $\rho > 0$, so P and M go to RM with positive probability. Within a Regime we also distinguish between pure or mixed strategies, and between cases where debt limits bind or are slack.

In what follows we ignore nongeneric outcomes – 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 τ is a best response, but that would not be true after a small change in parameters. Then we have the following result:

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

The details that constitute a proof are provided in Appendix B, while the results are illustrated in Fig. 1. (Appendix A provides results assuming indivisible Q, which are qualitatively similar.) The graphs come from examples with parameters listed in Appendix D, but the results are general. To interpret them, note that the solid borders separate regions with different Regimes, dashed borders separate pure and mixed strategies in a Regime, and dotted borders separate binding and slack debt limits in a Regime. The left panel has $\theta_{mc} = \theta_{pc}$ while the right panel has $\theta_{mc} > \theta_{pc}$ to show the impact of M and P having different RM bargaining power.

The first result to highlight is that this partition of parameter space into regions, each containing one SE (generically), demonstrates existence and uniqueness. Moreover, in terms of the endogenous pattern of exchange, it tells us *exactly what happens for all parameters.*[10]

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 emerge. Regime D obtains if $D_{pc} > \bar{D}_{pc}$ (light yellow region) since P is unconstrained in RM, or if $D_{pc} < \bar{D}_{pc}$ and we are below the boundary separating Regime D from I/B (dark yellow region) since P is constrained but M is more constrained.

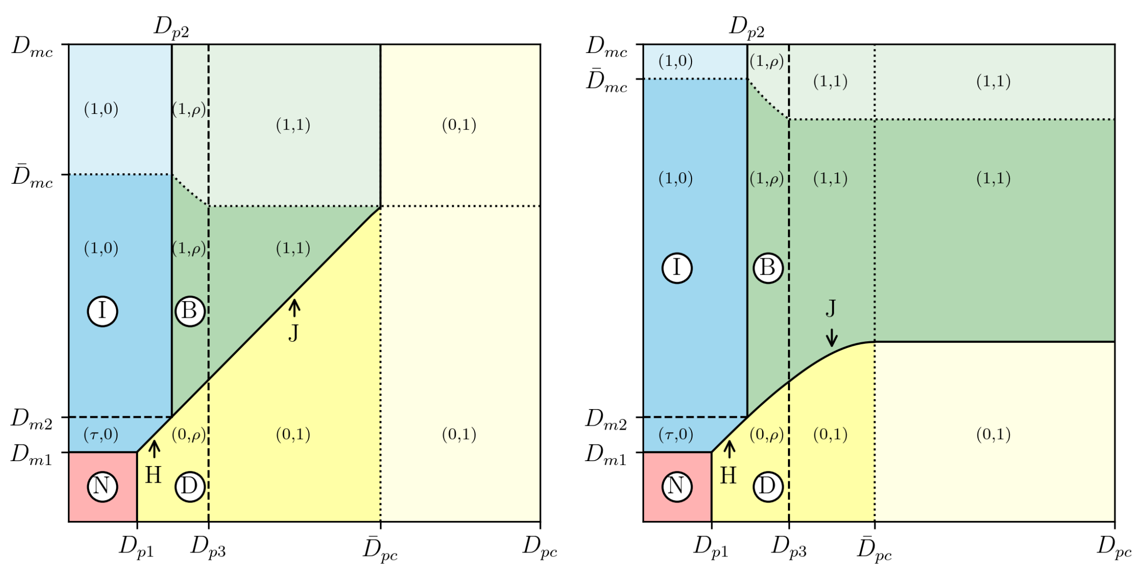

Figure 1: SE with Q divisible and $D_{pm} = \infty$; $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

[10]This is reminiscent of work in monetary economics determining endogenous transaction patterns – barter, credit, monetary exchange or combinations thereof (e.g., Kiyotaki and Wright 1989). In terms of the findings here, 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, but an interesting case study involves high-end purveyors of coffee, chocolate and tea that these days are buying direct from sources (Charles 2024). In the used-car market two-thirds of sales go through dealers (Li et al. 2025). In asset markets, trade in 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 2007). We also mention middlemen chains, like farmer to broker to distributor to retailer to consumer (Wright and Wong 2014).

Now suppose $D_{pc} < \bar{D}_{pc}$ and we are above that boundary. When D_{pc} is low, Regime I obtains since 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 is this: *M is active when D_{pc} is low and D_{mc} high.*[11] 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. Hence, *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., there would be no trade if intermediation were shut down by regulation or taxation. For higher D_{pc} we switch from Regime D to I, so M is not crucial for RM trade, 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 is for $\theta_{mc} = \theta_{pc}$. 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 may be active solely due to bargaining power.*

Additional economic insight comes from Figs. 2 and 3 showing the impact on various variables of changes in D_{mc} and D_{pc}. The left and right panels are for different values of other parameters. The curves are only drawn over relevant ranges – e.g., if M does not enter RM then Q_{mc} is not shown. In the top row of Fig. 2, M becomes active 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, so we switch from Regime D to I and increase N_s. In the right panel, when M starts going P does not stop, so we switch from Regime D to B

[11] This may not be too surprising, but at least it is precise, giving the exact cutoffs for different Regimes. Similarly, the classic Rubinstein-Wolinsky result, middlemen are active when they can contact consumers faster than producers, is not surprising but is precise in the same way.

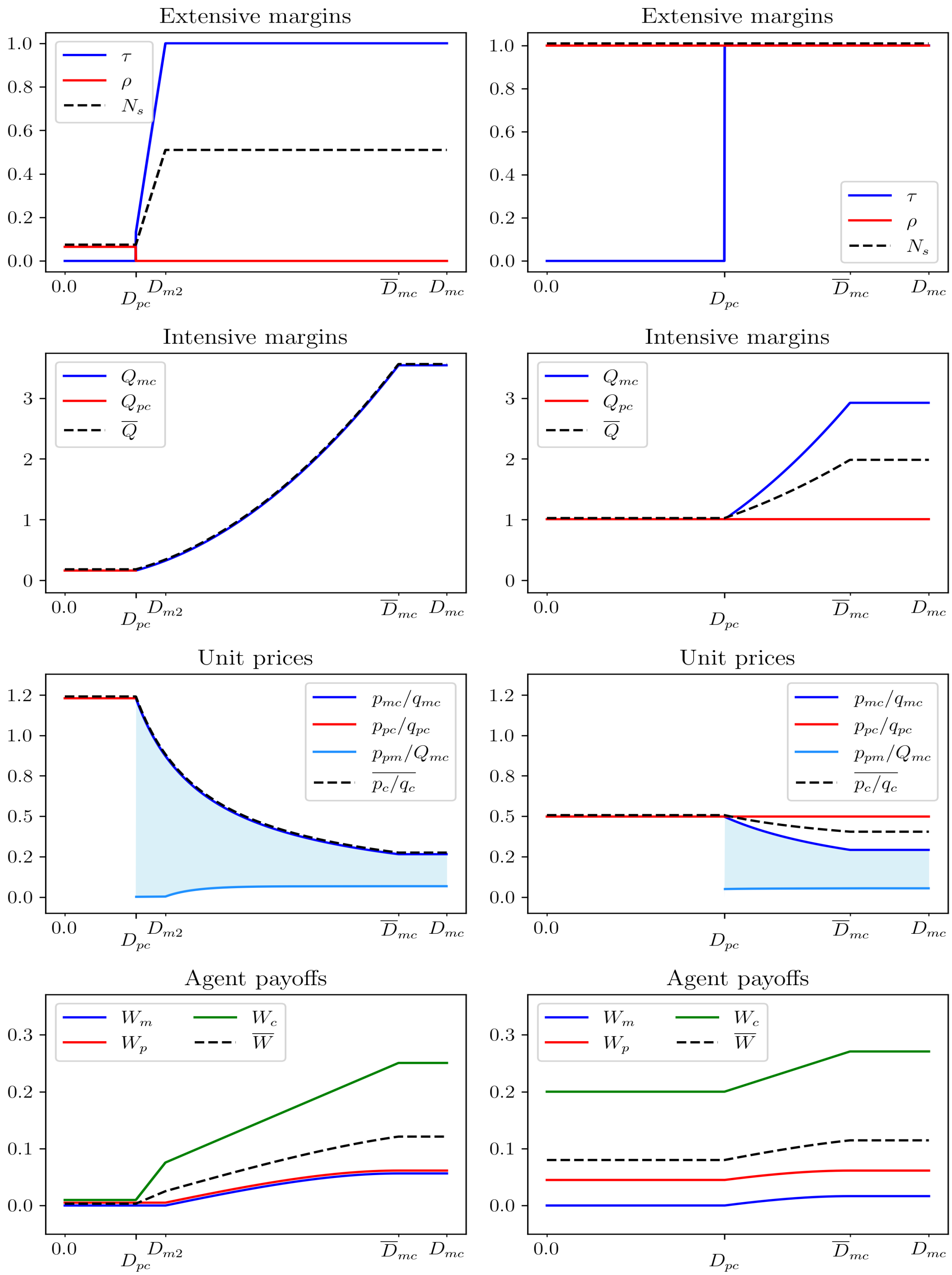

Figure 2: Effects of D_{mc} with Q divisible, $D_{pm} = \infty$ and D_{pc} low (left) or high (right).

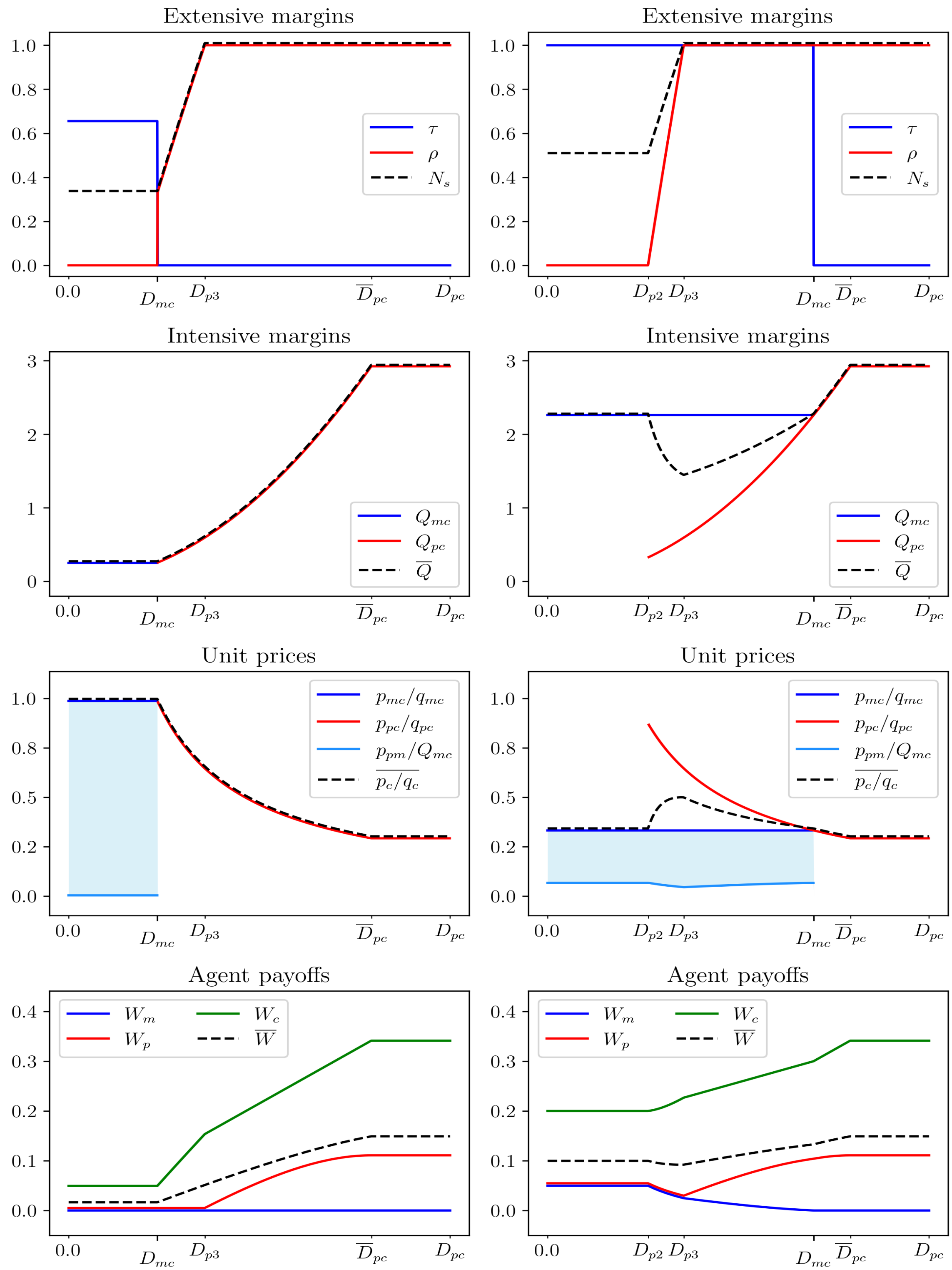

Figure 3: Effects of D_{pc} with Q divisible, $D_{pm} = \infty$ and D_{mc} low (left) or high (right).

but do not increase N_s since M simply crowds out P. Hence, *activity by middlemen may or may not increase the number of sellers.*

The second and third rows in Fig. 2 show 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 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. 3 fixes D_{mc} and varies D_{pc}, and there the average unit price 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 $D_{pc} \in (D_{p2}, D_{mc})$. Hence, *intermediation can generate dispersion in price and quantity*, and less obviously *average price and price dispersion can be nonmonotone in credit frictions.*[12]

Notice in Fig. 2 that payoffs can increase with D_{mc}, while in the right panel of Fig. 3 higher D_{pc} reduces payoffs of P and M, as well as the average payoff, over some range, but raises C's. In other examples, higher debt limits can lower C's payoff – the simplest case is for Q indivisible (Appendix A), so raising D means C pays more but does not get more. In general, *lower credit frictions can increase or decrease welfare.* While important, we do not dwell on welfare because there are already many discussions in the literature, and it is well known that the results depend on the details (see Gong and Wright 2024 and references therein).

It is true that for some parameters welfare is higher with middlemen than without – in the language of monetary theory, *for some parameters intermediation is essential.* In the mechanism-design approach to monetary economics (Wallace 2010) fiat currency is said to be essential when the set of incentive-feasible allocations is bigger with money than without it. Again, depending on parameters, middlemen may or may not be essential. This captures the age-old idea that they might perform valuable functions –

[12]This is not shown in the graphs, but one can reduce search frictions by raising the constant in the meeting technologies, and see that average price and price dispersion can go up or down. Apropos fn. 11, this may be surprising since some people appear puzzled about average price and price dispersion not falling in the data with improvements in search efficiency (see the survey by Gong et al. 2026 for more discussion and references). Our model does *not* imply average price and price dispersion must fall with improvements in search efficiency.

here, they facilitate deferred payment – or they might be in the market solely due to their ability to buy low and sell high because they have high bargaining power.

4 Endogenous Debt Limits and Dynamics

Next we endogenize the D's using ideas and methods from, e.g., Kehoe and Levine (1993) and Alvarez and Jermann (2000). One reason is that this allows us to reassess how the transaction pattern depends on parameters. Another reason is to see whether interesting dynamics can emerge. As in the baseline model we focus on $D_{pm} = \infty$, so WM credit is unconstrained but RM credit is limited by $D_{pc,t}$ and $D_{mc,t}$ (again see Appendix A for $D_{pm} < \infty$).

4.1 Endogenous Debt Limits

Buyers can renege on payments, but risk being punished by taking away future credit. Here C loses all future credit if caught reneging, but we also analyzed alternative punishments – e.g., if C reneges on P, C loses credit with P but not M – and the results were similar. For now Q is divisible, as in the baseline, but indivisible Q is also discussed below. Also, to keep things manageable, here P and M have the same RM bargaining power, $\theta_{mc} = \theta_{pc}$.

In the CM, if C reneges on the p_{ic} owed to $i \in \{p, m\}$ there is a proportional cost $\lambda_i p_{ic}$, with $\lambda_i \in [0, 1]$. Moreover, opportunistic agents are only caught, and hence are only punished, with probability $\mu_i \in [0, 1]$, where one interpretation of $\mu_i < 1$ is that debt payments are randomly monitored, like the IRS randomly audits taxpayers. Both λ_i and μ_i capture disincentives to misbehave, and are both included because they are known to have interesting implications in other work.[13]

Since C loses all future credit if caught reneging, on either P or M, the punishment

[13]In banking theory, μ is used by Gu et al. (2013b) and Huang (2026) to analyze who should be a banker and how many bankers we should have. In monetary economics, Kocherlakota (1998) shows that fiat currency cannot be welfare-enhancing (it is dominated by credit) if $\mu = 1$, but can be if $\mu < 1$. In finance, λ captures resources used up by opportunistic behavior as in the cash-diversion models of DeMarzo and Fishman (2007) and Biais et al. (2007).

payoff is the value of autarky, $V_{c,t}^D = (1-\beta)^{-1}[U(\ell) - \ell]$. Then the incentive condition for C to honor debt $p_{ic,t}$ is

$$V_{c,t}^C(0) - p_{ic,t} \geq -\lambda_i p_{ic,t} + \mu_i V_{c,t}^D + (1-\mu_i) V_{c,t}^C(0). \tag{16}$$

An endogenous debt limit $D_{ic,t}$ satisfies (16) with equality:

$$D_{ic,t} = R_i \left[V_{c,t}^C(0) - V_{c,t}^D \right], \tag{17}$$

where $R_i \equiv \mu_i/(1-\lambda_i)$ captures the i-specific ability to discourage default for $i = p, m$.

From (1) and (6), we have

$$V_{c,t}^C(0) - V_{c,t}^D = \beta \left[V_{c,t+1}^C(0) - V_{c,t+1}^D + \sum_{i \in \{p,m\}} \alpha_{ci,t+1} S_{ci,t+1} \right] \tag{18}$$

where $\alpha_{ci,t+1}$ is C's probability of trading with i, and $S_{ci,t+1}$ is the corresponding surplus C would lose from the punishment. Since all $\alpha_{ci,t}$ and $S_{ci,t}$ depend on $D_{ic,t}$, (17) reduces to the difference equations

$$D_{ic,t} = \beta \left[D_{ic,t+1} + R_i \sum_{j \in \{p,m\}} \alpha_{cj,t+1} \left(D_{pc,t+1}, D_{mc,t+1} \right) S_{cj,t+1} \left(D_{jc,t+1} \right) \right] \tag{19}$$

for $i = p, m$. Note the trading probabilities depend on both debt limits, while the surplus from trading with j depends only on j's debt limit.

In principle, the system of equations forms a bivariate system,

$$D_{pc,t} = \Delta_p \left(D_{pc,t+1}, D_{mc,t+1} \right) \tag{20}$$

$$D_{mc,t} = \Delta_m \left(D_{pc,t+1}, D_{mc,t+1} \right), \tag{21}$$

but conveniently (17) implies the D's are proportional: $D_{mc,t}/D_{pc,t} = R \equiv R_m/R_p$. So we can analyze a univariate system for $D_{pc,t}$ and then set $D_{mc,t} = R D_{pc,t}$, or vice versa.

In words, $D_{ic,t}$ is the most a debtor would pay at t given the path of future debt limits, described recursively in (19). A steady state of this system satisfies $D_{pc} = \Delta_p(D_{pc})$ and constitutes an SE with an endogenous D_{pc}. Other (nonnegative and bounded) solutions to (19) are dynamic equilibria, with time-varying debt limits. Appendix C presents the formulas for $\alpha_{cj,t}$ and $S_{cj,t}$, so we know (19) in closed form.

We start by analyzing the steady states:

Proposition 2: With $D_{pm} = \infty$, $D_{pc} = D_{mc} = 0$ is always a steady state and it entails no trade. This is the only steady state if the following conditions hold,

$$(1-\beta)\bar{D}_{pc} > \beta R_p \bar{\alpha}_{cs} \left[u(Q^*_{pc}) - \bar{D}_{pc} \right] \text{ for } R \leq 1 \tag{22}$$

$$(1-\beta)\bar{D}_{mc} > K(R_p, R_m) \text{ for } R > 1, \tag{23}$$

where

$$K(R_p, R_m) \equiv \beta \left\{ R_p \hat{\alpha}_{cp} (1/\theta_{pc} - 1)\bar{D}_{mc} + R_m \hat{\alpha}_{cm} \left[u(Q^*_{mc}) - \bar{D}_{mc} \right] \right\}, \tag{24}$$

while $\bar{\alpha}_{cs}$ is C's maximum trading probability when $\rho = 1$, and $\hat{\alpha}_{cp}$ and $\hat{\alpha}_{cm}$ are α_{cp} and α_{cm} evaluated at $(D_{pc}, D_{mc}) = (\bar{D}_{mc}/R, \bar{D}_{mc})$. Otherwise, if (22) and (23) do not hold, there are generically two steady states with trade.

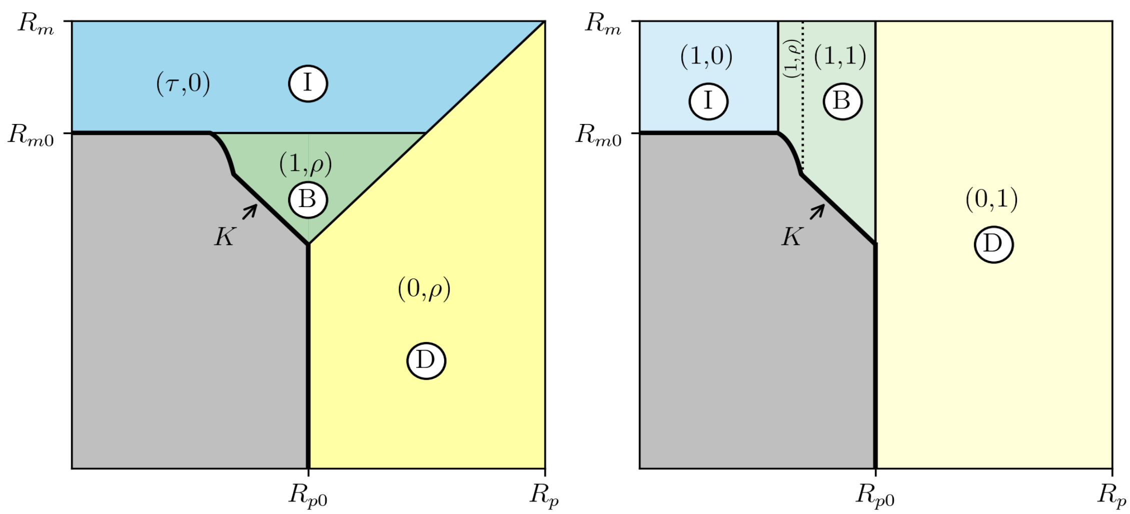

Figure 4: Middle (left) and top (right) steady states with $D_{pm} = \infty$.

When (R_p, R_m) is within the K boundary defined by Proposition 2 and shown as the gray region in Fig. 4, the unique outcome is $D_{pc} = D_{mc} = 0$. Ignoring that uninteresting case, there are three steady states: a low one with no credit, and two active ones – a middle steady state with some credit, and a top steady state with more credit. Note that the two active steady states can exhibit different Regimes.[14]

The left panel of Fig. 4 shows the Regimes that obtain in (R_p, R_m) space conditional on being in the middle steady state; the right panel shows the same thing conditional on being in the top steady state. On the left, Regime D obtains when $R_p \geq R_m$ since M has no advantage over P. When $R_m > R_p$, Regime I obtains if $R_m > R_{m0}$ and Regime B otherwise. Here all active regions have mixed entry. On the right, the lighter color indicates nonbinding debt limits, so the Regime only depends on D_{pc}.

Fig. 5 shows how active outcomes depend on R_m with solid lines for the top steady state and dashed lines for the middle one. The left panel has fixed R_p, so varying R_m changes R, while the right has fixed R, so R_p scales with R_m. When R_m is low, neither active steady state exists, since the frictions are too severe to support sellers' participation. As R_m increases the middle steady state contracts along both the extensive and intensive margins. When R_p is fixed, higher R_m corresponds to better enforcement by M. As P cannot provide enough credit, the top steady state stays in Regime B once R_m is big enough. When R is fixed, R_p and R_m rise at the same rate. Once D_{pc} becomes slack, the top steady state experiences disintermediation.[15]

4.2 Stochastic Cycles

Given that $D_{pc,t}$ and $D_{mc,t}$ are proportional, we focus on the system (19) for $D_{mc,t}$, as shown in Fig. 6 for two levels of R. Intuitively, more credit tomorrow incentivizes better behavior today, so Δ is increasing. When R is low, all four Regimes are possible (left panel); when R is high, only Regimes N and I are possible (right panel).

A path for $D_{mc,t}$ is a dynamic equilibrium when it is not a steady state and starts

[14] The reason there are two steady states with trade, rather than other even number, generically, is that S_{ci} is linear in D_{ic}, and the pivot $\Delta(D_{ic})$ for existence is greater than D_{ic} iff $\Delta'(D_{ic}) > 1$. See Appendix C for details. Also, below we drop the qualifier generically.

[15] There are also interesting patterns in quantities and prices but we do not want to get bogged down in details here. In terms of welfare, the effects are generally ambiguous. One can check that eliminating M by, e.g., regulation or taxation, hurts welfare, but raising R_m does not always help – see the bottom-right panel of Fig. 5.

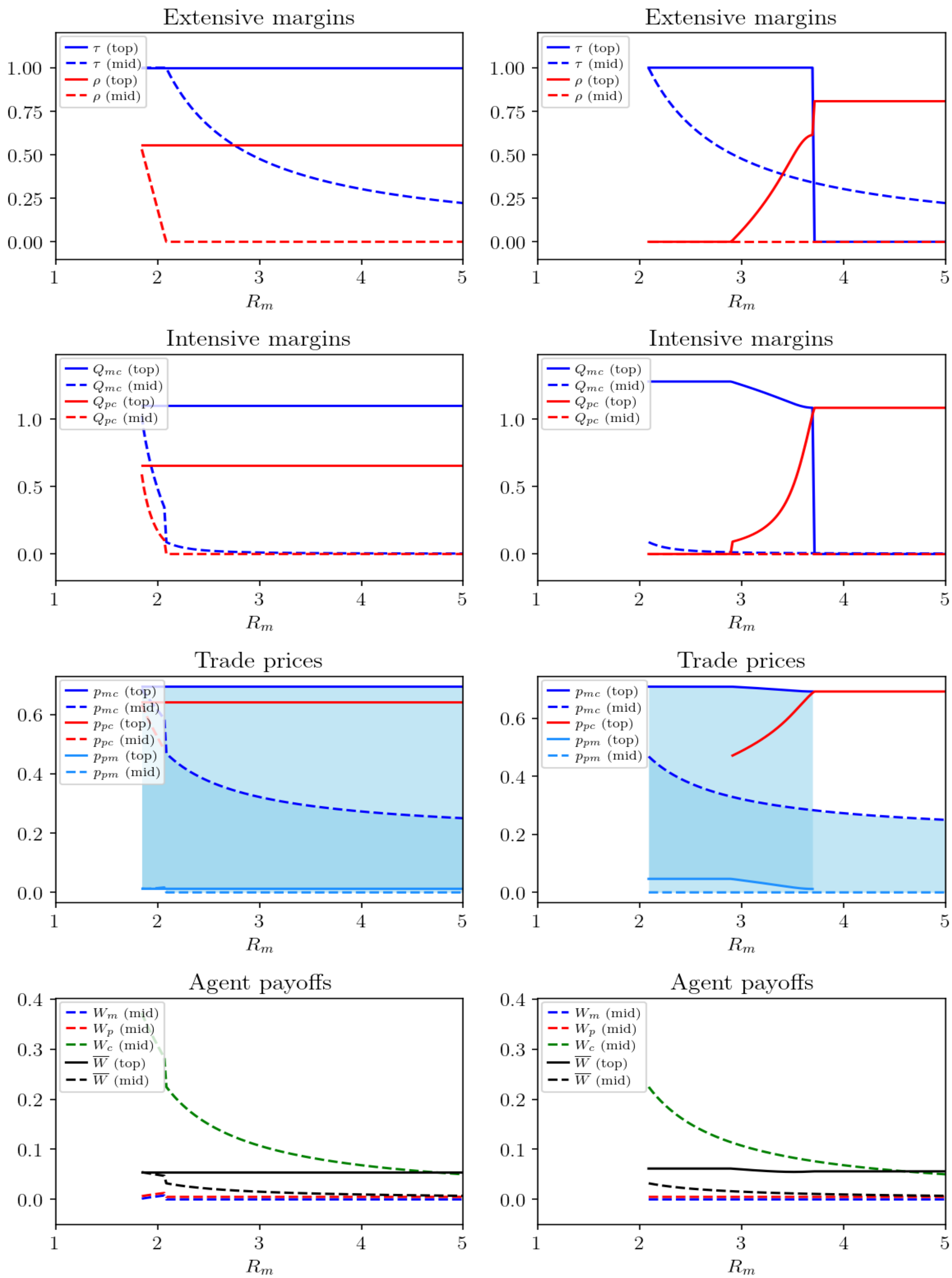

Figure 5: Effects of R_m, $D_{pm} = \infty$; fixed R_p (left) or fixed R (right).

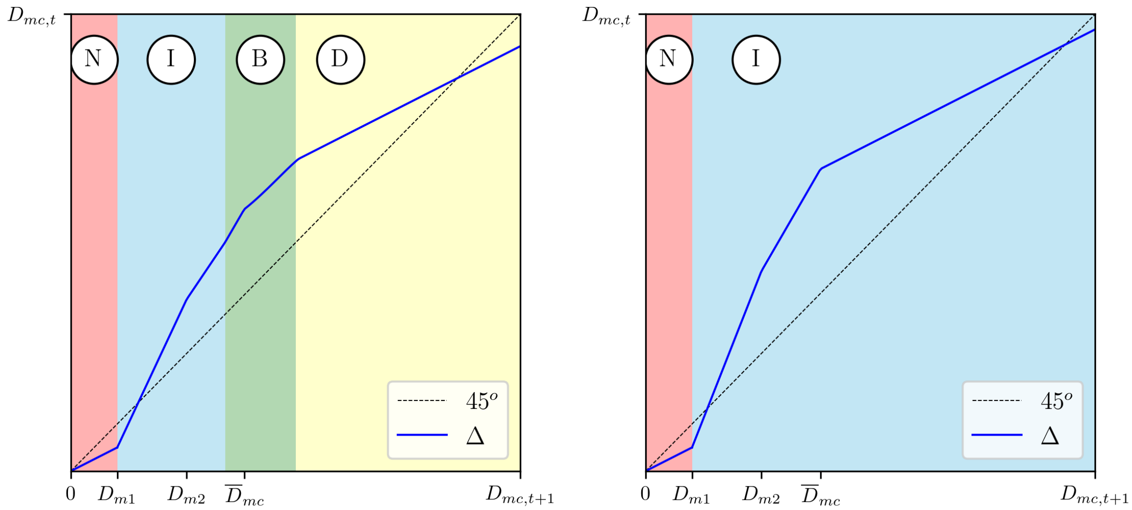

Figure 6: $D_{mc,t} = \Delta\left(D_{mc,t+1}\right)$ with $D_{pm} = \infty$; R low (left) or high (right).

anywhere between 0 and the top steady state, and all such equilibria converge to the middle steady state. These paths can entail Regime switching, but here switches are always toward the middle steady state – when Δ is monotone, there can be no deterministic cycles. However, there can be stochastic fluctuations, i.e., sunspot equilibria.

To pursue this, consider a random 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 might still affect endogenous variables. Here it suffices to consider a two-state process with time-invariant transition probabilities $\pi_j = \text{prob}(s_{t+1} = s_{-j}|s_t = s_j)$, and to focus on stationary sunspot equilibrium.

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

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

where the subscripts indicate variables now depend on the state $s \in \{1, 2\}$ but not the date. Then after some algebra, similar to reducing the steady state conditions to

$D_{mc} = \Delta(D_{mc})$ we can reduce the sunspot equilibrium conditions to

$$D_{mc,1} = (1-\pi_1)\Delta(D_{mc,1}) + \pi_1\Delta(D_{mc,2}) \tag{25}$$

$$D_{mc,2} = \pi_2\Delta(D_{mc,1}) + (1-\pi_2)\Delta(D_{mc,2}). \tag{26}$$

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

Following textbook methods (Azariadis 1993), we first solve (25)-(26) for the π's as functions of the D's:

$$\pi_1 = \frac{D_{mc,1} - \Delta(D_{mc,1})}{\Delta(D_{mc,2}) - \Delta(D_{mc,1})} \text{ and } \pi_2 = \frac{\Delta(D_{mc,2}) - D_{mc,2}}{\Delta(D_{mc,2}) - \Delta(D_{mc,1})}. \tag{27}$$

A pair $(D_{mc,1}, D_{mc,2})$ together with the probabilities in (27) is a sunspot equilibrium as long as $\pi_1, \pi_2 \in (0,1)$. Whenever there are three steady states, using Fig. 6 it is routine to show $\pi_1, \pi_2 \in (0,1)$ iff $D_{mc,1}$ is between 0 and the middle steady state while $D_{mc,2}$ is between the middle and top steady state. Hence, there are many equilibria where D_{mc} fluctuates randomly as a self-fulfilling prophecy.

Proposition 3: Assume Q is divisible, $D_{pm} = \infty$, and $D_{mc} = RD_{pc}$ are endogenous. When there are three steady states, there exist sunspot equilibria with D_{mc} fluctuating across any $D_{mc,1}$ between 0 and the middle steady state, and any $D_{mc,2}$ between the middle and top steady state, with transition probabilities given by (27).

While it is well known that a variety of models can have sunspot equilibria, the equilibria here are somewhat novel.[16] Here is the intuition for our results. First, the

[16] To explain what we mean by novel, consider some examples, and note that these remarks apply to the stochastic cycles here and the deterministic ones in Section 4.3. Gu et al. (2013a) get cycles and sunspot equilibria with endogenous credit but the method is different. Their dynamic system Δ has a steady state where it crosses the 45° line from above, not below. Then $\Delta' < -1$ implies cycles and sunspot equilibria (as is well known; more on this below). That is irrelevant here since $\Delta' > 0$. Moreover, their result depends on how agents determine the terms of trade (e.g., with Nash bargaining C's bargaining power must be low), while we have no such restriction. Myerson (2012) also gets credit cycles, using moral hazard, but the model is completely different. The same can be said for Azariadis and Kaas (2007,2013). It is also worth noting that Kiyotaki and Moore (1997) do not actually get endogenous fluctuations – they get amplification or propagation of fundamental shocks. Various monetary models yield cyclic and sunspot equilibria (Rocheteau and Wright 2013; Azariadis 1993), but that is driven by the idea that what you accept in trade depends on what others accept. As for dynamics in middlemen models, see Gu et al. (2025) and references therein, but note that they use different economics and mathematics (continuous-time bifurcation theory). In sum, while we do not claim to be doing anything "revolutionary," the results are not "already well known."

reason for multiple steady states with different debt limits is simple: 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 $\varepsilon > 0$, no matter how small ε is, and that makes $D = 0$ a steady state; but if they believe $D_t = \hat{D} > 0\ \forall t > t_0$ then taking away future credit can dissuade reneging at t_0, so there can be a steady state at $\hat{D} > 0$.

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 $D = \hat{D}$. Heuristically, this suggests there are equilibria fluctuating across $D_1 > 0$ and $D_2 \in (D_1, \hat{D})$, although technical conditions are needed to make it work (here, Δ should cross the 45° line from below).

These sunspot equilibria can involve random, recurrent Regime switches. To be clear, the dynamics are not due to the presence of middlemen *per se*, but to the self-referential nature of endogenous debt limits. Yet these fluctuations have implications for intermediation. Of course, whether middlemen accentuate or attenuate cycles is an old question (e.g., see Weill 2007 and references therein). In this specification, we can show intermediaries attenuate fluctuations on the extensive margin but amplify them on the intensive margin, in a precise sense, as we now discuss.

One can check that when $D_{mc,1}$ is low the constraint binds, and when $D_{mc,1}$ is high it may or may not bind. This implies $Q_{mc,1} < Q_{mc,2}$. Since D_{pc} is proportional to D_{mc}, we know $Q_{pc,1} < Q_{pc,2}$. Hence, when P brings lower Q_{pc} to RM, due to tighter credit conditions, M also brings lower Q_{mc}. There is causation beyond correlation. Since D_{pc} depends on S_{cm}, M can amplify cycles on the intensive margin. In some equilibria we switch between Regime I and B, so RM is always open, but only M participates when credit is tight while P and M participate when credit is loose. So there are nontrivial connections between volatility and intermediation, although in general the results depend on details.[17]

4.3 Deterministic Cycles

In Section 4.2 deterministic cycles cannot exist. We now show they can when we make Q indivisible, keeping the environment otherwise the same. One can check the results for steady states are similar to Proposition 2, and the D's are still proportional, so we can again analyze a univariate system. The difference is that now, with indivisible Q, it is possible to have $\Delta' < 0$, because higher D_{ic} means C pays more but does not get more, which does not happen when Q is divisible because then P produces less when D is lower.[18]

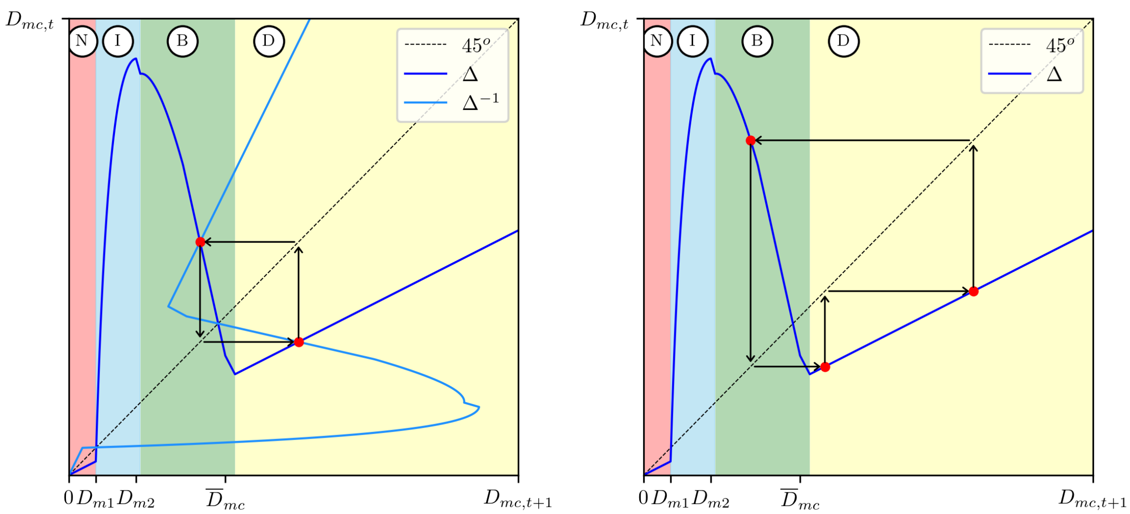

Figure 7: $D_{mc,t} = \Delta\left(D_{mc,t+1}\right)$ with $D_{pm} = \infty$; 2-cycle (left) and 3-cycle (right).

As Fig. 7 shows, sunspot equilibria like those constructed above still exist, since at the middle steady state Δ still crosses the 45° line from below. What is new with indivisible Q is that it is possible to get $\Delta' < -1$ at the top steady state. Then there are

[17] Suppose, e.g., that $D_{pm} = D_{mc} = \infty$. Then RM is always open with $\tau = 1$, and sunspot equilibria have M attenuating cycles on the intensive margin.

[18] In case it is not obvious, here is how cycles can 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 less inclined to default, so this period's endogenous D_{ic} is high. And vice versa if D_{ic} is high next period. Hence there is a tendency for D_{ic} to oscillate, similar to Gu et al. (2013a). 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 having S_{ci} decreasing in D_{ic} is necessary but not sufficient for $\Delta'(D_{ic}) < 0$.

2-cycles, as shown in the left panel, and hence sunspot equilibria by standard reasoning (Azariadis 1993). Indeed, we can find parameters that yield 3-cycles, fixed points of the third iterate of Δ, 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).

Hence indivisible Q can generate many kinds of equilibria, including deterministic and stochastic regime switching. We are agnostic about whether indivisible or divisible Q is a better assumption – presumably that depends on the application – but at least we know what happens if Q is indivisible.[19]

5 Conclusion

This paper studied environments with trade intermediated by middlemen and imperfect credit, where the latter is captured by debt limits that could be exogenous or endogenous. The contribution can be seen as putting payment frictions into models of intermediated trade, or putting middlemen into models of constrained credit. The results apply to markets for goods, inputs or assets, and to indivisible or divisible quantities. In a nutshell, we asked and answered this question: In an environment with limited commitment where agents can differ in their ability to enforce debt, what is the equilibrium outcome?

Exogenous debt limits implied existence and uniqueness, and we demonstrated exactly how parameters determine the pattern of exchange – in particular, when middlemen are active. With endogenous debt limits we showed there were multiple stationary equilibria, as well as equilibria with deterministic or stochastic fluctuations that are self-fulfilling prophecies. A message is that intermediation may or may not cause instability, but limited commitment provides an endogenous role for middlemen plus an endogenous source of dynamics.

[19] One issue with indivisible Q is that, as in other models with nonconvexities, there is an incentive to trade using lotteries: a seller cannot give a buyer $q < Q$, but a seller can give a buyer $q = Q$ with probability $\zeta < 1$ and they can bargain over ζ. Hence, allowing lotteries can make an indivisible-goods model look or work like a divisible goods model.

Future work might involve putting the framework to use in applications geared to particular kinds of markets or products.[20] Or it might involve designing tax or regulatory schemes to promote desirable outcomes. Another natural question is why intermediaries who provide credit often rely on third parties and why producers cannot do so (we thank a referee for raising this issue). One possibility is that different agents have different access to such third parties. Modeling this explicitly may be interesting. That said, classic credit cards issued by retailers like Sears and oil companies show that direct credit provision without third parties is not irrelevant.

Another possible extension is to endogenize monitoring (as Huang 2026 does in a different but related model). A related extension would have middlemen offering consumers higher credit than producers because middlemen know their consumers better (say, because they are local as suggested by the Associate Editor). Also, endogenizing the population – e.g., the number of producers or middlemen – has been shown to have interesting implications in other models (like Gu et al. 2025). Finally, introducing money is a good idea, since money and middlemen are alternative ways to deal with imperfect credit. There is room for much future work.

[20] Why does Apple give consumers a lower credit limit than third-party retailers like Amazon or Target? Why does Delta offer different payment plans than Expedia? Why do Toyota or Ford have stricter in-house financing than third-party financing companies that offer higher credit limits and more flexible terms? Why do carriers such as Verizon or T-Mobile have tighter financing limits than Best Buy or Walmart that extend credit through their store cards? These are interesting questions where formal models may be of some help.

Appendix A: Alternative Specifications

Section 3 focuses on divisible Q. Fig. 8 shows the analogous partition of (D_{pc}, D_{mc}) space when Q is indivisible, $\mathcal{Q} = \{Q\}$, using the same algorithm and parameters (Appendix D). The Regimes, cutoffs, and general logic are the same as for divisible Q: M is active when D_{pc} is low and D_{mc} is high; relaxing credit frictions can lead to disintermediation; and middlemen can be active solely due to their ability to enforce debt or due to bargaining power. Two differences stand out. First, the boundary separating Regime D from I/B is exactly the 45° line rather than concave, since with indivisible Q there is no marginal adjustment of quantity to blunt the comparison between D_{pc} and D_{mc}. Second, region N is larger than under divisibility (compare Fig. 1), confirming that divisibility encourages entry by letting sellers better cater to market conditions.

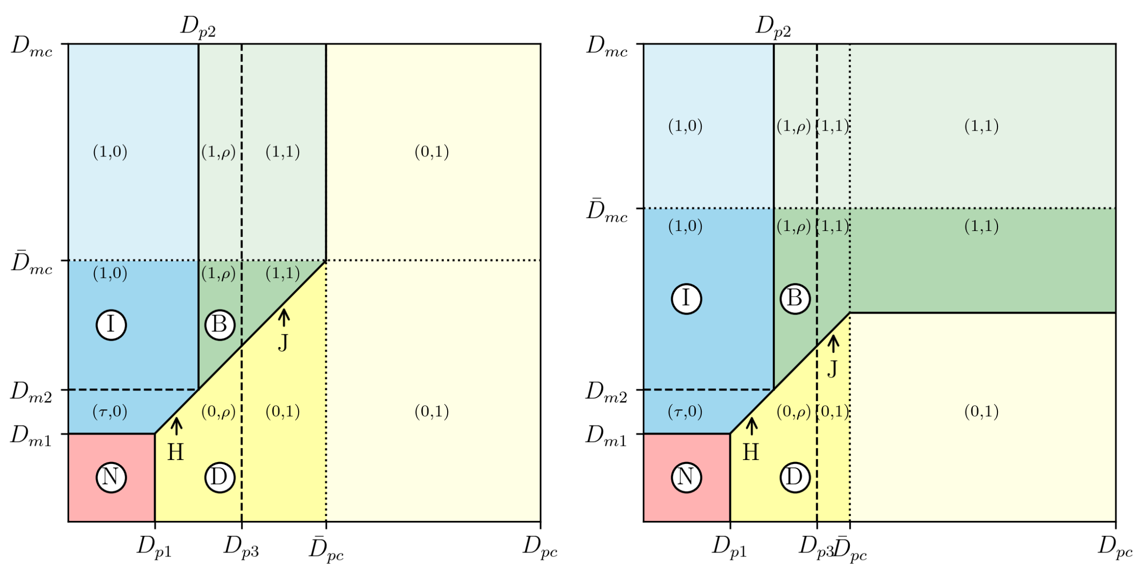

Figure 8: SE with Q indivisible and $D_{pm} = \infty$; $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

The baseline model has $D_{pm} = \infty$, so WM trade is unconstrained. Suppose now $D_{pm} < \infty$. Then τ is determined by (14), where both P and M need a positive surplus to trade, but if $S_{pm} + S_{mp} > 0$ a binding D_{pm} will not make $S_{mp} < 0$, so we only need to check $S_{pm} \geq 0$. One can show Proposition 1 still holds: SE exists uniquely for indivisible or divisible Q, although some details are slightly different.

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 P's opportunity cost of skipping RM, so all Regimes are possible. Either way, $D_{pm} < \infty$ is still tractable and does not change main messages.

Next suppose production occurs ex ante, before WM convenes. Let ϕ be the probability P produces, to be determined along with τ and ρ. Some details change (see the working paper at https://qiao-ziqi.github.io/asset/research/GQWW-working.pdf), but in terms of substance the main difference is that now there is a WM holdup problem since $c(Q)$ is sunk when P meets M. Still, SE exists uniquely, but as shown in Fig. 9 for indivisible Q, there are more cases since each element of (ϕ, τ, ρ) can be 0, 1 or mixed. Still the main insights hold.

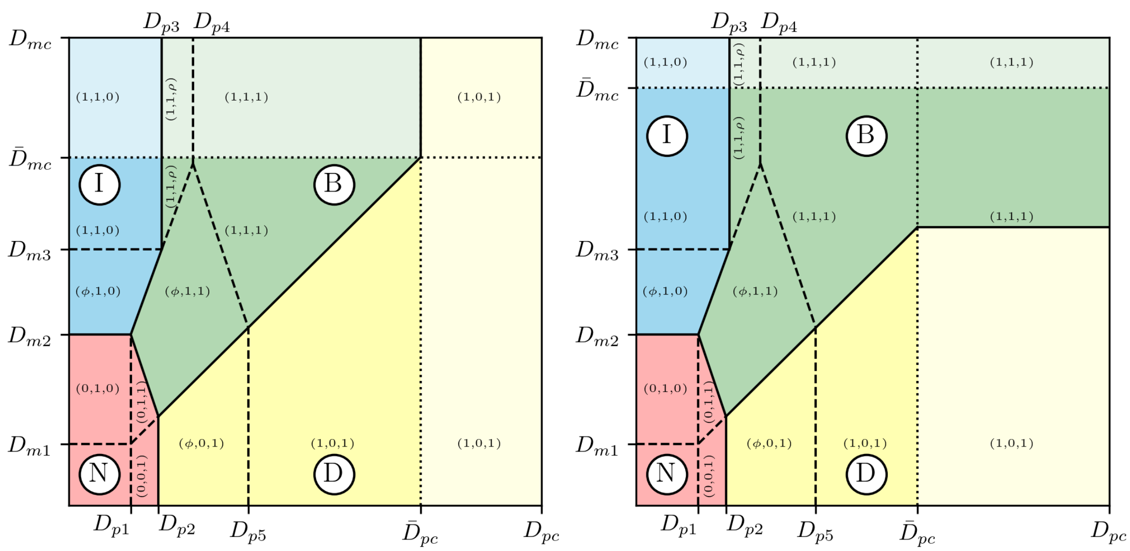

Figure 9: SE with ex ante prod. and $\theta_{mc} = \theta_{pc}$ (left) or $\theta_{mc} > \theta_{pc}$ (right).

Next we add scrap value: unsold RM output gets converted into CM numeraire according to $A(Q - q)$ with $A > 0$. This can increase the incentive for sellers to enter RM by reducing the risk of not meeting C. Also, after meeting C, it could affect the bargaining outcome through threat points. In fact, with Q divisible, $A > 0$ does not qualitatively affect the results. However, with Q indivisible in WM but divisible in RM, $A > 0$ can matter, and 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. In any case, one can check that SE still exists uniquely and the key results are similar to the baseline model.

Finally, in the spirit of Kiyotaki and Moore (1997), consider secured credit. Suppose C wants a car that can potentially be obtained in two ways: buy from P, where here P is not a producer, just someone with a car for sale; or buy from a dealer M. If C fails to make the required CM payment the seller can try to repossess the car, where the probability a type j seller succeeds is χ_j. The natural case is $\chi_m > \chi_p$. Here we revert to $D_{pm} = \infty$ and $A = 0$, and assume 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 $u(Q_{ic}) - p_{ic} \geq (1 - \chi_i)u(Q_{ic})$, so 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 $\chi_i \leq \theta_{ic}$. Hence, with indivisible Q, SE is the same as the baseline model with $D_{ic} = \chi_i u(Q)$. With divisible Q, $\chi_i > \theta_{ic}$ implies SE is the same but $\chi_i \leq \theta_{ic}$ implies $p_{ic} = \chi_i u(Q_{ic})$, where Q_{ic} solves $\alpha_{sc}\chi_i u'(Q_{ic}) = c'(Q_{ic})$. Still, SE exists uniquely, and dependence on parameters is similar to the baseline model with χ replacing D. However, now reductions in credit frictions, higher χ, raise p as well as p/Q since χ is effectively sellers' bargaining power. Again, this extension does not change the benchmark results.

Appendix B: Proof of Proposition 1

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) solve the dynamic programming equations for the V's taking p's as given (easy, since the system is linear); (4) use the bargaining solution to get the p's; (5) check the best response conditions, since the variables just constructed constitute an SE iff (13) and (15) hold; (6) repeat until exhausting possible strategy profiles. For divisible Q the procedure is the same, just add as a constraint the marginal condition for the choice of Q.

We now derive the borders in Figs. 1 and 8 directly. 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

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

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

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 $\bar{D}_{mc} = \theta_{mc}u\left(Q_{mc,1\rho}\right)$. 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 an 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 an 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 an 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 an 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 an SE when $D_{pc} \leq D_{p2}$ and $D_{mc} \geq D_{m2}$.

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

This partitions parameter space as shown in Figs. 8 and 1. Therefore, for all parameters there is one and only one SE. ■

Appendix C: Details for Proposition 2

We can write down the meeting rate and corresponding surplus in piecewise form. Note that $\alpha_{ci,t}$ depends on τ_t and ρ_t, which depend on $D_{ic,t}$ via (15) and (13), while $S_{ci,t}$ depends on $D_{ic,t}$ via (2) and (7). There are various (τ, ρ) pairs consistent with equilibrium, and we can group them into two scenarios. When $R \leq 1$, $D_{mc,t} \leq D_{pc,t}$ and $\tau = 0$; when $R > 1$, $D_{mc,t} > D_{pc,t}$ and $\tau > 0$. The former case is rather straightforward with $\alpha_{cm,t} = 0$ for all t. We lay out the latter case in which M has some advantage and is potentially active. Using the cutoffs from Fig. 1, for the pc meeting we have:

$$\alpha_{cp,t}\left(D_{pc,t}\right) = \begin{cases} 0 & \text{if } D_{pc,t} \leq D_{p2} \\ m_R(N_{s,t}/N_c, 1)(1 - \alpha_{pm}N_p/N_{s,t}) & \text{if } D_{p2} < D_{pc,t} < D_{p3} \\ \bar{\alpha}_{cs}(1 - \alpha_{pm}) & \text{if } D_{p3} \leq D_{pc,t} < \bar{D}_{pc} \\ [\bar{\alpha}_{cs}(1 - \alpha_{pm}), \bar{\alpha}_{cs}] & \text{if } D_{pc,t} = \bar{D}_{pc} \\ \bar{\alpha}_{cs} & \text{if } D_{pc,t} > \bar{D}_{pc} \end{cases} \tag{33}$$

$$S_{cp,t}\left(D_{pc,t}\right) = \begin{cases} 0 & \text{if } D_{pc,t} \leq D_{p2} \\ (1/\theta_{pc} - 1)D_{pc,t} & \text{if } D_{p2} < D_{pc,t} < \bar{D}_{pc} \\ u(Q^*_{pc}) - \bar{D}_{pc} & \text{if } D_{pc,t} \geq \bar{D}_{pc} \end{cases} \tag{34}$$

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 $m_R(1, N_c/N_{s,t})D_{pc,t} = c(D_{pc,t}/\theta_{pc}) + \kappa$. Note that the credit-limit cutoffs depend on both the arrival rate and the bargaining power, while the arrival rate itself depends on entry decisions and hence on $D_{pc,t}$.

For the mc meeting, we have:

$$\alpha_{cm,t}\left(D_{mc,t}\right)=\begin{cases}0 & \text{if } D_{mc,t}\leq D_{m1}\\ m_R(N_{m,t}/N_c,1) & \text{if } D_{m1}<D_{mc,t}<D_{m2}\\ m_R(\alpha_{mp}N_m/N_c,1) & \text{if } D_{m2}\leq D_{mc,t}<RD_{p2}\\ m_R(N_{s,t}/N_c,1)\alpha_{mp}N_m/N_{s,t} & \text{if } RD_{p2}\leq D_{mc,t}<RD_{p3}\\ \bar{\alpha}_{cs}\alpha_{pm} & \text{if } RD_{p3}\leq D_{mc,t}<R\bar{D}_{pc}\\ [0,\bar{\alpha}_{cs}\alpha_{pm}] & \text{if } D_{mc,t}=R\bar{D}_{pc}\\ 0 & \text{if } D_{mc,t}>R\bar{D}_{pc}\end{cases} \tag{35}$$

$$S_{cm,t}\left(D_{mc,t}\right)=\begin{cases}0 & \text{if } D_{mc,t}\leq D_{m1}\\ (1/\theta_{mc}-1)D_{mc,t} & \text{if } D_{m1}<D_{mc,t}<\bar{D}_{mc}\\ u(Q^*_{mc})-\bar{D}_{mc} & \text{if } D_{mc,t}\geq\bar{D}_{mc}\end{cases} \tag{36}$$

where $N_{m,t}$ satisfies M's entry condition $m_R(1,N_c/N_{m,t})D_{mc,t}=c(D_{mc,t}/\theta_{mc})+\kappa$ and $Q^*_{mc}=Q^*_{pc}$ given $\theta_{pc}=\theta_{mc}$. Different from $\alpha_{cp,t}$, $\alpha_{cm,t}$ is zero on both ends of $D_{mc,t}$, because when P is credit unconstrained, M does not participate without bargaining advantage. Also notice that both $\alpha_{cp,t}$ and $\alpha_{cm,t}$ are set-valued when P is just credit unconstrained, since any $\tau\in[0,1]$ can be supported at this point. Nonetheless, these probabilities always add up to $\bar{\alpha}_{cs}$ so that (19) is single-valued at that point.

Appendix D: Parameters for Figures

The matching function is $m_k(N_i,N_j)=a_kN_iN_j/(N_i+N_j)$ with $a_W=1$ for WM and $a_R=0.8$ for RM. The utility function is $u(q)=\omega_0q^{\omega_1}$. The cost function is $c(Q)=\psi_0Q^{\psi_1}$ with $(\psi_0,\psi_1)=(0.01,2)$, $\kappa=0.15$. Bargaining power in WM is $\theta_{pm}=0.5$. For indivisible Q, the quantity is fixed at $Q=2$.

In Section 3, $(\omega_0,\omega_1)=(1,0.5)$ and $N_p=N_m=N_c=1$. When RM bargaining powers equal, $\theta_{pc}=\theta_{mc}=0.5$; otherwise, $(\theta_{pc},\theta_{mc})=(0.4,0.6)$. Then Fig. 2 has $D_{pc}\in\{0.2,0.5\}$, Fig. 3 has $D_{mc}\in\{0.25,0.75\}$, and Fig. 9 has $D_{pc}\in\{0.2,0.5\}$ and $D_{mc}=0.75$.

In Section 4, $(N_p,N_m,N_c)=(3,3,1)$ and $\beta=0.5$. Figs. 4-6 have $\theta_{ic}=0.5$. Fig. 5 has $R_p=1.7$ (left) and $R=2.1$ (right). Fig. 6 has $R_p\in\{3,1\}$, $R_m\in\{4,5\}$, $(\omega_0,\omega_1)=(2,0.1)$. Fig. 7 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] C. Azariadis (1993) *Intertemporal Macroeconomics.*

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

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

[5] 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.

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

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

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

[9] S. Claessens and M.A. Kose (2017) "Macroeconomic implications of financial imperfections: a survey," BIS Working Paper 677.

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

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

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

[13] 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.

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

[15] N. Ferguson (2008) *The Ascent of Money: A Financial History of the World.*

[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, Z. Qiao and R. Wright (2026) "Middlemen in search equilibrium: A survey," *Elements in Money and Banking*, ed. C. Gu and J. Haslag.

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

[19] C. Gu, F. Mattesini, C. Monnet and R. Wright (2013a) "Endogenous credit cycles," *JPE* 121, 940-65.

[20] C. Gu, F. Mattesini, C. Monnet and R. Wright (2013b) "Banking: A new monetarist approach," *RES* 80, 636-62.

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

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

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

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

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

[26] A. Huang (2026) "On the number and size of banks: Efficiency and equilibrium," mimeo.

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

[28] J. Kareken and N. Wallace (1980) "Introduction," in *Models of Monetary Economies*, ed. J. Kareken and N. Wallace, Federal Reserve Bank of Minneapolis.

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

[30] J. Keynes (1936) *The General Theory of Employment, Interest and Money.*

[31] C. Kindleberger (1978) *Manias, Panics and Crashes: A History of Financial Crises.*

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

[33] N. Kiyotaki and R. Wright (1989) "On money as a medium of exchange," *JPE* 97, 927-54.

[34] N. Kiyotaki and R. Wright (1993) "A search-theoretic approach to monetary economics," *AER* 83, 63-77.

[35] N. Kocherlakota (1998) "Money is memory," *JET* 81, 232-251.

[36] R. Lagos and G. Rocheteau (2007) "Search in asset markets: Market structure, liquidity, and welfare," *AER* 97, 198-202.

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

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

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

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

[41] G. Lorenzoni (2008) "Inefficient credit booms," *RES* 75, 809-33.

[42] S. Lotz and C. Zhang (2026) "Supply chains," mimeo.

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

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

[45] H. Minsky (1992) *The Financial Instability Hypothesis*. Vol. 74, Levy Econ Institute WP.

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

[47] R. Myerson (2012) "A model of moral-hazard credit cycles," *JPE* 120, 847-78.

[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] E. Nosal, Y. Wong and R. Wright (2019) "Intermediation in markets for goods and markets for assets," *JET* 183, 876-906.

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

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

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

[53] G. Rocheteau and R. Wright (2013) "Liquidity and asset-market dynamics." *JME* 60, 275-94.

[54] A. Rubinstein and A. Wolinsky (1987) "Middlemen," *QJE* 102, 581-93.

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

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

[57] S. Shi (1996) "Credit and money in a search model with divisible commodities," *RES* 63, 627-52.

[58] R. Shiller (1988) "Causes of changing financial market volatility," *Proceedings, Federal Reserve Bank of Kansas City*, 1-32.

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

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

[61] K. Tretina and K. Little (2025) "When were credit cards invented?" https://finance.yahoo.com/personal-finance/credit-cards/article/when-were-credit-cards-invented-202145703.html.

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

[63] N. Wallace (1996) "A dictum for monetary theory," in *Foundations of Research in Economics: How Do Economists Do Economics?* ed. S. Medema and W. Samuels.

[64] N. Wallace (2010) "The mechanism-design approach to monetary theory," in *Handbook of Monetary Economics* vol. 3.

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

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

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

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

[69] M. Yang and Y. Zeng (2021) "Coordination and fragility in liquidity provision," mimeo.