Breaking Up With M: Cashless Limits Under Limited Commitment*

Tai-Wei Hu
Bristol University

Guillaume Rocheteau
University of California, Irvine

This version: March 2026.

Abstract

We examine the welfare implications of eliminating cash in nearly cashless economies. Our model features money and credit coexisting under limited commitment. The analysis yields four key insights. First, banks' market power is neither necessary nor sufficient for cash elimination to affect aggregate welfare, but it is sufficient to generate distributional effects. Second, the welfare consequences depend on the source of bank market power, bargaining strength versus information. Third, removing cash generates positive welfare gains if limited commitment is achieved through public record-keeping. Finally, cash elimination can create financial exclusion even when all agents are initially banked.

JEL Classification: E40, E50

Keywords: money, limited commitment, cashless limit

*We thank for their comments from Wei Cui, Russell Wong, Yidan Yan, Yu Zhu, and seminar participants at Renmin University of China, University of Bristol, University College of London, University of California at Irvine, University of Essex, University of Hawaii at Manoa, University of Paris Pantheon-Assas.

1 Introduction

Advances in financial innovation and digital payment technologies are pushing many economies toward cashlessness, prompting Rogoff (2014) to argue that "we may already live in the twilight of the paper currency era."[1] Against this backdrop, this paper asks a simple but fundamental question: does eliminating cash—a government-issued tangible means of payment—have implications for real economic outcomes and welfare in an economy that is nearly cashless, where cash usage has become quantitatively negligible?

Woodford (1998) answers in the negative. He argues that money has become both quantitatively negligible and analytically irrelevant, and can therefore be dispensed with altogether—a view that led a generation of economists and policymakers to adopt cashless models. Recent work by Lagos and Zhang (2022) and Lagos (2025), however, challenges this conclusion by showing that when financial intermediaries possess market power, the mere availability of fiat money can affect real economic outcomes. In their environment, eliminating cash *reduces* aggregate welfare.

In order to resolve this debate over the role of money in nearly cashless economies, one needs a model that features the coexistence of fiat money—government-provided means of payment—and competing inside money—bank-issued liabilities—where the latter can substitute for the former. This coexistence, however, poses a theoretical challenge: the very frictions that sustain the value of fiat money—such as limited enforcement or the absence of record-keeping—also tend to undermine the feasibility of inside money (e.g., Kocherlakota, 1998). Yet these frictions must be modeled explicitly to account for the transition toward cashlessness, including advances in payment technologies.

In this paper, we treat *limited commitment* as the fundamental friction that renders credit market imperfect and thereby rationalizes the joint use of fiat money and credit in equilibrium.[2] To mitigate limited commitment, we introduce explicit bank-operated enforcement and monitoring technologies. Because bank market power is central to the debate, we model the market structure and informational frictions that govern the formation of banking relationships.

The distributional relevance of cash under complete information Throughout the paper, we abstract from two common motivations for eliminating cash—namely, tax evasion

[1] For some empirical evidence, see Section 1.1.

[2] According to Kiyotaki and Moore (2002), "placing a limitation on the degree of commitment (...) is the right starting point for a theory of money." In Woodford (1998) and Lagos and Zhang (2022), the ability of a subset of agents to use credit to pay for a subset of goods is exogenous, and credit is frictionless in the sense that loan repayments are perfectly enforced.

and its use in criminal activities—in order to focus squarely on the role of cash in the cashless limit.[3] Consequently, any welfare gains from removing cash in our framework do not arise from suppressing welfare-reducing activities that are assumed to be financed with cash.

We begin by developing a simple, competitive monetary model in the spirit of Townsend (1980) and Rocheteau and Wright (2005), in which all markets are competitive and open sequentially. Agents are divided into two groups: those who receive positive labor endowments every period and those who do so only every other period. We refer to the former as sellers and to the latter as buyers. Buyers acquire means of payment to finance consumption in periods in which they do not work. Moreover, they are heterogeneous in their liquidity needs allowing us to discuss financial exclusion.

The banking sector is described as an over-the-counter market where non-financial agents and banks form lending relationships, as in Bethune et al. (2022), and where banks have access to costly commitment and enforcement technologies. The terms of the lending contracts—the interest rate and loan size—are negotiated between buyers and banks. In equilibrium, the share of buyers that are financially included and trade using banks' inside money is endogenously determined.[4] As the cost of the commitment technology vanishes, aggregate real money balances and financial exclusion converge to zero—the economy becomes cashless. At the cashless limit, when cash is removed, aggregate welfare remains constant but banks are better off while consumers are worse off.

TAKE-AWAY #1A. *In near-cashless economies with costly external enforcement and complete information eliminating cash is neutral for aggregate welfare, yet it induces redistributional effects whenever banks have positive bargaining power.*

To underscore the significance of market structure, we consider a variant of the environment in which banks post contracts while buyers direct their search, as modeled in the competitive search literature. In this setting, banks' market power is endogenous with the frictions that generate financial exclusion. As entry costs converge to zero, all buyers obtain access to credit, and the economy approaches the cashless limit. In that limit, banks' endogenous bargaining power vanishes, so the removal of cash has no effect on real allocations or welfare, thereby restoring Woodford's (1998) irrelevance result.

TAKE-AWAY #1B. *In near-cashless economies with costly external enforcement, directed search, and endogenous bank entry, removing cash is neutral with respect to both welfare and*

[3] The welfare implications of reducing or deterring cash usage, motivated by concerns that cash transactions facilitate tax evasion and other illicit activities, are studied, for example, in Williamson (2012) and Rogoff (2016).

[4] Freeman and Kydland (2000) adopts a similar approach to endogenize the fraction of credit trades. However, in their model there are hterogenous goods and a buyer has to pay a fixed cost to use credit for each different good. We focus on heterogeneous liquidity needs of buyers instead.

distribution.

The aggregate relevance of cash under private information We explore the sources of banks' market power by letting banks set contract terms but by allowing buyers' borrowing needs to be privately known, generating informational rents that constrain banks' power. As the fixed cost of the bank's commitment technology converges to zero, all agents gain access to credit—there is no financial exclusion—and the economy approaches a cashless limit. However, there is credit rationing in the sense that banks offer lower debt limits as compared to the complete-information benchmark due to incentive compatibility. Paradoxically, although the demand for cash becomes arbitrarily small, eliminating the availability of cash to buyers induces banks to exclude a positive measure of agents from credit and to reduce loan sizes for others. Moreover, there is more distortion of credit rationing when cash is removed as banks extract more rents.

TAKE-AWAY #2. *In near-cashless economies with monopolistic banks, privately informed consumers, and costly external enforcement, the removal of cash reduces aggregate welfare by generating financial exclusion and exhacerbating credit rationing, while increasing bank profits.*

We show that the source of market power—whether it stems from bargaining power or informational advantages—is fundamental in determining the effects of cash removal. We start with a situation in which banks possess neither bargaining power nor information about buyers' liquidity needs. We then increase banks' market power in two stages: first, by allowing banks to set the terms of loan contracts; second, by granting them information. For each scenario, we evaluate the welfare loss associated with removing cash.

TAKE-AWAY #3. *In near-cashless economies, removing cash when banks have no bargaining power and no information is neutral. If banks gain bargaining power while remaining uninformed, cash removal reduces aggregate welfare. If banks have full bargaining power and information, cash removal has no effect on aggregate welfare.*

The detrimental role of cash under public monitoring Our second approach to limited commitment focuses on self-enforcement. We introduce a costly record-keeping technology operated by banks that generates reputational losses upon default. As in Kehoe and Levine (1993) and Alvarez and Jermann (2000), monitored agents can borrow up to an endogenous debt limit equal to the utility loss they would suffer if excluded from future credit relationships. Only a subset of agents chooses to incur the cost of accessing this monitoring arrangement due to the heterogeneity in buyers' liquidity needs.

The elimination of cash affects welfare in two ways. First, it increases the value of being banked and relaxes the incentive constraints that govern debt limits, thereby raising aggregate welfare. Second, it weakens buyers' bargaining position and shifts surplus toward banks, which can reduce buyers' welfare.

TAKE-AWAY #4. *In near-cashless economies with costly record-keeping, assuming the inflation rate is not too high, the elimination of cash raises aggregate welfare regardless of whether banks possess market power. A fraction of buyers can be made worse off if banks have some positive bargaining power.*

In summary, our framework yields the following insights. First, banks' market power is neither necessary nor sufficient for fiat money to remain economically relevant—in terms of aggregate welfare—even when its real value is quantitatively negligible. Second, the welfare consequences of eliminating cash do not vary monotonically with the degree of banks' market power. Third, these effects need not be negative. Finally, in economies where both financial exclusion and money demand are nearly absent, removing cash can itself generate financial exclusion.

1.1 Empirical evidence

Since our analysis focuses on economies that are already nearly cashless, this section reviews empirical evidence on the evolution of different countries toward cashlessness. First, there exist substantial cross-country differences in the use of cash. In 2023, the ratio of cash in circulation to GDP was approximately 23 percent in Japan, 8 percent in the United States, and less than 1 percent in Sweden, the archetype of a near-cashless economy (left panel of Figure 1).[5] More broadly, across both advanced and developing economies, cash in circulation averages about 7.5 percent of GDP and does not exhibit a sustained downward trend. In Japan, for example, the ratio increased from 18.8 percent in 2014 to 23 percent in 2022.

Cash in circulation is an imperfect proxy for the use of cash as a means of payment (Khiaonarong and Humphrey, 2023).[6] A more informative indicator is the ratio of cash withdrawals to GDP, reported in the right panel of Figure 1. By this measure, reliance on cash has declined markedly over the past decade. In advanced economies, cash withdrawals fell from 13 percent of GDP in 2013 to 6 percent in 2023. In developing economies, they declined from 31 percent to 15 percent over the same period. This downward trend is

[5] This data can be found here: https://data.bis.org/topics/CPMI_CT/tables-and-dashboards

[6] In particular, large-denomination banknotes are often held as a store of value rather than used for transactions (Di Iorio, Kosse, and Mustafi, 2025).

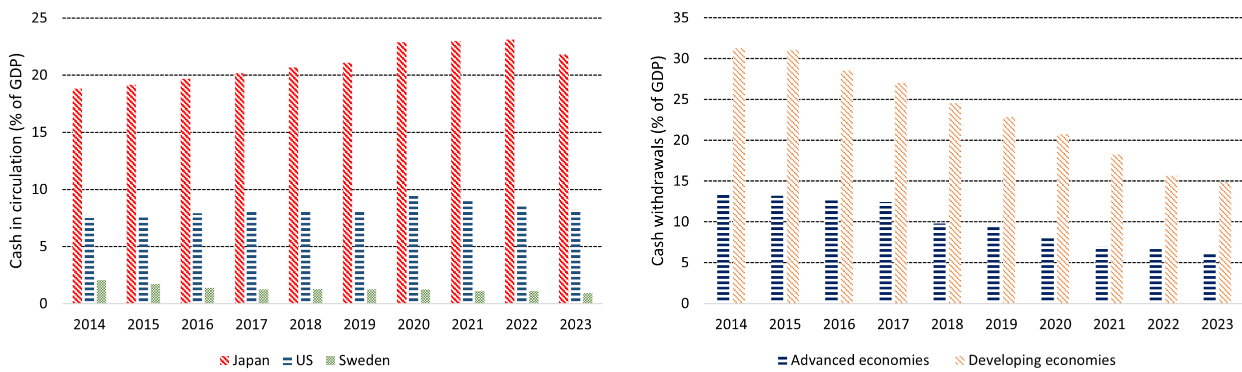


Figure 1: Left: Evolution of the cash in circulation in different countries. Right: Cash withdrawals. Source: Di Iorio, Kosse and Mustafi (2025).

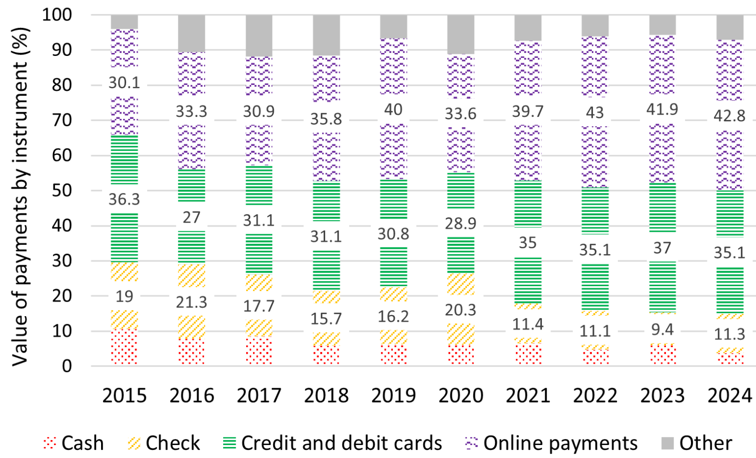
observed in most countries, with a few notable exceptions, such as Mexico (Khiaonarong and Humphrey, 2023).

In our model, we will allow cash to compete against inside money provided by banks. Therefore, we now look at the evolution of consumers' use of competing means of payment. The U.S. Diary of Consumer Payment Choice shows a substantial shift in payment behavior over recent years (Bayeh et al., 2025). The share of the number of payments made in cash declined from 33 percent in 2015 to 14 percent in 2024, while the share made by credit card increased from 18 percent to 35 percent, making credit cards the most frequently used payment method. Debit card usage has remained relatively stable at approximately 30 percent.

Figure 2 reports the value of payments by type of payment instruments. Cash represented 10 percent of payments in 2015 and less than 4 percent in 2024. Payment by cards represent about 35 percent of all payment and online payments (bank account number payment and online banking bill payment) 42 percent. By that metric, the US economy is nearly cashless. Despite the gradual disappearance of cash, consumers keep carrying on-person cash, on average \$67, and more than 92% of respondents stated that they had no plans to stop using cash.

Our model will emphasize heterogeneity in terms of liquidity needs across consumers and will endogenize financial inclusion. Evidence from the Diary of Consumer Payment Choice indicates that the decline in cash usage is substantially more pronounced among younger individuals suggesting that the aggregate decline in cash usage is likely to persist (Bayeh et al., 2025). The survey also documents systematic differences in payment behavior across income groups. Low-income households use cash for a larger share of their transactions. Unbanked consumers—who are disproportionately concentrated among low-income

Diary Consumer Payment
Choice



$_2$.pdf

Figure 2: Dollar value of payments by type of payment instrument. Source: Survey and Diary of Consumer Payment Choice

households—conduct approximately two thirds of their transactions in cash.

Since a fundamental component of our model is the structure of the bank loan market, we briefly review evidence on banks' market power in deposit and lending markets and its implications for the transmission of monetary policy. On the deposit side, Drechsler, Savov, and Schnabl (2017) provide evidence of market power by documenting imperfect pass-through of the federal funds rate to deposit rates and demonstrating that this pass-through declines with market concentration. On the lending side, Scharfstein and Sunderam (2017) find that the transmission from MBS yields to mortgage rates weakens as mortgage market concentration rises. Wang et al. (2022) quantify the role of bank market power in monetary policy transmission and conclude that it accounts for a substantial share of the overall pass-through to borrowers. Using U.S. data, Head et al. (2025) provide additional evidence of market power by documenting significant within-market dispersion in loan and deposit interest rates.

1.2 Literature

Formalizing environments where money and credit coexist is an integral part of monetary theory, e.g., in the context of the New-Monetarist literature, Gu, Mattesini, and Wright (2016) and the references therein. Surveys include Lagos, Rocheteau, and Wright (2017) and Rocheteau and Nosal (2017).

Lagos and Zhang (2022) provide such a model where banks' primary role is to connect sellers with credit-buyers. A key feature of their model is a holdup problem arising from irreversible production decisions. Our environment is related in its description of the market structure but it specializes on the market for lending relationships, as in Bethune et al. (2022). It does not feature holdup problems and it considers different information structures and trading mechanisms, including the mechanism design problems of monopolist banks. Moreover, we relax the assumption of perfect enforcement and we formalize financial exclusion and endogenous participation to the credit market along the lines of Freeman and Kydland (2000), Li (2011), and Rocheteau and Nosal (2017, Chapter 8.3).

The market for bank loans is described as an over-the-counter market with search and bargaining, as in Duffie, Garleanu, and Pedersen (2005) in the context of financial assets. For a review of this literature, see Hugonnier, Lester, and Weill (2025).

The idea that central bank money can operate as an outside option disciplining banks' market power appears in the literatures on CBDC (e.g., Anfolfatto, 2021; Chiu et al., 2023), bank lending (e.g., Rocheteau, Wright, and Zhang, 2018; Head et al., 2025), and the deposits channel of monetary policy (e.g., Choi and Rocheteau, 2023).

Our model introduces limited commitment as the fundamental friction to make money valuable. Banks mitigate this friction by engaging in costly monitoring to supply buyers with credit instruments that function as means of payment, in the spirit of the literature on delegated monitoring pioneered by Diamond (1984) and surveyed by Freixas and Rochet (1997, Section 2.4). We show that the welfare consequences from eliminating cash depend on the source of market power and the informational rents that arise in equilibrium. The version of our model with monopolistic banks solving a mechanism design problem is related to Choi and Rocheteau (2023), except that we focus on the credit market rather than the deposit market.

In order to establish that market power is not necessary for the relevance of cash at the cashless limit, we adopt an approach with self-enforcing credit under a record-keeping technology, as in Kehoe and Levine (1993). Applications within New Monetarist environments include Cavalcanti and Wallace (1999), Sanches and Williamson (2010), Gu et al. (2013a,b), and Bethune, Hu, and Rocheteau (2018a,b), among others. Kocherlakota and Wallace (1998) addresses a related question to ours by formalizing technological advances in payments within a random-matching economy where money coexists with a public record of all past actions that is updated with a lag. In comparison to these papers, we obtain the cashless limit by introducing heterogeneity in credit access across agents and by introducing bank market power, so that the version with no bank market power is a special case. We show that the welfare effect of removing cash reverses relative to the environment with externally enforceable credit.

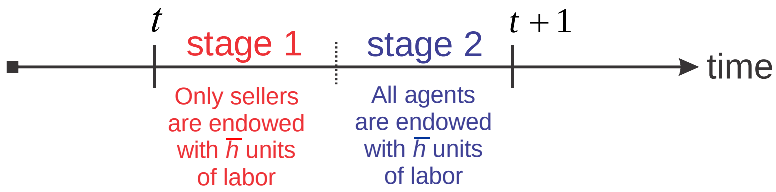


Figure 3: Description of time and endowments

2 Environment

We formalize a competitive monetary economy in the spirit of Townsend (1980) and Rocheteau and Wright (2005) with long-term lending relationships as in Bethune et al. (2022). Time is discrete, indexed by $t \in \mathbb{N}_0$. Each period consists of two stages, $\tau \in \{1, 2\}$, as illustrated in Figure 3, in which a good is traded in a perfectly competitive market. The economy features a continuum of nonfinancial agents divided into two groups, $\varkappa \in \{1, 2\}$, of unit measure each. As will become clear later, type-1 agents—who receive an endowment in only one of the two stages—play the central role in determining the demand for payment instruments. We will label the type-1 agents "*buyers*" and the type-2 agents "*sellers*" to reflect their trading behavior in stage 1.

Agents' lifetime preferences over consumption and work are represented by

$$\mathbb{E}_0 \sum_{t=1}^{+\infty} \beta^t \mathcal{U}_t \;\text{ with }\; \mathcal{U}_t = \sum_{\tau=1}^{2} \left[\varepsilon_\tau u(c_{t,\tau}) - h_{t,\tau}\right],$$

where $\mathcal{U}_t$ is the period utility, $\beta \equiv (1 + \rho)^{-1} \in (0, 1)$ is the discount factor, $c_{t,\tau}$ denotes consumption in stage $\tau \in \{1, 2\}$ of period t, and $h_{t,\tau}$ denotes labor supply. The utility function, $u(c)$, is twice continuously differentiable and satisfies $u(0) = 0$, $u'(c) > 0$, $u''(c) < 0$, and $u'(0) = +\infty$.

The utility-weight, $\varepsilon_\tau \in \mathbb{R}^+$, is normalized to one in stage 2 ($\varepsilon_2 = 1$) for all agents and in stage 1 ($\varepsilon_1 = 1$) for sellers. In contrast, buyers are heterogeneous in terms of ε_1 which is drawn at time 0 from a continuous distribution, $F(\varepsilon)$, with density $f(\varepsilon)$ and support $[0, \bar{\varepsilon}]$. The assumptions regarding utility weights are summarized in Table 1. We denote by c^*_ε the unique solution to $\varepsilon u'(c^*_\varepsilon) = 1$ for all $\varepsilon \in \mathbb{R}^+$.

agent's ype → / utility weight ↓	$\varkappa = 1$ (buyers)	$\varkappa = 2$ (sellers)
Stage 1: ε_1	$\sim F(\varepsilon)$	=1
Stage 2: ε_2	=1	=1

Table 1: Utility weights

In each period, there is a technology to transform labor into consumption goods one for one. Each agent of type $\varkappa$ can supply up to $\zeta_\tau^\varkappa \bar{h}$ units of labor in stage τ, where $\zeta_\tau^\varkappa \in \{0, 1\}$. Assumptions regarding $\zeta_\tau^\varkappa$ are summarized in Table 2. For buyers, $\zeta_1^1 = 0$ and $\zeta_2^1 = 1$. Thus, buyers can only work in stage 2. By contrast, for sellers, $\zeta_1^2 = \zeta_2^2 = 1$. So, sellers can work in both stages. We assume that $\bar{h}$ is sufficiently large so that the constraint, $h_{t,\tau} \leq \zeta_\tau^\varkappa \bar{h}$ never binds in equilibrium whenever $\zeta_\tau^\varkappa = 1$.

agent's type → / endowment ↓	$\varkappa = 1$ (buyers)	$\varkappa = 2$ (sellers)
Stage 1: $\zeta_1^\varkappa$	=0	=1
Stage 2: $\zeta_2^\varkappa$	=1	=1

Table 2: Labor endowments

We assume that agents cannot borrow across periods because of limited commitment. However, a subset of agents can obtain intra-period loans from banks. Banks are agents with linear utility over stage-2 consumption that have access to costly technologies to commit and enforce. In stage 1, banks issue their own IOUs in exchange for buyers' IOUs. Bank-issued IOUs can circulate as means of payment—they are inside money. Thus, banks issue liquid IOUs as liabilities, interpreted as deposits, and hold illiquid IOUs as assets, corresponding to loans to nonfinancial agents, the buyers. Figure 4 represents the trades between banks, buyers, and sellers in stages 1 (left panel) and 2 (right panel). An alternative and equivalent interpretation is that buyers can directly issue IOUs to sellers, while banks act as guarantors of those obligations in exchange of an intermediation fee. We adopt this interpretation in what follows, thereby restricting attention to a single type of debt.

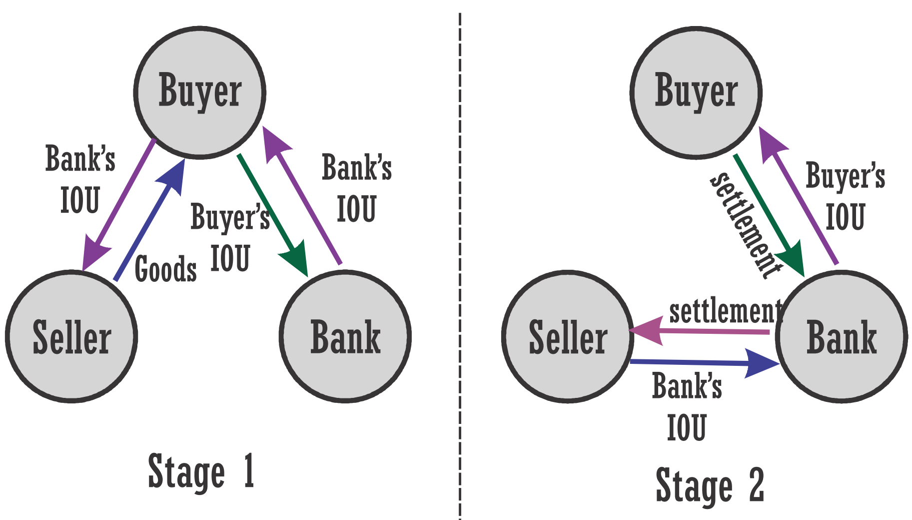


Figure 4: Bank credit

Relationships between banks and buyers are bilateral and are formed in stage 2 of the initial period, $t = 0$, through one-to-one random matching. In order to commit to honor their obligations, banks incur a fixed cost every period from $t = 1$ onward equal to $k > 0$ in terms of stage-2 consumption good. In addition, there is a cost for banks to monitor loans and enforce repayments.[7] In our model, the per-period cost of monitoring a loan of size d (expressed in stage-2 good) is a real resource cost equal to σd, with $\sigma > 0$.[8] Without loss, we assume that sellers do not have access to banks hence cannot issue debt.[9]

Fiat money takes the form of an intrinsically useless asset issued by the government that is storable at no cost. Its supply, M_t, grows at rate π, i.e., $M_{t+1} = (1 + \pi)M_t$, through lump-sum transfers (or taxes if $\pi < 0$) to buyers. We denote $1 + i \equiv (1 + \rho)(1 + \pi)$ and we assume $i > \sigma$.

Discussion of the environment

We briefly discuss several assumptions and interpretations of the environment. The model features two means of payment: outside money and private debt. We identify cash with outside money—a tangible government-issued asset that is not a liability of private agents. Because it can facilitate transactions without enforcement or monitoring technologies, it is immune to limited-commitment frictions. By contrast, credit arrangements are subject to limited commitment and therefore require some form of monitoring or record-keeping to operate. Digital means of payment likewise rely on record-keeping technologies, although their programmability—for example, through smart contracts—may mitigate the severity of limited-commitment frictions.

In addition to cash, a variety of payment methods are available in practice. However, as documented in Section 1.1, credit cards are the most widely used payment method in the United States. For this reason, we focus on credit as the main alternative to cash. We interpret credit payments in the model broadly as representing both credit and debit card transactions in practice. Both forms of payment are typically provided by banks and involve

[7] In Freixas and Rochet (1997, Section 2.4), monitoring activities include screening projects, preventing opportunistic behavior, and punishing or auditing. In Holmstrom and Tirole (1997), in the context of corporate finance, banks monitor loans to prevent moral hazard by firms, e.g., by inspecting a firm's potential cash flow and its balance sheet position and by verifying that the firm conforms with covenants of the financial contract.

[8] In principle, the enforcement cost could depend on all the dimensions of the contract, including the intermediation fees paid by buyers to banks. Here, we assume that d summarizes the size of the buyers' obligations and abstract from the fees, but they could be added to the enforcement cost without changing any of the results. For a specification with a strictly convex cost, see Bethune et al. (2022).

[9] Allowing sellers to commit intertemporally would give them incentives to issue liabilities that could circulate as means of payment in stage 1, effectively turning them into banks.

some degree of monitoring. To reinterpret the model more precisely in terms of debit cards, one could follow the demand-deposit framework of Choi and Rocheteau (2022). In that environment, banks have access to a technology that allows them to invest deposits at a gross real interest rate R_d. The opportunity cost of deposits is $\sigma \equiv (\beta^{-1} - R_d)/R_d$, which corresponds to the rate-of-return differential between an illiquid asset and demand deposits. The remainder of the analysis would proceed unchanged.

We introduce heterogeneity in liquidity needs by assuming that buyers differ in their stage-1 preferences. In Section 1.1, we presented evidence from the Diary of Consumer Payment Choice showing that differences in cash usage are correlated with income. Accordingly, one could instead assume that buyers differ in their stage-2 labor endowment, $\bar{h}$, and focus on equilibria in which the constraint $h \leq \bar{h}$ binds.[10] This alternative specification would generate heterogeneity in the marginal rate of substitution between stage-1 and stage-2 consumption, similar to what arises in our framework. We adopt the current formulation because it is more tractable.

Finally, we assume that banks have access to a commitment technology at some cost. Alternatively, repayment of banks' liabilities could be self-enforced in the presence of public monitoring, as in Cavalcanti and Wallace (1999) and Gu et al. (2013a). We introduce such a reputational mechanism on the buyers' side in Section 5. In Appendix D we consider the reputational mechanism on the banks' side. A fixed cost of banking activities is nevertheless required to generate endogenous financial exclusion.

3 Complete information: The distributional relevance of cash

We restrict our attention to stationary equilibria where the rate of return of money and the rate of return of private debt are constant across stages. The gross real rate of return of money from stage $\tau \in \{1, 2\}$ to $\tau + 1$ (modulo 2) is denoted R^m_τ. The gross real rate of return of bonds (buyers' IOUs guaranteed by banks) from stage 1 to stage 2 is denoted R^b_1. The price of money in stage $\tau \in \{1, 2\}$ of period $t \in \mathbb{N}_0$, expressed in the consumption good, is denoted $\upsilon_{t,\tau}$. In a steady-state monetary equilibrium, $\upsilon_{t,\tau} M_t = \upsilon_{t+1,\tau} M_{t+1}$, which implies the gross rate of return of money across periods is $R^m_1 R^m_2 = \upsilon_{t+1,\tau}/\upsilon_{t,\tau} = M_t/M_{t+1} = (1 + \pi)^{-1}$. For now we assume that there is complete information between a buyer and a bank.

[10] For such an approach in an OLG model, see, e.g., Banet and Lebeau (2025).

3.1 Value functions

In this section, we characterize the value functions taking borrowing limits as given. We first analyze stage 2 and then solve backward for stage 1. A subsequent section endogenizes banking status and borrowing limits by modeling the negotiation between a buyer and a bank at time $t = 0$.

Agents' problems in stage 2 We denote $W^{\varkappa}(\omega, \varepsilon, \underline{b})$ the stage-2 value function of an agent of type $\varkappa \in \{1, 2\}$ with wealth ω (composed of money and bonds) expressed in the consumption good, a stage-1 utility weight, ε, and a within-period borrowing limit of $-\underline{b}$. (Equivalently, $\underline{b}$ is the lower bound on bond holdings.) It solves the following Bellman equation:

$$W^{\varkappa}(\omega, \varepsilon, \underline{b}) = \max_{c,h,m'} \{u(c) - h + \beta V^{\varkappa}(m', \varepsilon, \underline{b})\} \tag{1}$$

$$\text{s.t. } c + \frac{m'}{R_2^m} = h + \omega + T^{\varkappa} \tag{2}$$

$$h \leq \bar{h}, \tag{3}$$

where $V^{\varkappa}(m, \varepsilon, \underline{b})$ is the value of an agent of type $(\varkappa, \varepsilon)$ holding m real balances at the beginning of stage 1 with borrowing limit of $-\underline{b}$ in stage 1. According to (1), the agent chooses its consumption, work, and next-period real money holdings, m', in order to maximize its utility of consumption net of disutility of work plus its continuation value in the next period. From the budget identity, (2), the agent finances its consumption and next-period money holdings with its labor income, current wealth, and transfer from the government, $T^{\varkappa}$. Note that in order to hold m' real balances in the next stage, the agent must hold m'/R_2^m in the current stage. By assumption, agents are unable to commit across periods, so there is no debt issuance in stage 2. According to (3), agents cannot work more than $\bar{h}$.

If the constraint, $h \leq \bar{h}$, does not bind in stage 2, which is our maintained assumption, then the value function in stage 2 is linear in wealth,

$$W^{\varkappa}(\omega, \varepsilon, \underline{b}) = \omega + W^{\varkappa}(0, \varepsilon, \underline{b}), \quad \forall \varkappa \in \{1, 2\} \text{ and } \forall \varepsilon \in \mathbb{R}^+. \tag{4}$$

The optimal stage-2 consumption is $c = c_1^*$. As in Rocheteau and Wright (2005, Lemma 2), sellers have no demand for money, $m' = 0$. The buyer's choice of real balances is:

$$m_\varepsilon(\underline{b}) \in \arg\max_{m' \geq 0} \left\{ -\frac{m'}{R_2^m} + \beta V^{\varkappa}(m', \varepsilon, \underline{b}) \right\}. \tag{5}$$

The optimal choice of real balances is independent of ω. We denote $m_\varepsilon^c \equiv m_\varepsilon(0)$ the real balances of an unbanked buyer.

Agents' problems in stage 1 The stage-1 value function of an agent of type $(\varkappa, \varepsilon)$ satisfies

$$V^{\varkappa}(m, \varepsilon, \underline{b}) = \max_{c,h,m',b'} \{\varepsilon u(c) - h + W^{\varkappa}(m' + b', \varepsilon, \underline{b})\} \tag{6}$$

$$\text{s.t. } c + \frac{m'}{R_1^m} + \frac{b'}{R_1^b} = h + m \tag{7}$$

$$h \leq \zeta_1^{\varkappa} \bar{h} \text{ and } b' \geq \underline{b}. \tag{8}$$

Recall that for a seller ($\varkappa = 2$) in stage 1, $\varepsilon = 1$. The interpretation of this Bellman equation is similar to the one for $W^{\varkappa}$ with the following differences. Agents can issue IOUs that are redeemed in stage 2 but are subject to the borrowing constraint $b' \geq \underline{b}$. If the agent is in a banking relationship, $\underline{b} = -d_{\varepsilon}$, where $d_{\varepsilon} \geq 0$ represents the debt limit, and $\underline{b} = 0$ otherwise. In addition, only sellers can work in stage 1, $\zeta_1^2 = 1$.

If the constraint, $h \leq \bar{h}$, does not bind for sellers then $V^2(m) = m + V^2(0)$ and the optimal choice of consumption is $c = c_1^*$. Moreover, sellers choose their money (m') and bond (b') holdings to maximize $(R_1^m - 1)\, m'$ and $\left(R_1^b - 1\right) b'$. Hence, in any equilibrium where sellers acquire money and bonds (in exchange for their output) in stage 1,

$$R_1^m = R_1^b = 1 \text{ and } R_2^m = \frac{1}{1+\pi}. \tag{9}$$

In the absence of discounting across stages, the gross rates of return of money and bonds in stage 1 are equal to one. The gross rate of return of money across periods is the reciprocal of the gross growth rate of money supply, a result following from stationarity.

Using that $\zeta_1^1 = 0$, the value function of buyers in stage 1 is equal to

$$V^1(m, \varepsilon, \underline{b}) = \mathcal{S}(m - \underline{b}, \varepsilon) + m + W^1(0, \varepsilon, \underline{b}), \quad \varepsilon \in [0, \bar{\varepsilon}], \tag{10}$$

where the stage-1 trade surplus is

$$\mathcal{S}(a, \varepsilon) = \max_{c \leq a} \{\varepsilon u(c) - c\}. \tag{11}$$

If $-\underline{b} \geq c_{\varepsilon}^*$, then the buyer's stage-1 value function is linear and consumption is equal to c_{ε}^*. Otherwise, the value function is strictly concave for all $m < c_{\varepsilon}^* + \underline{b}$. In that case, the liquidity constraint, $c \leq m - \underline{b}$, binds and the derivative of the value function is $\partial V^{\varkappa}(m, \varepsilon, \underline{b})/\partial m = \varepsilon u'(m - \underline{b})$.

From (5) and (10), the buyer's choice of real money balances is

$$m_{\varepsilon}(\underline{b}) = \arg\max_{m \geq 0} \{-(1+i)\, m + \varepsilon u(m - \underline{b})\}. \tag{12}$$

The quantity $i \equiv (1+\pi)(1+\rho)-1$ is the (opportunity) cost of holding money across periods. It can also be interpreted as the nominal interest rate of an illiquid bond that cannot serve as means of payment in stage 1. Buyers, who can borrow up to $-\underline{b} \geq 0$, holds positive real balances if $1+i < \varepsilon u'(-\underline{b})$. In particular, if $\underline{b}=0$, it is optimal to hold real money balances for any $\varepsilon > 0$.

3.2 Value of commitment

We now compute the value for buyers to form a relationship with a bank at time $t=0$. From the Bellman equations (1) and (10), the lifetime expected utility of a buyer entering stage 2 with no wealth, expressed in flow, is

$$\rho W^1(0,\varepsilon,\underline{b}) = \max_{m' \geq 0} \{-im' + \mathcal{S}(m' - \underline{b}, \varepsilon)\} + (1+\rho)\left[T + u(c_1^*) - c_1^*\right], \quad \varepsilon \in [0, \bar{\varepsilon}]. \tag{13}$$

The first term on the right side is the stage-1 surplus net of the cost of holding real money balances. The second term is the stage-2 surplus augmented with the government transfer. From (13), the gain from having access to a credit line of size d_ε is

$$\mathcal{G}_\varepsilon(d_\varepsilon) \equiv \rho\left[W^1(0,\varepsilon,-d_\varepsilon) - W^1(0,\varepsilon,0)\right].$$

Lemma 1 ***(Value of commitment.)*** *The buyer's utility gain from forming a relationship with a bank at time 0 is $\mathcal{G}_\varepsilon$ solution to*

$$\mathcal{G}_\varepsilon(d_\varepsilon) = \max_{m^b \geq 0}\left\{-im^b + \mathcal{S}(m^b + d_\varepsilon, \varepsilon)\right\} - \max_{m^c \geq 0}\left\{-im^c + \mathcal{S}(m^c, \varepsilon)\right\}, \tag{14}$$

or, equivalently,

$$\mathcal{G}_\varepsilon(d_\varepsilon) = \begin{cases} id_\varepsilon & \text{if} \quad d_\varepsilon \leq m_\varepsilon(0) \\ \mathcal{S}(d_\varepsilon,\varepsilon) - \max_{m\geq 0}\left[-im + \mathcal{S}(m,\varepsilon)\right] & \quad d_\varepsilon > m_\varepsilon(0). \end{cases} \tag{15}$$

The determination of the gain, $\mathcal{G}_\varepsilon(d_\varepsilon)$, is illustrated graphically in Figure 5. It is the difference between the net surplus from trade when buyers have and have not access to a debt limit, d_ε, and real money balances are chosen optimally. From (15), if d_ε is lower than $m_\varepsilon^c \equiv m_\varepsilon(0)$ then a buyer with access to credit saves the opportunity cost associated with holding d_ε in the form of money. If d_ε is larger than $m_\varepsilon(0)$, then the buyer saves $im_\varepsilon(0)$ and raises her surplus by $\mathcal{S}(d_\varepsilon,\varepsilon) - \mathcal{S}(m_\varepsilon(0),\varepsilon)$.

3.3 Endogenous debt limits

Relationships between banks and buyers are established at the beginning of stage 2 of period $t=0$ through random matching. These relationships, if they are formed, last forever.[11] The

[11] We could assume that the relationships can be formed in any period and are destroyed probabilistically, as in Bethune et al. (2022), but it would not change the main insights.

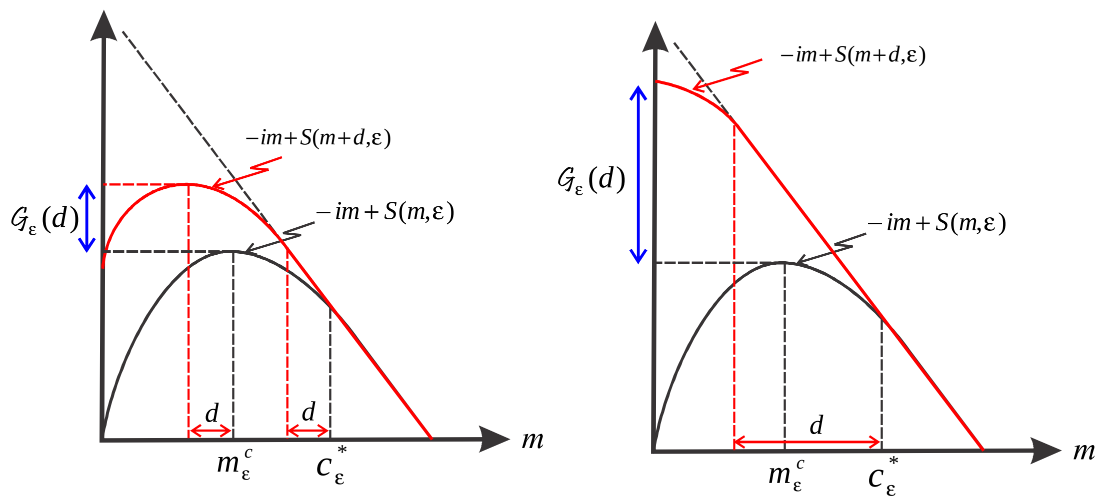


Figure 5: Value to the buyer of access to credit. Left panel: $d < m^c_\varepsilon$. Right panel: $d > m^c_\varepsilon$.

negotiation between the bank and the buyer takes place under complete information. The terms of the contract are composed of the total revenue of the bank, Φ_ε, and a debt limit, d_ε. The payment Φ_ε is payoff-equivalent to a per period fee (in terms of stage-2 consumption), $\phi_\varepsilon = \rho\Phi_\varepsilon$, starting from period 1 onward. Assuming there are gains from trade, the outcome of the negotiation is given by the generalized Nash solution, with the bank's bargaining power denoted by $\theta \in [0,1]$, i.e.,

$$(d_\varepsilon, \phi_\varepsilon) = \arg\max_{\phi,d} \left\{\mathcal{G}_\varepsilon(d) - \phi\right\}^{1-\theta} \left\{\phi - \sigma d - k\right\}^{\theta}, \tag{16}$$

where $\mathcal{G}_\varepsilon(d)$ is given by (14). The surplus of the bank is equal to the difference between the intermediation fee and the banking costs. The surplus of the buyer is equal to the increase in utility from being banked net of the intermediation fee, $\mathcal{G}_\varepsilon - \phi$. Assuming the joint surplus is positive, the assumption $i > \sigma$ implies that banked buyers do not hold money, and

$$d_\varepsilon = \arg\max_{d\geq 0} \left\{\mathcal{S}(d,\varepsilon) - \sigma d\right\}. \tag{17}$$

The loan size (or debt limit), which maximizes the joint surplus, is independent of banks' bargaining power. From (16), the banking fee is equal to

$$\phi_\varepsilon = \sigma d_\varepsilon + k + \theta\left\{\left[\mathcal{S}(d_\varepsilon,\varepsilon) - \sigma d_\varepsilon - k\right] - \left[\mathcal{S}(m^c_\varepsilon,\varepsilon) - i m^c_\varepsilon\right]\right\}, \tag{18}$$

where $m^c_\varepsilon \equiv m_\varepsilon(0)$ represents the real money holdings of an unbanked buyer. The payment to the bank is equal to the cost of monitoring loans, σd_ε, the cost of the commitment technology, k, plus a fraction θ of the match surplus, $\mathcal{G}_\varepsilon - \sigma d_\varepsilon - k$. We can reformulate the terms of the contract in terms of a lending rate defined as the ratio $r^\ell_\varepsilon \equiv \phi_\varepsilon / d_\varepsilon$.

It is easily checked that the match surplus is increasing in ε. It is negative and equal to $-k$ when $\varepsilon = 0$ and it grows unbounded as $\varepsilon \to +\infty$. Hence, there is a reservation value, $\varepsilon_R > 0$, above which lending relationships are formed. This threshold solves

$$\max_{d \geq 0} [\mathcal{S}(d, \varepsilon_R) - \sigma d] - \max_{m \geq 0} [-im + \mathcal{S}(m, \varepsilon)] = k. \tag{19}$$

It is determined graphically in Figure 6. The threshold, ε_R, is decreasing in i and increasing in k. It approaches 0 as k tends to 0.

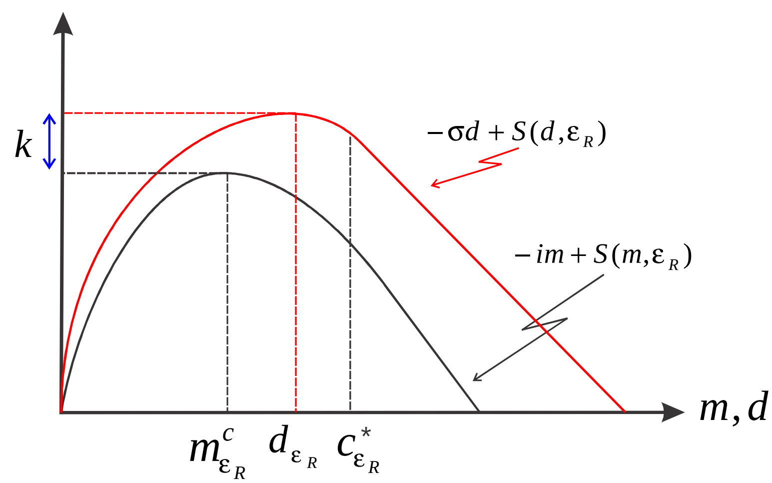


Figure 6: Determination of the exclusion threshold, ε_R

3.4 The value of money

Given ε_R, aggregate real balances in stage 1 are equal to

$$\text{M} = \int_0^{\varepsilon_R} m^c_\varepsilon dF(\varepsilon). \tag{20}$$

By market clearing, $\text{M} = \upsilon_{1,t} M_t$, so that the value of money is given by

$$\upsilon_{t,1} = \upsilon_{t,2} = \frac{\text{M}}{M_t} \quad \forall t \in \mathbb{N}_0. \tag{21}$$

An equilibrium can be defined as an exclusion threshold, ε_R, debt limits, d_ε, individual real money holdings for unbanked buyers, $m^c_\varepsilon \equiv m_\varepsilon(0)$, aggregate real money balances, M, and a time path for the value of money, $\upsilon_{t,1} = \upsilon_{t,2}$, solutions to (12), (17), (19), (20), and (21).

Proposition 1 ***(Existence and uniqueness of monetary equilibrium.)*** *For all $i > 0$, there exists a steady-state monetary equilibrium and it is unique. If $k < \mathcal{G}_{\bar{\varepsilon}}(d_{\bar{\varepsilon}}) - \sigma d_{\bar{\varepsilon}}$, money and credit coexist in equilibrium.*

Due to the fixed cost of banking and the assumption that the continuous support of $F(\varepsilon)$ includes 0, there always exist a demand for cash from buyers with low liquidity needs. Moreover, as long as k is not too large, there is a positive measure of buyers who form a relationship with a bank and can issue debt.

3.5 Cashless limits and the distributional relevance of cash

We consider a sequence of economies indexed by $n \in \mathbb{N}$ that differ in the cost of the commitment technology, k_n. For each k_n in the sequence, we associate the corresponding stationary equilibrium as characterized above. We say that the sequence of stationary equilibria converges when the associated equilibrium allocations and real money balances converge.

Definition 1 ***(Cashless limit.)*** *Consider a sequence of economies, $\{k_n\}_{n=1}^{+\infty}$, and a convergent sequence of monetary equilibria represented by $\{\mathrm{M}_n\}_{n=1}^{+\infty}$ where each $\mathrm{M}_n > 0$ are the equilibrium real money balances associated with the economy k_n. The limit is said to be cashless if M_n tends to 0 as n tends to $+\infty$.*

According to this definition, a monetary economy is cashless if it is the limit of a sequence of monetary equilibria such that aggregate real money balances approach 0.

Lemma 2 ***(Necessary and sufficient conditions for a cashless limit.)*** *Consider a sequence of economies, $\{k_n\}_{n=1}^{+\infty}$. It admits a cashless limit if and only if $\lim_{n\to+\infty} k_n = 0$.*

From Lemma 2, the economy becomes cashless as the cost of commitment, k, vanishes. At the limit, the threshold ε_R converges to zero, so that all buyers are banked. Credit supply within each relationship is sufficiently abundant that buyers have no incentive to hold money to supplement the liquidity provided by bank loans. In summary, at the cashless limit, there is no financial exclusion and no demand for cash.

In the following, we compare the cashless limit associated with $\{k_n\}_{n=1}^{+\infty}$ to the equilibrium in the economy without cash, the *no-cash economy*, under the cost $k_{+\infty} = \lim_{n\to+\infty} k_n = 0$. In an economy without cash, d_ε is still given by (17) and the intermediation fee is given by (18) where $m_\varepsilon^c = 0$, i.e.,

$$\phi_\varepsilon^0 = \sigma d_\varepsilon + k + \theta \left[\mathcal{S}(d_\varepsilon, \varepsilon) - \sigma d_\varepsilon - k\right]. \tag{22}$$

Proposition 2 ***(Relevance of cash at the cashless limit.)*** *Consider the limit of a sequence of economies, $\{k_n\}_{n=1}^{+\infty}$, that satisfies $\lim_{n\to+\infty} k_n = 0$. Suppose cash is removed.*

1. *Irrelevance of cash. If $\theta = 0$, allocations are unaffected. For all $\theta \geq 0$, aggregate welfare is unchanged.*

2. <u>Distributional relevance of cash</u>. *If $\theta > 0$, buyers are worse-off while banks are better off.*

Proposition 2 shows that if $\theta = 0$, i.e., banks have no market power, a monetary equilibrium that is nearly cashless has approximately the same allocation and welfare as an equilibrium in the no-cash economy. In that case, allocations are unaffected if cash is removed.

Even when $\theta > 0$, loan sizes d_ε are unchanged when cash is removed. As a result, consumption and aggregate welfare are unaffected, implying that—from an aggregate welfare perspective—cash is irrelevant at the cashless limit. By contrast, the prices of lending relationships, ϕ_ε, do respond to the removal of cash and increase. Hence, while cash is neutral for aggregate welfare, it remains fundamental for the distribution of surpluses between nonfinancial and financial agents. The availability of cash as a means of financing stage-1 consumption improves buyers' outside options, strengthens their bargaining positions vis-à-vis banks, and limits the fees or lending rates banks can charge. Consequently, buyers achieve higher welfare in the cashless limit than in an otherwise identical no-cash economy.

3.6 Competitive search in the credit market

We have seen in Proposition 2 that banks' bargaining power is fundamental for the relevance of cash at the cashless limit, yet it is taken to be constant as the economy progresses toward cashlessness. Here, we endogenize bargaining power by adopting the competitive search framework of Moen (1997). In this environment, banks post loan contracts and buyers direct their search across frictional submarkets, with free entry of bank agents.

We treat the cost of investing in a commitment technology, $K = k/\rho$, as an entry cost at time 0. The matching probability of a buyer with a bank in the initial period, $\alpha(q)$, depends on tightness in the credit market, defined as the ratio of banks to buyers, q (for queue length). The function, $\alpha(q)$, is increasing and concave, with $\alpha(0) = 0$, $\alpha(+\infty) = 1$, $\alpha'(0) = 1$, $\alpha'(+\infty) = 0$. Matches are bilateral so that the matching probability of a bank is $\alpha(q)/q$. We denote $\theta(q) \equiv \alpha'(q)q/\alpha(q)$ the elasticity of the matching function. (We use this notation to anticipate on the result that this elasticity will be equal to the bank's share in the gains from trade.)

Markets are segmented across buyers' types. Let $\mathcal{E}^b$ denote the set of buyer types with an active submarket. For each $\varepsilon \in \mathcal{E}^b$, there are banks posting a contract composed of an intermediation fee, ϕ_ε, a loan size, d_ε, and a market tightness, q_ε. The optimal contract solves:

$$(q_\varepsilon, d_\varepsilon, \phi_\varepsilon) = \arg\max \alpha(q) \left\{ [\mathcal{S}(d, \varepsilon) - \phi] - \max_{m^c \geq 0} [-im^c + \mathcal{S}(m^c, \varepsilon)] \right\} \text{ s.t. } k = \frac{\alpha(q)}{q}(\phi - \sigma d). \tag{23}$$

It maximizes the expected surplus of buyers subject to the zero profit condition of banks. From the first-order conditions, the endogenous share of the bank in the gains from trade is equal to the elasticity of the matching function, $\theta(q)$, in accordance with the Hosios condition. (See the supplementary Appendix C for details.) Moreover, the set $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$, where ε_R is still given by (19).

Aggregate real money balances are equal to

$$\mathrm{M} = \int_0^{\varepsilon_R} m_\varepsilon^c \mathrm{d}F(\varepsilon) + \int_{\varepsilon_R}^{\bar{\varepsilon}} [1 - \alpha(q_\varepsilon)] \, m_\varepsilon^c \mathrm{d}F(\varepsilon). \tag{24}$$

If $\varepsilon > \varepsilon_R$, buyers remain unbanked with probability $1 - \alpha(q_\varepsilon)$. Using that the optimal contract in (23) is uniquely determined, it can be shown that there exists a unique steady-state monetary equilibrium.

Proposition 3 ***(The irrelevance of cash under competitive search.)*** *Consider a sequence of monetary economies characterized by* $\{k_n\}_{n=0}^{+\infty} \in (\mathbb{R}^+)^{\mathbb{N}}$ *such that* k_n *decreases with* n*, and* $\lim_{n \to +\infty} k_n = 0$.

1. *The limit of monetary equilibria is cashless,* $\mathrm{M}_n \to 0$.

2. *At the cashless limit, banks' bargaining power is* $\theta(q_{\varepsilon,n}) \to 0$ *for all* ε*. Removing cash has not effect on allocations and welfare.*

For the economy to become cashless, the probability that a buyer forms a relationship with a bank must converge to one. When this occurs, the elasticity of the matching function, $\theta(q)$, which is also banks' endogenous bargaining power, tends to zero. Since banks have no bargaining power at the cashless limit, the role of cash as an outside option vanishes and removing cash has no welfare consequences. This result indicates that the relevance of cash at the cashless limit depends fundamentally on market structure, including the pricing mechanism and barriers to entry. Moreover, the analysis highlights the importance of understanding how the disappearance of cash and changes in banks' bargaining power can be both endogenous responses to fundamental changes in credit and payment markets.

4 Private information: The general relevance of cash

So far, we have assumed that the negotiation between the bank and the buyer was taking place under complete information. More realistically, banks set the terms of the loan contracts ($\theta = 1$) without knowing buyers' preferences and their needs for liquidity. It is this informational asymmetry that limit banks' market power and allows buyers to capture some rents.

A key determinant of the size of informational rents is the hazard rate of the distribution $F(\varepsilon)$ defined as $f(\varepsilon)/[1-F(\varepsilon)]$. We assume it is strictly increasing so that the outcome of the optimal mechanism is separating. As we will see later, the so-called *virtual valuation*—the value that is effectively used by banks to compute the match surplus—is equal to

$$v(\varepsilon) \equiv \varepsilon - \frac{1 - F(\varepsilon)}{f(\varepsilon)}. \tag{25}$$

It is strictly increasing, and $v(\varepsilon)/\varepsilon$ is strictly increasing as well.

4.1 Optimal contracts

At time 0, when matched with a buyer, the banker offers a menu of contracts, $\{(d_\varepsilon, \phi_\varepsilon) : \varepsilon \in \mathcal{E}^b\}$, where each contract specifies a loan size (or debt limit) and banking fees for all ε in the set $\mathcal{E}^b \subset [0, \bar{\varepsilon}]$.[12] We focus on equilibria in which any buyer of type $\varepsilon \in \mathcal{E}^b$ selects the contract in the menu corresponding to her type while any buyer of type $\varepsilon \notin \mathcal{E}^b$ rejects all the contracts. The per-period payoff to the banker associated with this menu is

$$\int_{\mathcal{E}^b} (\phi_\varepsilon - \sigma d_\varepsilon - k) \, \mathrm{d}F(\varepsilon). \tag{26}$$

The menu must satisfy individual rationality (IR), which means that for all $\varepsilon \in \mathcal{E}^b$, the gain from participating in the mechanism, $\mathcal{G}_\varepsilon$, must be equal or greater than the fee charged by the bank, ϕ_ε, that is,

$$\max_{m \geq 0} \{\mathcal{S}(m + d_\varepsilon, \varepsilon) - im - \phi_\varepsilon\} \geq U^m(\varepsilon, i) \equiv \max_{m \geq 0} [-im + \mathcal{S}(m, \varepsilon)]. \tag{27}$$

Thus, (27), specifies that the utility of the buyer net of fee must be greater than the utility from being unbanked. As before, we use m^c_ε to denote the optimal m in $U^m(\varepsilon, i)$. The

[12] Here we assume that banks commit to the contract terms for the entire duration of the relationship. In principle, the bank can learn about the buyer's liquidity needs in later stages and may want to reset the terms. One alternative assumption is to have banks and buyers rematched randomly before that learning happens. We also discuss the implications of financial innovations that allow banks to learn the buyer types more effectively at the end of this section.

optimal menu must also satisfies incentive-compatibility (IC) constraints requiring that, for any $\varepsilon,\ \varepsilon' \in \mathcal{E}^b$,

$$\max_{m \geq 0} \{\mathcal{S}(m + d_\varepsilon, \varepsilon) - im - \phi_\varepsilon\} \geq \max_{m \geq 0} \{\mathcal{S}(m + d_{\varepsilon'}, \varepsilon) - im - \phi_{\varepsilon'}\}, \tag{28}$$

and for any $\varepsilon \notin \mathcal{E}^b$ and any $\varepsilon' \in \mathcal{E}^b$,

$$U^m(\varepsilon, i) \geq \max_{m \geq 0} \{\mathcal{S}(m + d_{\varepsilon'}, \varepsilon) - im - \phi_{\varepsilon'}\}. \tag{29}$$

No buyer would be better off by choosing a contract that is not intended for her type.

We simplify the problem by establishing that, under the optimal contract, buyers of types in $\mathcal{E}^b$ never hold money. The intuition for this result is that banks can always expand loan sizes to replicate the payment capacity provided by money holdings. This substitution increases bank profits because the opportunity cost of money, i, is larger than the monitoring and enforcement costs of loans, σ. Moreover, we demonstrate in the proof that this substitution can be carried out without violating incentive compatibility. As a result, the optimal menu of contracts must satisfy $d_\varepsilon \geq m^c_\varepsilon$ for all $\varepsilon \in \mathcal{E}^b$. Using this observation and texbook mechanism design results, the IC constraints can be reduced to the conditions that $(U^b)'(\varepsilon) = u(d_\varepsilon)$ and d_ε is increasing in ε, where

$$U^b(\varepsilon) \equiv \varepsilon u(d_\varepsilon) - d_\varepsilon - \phi_\varepsilon, \quad \forall \varepsilon \in \mathcal{E}^b, \tag{30}$$

is the buyer's per-period gain from the credit contract. This reformulation of the IC constraints allows us to rewrite the bank's objective function in terms of the virtual valuation, (25), by substituting ϕ_ε from (30) into (26) and by using integration by parts, as shown in the following lemma.

Lemma 3 ***(Optimal supply of credit.)*** *The set of buyers that are served by the optimal mechanism takes the form $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$. The loan sizes, $\{d_\varepsilon : \varepsilon \in [\varepsilon_R, \bar{\varepsilon}]\}$, and the threshold, ε_R, solve*

$$\max_{\varepsilon_R, d_\varepsilon \geq m^c_\varepsilon} \int_{\varepsilon_R}^{\bar{\varepsilon}} \{\upsilon(\varepsilon)u(d_\varepsilon) - (1+\sigma)\, d_\varepsilon - k - U^m(\varepsilon_R, i)\}\, \mathrm{d}F(\varepsilon). \tag{31}$$

From (31), banks choose their supply of credit—loan sizes, d_ε, and exclusion threshold, ε_R—to maximize the expected virtual surplus defined as the buyer's virtual utility net of the costs of supplying loans and net of the buyer's payoff at the exclusion threshold. The constraint, $d_\varepsilon \geq m^c_\varepsilon$, embodies individual rationality by ensuring that participation in banking provides weakly greater liquidity than the outside option of remaining unbanked. The first-order condition with respect to ε_R is

$$[\upsilon(\varepsilon_R)u(d_{\varepsilon_R}) - (1+\sigma)\, d_{\varepsilon_R}] - \left[\upsilon(\varepsilon_R)u\left(m^c_{\varepsilon_R}\right) - (1+i)\, m^c_{\varepsilon_R}\right] = k. \tag{32}$$

This condition is identical to the one obtained under complete information, (19), where the valuation, ε, has been replaced by the virtual valuation, $\upsilon(\varepsilon)$.

The first-order condition with respect to the loan size, assuming the constraint $d_\varepsilon \geq m^c_\varepsilon$ does not bind, is

$$\upsilon(\varepsilon)u'(d_\varepsilon) = 1 + \sigma. \tag{33}$$

This unconstrained solution, denoted by d^*_ε, is increasing in ε and it is equal to 0 when $\upsilon(\varepsilon) = 0$, which corresponds to a positive ε. From (12) and (33), the distance between d_ε and m^c_ε can be measured by

$$u'(d_\varepsilon) - u'(m^c_\varepsilon) = \frac{1}{\varepsilon}\left(\frac{1+\sigma}{\upsilon(\varepsilon)/\varepsilon} - 1 - i\right).$$

Using the fact that $\upsilon(\varepsilon)/\varepsilon$ is increasing in ε from 0 when $\upsilon(\varepsilon) = 0$ to one when $\varepsilon = \bar{\varepsilon}$, and the assumption $i > \sigma$, there is a threshold for ε, $\tilde{\varepsilon} < \bar{\varepsilon}$, above which the participation constraint, $d_\varepsilon \geq m^c_\varepsilon$, is slack. In light of these observations, the following proposition characterizes the steady-state monetary equilibrium.[13]

Proposition 4 ***(Optimal menu of lending contracts.)*** *Suppose that $k > 0$. There is a unique monetary equilibrium.*

1. *If $k < \widetilde{k}(i) \equiv (i-\sigma)m^c_{\tilde{\varepsilon}}$, where $\tilde{\varepsilon} \in (0, \bar{\varepsilon})$ is the unique solution to*

$$\upsilon(\tilde{\varepsilon}) = \left(\frac{1+\sigma}{1+i}\right)\tilde{\varepsilon}, \tag{34}$$

the optimal menu of contracts offered by each bank is

$$\begin{aligned}(d_\varepsilon, \phi_\varepsilon) &= (m^c_\varepsilon, im^c_\varepsilon) && \forall \varepsilon \in [\varepsilon_R, \tilde{\varepsilon}] && (35)\\ &= \left(d^*_\varepsilon, id_{\tilde{\varepsilon}} + \int_{\tilde{\varepsilon}}^{\varepsilon} \frac{\partial \mathcal{S}[d(x),x]}{\partial d} d'(x)\mathtt{d}x\right) && \forall \varepsilon \in (\tilde{\varepsilon}, \bar{\varepsilon}] && (36)\end{aligned}$$

where ε_R is the unique solution to (32) with $d_{\varepsilon_R} = m^c_{\varepsilon_R}$, i.e.,

$$k = (i-\sigma)m^c_{\varepsilon_R}. \tag{37}$$

2. *If $k \geq \widetilde{k}(i)$, then $d_\varepsilon = d^*_\varepsilon$ for all $\varepsilon \in [\varepsilon_R, \bar{\varepsilon}]$, and ε_R is the unique solution to (32) with $d_{\varepsilon_R} = d^*_{\varepsilon_R}$.*

[13] Note that in (36) we use $d(x)$ to denote d_x to emphasize that it is a function of the buyer type, and $d'(x)$ denotes its derivative.

For k sufficiently small, there are banked buyers for whom the IR constraint binds, $\varepsilon_R < \tilde{\varepsilon}$. For all ε above $\tilde{\varepsilon}$, $d_\varepsilon \geq m_\varepsilon^c$ is slack and the loan size maximizes the joint *virtual* surplus. The fact that the virtual valuation is used instead of the actual valuation, ε, implies that the loan size is lower than under complete information, thereby generating a welfare loss.

For all ε below the threshold, $\tilde{\varepsilon}$, the constraint, $d_\varepsilon \geq m_\varepsilon^c$, binds. In that case, the loan size is equal to the real money balances chosen by the same buyer without access to credit. Since the unbanked buyer has to bear the opportunity cost, i, for each unit of money she holds, a bank fee equal to id_ε satisfies the buyer's participation constraint at equality while generating a positive profit, $(i-\sigma)d_\varepsilon$, for the bank.

Finally, the exclusion threshold, ε_R, is determined by the marginal type for which this profit exceeds the fixed cost, k. Buyers below that type are excluded and use cash. Buyers of types between ε_R and $\tilde{\varepsilon}$ receive no informational rent while buyers above $\tilde{\varepsilon}$ receive some informational rent. Crucially, in both cases the loan size is distorted as compared to the complete-information case, in a way that we interpret as credit rationing.

When k is sufficiently high, all banked buyers face slack IR constraints, $\varepsilon_R > \tilde{\varepsilon}$. Thus, buyers are either excluded from accessing credit, for types $\varepsilon < \varepsilon_R$, or they do have access to credit and enjoy some informational rent, for types $\varepsilon > \varepsilon_R$. The following corollary to Proposition 4 characterizes the bank contract in an economy without cash.

Corollary 1 ***(Optimal menu of contracts in the absence of cash.)*** *In the no-cash economy,* $d_\varepsilon = d_\varepsilon^* > 0$ *for all* $\varepsilon > \varepsilon_R^0$*, where the exclusion threshold,* ε_R^0*, is the unique solution to*

$$\max_{d\geq 0}\left\{\upsilon(\varepsilon_R^0)u\left(d\right)-(1+\sigma)d\right\}=k. \tag{38}$$

Since the left-side of (32) is strictly positive for all $\varepsilon_R > 0$ while the left-side of (38) is strictly negative for ε_R^0 small, it follows that $\varepsilon_R < \varepsilon_R^0$ for k small, that is, there is more financial exclusion in the no-cash economy.

4.2 On the relevance of cash under private information

The remainder of this section examines the relevance of cash in the cashless limit. The following proposition evaluates the welfare cost from removing cash as the fixed cost of banking, k, approaches 0.

Proposition 5 ***(Relevance of cash when liquidity needs are private information.)*** *Consider a sequence of economies characterized by* $\{k_n\}_{n=1}^{+\infty}$ *such that* $\lim_{n\to+\infty} k_n = 0$.

1. *As $n \to +\infty$, the economy becomes cashless, $\text{M}_n \to 0$, and there is no financial exclusion, $\varepsilon_{R,n} \to 0$.*

2. *Consider the economy at the cashless limit. If cash is removed:*

   (a) *Aggregate welfare falls;*

   (b) *Financial exclusion increases;*

   (c) *All buyers are strictly worse off.*

As the fixed cost of credit, k, goes to zero, the exclusion threshold in the monetary economy, $\varepsilon_R(i)$, converges to zero. Consequently, all buyers are banked—there is no financial exclusion—and the demand for real money balances vanishes. Start from this economy and remove cash. Surprisingly, the exclusion threshold in the no-cash economy, ε_R^0, jumps up and becomes strictly positive. As a result, removing cash from an economy that does not use it in equilibrium generates financial exclusion. In Figure 7, we represent the debt limits at the frictionless limit (d_ε) and in the no-cash economy (d_ε^0). For all $\varepsilon < \tilde{\varepsilon}$, they are strictly lower in the no-cash economy. As a result, all buyers are worse off: their payment capacity declines, and they incur higher fees from banks. Aggregate welfare falls as well.[14] This result holds not only at the limit where $k = 0$, but also for k strictly positive but sufficiently small.

The key friction that explains the relevance of cash at the cashless limit is the presence of informational rents accruing to buyers with high valuations. These rents reduce the market power of banks and they distort the allocation of credit. Although cash is not used in equilibrium, the fact that it is an option for buyers in case they reject banks' offers disciplines the ability of banks to extract rents through fees and interest payments. When that option is removed, banks can reduce the informational rents of high-liquidity-needs buyers by excluding low-liquidity-needs buyers.

4.3 Non-monotone effects of bank market power on welfare

We now consider changes in banks' bargaining power and the information structure that can be mapped into a gradual increase in banks' market power. These changes are illustrated in Figure 8. Start with a situation where banks have no market power: they have no information and no bargaining power, $\theta = 0$. In that case, buyers make take-it-or-leave-it offers to banks.

[14] Lagos and Zhang (2022) also show a welfare loss of cash removal at the cashless limit, though for a different reason. In their model, vendors depend on banks to reach a competitive market where goods are traded for buyers' IOUs. Since production is ex ante and banks have bargaining power, a hold-up problem emerges. The availability—actual or potential—of cash alleviates this by giving vendors an outside option for selling their inventories.

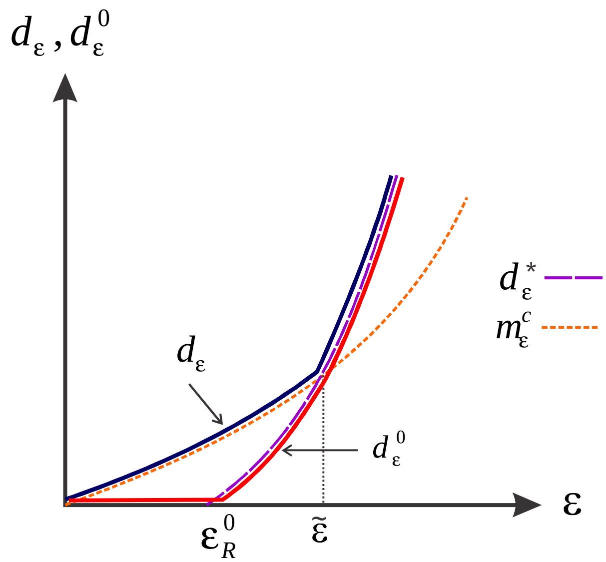


Figure 7: Debt limits at the cashless limit (d_ε) and in the no-cash economy (d_ε^0)

At the cashless limit, removing cash has no effect on allocations or prices—cash is irrelevant. This is also true in the competitive search framework of Section 3.6. These environments can be interpreted as instances of cash irrelevance in the sense of Woodford (1998).

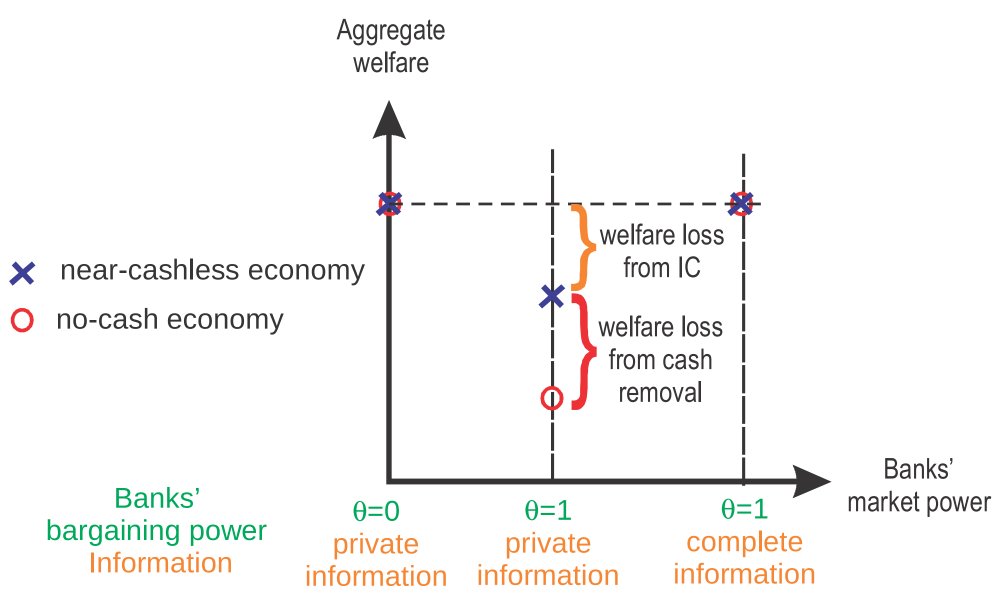


Figure 8: Welfare and market power in near-cashless and no-cash economies

Next, we introduce bank market power by increasing θ from 0 to 1 while keeping the informational environment unchanged—buyers remain privately informed about their borrowing needs. Banks now make take-it-or-leave-it offers and are therefore able to extract a positive surplus from their relationships with buyers. However, they do not possess full market power, since buyers continue to capture informational rents. At the cashless limit, eliminating cash generates a welfare loss. Moreover, even when cash is available, aggregate welfare is lower than in the no-market-power benchmark, reflecting the distortions induced by the incentive-compatibility constraints.

Finally, we further increase banks' market power by assuming that they are fully informed about buyers' preferences while keeping their bargaining strength unchanged at $\theta = 1$; that is, banks set the terms of lending contracts unilaterally. In this environment, banks extract the entire surplus from their relationships with buyers, so that their market power is maximal. Although eliminating cash at the cashless limit entails a redistribution of surplus between buyers and banks, it does not generate any loss in aggregate welfare.

These results indicate that the origins of bank market power are fundamental for assessing the overall welfare effects of eliminating cash. For example, if innovations in banking (e.g., FinTech) improve lenders' information, they may increase aggregate welfare by alleviating credit rationing while simultaneously mitigating the welfare costs associated with removing cash.

In the literature, there are proposals to eliminate cash motivated by its use for criminal activities (e.g., Rogoff, 2016). Suppose that there exists a social welfare loss, $\mathcal{L}(i)$, associated with the use of cash for criminal or illegal activities, and assume that this loss is invariant to the degree of bank market power. When banks have either no market power or full market power, eliminating cash generates a welfare gain equal to $\mathcal{L}(i)$. In the case of full market power, any distributional effects arising from the removal of cash can potentially be offset through alternative policy instruments, such as a tax on bank profits.

By contrast, if bank market power is partial and arises solely from bargaining strength, eliminating cash may be socially harmful despite the gain, $\mathcal{L}(i)$. In this case, the additional distortions to credit supply induced by information friction may outweigh the welfare gains from curbing cash-financed illegal activities.

In the next section we consider an alternative mechanism that mitigates the limited-commitment problem under which cash remains relevant even if banks enjoy no market power.

5 Imperfect record-keeping: The detrimental role of cash

In this section, we formalize buyers' limited commitment following Kehoe and Levine (1993) and Alvarez and Jermann (2000).[15] Banks lack the technology to directly enforce loan repayment but have access to buyers' credit histories. Repayment is therefore self-enforced through reputational incentives. As illustrated in Figure 9, any buyer matched with a banker can exchange her own IOU for a bank-issued IOU, which can be interpreted as inside money. A bank accepts a buyer's IOU only if the buyer's credit history contains no record of default. In exchange for providing liquidity transformation, banks charge an endogenous fee determined through bilateral negotiation. As in the previous sections, we adopt an equivalent alternative interpretation according to which buyers directly issue IOUs to sellers but the IOUs are guaranteed by the banks.

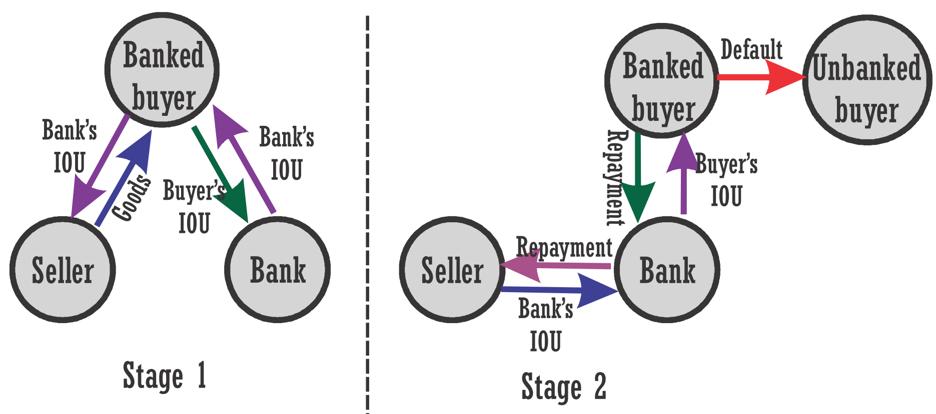


Figure 9: Credit under public record keeping

We assume that banks and buyers are randomly matched in stage 2 of every period. Each match lasts until stage 2 of the following period.[16] Upon formation of the match, in stage 2, the banking fee is paid prior to the extension of credit. Figure 10 illustrates the timing of events within the bank-buyer relationship.

Banks incur a fixed cost, k, whenever they issue IOUs. This cost may reflect, for example, the expense of maintaining a commitment technology that enables the bank to honor its own

[15] This formulation is consistent with a large strand of monetary theory that emphasizes the technological role of money as a record-keeping device (e.g., Ostroy and Starr, 1990; Kocherlakota, 1998).

[16] We adopt this assumption to capture the need of a record of buyers' credit histories that banks can share and access. It also facilitates the description of the renegotiation of the loan contracts every period. Alternatively, we could assume that matching occurs only in stage 2 of the initial period, $t = 0$, thereby formalizing a repeated game between a buyer and a bank. Under this formulation, a public record of buyers' histories would be unnecessary, provided agents have perfect recall of their own past actions. Nevertheless, the equilibrium characterized below would also constitute an equilibrium of that repeated game.

obligations or the cost of accessing buyers' credit histories. The fixed cost is incurred at the time the loan is repaid, in stage 2.

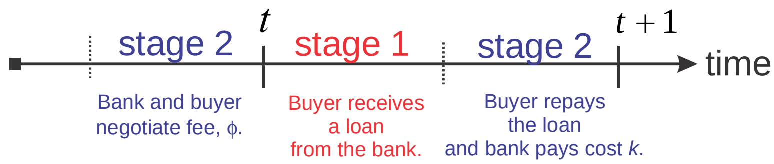


Figure 10: Timing of the bank-buyer lending relationship

5.1 Bellman equations

In the second stage of each period, buyers contract with banks to obtain a line of credit in the subsequent period. The contract specifies an intermediation fee, ϕ_ε, in exchange for a debt limit, d_ε. The debt limit of a type-ε buyer, d_ε, is an equilibrium object determined endogenously based on the buyer's credit history and taken as given by agents. We restrict attention to equilibria in which a buyer who defaults on her loan obligations at any point is permanently excluded from future credit. This equilibrium selection is consistent with the "not-too-tight solvency constraint" in Alvarez and Jermann (2000). By contrast, a buyer who does not pay the banking fee is excluded from credit for only one period. The interpretation is that the fee is paid upfront to secure access to credit in the subsequent period; failure to pay is therefore not treated as default but rather as a decision not to enter the loan contract.

While the Bellman equations closely resemble their earlier counterparts, subtle differences regarding the timing of fees justify rewriting those for the buyers. The stage-2 Bellman equation of the buyer is:

$$W^1(\omega,\varepsilon) = \max_{c,h,m'} \left\{u(c) - h + \beta V^1(m',\varepsilon,\underline{b}_\varepsilon)\right\} \tag{39}$$
$$\text{s.t. } c + (1+\pi)\,m' + \phi_\varepsilon = h + \omega + T^1 \text{ and } h \leq \bar{h}.$$

The buyer who enters stage 2 negotiates with a bank access to a line of credit in exchange for an intermediation fee, ϕ_ε. The latter enters the budget constraint. If the negotiation fails, and access to credit is denied, then $\underline{b}_\varepsilon = \phi_\varepsilon = 0$. Otherwise, $\underline{b}_\varepsilon = -d_\varepsilon$ and $\phi_\varepsilon \geq 0$. By the same reasoning as in Section 3, assuming $h \leq \bar{h}$ does not bind, W^1 is linear, that is, $W^1(\omega,\varepsilon) = \omega + W^1(0,\varepsilon)$.

Lemma 4 *The buyer's surplus from being banked equal to*

$$-(1+\rho)\phi_\varepsilon + \mathcal{G}_\varepsilon(d_\varepsilon; i) \tag{40}$$

where

$$\mathcal{G}_\varepsilon(d_\varepsilon; i) \equiv \max_{m' \geq 0} \{-im' + \mathcal{S}(m' + d_\varepsilon, \varepsilon)\} - \{-im^c_\varepsilon + \mathcal{S}(m^c_\varepsilon, \varepsilon)\}.$$

The surplus is equal to the value of having access to credit, $\mathcal{G}_\varepsilon(d_\varepsilon; i)$, which is the same as in Lemma 1, net of the capitalized fee, $(1+\rho)\phi_\varepsilon$. The surplus of the bank is equal to the capitalized fee net of the cost of the monitoring technology,

$$(1+\rho)\phi_\varepsilon - k.$$

We determine $\hat{\phi}_\varepsilon \equiv (1+\rho)\phi_\varepsilon$ by applying the generalized Nash solution where the bank's bargaining power is θ. The bank and the buyer reach an agreement if the joint surplus is positive, $\mathcal{G}_\varepsilon(d_\varepsilon) - k \geq 0$. In that case, the intermediation fee is equal to

$$\hat{\phi}_\varepsilon = k + \theta \left[\mathcal{G}_\varepsilon(d_\varepsilon; i) - k\right]. \tag{41}$$

It is equal to the fixed cost of credit, k, plus a fraction θ of the overall gains from trade.

5.2 Endogenous debt limits

We substitute $\hat{\phi}_\varepsilon$ by its expression given by (41) into (39) and use the expression for $U^m(\varepsilon, i)$ from (27) to reexpress the value function, $W^1(0, \varepsilon)$, as

$$\begin{aligned} \rho W^1(0,\varepsilon) &= (1+\rho)\left[T^1 + u(c_1^*) - c_1^*\right] + U^m(\varepsilon, i) \\ &\quad + (1-\theta)\max\{\mathcal{G}_\varepsilon(d_\varepsilon; i) - k, 0\}. \end{aligned} \tag{42}$$

The first term on the right side corresponds to the stage-2 utility while the second term corresponds to the stage-1 utility net of the cost of holding real money balances. The last term corresponds to the surplus enjoyed by the buyer from having access to credit: it is a fraction $1-\theta$ of the total gains from trade. If the buyer does not have access to credit, her lifetime utility is

$$\rho \tilde{W}^1(0,\varepsilon) = (1+\rho)\left[T^1 + u(c_1^*) - c_1^*\right] + U^m(\varepsilon, i). \tag{43}$$

The debt limit is defined as the highest loan size that the buyer would willingly repay in stage 2. It solves

$$d_\varepsilon = W^1(0,\varepsilon) - \tilde{W}^1(0,\varepsilon),$$

where the left side is the utility gain from not repaying one's debt in stage 2 and the right side is the lifetime utility loss from being excluded from future credit in case of default. Equivalently, from (42) and (43), d_ε solves

$$\rho d_\varepsilon = (1-\theta)\left[\mathcal{G}_\varepsilon(d_\varepsilon; i) - k\right]. \tag{44}$$

Equation (44) is represented graphically in Figure 11 where the left side is the blue line and the right side is the red or orange upward-sloping curve. The right side is linear with slope equal to $(1-\theta)i$ when d_ε is smaller than m^c_ε. It becomes horizontal for all d_ε larger than c^*_ε. Hence, if $k=0$, assuming $\rho < (1-\theta)i$, there is a unique $d_\varepsilon > 0$ solution to (44). This positive solution is larger than m^c_ε, which means that banked buyers have no incentive to hold money. It increases with ε and it goes to 0 as ε goes to 0. As k rises above 0, the right side shifts downward. As a result, (44) admits generically zero or two positive solutions. If there are multiple positive solutions, we focus on the largest one as it is the one that generates the highest welfare. We define the set of buyers who have access to credit as

$$\mathcal{E}^b \equiv \{\varepsilon \in [0, \bar{\varepsilon}] : \mathcal{G}_\varepsilon(d_\varepsilon; i) \geq k\}. \tag{45}$$

We show in the Appendix (see Lemma 6) that $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$ for some $\varepsilon_R > 0$.

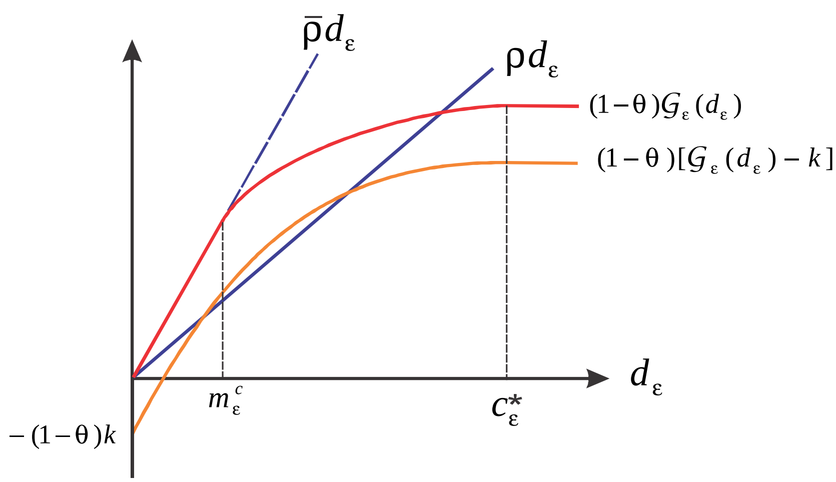


Figure 11: Determination of the debt limit

We define an equilibrium as a list of debt limits, (d_ε), the largest solutions to (44), a set of buyers with access to credit, $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$, solution to (45), and aggregate real money balances,

$$\text{M} = \int_0^{\varepsilon_R} m^c_\varepsilon \text{d}F(\varepsilon). \tag{46}$$

The following proposition establishes the existence and uniqueness of a steady-state monetary equilibrium, and the existence of a cashless limit as k converges to zero. Here we consider parameters ρ and i as primitives and take π as endogenously determined by those paraemters according to $\pi = (i-\rho)/(1+\rho)$.[17]

[17] Of course, we can also take ρ and π as primitives and have $i = (1+\rho)(1+\pi)-1$. In the proof we show that we can express all results in terms of π as well.

Proposition 6 ***(Monetary equilibria under limited commitment.)*** *Consider a sequence of economies characterized by $\theta < 1$ and $\{k_n\}_{n=1}^{+\infty}$. Define*

$$\overline{\rho}(i, \theta) = (1 - \theta)i. \tag{47}$$

Suppose that $\rho < \overline{\rho}(i, \theta)$.

1. ***Existence and uniqueness**. For all $n \in \mathbb{N}$, there is a unique monetary equilibrium. Moreover, for any $k_n > 0$ sufficiently small, there is a unique $\varepsilon_{R,n} \in (0, \bar{\varepsilon})$.*

2. ***Cashless limits**. If $\lim_{n \to +\infty} k_n = 0$, then, as $n \to +\infty$, the economy becomes cashless, $\mathrm{M}_n \to 0$ and $\varepsilon_{R,n} \to 0$.*

In the presence of a positive fixed cost, k, there always exists a monetary equilibrium. Indeed, there are no gains from trade between banks and buyers with low ε, i.e., buyers with low liquidity needs are excluded from credit trades and rely exclusively on cash.

The condition $\rho < \overline{\rho} \equiv (1 - \theta)i$ guarantees the existence of a monetary equilibrium with positive debt limits when $k = 0$. Note that the condition $\rho < \overline{\rho}$ implicitly requires $\pi > 0$. The threshold, $\bar{\rho}$, is represented graphically in Figure 11 by the slope of the dashed blue line, which coincides with the linear segment of $(1 - \theta)\mathcal{G}_\varepsilon(d_\varepsilon; i)$ for $d_\varepsilon < m_\varepsilon^c$.

The threshold $\overline{\rho} \equiv (1 - \theta)i$ decreases with θ, and the intuition is that as the bank's bargaining power increases, the surplus accruing to the buyer from being banked declines, thereby reducing the loss associated with exclusion from credit. At the extreme, if banks have full bargaining power, $\theta = 1$, buyers obtain no surplus from future access to credit and therefore have no incentive to repay their debts, $d_\varepsilon = 0$.

On the other hand, the threshold $\bar{\rho}$ is increasing in i, as illustrated in Figure 12. Intuitively, higher values of i increase the opportunity cost of holding money. As a result, the penalty from being unbanked becomes more severe, which relaxes the incentive constraint and supports higher debt limits. Recall that $\mathcal{G}_\varepsilon(d_\varepsilon; i) = id_\varepsilon$ for d_ε small. Thus, if i is low, the gain from defaulting, ρd_ε, increases more rapidly with d_ε than the gain from repaying and maintaining access to credit, $(1-\theta)\,(id_\varepsilon - k)$. In other words, for low i, the reputational mechanism is not strong enough to support the emergence of self-enforced credit.

Provided the condition for positive debt limits holds, $\rho < \overline{\rho}$, ε_R converges to zero as $k \to 0$. Consequently, in the limit, all buyers are banked and the economy becomes cashless.

Proposition 7 ***(Relevance of cash at the cashless limit.)*** *Consider a sequence of economies, $\{k_n\}_{n=1}^{+\infty}$, featuring $\rho < \overline{\rho}(i, \theta)$ and $\lim_{n \to +\infty} k_n = 0$.*

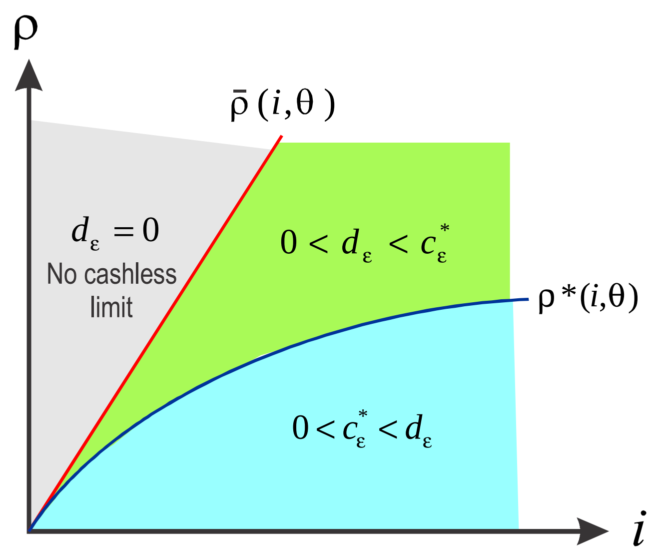


Figure 12: Thresholds for ρ below which debt limits are positive, $d_\varepsilon > 0$, and borrowing constraints are slack, $c^*_\varepsilon < d_\varepsilon$, when $k = 0$.

1. ***Relevance of cash****. At the limit, there is a $\rho^* < \bar{\rho}$ such that for all $\rho \in (\rho^*, \bar{\rho})$, removing cash increases the aggregate welfare.*

2. ***Distributional effects of cash removal****. Suppose $\theta \in (0, 1)$. There exists $\rho^1 \in (\rho^*, \bar{\rho}]$ such that for all $\rho \in (\rho^*, \rho^1)$, removing cash at the cashless limit strictly reduces the welfare of a subset of buyers.*

If the rate of time preference is below a cutoff, ρ^*, the debt limit is sufficiently large to finance the first-best level of consumption for all buyers, $d_\varepsilon \geq c^*_\varepsilon$ for all ε. This cutoff is represented graphically, for a given ε, by the slope of the lower dashed line in Figure 13. In the case of CRRA preferences, $u(c) = c^{1-a}/(1-a)$ with $a \in (0, 1)$,

$$\rho^* = (1-\theta)\left(\frac{a}{1-a}\right)\left[1 - \frac{1}{(1+i)^{\frac{1-a}{a}}}\right]. \tag{48}$$

We represent ρ^* as a function of i in the right panel of Figure **??**. It is independent of ε. Thus, if the borrowing constraint is slack for one type of buyer, then all buyers are unconstrained. In this case, removing cash at the cashless limit relaxes debt limits but leaves stage-1 consumption unchanged, so aggregate welfare is unaffected.

By contrast, if $\rho^* < \rho < \bar{\rho}$, as illustrated in Figure 13, then $d_\varepsilon < c^*_\varepsilon$ for some ε. When cash is removed, the stage-1 consumption of previously liquidity-constrained buyers, and hence aggregate welfare, increase. Graphically, the dash orange curve representing the reputational loss from defaulting, $(1-\theta)\mathcal{G}_\varepsilon(d_\varepsilon; i)$, shifts upward.

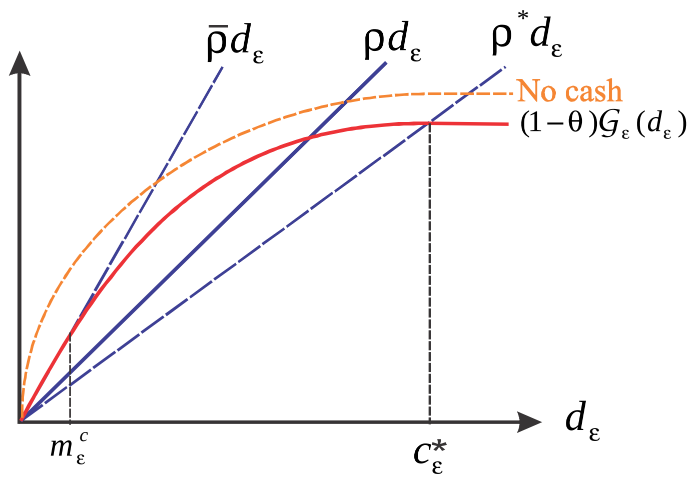


Figure 13: Effect of cash removal at the cashless limit on debt limits

An important implication of Proposition 7 is that money can affect welfare in the cashless limit even when banks have no bargaining power, $\theta = 0$. Moreover, relative to Section 4, the sign of the welfare effect is reversed. When credit relies on costly enforcement, eliminating cash can reduce welfare; when credit is self-enforced through public monitoring, eliminating cash can increase welfare.

The common thread is that cash provides insurance against financial exclusion, but the way this insurance operates differs across environments. With external enforcement, money gives buyers a fall-back option if a bank refuses to establish a relationship. This outside option disciplines banks and mitigates distortions arising from market power. When $\theta = 0$, this channel is inactive: fees are zero and debt limits are efficiently determined by the enforcement technology.

With self-enforcement, money instead relaxes borrowers' incentive constraints by offering a fall-back option after default and exclusion. Removing this option raises the cost of default, strengthens the reputational mechanism sustaining credit, and allows larger debt limits to be supported, thereby increasing aggregate welfare. Importantly, this mechanism operates even in the absence of banks' market power.

When banks possess bargaining power, $\theta > 0$, the removal of cash raises aggregate welfare, yet buyers need not be better off. From (42), a buyer's stage-1 utility consists of two components: the statu-quo payoff, $U^m(\varepsilon, i)$, and the surplus from accessing credit, $(1-\theta)\mathcal{G}_\varepsilon(d_\varepsilon; i)$. Eliminating cash—equivalently, setting $i = +\infty$—drives the former to zero. The latter increases for buyers whose initial debt limit lies below the first-best level, as their borrowing constraint relaxes. When $\rho \leq \rho^*$, only the first effect operates and thus buyers are made worse off by the removal of cash even though aggregate welfare does not change.

As ρ rises slightly above ρ^*, buyers' borrowing constraints start binding. As a result, buyers' welfare falls while at the same time aggregate welfare rises.

5.3 Discussion

In this section, a public-record mechanism provides buyers with incentives to repay their debts. By contrast, banks are assumed to have access to a technology that allows them to commit fully to the repayment of their IOUs, albeit at a cost. This assumption can be relaxed by allowing banks to build a reputation through public monitoring instead. This is the approach taken, for instance, in Cavalcanti and Wallace (1999) and Gu et al. (2013a), where the repayment of banks' own liabilities is self-enforced. The qualitative implications would resemble those we obtained here: in particular, removing cash in the cashless limit generates a first-order welfare gain by relaxing incentive constraints. An important difference, however, is that some degree of market power becomes necessary to provide banks with incentives to honor their obligations. See Appendix D for formal details.

Our result that eliminating cash can improve aggregate welfare when buyers are subject to self-enforcement hinges on the assumption that cash transactions are independent of default histories in the credit market. This assumption is justified by the fact that either sellers cannot observe buyers' credit histories or, even if they can, they have no incentive to refuse trade. If agents are ex ante identical, with their role as buyer or seller determined stochastically, and if all goods-market transactions are recorded, then it would be possible to sustain an equilibrium in which agents with a credit incident are excluded even from cash trade and are forced into full autarky. In that case, the removal of cash would have no effect on welfare. One may conjecture that such record-keeping could be enabled by technological advances in payment systems—such as blockchain—so that the use of digital cash is conditioned on past behavior.

6 Conclusion

In this paper, we examined economies in which money and credit coexist under limited commitment. We constructed sequences of such economies that converge to a cashless limit, where aggregate real money balances approach zero. Our central question was: What are the welfare consequences of completely removing cash from these economies? We showed that the answer depends on the mechanisms mitigating agents' limited commitment, the informational frictions present, and the market structure of the credit sector.

When credit is sustained by a costly enforcement technology operated by banks with market power, removing cash leaves aggregate welfare unchanged but redistributes surplus by weakening consumers' outside options. When consumers are heterogeneous and their preferences are private information, removing cash reduces aggregate welfare and induces financial exclusion. By contrast, when credit is supported by a record-keeping technology, removing cash raises welfare provided that banks possess little or no bargaining power. The distributional effect associated with cash removal remains if banks have market power.

These findings indicate that money near-cashless economies remains relevant and its role depends on the interplay between limited commitment, informational frictions, and market power. In practice, the welfare implications of removing cash may also hinge on the composition of credit instruments—such as the relative prevalence of secured versus unsecured or intermediated versus disintermediated credit.

The most direct interpretation of money in our model is physical cash issued by the government. Money need not be quantitatively important, nor even actively held, in order to perform its role. It suffices that it be available as an outside option that agents can substitute for the means of payment provided by financial intermediaries. Consistent with this implication, European Union has ruled that paper money should in principle be accepted by all businesses.[18] Our results also suggest that money could be entirely nonexistent, provided that the government stands ready to supply it on demand at a committed price or rate of return. As an increasing share of transactions takes place on online platforms requiring digital forms of payment, the government could instead commit to supplying a digital currency (CBDC). The mere availability of such an instrument would discipline banks' market power in the same way as physical cash (Andolfatto, 2021). However, our model also suggests that whether this would be a good alternative will depend on the society's rate of financial literacy in digital payments–if some group still find it too cumbersom to use, financial exclusion will remain a significant concern, especially for the more vulnerable groups.

[18] In the article, "Why Europe is rediscovering the virtues of cash" (Economist, Jan 2026), it is reported that "In 2021 a ruling from the top EU court affirmed that paper money should in principle be accepted [...] ministers from the EU's 27 member states reiterated their desire to ban businesses from refusing notes and coins." The article also cites financial inclusion as an important rationale for this legislation, as "some poor people struggle to open bank accounts at all."

Appendix A: Proofs of lemmas and propositions

Proof of Lemma 1. The equation (14) follows immediate from (13) and the definition of $\mathcal{G}_\varepsilon$. Iif $d_\varepsilon \leq m_\varepsilon(0)$, then

$$\max_{m^b \geq 0}\{-im^b + \mathcal{S}(m^b + d_\varepsilon, \varepsilon)\} = -i[m_\varepsilon(0) - d_\varepsilon] + \mathcal{S}[m_\varepsilon(0), \varepsilon],$$

where we used that $m_\varepsilon(-d_\varepsilon) = m_\varepsilon(0) - d_\varepsilon$. Similarly,

$$\max_{m^c \geq 0}\{-im^c + \mathcal{S}(m^b, \varepsilon)\} = -im_\varepsilon(0) + \mathcal{S}[m_\varepsilon(0), \varepsilon]$$

and hence $\mathcal{G}_\varepsilon = id_\varepsilon$. If $d_\varepsilon > m_\varepsilon(0)$, then the buyer does not hold any real money balances, and hence $\max_{m^b \geq 0}\{-im^b + \mathcal{S}(m^b + d_\varepsilon, \varepsilon)\} = \mathcal{S}(d_\varepsilon, \varepsilon)$. Thus, we obtain (15). ■

Proof of Proposition 1. The equilibrium has a recursive structure. From (12), we determine $m_\varepsilon(0)$ for all ε. Given $m_\varepsilon(0)$, $\mathcal{G}_\varepsilon$ is uniquely determined from (15) for all ε. With d_ε determined by (17), $d_\varepsilon > m_\varepsilon(0)$ for all ε as $\sigma < i$. Moroever, the left-side of (19) is strictly increasing in ε, since by the Envelope Theorem, the derivative of the left-side w.r.t. ε is $u(d_\varepsilon) - u[m_\varepsilon(0)] > 0$. As a result, there is a unique $\varepsilon_R > 0$. We obtain aggregate real money balances, M, from (20) and the value of money, $\upsilon_{t,1} = \upsilon_{t,2}$, from (21). As a result, M> 0, the equilibrium is monetary. This closes the determination of a steady-state monetary equilibrium, which is unique by construction. Finally, if $k < \mathcal{G}_{\bar{\varepsilon}}(d_{\bar{\varepsilon}}) - \sigma d_{\bar{\varepsilon}} = \max_{d \geq 0}[\mathcal{S}(d, \bar{\varepsilon}) - \sigma d] - \max_{m \geq 0}[\mathcal{S}(m, \bar{\varepsilon}) - im]$, then $\varepsilon_R < \bar{\varepsilon}$ and hence the equilibrium features some credit. ■

Proof of Lemma 2. Sufficiency ($\Leftarrow$). If $k_{+\infty} = 0$, then $\varepsilon_{R,n} \to 0$, and hence almost all buyers have access to credit at the limit. By (17), $d_\varepsilon > m_\varepsilon(0)$ for all ε. Hence, $m_\varepsilon(\underline{b}_\varepsilon) = m_\varepsilon(-d_\varepsilon) = 0$ for all ε and, from (20), M$= 0$.

Necessity ($\Rightarrow$). In order for $\text{M}_n \to 0$, almost all buyers must have access to credit at the limit, otherwise $m_\varepsilon(\underline{b}_\varepsilon) = m_\varepsilon(0) > 0$ for a subset of $[0, \bar{\varepsilon}]$ with positive measure. Since $\max_{d \geq 0}[\mathcal{S}(d, \bar{\varepsilon}) - \sigma d] - \max_{m \geq 0}[\mathcal{S}(m, \bar{\varepsilon}) - im] < \mathcal{S}(c^*_\varepsilon, \varepsilon)$ and since $\mathcal{S}(c^*_\varepsilon, \varepsilon) \to 0$ as $\varepsilon \to 0$, if k_n stays away from zero as $n \to +\infty$, then $\varepsilon_{R,n}$ determined by (19) stays away from zero as well, that is, $\varepsilon_{R,\infty} > 0$, and $m_\varepsilon(\underline{b}_\varepsilon) = m_\varepsilon(0) > 0$ for all $\varepsilon \in (0, \varepsilon_{R,\infty})$, a contradiction to cashlessness. So it requires $k_n \to 0$. ■

Proof of Proposition 2. Part 1. Note that d_ε is the same with or without cash. Suppose that $\theta = 0$. Then, from (18) and (22) we have that $\phi_\varepsilon = \phi^0_\varepsilon$ for all ε. Hence, at the cashless limit and in the absence of cash, the stage-1 consumption of buyers is $c = d_\varepsilon$, which is financed by a loan repaid in stage 2. The stage-2 consumption is c_1^*. Thus, allocations are identical in both economies.

Part 2. Aggregate welfare depends only on individual consumption levels. In stage-2, the consumption of each agent is c_1^*. In stage-1, the consumption of a buyer of type ε is equal to d_ε solution to (17) whether the economy is at the cashless limit or it is without cash. Hence, aggregate welfare is not affected by the presence of cash. However, buyer's welfare is lower, and banks' welfare is higher, in the absence of cash since $\phi_\varepsilon^0 > \phi_\varepsilon$ for any $\theta > 0$. ■

Proof of Lemma 3 and Proposition 4. The main argument consists of two parts. Part 1 deals with optimal contract for $k = 0$, and show that in this case $\mathcal{E}^b = [0, \bar{\varepsilon}]$. Full participation allows for a simple characterization of the IC constraint, and we show that the optimal contract has a two-tier structure. Part 2 considers the case where $k > 0$.

Part 1. Now we characterize the optimal contract under $k = 0$. Lemma 7 in Appendix B establishes that under optimal contract no buyers hold real balances and hence we have $d_\varepsilon \geq m_\varepsilon^c$. Moreover, Lemma 8 in Appendix B shows that it is without loss of generality to consider contracts where all types participate. Thus, it is without loss of generality to consider menus that has an offer intended for each type $\varepsilon \in [0, \bar{\varepsilon}]$. As a result, the IC constraint (28) implies a simpler necessary condition:

$$\varepsilon \in \arg\max_{\varepsilon' \in \mathcal{E}^b} \{-\phi_{\varepsilon'} + \varepsilon u(d_{\varepsilon'}) - d_{\varepsilon'}\}, \quad \forall \varepsilon \in [0, \bar{\varepsilon}], \tag{49}$$

Our approach is to consider (49) only to solve for the optimal contract, and then show that the optimal contract we solved for satisfies (28). Note that (49) is equivalent to

$$\varepsilon\{u(d_\varepsilon) - u(d_{\varepsilon'})\} \geq \phi_\varepsilon - \phi_{\varepsilon'} + (d_\varepsilon - d_{\varepsilon'}) \geq \varepsilon'\{u(d_\varepsilon) - u(d_{\varepsilon'})\} \tag{50}$$

for any $\varepsilon \geq \varepsilon' \in [0, \bar{\varepsilon}]$, and hence d_ε increases with ε. Because all monotone functions are differentiable almost everywhere, d_ε is also differentiable almost everywhere as well. Since buyers' real money balances are zero in equilibrium, their equilibrium utilities from the contract is given by $U^b(\varepsilon)$ as in (30). Hence the fee is given by

$$\phi_\varepsilon = \varepsilon u(d_\varepsilon) - d_\varepsilon - U^b(\varepsilon). \tag{51}$$

Note that

$$\begin{aligned} U^b(\varepsilon') - U^b(\varepsilon) &= [\varepsilon' u(d_{\varepsilon'}) - d_{\varepsilon'} - \phi_{\varepsilon'}] - [\varepsilon u(d_\varepsilon) - d_\varepsilon - \phi_\varepsilon] \\ &= \phi_\varepsilon - \phi_{\varepsilon'} + (d_\varepsilon - d_{\varepsilon'}) - \varepsilon\{u(d_\varepsilon) - u(d_{\varepsilon'})\} + (\varepsilon' - \varepsilon) u(d_{\varepsilon'}) \leq (\varepsilon' - \varepsilon) u(d_{\varepsilon'}), \end{aligned}$$

where the inequality follows from (50); similarly, $U^b(\varepsilon') - U^b(\varepsilon) \geq (\varepsilon' - \varepsilon)\, u\,(d_\varepsilon)$. Since $u\,(d_\varepsilon)$ is bounded, by taking limits as $\varepsilon' \to \varepsilon$, U^b is continuous. The following lemma characterizes (49). We omit the proof as it is standard (see, for example, Mas Colell et al. (1995) Proposition 23.D.2).

Lemma 5 *A menu* $\{(d_\varepsilon, \phi_\varepsilon), \varepsilon \in [0, \bar{\varepsilon}]\}$ *satisfies (49) if and only if* $(U^b)'(\varepsilon) = u(d_\varepsilon)$ *and* d_ε *is increasing in* ε *for all* $\varepsilon \in [0, \bar{\varepsilon}]$.

Now we solve the optimal contracting problem under the IC constraints given by Lemma 5. As mentioned, we will verify that the optimal contract satisfies the original IC at the end. First define $\nu(\varepsilon) \equiv U^b(\varepsilon) - U^m(\varepsilon)$. Then, by the Envelope Theorem, we have

$$\nu'(\varepsilon) = u(d_\varepsilon) - u(m^c_\varepsilon). \tag{52}$$

Moreover, the IR can be written as $\nu(\varepsilon) \geq 0$. Note that $U^b(0) = 0 = U^m(\varepsilon)$ because of IR and hence $\nu(0) = 0$. From Lemma 7, $d_\varepsilon \geq m^c_\varepsilon$. Note that this inequality, together with (52) and $\nu(0) = 0$, immediately implies IR for all $\varepsilon > 0$. Since $k = 0$, the profits to the bank are

$$\Pi = \int_0^{\bar{\varepsilon}} (\phi_\varepsilon - \sigma d_\varepsilon)\, \mathrm{d}F(\varepsilon),$$

which, using (51), are also equal to

$$\Pi = \int_0^{\bar{\varepsilon}} [-\nu(\varepsilon) - U^m(\varepsilon) - (1+\sigma)d_\varepsilon + \varepsilon u(d_\varepsilon)]\, \mathrm{d}F(\varepsilon).$$

Thus, we can write the problem of the bank as follows:

$$\int_0^{\bar{\varepsilon}} [-\nu(\varepsilon) - U^m(\varepsilon) - (1+\sigma)d_\varepsilon + \varepsilon u(d_\varepsilon)]\mathrm{d}F(\varepsilon), \tag{53}$$

$$\text{subject to} \quad \nu'(\varepsilon) = u(d_\varepsilon) - u(m^c_\varepsilon), \tag{54}$$

$$d_\varepsilon \geq m^c_\varepsilon. \tag{55}$$

Now, integrating (54) from 0 to ε and using that $\nu(0) = 0$, we obtain

$$\nu(\varepsilon) = \int_0^{\varepsilon} [u(d_x) - u(m^c_x)]\mathtt{d}x = \int_0^{\varepsilon} u(d_x)\mathtt{d}x - U^m(\varepsilon),$$

and hence $\nu(\varepsilon) + U^m(\varepsilon) = \int_0^\varepsilon u(d_x)\mathtt{d}x$. Using integration by parts and the fact that $[\varepsilon - \upsilon(\varepsilon)]f(\varepsilon) = 1 - F(\varepsilon)$,

$$\int_0^{\bar{\varepsilon}} [\nu(\varepsilon) + U^m(\varepsilon)]\mathrm{d}F(\varepsilon) = F(\bar{\varepsilon}) \int_0^{\bar{\varepsilon}} u(d_\varepsilon)\mathtt{d}\varepsilon - \int_0^{\bar{\varepsilon}} F(\varepsilon)u(d_\varepsilon)\mathtt{d}\varepsilon = \int_0^{\bar{\varepsilon}} [\varepsilon - \upsilon(\varepsilon)]u(d_\varepsilon)f(\varepsilon)\mathtt{d}\varepsilon,$$

and the objective function in (53) becomes

$$\begin{aligned}
&\int_0^{\bar{\varepsilon}} [-\nu(\varepsilon) - U^m(\varepsilon) - (1+\sigma)d_\varepsilon + \varepsilon u(d_\varepsilon)]\mathrm{d}F(\varepsilon) \\
= & -\int_0^{\bar{\varepsilon}} [\varepsilon - \upsilon(\varepsilon)]u(d_\varepsilon)f(\varepsilon)\mathtt{d}\varepsilon + \int_0^{\bar{\varepsilon}} [-(1+\sigma)d_\varepsilon + \varepsilon u(d_\varepsilon)]f(\varepsilon)\mathtt{d}\varepsilon \\
= & \int_0^{\bar{\varepsilon}} [\upsilon(\varepsilon)u(d_\varepsilon) - (1+\sigma)\, d_\varepsilon]\mathtt{d}F(\varepsilon).
\end{aligned}$$

Thus, the optimal contract solves

$$\max \int_0^{\bar{\varepsilon}} [\upsilon(\varepsilon)u(d_\varepsilon) - (1+\sigma)\, d_\varepsilon]\mathtt{d}F(\varepsilon) \text{ subject to } d_\varepsilon \geq m_\varepsilon^c.$$

The Lagrangian for the above problem is

$$\upsilon(\varepsilon)u(d_\varepsilon) - (1+\sigma)\, d_\varepsilon + \lambda(\varepsilon)(d_\varepsilon - m_\varepsilon^c),$$

where $\lambda(\varepsilon)$ is the multiplier. The solution then satisfies the following FOC:

$$\upsilon(\varepsilon)u'(d_\varepsilon) - (1+\sigma) + \lambda(\varepsilon) = 0,\ \lambda(\varepsilon)(d_\varepsilon - m_\varepsilon^c) = 0,\ \lambda(\varepsilon) \geq 0. \tag{56}$$

These admit a unique solution:

$$\lambda(\varepsilon) = -\upsilon(\varepsilon)u'(m_\varepsilon^c) + (1+\sigma) \text{ and } d_\varepsilon = m_\varepsilon^c \text{ for all } \varepsilon < \tilde{\varepsilon}, \tag{57}$$

$$\lambda(\varepsilon) = 0 \text{ and } u'(d_\varepsilon) = \frac{1+\sigma}{\upsilon(\varepsilon)} \text{ for all } \varepsilon \geq \tilde{\varepsilon}, \tag{58}$$

where $\tilde{\varepsilon}$ is determined by

$$u'(m_\varepsilon^c) = \frac{1+i}{\varepsilon} = \frac{1+\sigma}{\upsilon(\varepsilon)}. \tag{59}$$

Since $\upsilon(\varepsilon)/\varepsilon$ is increasing,

$$\lambda(\varepsilon) = -\upsilon(\varepsilon)u'(m_\varepsilon^c) + (1+\sigma) = -\upsilon(\varepsilon)\frac{(1+i)}{\varepsilon} + (1+\sigma) > 0$$

for all $\varepsilon < \tilde{\varepsilon}$. Moreover, by (51), $\phi_\varepsilon = -U^b(\varepsilon) - d_\varepsilon + \varepsilon u(d_\varepsilon)$ and hence $\phi_\varepsilon = id_\varepsilon > 0$ for all $\varepsilon \leq \tilde{\varepsilon}$. Finally,

$$\phi_\varepsilon = -U^m(\tilde{\varepsilon}, i) - \int_{\tilde{\varepsilon}}^{\varepsilon} u[d(x)]\mathrm{d}x - d(\varepsilon) + \varepsilon u[d(\varepsilon)] = id(\tilde{\varepsilon}) + \int_{\tilde{\varepsilon}}^{\varepsilon} \frac{\partial \mathcal{S}[d(x), x]}{\partial d} d'(x)\mathrm{d}x, \tag{60}$$

where in the last equality we use integration by parts, and note that we use $d(\varepsilon)$ ($d(x)$) to denote d_ε (d_x) here and $d'(\varepsilon)$ for its derivative. Note that the optimal contract then coincide with thsoe given by (35)-(36). To conclude our proof of the optimal contract for $k = 0$, Lemma 9 in Appendix B shows that the optimal solution (57) satisfies the original IC, (28).

Part 2. Now, consider the case where $k > 0$. Given a set $\mathcal{E}^b$ of buyer types served by the bank, and a menu of contracts, $\{(d_\varepsilon, \phi_\varepsilon) : \varepsilon \in \mathcal{E}^b\}$, define $U^b(\varepsilon)$ as in (30) for $\varepsilon \in \mathcal{E}^b$ and define $U^b(\varepsilon) = U^m(\varepsilon)$ for $\varepsilon \in \mathcal{E}^c = [0, \bar{\varepsilon}] - \mathcal{E}^b$. We first claim that, if the menu satisfies IC and IR, then $(U^b)'(\varepsilon) = u(d_\varepsilon)$ if $\varepsilon \in \mathcal{E}^b$ and $(U^b)'(\varepsilon) = u(m^c_\varepsilon)$ if $\varepsilon \in \mathcal{E}^c$. Note that, by Lemma 7 (which does not need the assumption $k = 0$), we can restrict attention to menus such that $d_\varepsilon \geq m^c_\varepsilon$ for all $\varepsilon \in \mathcal{E}^b$. Now, for each $\varepsilon \notin \mathcal{E}^b$, consider the contract $(d_\varepsilon, \phi_\varepsilon) = (m^c_\varepsilon, im^c_\varepsilon)$, which gives the same payoff as $U^m(\varepsilon)$ and by IR, for any other type, it gives a lower payoff to a buyer of any other type. Following the same logic as in the proof of Lemma 8, the expanded menu also satisfies IC. Using Lemma 5 to the expanded menu, we obtain that $(U^b)'(\varepsilon) = u(d_\varepsilon)$ if $\varepsilon \in \mathcal{E}^b$ and $(U^b)'(\varepsilon) = u(m^c_\varepsilon)$ if $\varepsilon \notin \mathcal{E}^c$.

Again, if we let $\nu(\varepsilon) = U^b(\varepsilon) - U^m(\varepsilon)$, then $\nu'(\varepsilon) = u(d_\varepsilon) - u(m^c_\varepsilon) \geq 0$, that is, ν is increasing. Thus, either $\nu(\varepsilon) = 0$ for all $\varepsilon \in [0, \bar{\varepsilon}]$, or for some $\varepsilon_0 \geq 0$ such that $\nu(\varepsilon) = 0$ for all $\varepsilon \leq \varepsilon_0$ and $\nu(\varepsilon) > 0$ for all $\varepsilon > \varepsilon_0$. That is, $\mathcal{E}^b$ includes the interval $[\varepsilon_0, \bar{\varepsilon}]$, and, for any $\varepsilon < \varepsilon_0$ but is served by the bank, the buyer's utility is $U^m(\varepsilon)$ and $d_\varepsilon = m^c_\varepsilon$. This also implies that $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$ for some ε_R.

Using (51), the profit becomes

$$\int_{\varepsilon_R}^{\bar{\varepsilon}} [\phi_\varepsilon - \sigma d_\varepsilon - k] \mathrm{d}F(\varepsilon) = \int_{\varepsilon_R}^{\bar{\varepsilon}} [-U^b(\varepsilon) + \varepsilon u(d_\varepsilon) - d_\varepsilon - \sigma d_\varepsilon - k] \mathrm{d}F(\varepsilon)$$
$$= \int_{\varepsilon_R}^{\bar{\varepsilon}} \{\upsilon(\varepsilon) u(d_\varepsilon) - (1+\sigma) d_\varepsilon - k\} \mathrm{d}F(\varepsilon) - U^m(\varepsilon_R)[1 - F(\varepsilon_R)]$$

Now, using the fact that $(U^b)'(\varepsilon) = u(d_\varepsilon)$, the boundary condition $U^b(\varepsilon_R) = U^m(\varepsilon_R)$, and integration by parts, we have

$$\int_{\varepsilon_R}^{\bar{\varepsilon}} U^b(\varepsilon) \mathrm{d}F(\varepsilon) = U^b(\bar{\varepsilon}) F(\bar{\varepsilon}) - U^b(\varepsilon_R) F(\varepsilon_R) - \int_{\varepsilon_R}^{\bar{\varepsilon}} F(\varepsilon) u(d_\varepsilon) \mathrm{d}\varepsilon$$
$$= U^b(\varepsilon_R) + \int_{\varepsilon_R}^{\bar{\varepsilon}} u(d_\varepsilon) \mathrm{d}\varepsilon - \int_{\varepsilon_R}^{\bar{\varepsilon}} F(\varepsilon) u(d_\varepsilon) \mathrm{d}\varepsilon - U^b(\varepsilon_R) F(\varepsilon_R)$$
$$= \int_{\varepsilon_R}^{\bar{\varepsilon}} \left\{ \frac{[1 - F(\varepsilon)]}{f(\varepsilon)} u(d_\varepsilon) + U^m(\varepsilon_R) \right\} f(\varepsilon) \mathrm{d}\varepsilon.$$

Thus, the bank profit becomes

$$\Pi = \int_{\varepsilon_R}^{\bar{\varepsilon}} \left\{ \left(\varepsilon - \frac{[1 - F(\varepsilon)]}{f(\varepsilon)} \right) u(d_\varepsilon) - (1+\sigma) d_\varepsilon - U^m(\varepsilon_R) - k \right\} f(\varepsilon) \mathrm{d}\varepsilon$$
$$= \int_{\varepsilon_R}^{\bar{\varepsilon}} \{\upsilon(\varepsilon) u(d_\varepsilon) - (1+\sigma) d_\varepsilon - U^m(\varepsilon_R) - k\} f(\varepsilon) \mathrm{d}\varepsilon.$$

We claim then the optimal contract maximize Π over ε_R and d_ε subject to $d_\varepsilon \geq m^c_\varepsilon$. We have seen the necessity of the constraint from Lemma 7, and we can show that the optimal solution also satisfies the global IC, (28) and (29), following the same arguments as in Lemma 9. This proves Lemma 3.

Now, if $\varepsilon_R < \tilde{\varepsilon}$, then the solution is exactly the same as (57) and (58); otherwise, we have (58) only. Given the optimal d_ε,

$$\Pi'(\varepsilon_R) = k - \left\{ \upsilon(\varepsilon_R) \left[u(d_{\varepsilon_R}) - u(m^c_{\varepsilon_R}) \right] - \left[(1+\sigma) d_{\varepsilon_R} - (1+i) m^c_{\varepsilon_R} \right] \right\}.$$

Thus, if $k \geq \tilde{k}$, the solution to $\Pi'(\varepsilon_R) = 0$ satisfies $\varepsilon_R \geq \tilde{\varepsilon}$. If $k < \tilde{k}$, then $\Pi'(\varepsilon_R) = k - (i - \sigma) m^c_{\varepsilon_R} = 0$ determines the optimal ε_R. This prove Proposition **4.** ∎

Proof of Corollary 1. In the no cash economy, the IR is simply $U^b(\varepsilon) \geq 0$, since the only outside option for a buyer is no-trade. Note that since there is no cash, the IC constraint is also essentially the same as (49), except for the case that for $\varepsilon < \varepsilon'$, it is possible that for type-ε to consider the contract intended for type-ε', it maybe the case that $c^*_\varepsilon < d_\varepsilon$ and hence the payoff from that contract is $\varepsilon u(c^*_\varepsilon) - d_\varepsilon$, but this IC will be satisfied in the optimal contract and hence is not binding. Moreover, even though the optimal contract may exclude some buyers, the IC constraint, (49), still needs to hold for all types of buyers, as we can regard buyers who are excluded as having contract $(d_\varepsilon, \phi_\varepsilon) = (0, 0)$. Then, using the earlier argument around (50) and Lemma 5, it follows that IC is equivalent to $(U^b)'(d_\varepsilon) = u(d_\varepsilon)$ for all $\varepsilon \in [0, \bar{\varepsilon}]$ and that d_ε is increasing in ε. This implies that the optimal set $\mathcal{E}^b$ is an interval $[\varepsilon_R, \bar{\varepsilon}]$ with ε_R to be determined below, but $U^b(\varepsilon) = 0$ for all $\varepsilon \leq \varepsilon_R$.

Now, the bank's profit is then

$$\int_{\varepsilon_R}^{\bar{\varepsilon}} (\phi_\varepsilon - \sigma d_\varepsilon - k) \mathrm{d}F(\varepsilon) = \int_{\varepsilon_R}^{\bar{\varepsilon}} [\upsilon(\varepsilon) u(d_\varepsilon) - (1+\sigma) d_\varepsilon - k] \mathrm{d}F(\varepsilon),$$

where we used (30), $(U^b)'(\varepsilon) = u(d_\varepsilon)$, and integration by parts with the condition $U^b(\varepsilon_R) = 0$.

Thus, the bank's problem is to choose ε_R to maximize

$$\Pi(\varepsilon_R) = \max_{d_\varepsilon \geq 0} \int_{\varepsilon_R}^{\bar{\varepsilon}} [\upsilon(\varepsilon) u(d_\varepsilon) - (1+\sigma) d_\varepsilon - k] \mathrm{d}F(\varepsilon). \tag{61}$$

The solution is then $d^0_\varepsilon = d^*_\varepsilon$ given by (33) whenever $\varepsilon > \varepsilon^0_R$, where ε^0_R solves

$$\upsilon(\varepsilon^0_R) u(d^*_{\varepsilon^0_R}) - (1+\sigma) d^*_{\varepsilon^0_R} = k,$$

as stated in (38). Note that this solution coincides with the one in Proposition 4 as $i = \infty$. When $i = \infty$, $\tilde{k} = 0$ and hence the case $k \geq \tilde{k}$ applies and hence $d_\varepsilon = d^*_\varepsilon$ and ε_R satisfies (32) with $m^c_\varepsilon = 0$. ∎

Proof of Proposition 5. We consider the case where $k < \widetilde{k}$ so that $\varepsilon_R < \tilde{\varepsilon}$. We measure aggregate welfare as follows:

$$
\begin{aligned}
\mathcal{W}(i) \quad \equiv \quad & 3\left[u(c_1^*) - c_1^*\right] + \int_0^{\varepsilon_R(i)} \{\varepsilon u[m_\varepsilon^c(i)] - m_\varepsilon^c(i)\}\, \mathtt{d}F(\varepsilon) \qquad (62) \\
& + \int_{\varepsilon_R(i)}^{\tilde{\varepsilon}(i)} \{\varepsilon u[m_\varepsilon^c(i)] - (1+\sigma) m_\varepsilon^c(i) - k\}\, \mathtt{d}F(\varepsilon) + \int_{\tilde{\varepsilon}(i)}^{\bar{\varepsilon}} \{\varepsilon u\left(d_\varepsilon^*\right) - (1+\sigma) d_\varepsilon^* - k\}\, \mathtt{d}F(\varepsilon).
\end{aligned}
$$

The first term on the right side of (62) is explained as follows. In stage 2, both buyers and sellers who are unconstrained by their labor endowments consume c_1^* while in stage 1 all sellers consume c_1^*, which generates a total surplus equal to $3\left[u(c_1^*) - c_1^*\right]$. The second term corresponds to the surplus from the stage-1 consumption of unbanked buyers. The third and fourth terms are the sum of the surpluses from the stage-1 consumption of banked buyers where we take into account the enforcement costs incurred by banks.

Part 1. Now, for k small, the optimal contract in Proposition 4 depends on k_n only through $\varepsilon_{R,n}$, which by (37) converges to zero as $k_n \rightarrow 0$. Hence, for the economy indexed by n, the aggregate real balances are given by

$$
\mathrm{M}_n = \int_0^{\varepsilon_{R,n}} m_\varepsilon^c \mathrm{d}F(\varepsilon) \rightarrow 0
$$

as $k_n \rightarrow 0$, and the allocations converge. Hence, a cashless limit exists.

Part 2. Now, as $k_n \rightarrow 0$, the exclusion threshold $\varepsilon_{R,n}^0$ in the no-cash economy that solves (38) with $k = k_n$ converges to $\varepsilon_{R,\infty}^0 > 0$ that solves $v(\varepsilon_{R,\infty}^0) = 0$. In contrast, as noted earlier, the exclusion threshold $\varepsilon_{R,n}$ that solves (37) in the monetary equilibrium converges to zero, and hence $\varepsilon_{R,\infty} = 0 < \varepsilon_{R,\infty}^0$. So, for n sufficiently large, the share of unbanked buyers is larger in the no-cash economy (corresponding to the case where cash has been removed) than in the monetary equilibrium.

This also implies that, for $k = k_n$,

$$
\mathcal{W}(\infty) \equiv 3\left[u(c_1^*) - c_1^*\right] + \int_{\varepsilon_{R,n}^0}^{\bar{\varepsilon}} \{\varepsilon u\left(d_\varepsilon^*\right) - (1+\sigma) d_\varepsilon^* - k_n\}\, \mathtt{d}F(\varepsilon).
$$

For n sufficiently large or, equivalently, for k_n small, the difference $\mathcal{W}(i) - \mathcal{W}(\infty)$ is then

$$
\begin{aligned}
\mathcal{W}(i) - \mathcal{W}(\infty) \quad = \quad & \int_0^{\varepsilon_{R,n}(i)} \{\varepsilon u[m_\varepsilon^c(i) - m_\varepsilon^c(i)\}\, \mathtt{d}F(\varepsilon) + \int_{\varepsilon_{R,n}(i)}^{\varepsilon_{R,n}^0} \{\varepsilon u[m_\varepsilon^c(i)] - (1+\sigma) m_\varepsilon^c(i) - k_n\}\, \mathtt{d}F(\varepsilon) \\
& + \int_{\varepsilon_{R,n}^0}^{\tilde{\varepsilon}(i)} \{[\varepsilon u\left(m_\varepsilon^c(i)\right) - (1+\sigma) m_\varepsilon^c(i)] - [\varepsilon u\left(d_\varepsilon^*\right) - (1+\sigma) d_\varepsilon^*]\}\, \mathtt{d}F(\varepsilon).
\end{aligned}
$$

By the definition of m_ε^c, the first term is positive. The second term is also positive because

$$
\varepsilon u[m_\varepsilon^c(i)] - (1+\sigma) m_\varepsilon^c(i) > (i - \sigma) m_\varepsilon^c(i) > k_n
$$

for all $\varepsilon > \varepsilon_{R,n}(i)$ by (37) and using the fact that $U^m(\varepsilon, i)$ increases with ε. For the third term, note that for $\varepsilon \in (\varepsilon^0_{R,n}, \tilde{\varepsilon})$,

$$u'(d^*_\varepsilon) = \frac{1+\sigma}{\upsilon(\varepsilon)} > \frac{1+i}{\varepsilon} = u'(m^c_\varepsilon),$$

where the inequality follows from $\varepsilon < \tilde{\varepsilon}$, and the facts that the inequality becomes an equality at $\varepsilon = \tilde{\varepsilon}$ by (34) and that $\upsilon(\varepsilon)/\varepsilon$ is strictly increasing. Moreover, if we let $\widehat{d}_\varepsilon$ solve

$$\varepsilon u'(\widehat{d}_\varepsilon) = 1 + \sigma,$$

then $d^*_\varepsilon \leq m^c_\varepsilon < \widehat{d}_\varepsilon$ for all $\varepsilon < \tilde{\varepsilon}$. Since $[\varepsilon u\,(d) - (1+\sigma)d]$ strictly increases in d up to $\widehat{d}_\varepsilon$, the third term is also positive. The result also holds for $k_n \to 0$.

Finally, note that, at the limit $k = 0$, buyer's life-time expected utility is proportional to

$$\rho W^1(0, \varepsilon, -d_\varepsilon) - \phi_\varepsilon = (1+\rho)[T + u(c^*_1) - c^*_1] + U^b(\varepsilon),$$

in both the cashless limit and the no-cash economy, and in both cases we have

$$U^b(x) = \int_0^x u[d(\varepsilon)]\mathrm{d}\varepsilon,$$

where $d(\varepsilon)$ denotes d_ε, the debt limit under the optimal contracts. Since we have proved that the debt limit is strictly higher under the cashless limit than under the no-cash economy for all $\varepsilon \leq \widetilde{\varepsilon}$ and weakly so for ε larger, it follows that all buyers are strictly worse off when cash is removed. ■

Proof of Lemma 4. The buyer's surplus from reaching an agreement with a bank is

$$\max_{m'\geq 0}\left\{-\phi_\varepsilon - (1+\pi)\,m' + \beta V^1(m', \varepsilon, -d_\varepsilon)\right\} - \max_{m^c \geq 0}\left\{-(1+\pi)\,m^c + \beta V^1(m^c, \varepsilon, 0)\right\}, \quad (63)$$

where, by the same reasoning as in Section 3, the stage-1 value function solves

$$V^1(m, \varepsilon, -d_\varepsilon) = \mathcal{S}(m + d_\varepsilon, \varepsilon) + m + W^1(0, \varepsilon).$$

According to (63), if there is an agreement, the buyer pays ϕ_ε, which grants her a positive debt limit, d_ε, in the following period. The buyer can supplement her borrowing capacity, d_ε, by accumulating m' real balances. In case of disagreement, the buyer does not receive a line of credit, $\underline{b}_\varepsilon = 0$. After substituting $V^1(m, \varepsilon, -d_\varepsilon)$ by its expression and multiplying by $1 + \rho = 1/\beta$, we obtain (40). ■

Proof of Proposition 6. We start with the following lemma.

Lemma 6 *Suppose that* $\theta < 1$. *For each* $\varepsilon \in (0, \bar{\varepsilon})$, *(44) admits a largest solution with* $d_\varepsilon > m^c_\varepsilon > 0$ *for all* k *sufficiently small if and only if* $\rho < \overline{\rho}(\pi, \theta)$. *This solution is increasing in* ε. *The set of banked buyers is* $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$ *for some* $\varepsilon_R > 0$.

Proof. From the definition of $\mathcal{G}_\varepsilon$ following (15),

$$\mathcal{G}'_\varepsilon(d) = \begin{cases} i, & \text{if } d \leq m^c_\varepsilon \\ \varepsilon u'(d) - 1, & \text{otherwise.} \end{cases} \tag{64}$$

This implies that $\mathcal{G}'_\varepsilon(d) \leq i$ for all d. Thus, if $\rho \geq (1-\theta)i$, for any $k > 0$,

$$(1-\theta)[\mathcal{G}_\varepsilon(d) - k] < \rho d$$

for all $d > 0$ and hence (44) admits no positive solution.

Conversely, suppose that $\rho < (1-\theta)i$ holds, that is, $\rho < \overline{\rho}$, and let $\varepsilon > 0$ be given. Then, $(1-\theta)\mathcal{G}'_\varepsilon(0) = (1-\theta)i > \rho$ and $(1-\theta)[\mathcal{G}_\varepsilon(0) - k] = -(1-\theta)k < 0$, as illustrated in Figure 11. For k small, $\mathcal{G}_\varepsilon(m^c_\varepsilon) > k$ and hence there exist two values of $d > 0$ such that (44) holds with equality, and the larger one has $d > m^c_\varepsilon$.

Now, note that, since $i = \pi + \rho + \pi\rho$, if we take ρ and π as primitives, $\rho < i(1-\theta)$ is equivalent to $\rho < \overline{\rho}$ with $\overline{\rho}$ given by

$$\overline{\rho}(\pi, \theta) = \begin{cases} 0 \\ \frac{(1-\theta)\pi}{\theta - (1-\theta)\pi} \\ +\infty \end{cases} \text{if } \begin{cases} \pi \leq 0 \\ 0 < \pi < \frac{\theta}{1-\theta} \\ \pi \geq \frac{\theta}{1-\theta}. \end{cases} \tag{65}$$

Note that $\rho < \overline{\rho}$ holds only if $\pi > 0$ (as $\overline{\rho} = 0$ for $\pi \leq 0$) and that since $\frac{\rho}{1+\rho} < 1$, $\rho < \overline{\rho}$ always holds if $\pi > \theta/(1-\theta)$.

Finally, we show monotonicity. Let $k > 0$ be given. For any d, and any $\varepsilon > \varepsilon' > 0$, $\mathcal{G}_\varepsilon(d) > \mathcal{G}_{\varepsilon'}(d)$ for all $d > m_{\varepsilon'}$. Thus, if there is a solution to (44) under ε', there is also a solution to (44) under ε with $d_{\varepsilon'} < d_\varepsilon$, where $d_{\varepsilon'}$ and d_ε denote the largest solutions for ε' and ε, respectively. Moreover, we have

$$\mathcal{G}_{\varepsilon'}(d_{\varepsilon'}) < \mathcal{G}_{\varepsilon'}(d_\varepsilon) < \mathcal{G}_\varepsilon(d_\varepsilon).$$

■

Now we prove Proposition 6.

Part 1. Let $\widehat{d}_{\bar{\varepsilon}}$ be such that $\mathcal{G}'_{\bar{\varepsilon}}(\widehat{d}_{\bar{\varepsilon}}) = \rho/(1-\theta) < i$ and let

$$\overline{k} = \mathcal{G}_{\bar{\varepsilon}}(\widehat{d}_{\bar{\varepsilon}}) - \frac{\rho}{1-\theta}\widehat{d}_{\bar{\varepsilon}},$$

that is, $\rho\widehat{d}_{\bar{\varepsilon}} = (1-\theta)[\mathcal{G}_{\bar{\varepsilon}}(\widehat{d}_{\bar{\varepsilon}}) - \overline{k}]$. Note that $\widehat{d}_{\bar{\varepsilon}} > m^c_{\bar{\varepsilon}}$. Thus, $(1-\theta)[\mathcal{G}_{\bar{\varepsilon}}(d) - \overline{k}] \leq \rho d$ for all d and the curve $(1-\theta)[\mathcal{G}_{\bar{\varepsilon}}(d) - \overline{k}]$ and the line ρd are tangent at $d = \widehat{d}_{\bar{\varepsilon}}$. If $k_n > \overline{k}$, there is no positive solution to (44) under $\bar{\varepsilon}$ that is larger than $m^c_{\bar{\varepsilon}}$, and, by monotonicity, this holds for all other buyers as well, and hence no buyer is banked and the unique monetary equilibrium is such that ε-buyer holds real balances of $m^c_{\bar{\varepsilon}}$. If $k_n < \bar{k}$, then the largest solution to (44) under $\bar{\varepsilon}$ satisfies $d_{\bar{\varepsilon}} > m^c_{\bar{\varepsilon}}$. Hence, buyers of type $\bar{\varepsilon}$ enters the credit market if and only if $k_n < \bar{k}$. Morerover, if $k_n < \bar{k}$, there is a $\varepsilon_{R,n} < \bar{\varepsilon}$ so that buyers of types below use cash.

Part 2. First, we show that $\varepsilon_{R,n}$ converges to zero as $k_n \to 0$. To see this, we show that for each $\varepsilon > 0$,

$$\lim_{k_n \to 0} \mathcal{G}_\varepsilon(d_\varepsilon) > 0. \tag{66}$$

Note that d_ε exists for k sufficiently small and, once it exists, it is continuous and is decreasing in k. At $k = 0$, d_ε solves

$$\mathcal{G}_\varepsilon(d_\varepsilon) = \frac{\rho}{1-\theta} d_\varepsilon,$$

and since $\mathcal{G}_\varepsilon$ is concave and $\rho < \bar{\rho}$, $d_\varepsilon > m_\varepsilon$. This proves (66) and hence $\varepsilon_{R,n} \to 0$ as $k_n \to 0$. So, $\mathrm{M}_n \to 0$. ■

Proof of Proposition 7. Part 1.Note that $d_\varepsilon \geq c^*_\varepsilon$ at the limit $k = 0$ if

$$\rho \leq \rho^*_\varepsilon(i) \equiv \frac{(1-\theta)[\mathcal{S}(c^*_\varepsilon, \varepsilon) - U^m(\varepsilon, i)]}{c^*_\varepsilon}. \tag{67}$$

Note that $\rho^*_\varepsilon(+\infty)$ is well-defined as $U^m(\varepsilon, \infty) = 0$.

Let

$$\rho^*(i) = \inf\left\{\rho^*_\varepsilon(i) : \varepsilon \in (0, \bar{\varepsilon}]\right\}. \tag{68}$$

Note that ρ^*_ε is continuous in ε. We now show that $\rho^*(i)$ defined by (68) is strictly smaller than $\bar{\rho}$. Now, fix some $\varepsilon > 0$. At $\rho = \bar{\rho}$, as illustrated by Figure 13, any d satisfying (44) has $d \leq m^c_\varepsilon$. But $\bar{\rho} \leq \rho^*(i)$ implies that, using the fact that d_ε decreases with ρ, $d_\varepsilon \geq c^*_\varepsilon$ at $\bar{\rho}$, a contradiction. Then, for any $\rho \in (\rho^*(i), \bar{\rho})$, a cashless limit exists, and, since d_ε is continuous in ε, there is an interval $[\varepsilon_0, \varepsilon_1]$ such that $d_\varepsilon < c^*_\varepsilon$ for all $\varepsilon \in [\varepsilon_0, \varepsilon_1]$. Now, the aggregate welfare at the limit $k = 0$ is given by

$$\mathcal{W}(i) \equiv 3\left[u(c^*_1) - c^*_1\right] + \int_0^{\bar{\varepsilon}} \{\varepsilon u[d_\varepsilon(i)] - d_\varepsilon(i)\}\, \mathtt{d}F(\varepsilon),$$

where the debt limits under the cashless limit by $d_\varepsilon(i)$, which solves (44). Note that the debt limits under the no-cash economy is given by $d_\varepsilon(+\infty)$. Note that since $U^m(\varepsilon; +\infty) = 0$ for all ε, $\mathcal{G}_\varepsilon(d; i) < \mathcal{G}_\varepsilon(d; +\infty)$ for all d and for all ε. It follows that $d_\varepsilon(i) \leq d_\varepsilon(+\infty)$ and strictly so if $d_\varepsilon(i) < c^*_\varepsilon$. Thus, if we remove cash, the debt limit d_ε increases for all ε and strictly so for all $\varepsilon \in [\varepsilon_0, \varepsilon_1]$ since $\rho > \rho^*(i)$. As a result, this increases welfare.

Now, if we were to treat ρ and π as primitives (instead of ρ and i), note that since

$$\frac{\partial}{\partial i} U^m(\varepsilon, i) = -m^c_\varepsilon < 0 \text{ and } \frac{\partial^2}{\partial i^2} U^m(\varepsilon, i) = -\frac{\partial}{\partial i} m^c_\varepsilon > 0,$$

$\rho^*_\varepsilon(i)$ is a strictly concave function in i and, for a given $\pi > 0$, a strictly concave function and increasing function in ρ as well, noting that $i = \rho(1+\pi) + \pi$, with $\rho^*_\varepsilon(\pi) > 0$. Thus, for any $\pi > 0$, we can also express ρ^*_ε as a function of π.

<u>Part 2</u>. Let $\rho_0 = \rho^*(i)$. From (42), at the cashless limit $k = 0$, a type-ε buyer's life time utility is

$$\rho W^1(0, \varepsilon) = (1+\rho)\left[T^1 + u(c_1^*) - c_1^*\right] + U^m(\varepsilon, i) + (1-\theta)\mathcal{G}_\varepsilon(d_\varepsilon; i). \tag{69}$$

Hence, the difference under the cashless limit and the no-cash economy is given by

$$\begin{aligned} & (1-\theta)[\mathcal{G}_\varepsilon(d_\varepsilon(i); i) - \mathcal{G}_\varepsilon(d_\varepsilon(+\infty); +\infty)] + U^m(\varepsilon, i) \\ = \ & (1-\theta)[\mathcal{S}(d_\varepsilon(i), \varepsilon) - \mathcal{S}_\varepsilon(d_\varepsilon(+\infty), \varepsilon)] + \theta U^m(\varepsilon, i), \end{aligned}$$

where $d_\varepsilon(i)$ is the debt limit under the cashless limit and $d_\varepsilon(\infty)$ is that under the no-cash economy. At $\rho = \rho_0$, as illustrated in Figure 13 (note that $\rho_0 \leq \rho^*_\varepsilon(i)$), $c^*_\varepsilon \leq d_\varepsilon(i) < d_\varepsilon(+\infty)$ and hence $\mathcal{S}(d_\varepsilon(i), \varepsilon) = \mathcal{S}_\varepsilon(d_\varepsilon(+\infty), \varepsilon)$ for all ε. Thus, the first term is zero but the second term is positive, and so removing cash decreases all buyers' utility. By continuity, for a range of ρ's larger, removing cash decreases (at least some) buyer's surpluses. But it increases aggregate welfare as in Part 1. ■

Appendix B. Lemmas for the proof of Lemma 3 and Proposition 4

Here we give three lemmas that used in our main proof of Lemma 3 and Proposition 4 and their proofs.

Lemma 7 *Let $k \geq 0$ be given. For any optimal contract, a banked buyer's optimal choice of real money balances is $m = 0$.*

Proof. Suppose that under the menu of contracts, $\{(d_\varepsilon, \phi_\varepsilon), \varepsilon \in \mathcal{E}^b\}$, the set of types that are not excluded and whose money holdings are strictly positive, $\Lambda \subset \mathcal{E}^b$, has a positive measure. Note that if a type-ε buyer holds some money, her optimal real balances are equal to $m_\varepsilon^c - d_\varepsilon$. Define an alternative menu of contracts, $\left\{\left(\hat{d}_\varepsilon, \hat{\phi}_\varepsilon\right), \varepsilon \in \mathcal{E}^b\right\}$, as follow. Let $\left(\hat{d}_\varepsilon, \hat{\phi}_\varepsilon\right) = (d_\varepsilon, \phi_\varepsilon)$ for all $\varepsilon \notin \Lambda$. For each $\varepsilon \in \Lambda$, let $\hat{d}_\varepsilon = m_\varepsilon^c$ and $\hat{\phi}_\varepsilon = \phi_\varepsilon + i\left(m_\varepsilon^c - d_\varepsilon\right)$, and hence

$$\hat{\phi}_\varepsilon - \sigma m_\varepsilon^c = \phi_\varepsilon + (i-\sigma)m_\varepsilon^c - id_\varepsilon = \phi_\varepsilon - \sigma d_\varepsilon + (i-\sigma)\left(m_\varepsilon^c - d_\varepsilon\right) > \phi_\varepsilon - \sigma d_\varepsilon,$$

since $\sigma < i$ and $m_\varepsilon^c - d_\varepsilon > 0$ for $\varepsilon \in \Lambda$. In other words, the bank makes a strictly higher profit under the alternative menu of contracts.

We need to check IC for $\left\{\left(\hat{d}_\varepsilon, \hat{\phi}_\varepsilon\right), \varepsilon \in \mathcal{E}^b\right\}$. We claim that, for any $\varepsilon \in \Lambda$, and for any ε', if an ε'-buyer chooses the contract $\left(\hat{d}_\varepsilon, \hat{\phi}_\varepsilon\right)$, her payoff will be weakly lower than the payoff from choosing $(d_\varepsilon, \phi_\varepsilon)$, which gives her a lower payoff than that from choosing $(d_{\varepsilon'}, \phi_{\varepsilon'})$ (or using cash if $\varepsilon' \notin \mathcal{E}^b$). This immediately implies IC for the alternative contract, given that it holds under the original one. Suppose that $\hat{d}_\varepsilon > d_\varepsilon \geq m_{\varepsilon'}^c$; the other cases are similar. In this case the ε'-buyer will choose $m = 0$ under the contracts $\left(\hat{d}_\varepsilon, \hat{\phi}_\varepsilon\right)$ and $(d_\varepsilon, \phi_\varepsilon)$. We have

$$\begin{aligned} -\hat{\phi}_\varepsilon + \mathcal{S}\left(\hat{d}_\varepsilon; \varepsilon'\right) &= -\phi_\varepsilon - i\left(\hat{d}_\varepsilon - d_\varepsilon\right) + \mathcal{S}\left(\hat{d}_\varepsilon; \varepsilon'\right) \\ &= -\phi_\varepsilon - i\hat{d}_\varepsilon + \mathcal{S}\left(\hat{d}_\varepsilon; \varepsilon'\right) + id_\varepsilon < -\phi_\varepsilon + \mathcal{S}\left(d_\varepsilon; \varepsilon'\right), \end{aligned}$$

where we used the definition of $\hat{\phi}$ to obtain the first inequality, and we obtain the last inequality by concavity of u and the fact that $\hat{d}_\varepsilon > d_\varepsilon \geq m_{\varepsilon'}^c = \arg\max_m[-im + \mathcal{S}(m; \varepsilon')]$, which implies that $\mathcal{S}(d_\varepsilon; \varepsilon') - id_\varepsilon > -i\hat{d}_\varepsilon + \mathcal{S}\left(\hat{d}_\varepsilon; \varepsilon'\right)$. Since under the alternative menu of contracts the bank receives a strictly higher fee for a positive measure set of buyers, it is a profitable deviation. ■

Lemma 8 *Suppose that $k = 0$. It is optimal to set $\mathcal{E}^b = [0, \bar{\varepsilon}]$.*

Proof. Suppose that $\mathcal{E}^c \equiv [0, \bar{\varepsilon}] - \mathcal{E}^b$ has a positive measure. Construct an alternative menu of contracts, $\left\{\left(\tilde{d}_\varepsilon, \tilde{\phi}_\varepsilon\right), \varepsilon \in [0, \bar{\varepsilon}]\right\}$, as follows. Keep the original contracts for all $\varepsilon \in \mathcal{E}^b$, and set $\tilde{d}$ and $\tilde{\phi}$ according to $\widetilde{d}_\varepsilon = m^c_\varepsilon$ and $\widetilde{\phi}_\varepsilon = i\widetilde{d}_\varepsilon$ for all $\varepsilon \in \mathcal{E}^c$; it follows that $U^b(\varepsilon) = U^m(\varepsilon; i)$. Then all buyers are willing to participate. To ensure incentive compatibility, we only need to show that any buyer of type $\varepsilon' \neq \varepsilon \in \mathcal{E}^c$, the buyer does not want to mimic type-ε. The payoff of doing so is

$$\mathcal{S}\left(m^c_\varepsilon, \varepsilon'\right) - im^c_\varepsilon < \mathcal{S}\left(m^c_{\varepsilon'}, \varepsilon'\right) - im^c_{\varepsilon'} = U^m(\varepsilon') \leq U^b(\varepsilon'),$$

where the first inequality follows from the definition of m^c_ε as the maximizer of $\mathcal{S}(m, \varepsilon) - im$ and the last from the participation constraint. The bank makes a positive profit for all types from $\mathcal{E}^c$, while making the same profit from all types from $\mathcal{E}^b$ and hence this is a profitable deviation. ■

Lemma 9 *The optimal solution (57) satisfies the original IC, (28).*

Proof. It follows from (56) and (60) that d_ε and ϕ_ε are both strictly increasing and continuous in ε. Here, we prove that the IC constraint holds for case of a buyer of type $\varepsilon > \tilde{\varepsilon}$ and hence $d_\varepsilon = d^*_\varepsilon$ satisfies $m^c_\varepsilon < d_\varepsilon < c^*_\varepsilon$. The other case is omitted as it follows a similar logic.

To check IC for the buyer of type ε, we need to show that the buyer does not get a strictly better payoff by accepting a contract intended for some other type ε'. We consider three regions where ε' may be on.

To define the regions, let $\varepsilon_0 < \varepsilon$ be the type such that $d_{\varepsilon_0} = m^c_\varepsilon$ and let ε_1 be the type such that $d_{\varepsilon_1} = c^*_\varepsilon$. Note that ε_1 may be outside the support of F. Note also that $\varepsilon_0 > \tilde{\varepsilon}$ since $\varepsilon > \tilde{\varepsilon}$. So now we consider three cases: (1) $\varepsilon' > \varepsilon_1$; (2) $\varepsilon' \in [\varepsilon_0, \varepsilon_1]$; (3) $\varepsilon' < \varepsilon_0$. We begin with case (2).

Case (2): $\varepsilon' \in [\varepsilon_0, \varepsilon_1]$. By choosing the contract $(d_{\varepsilon'}, \phi_{\varepsilon'})$, the buyer of type ε will not hold any real balances since $d_{\varepsilon'} \geq m^c_\varepsilon$. Thus, the IC constraint, (28), is equivalent to the necessary condition, (49), which, by Lemma 5, is satisfied by our solution. Moreover, this also implies that the buyer's payoff from her intended contract, $(d_\varepsilon, \phi_\varepsilon)$, is strictly higher than from the contracts, $\left(d_{\varepsilon_0}, \phi_{\varepsilon_0}\right)$ and $\left(d_{\varepsilon_1}, \phi_{\varepsilon_1}\right)$, provided that the latter is within the support.

Case (1): $\varepsilon' > \varepsilon_1$. In this case, $d_{\varepsilon'} > d_{\varepsilon_1} = c^*_\varepsilon$. Thus, the payoff from the contract $(d_{\varepsilon'}, \phi_{\varepsilon'})$ is

$$\mathcal{S}\left(d_{\varepsilon'}, \varepsilon\right) - \phi_{\varepsilon'} < \mathcal{S}\left(d_{\varepsilon_1}, \varepsilon\right) - \phi_{\varepsilon_1} < \mathcal{S}\left(d_\varepsilon, \varepsilon\right) - \phi_\varepsilon,$$

where the first inequality follows from the fact that $\mathcal{S}\left(d_{\varepsilon'}, \varepsilon\right) = \mathcal{S}\left(d_{\varepsilon_1}, \varepsilon\right)$ and, because ϕ is strictly increasing, $\phi_{\varepsilon'} > \phi_{\varepsilon_1}$, and the second from the argument in case (2).

Case (3): $\varepsilon' < \varepsilon_0$. There are two subcases, depending on whether $\varepsilon' > \tilde{\varepsilon}$ or not.

First, suppose that $\varepsilon' \leq \tilde{\varepsilon} < \varepsilon_0$. In this case, $(d_{\varepsilon'}, \phi_{\varepsilon'}) = (m^c_{\varepsilon'}, im^c_{\varepsilon'})$ and $m^c_{\varepsilon'} < m^c_{\varepsilon}$. Thus, if the buyer were to choose this contract, she would supplement the debt limit with $m^c_{\varepsilon} - m^c_{\varepsilon'}$ units of real balances, and would have payoff

$$-i[m^c_{\varepsilon} - m^c_{\varepsilon'}] + \mathcal{S}\left(m^c_{\varepsilon}, \varepsilon\right) - im^c_{\varepsilon'} = U^m(\varepsilon),$$

and the result follows from IR.

Second, suppose that $\varepsilon' > \tilde{\varepsilon}$. We show that the type-$\varepsilon$ buyer is better off under the contract $(d_{\varepsilon_0}, \phi_{\varepsilon_0})$ than under $(d_{\varepsilon'}, \phi_{\varepsilon'})$, and hence IC is satisfied as the buyer prefers her intended contract to $(d_{\varepsilon_0}, \phi_{\varepsilon_0})$. Note that $d_{\varepsilon_0} = m^c_{\varepsilon}$ by the definition of ε_0. Under the contract $(d_{\varepsilon'}, \phi_{\varepsilon'})$ the buyer's real balance holding is $m^c_{\varepsilon} - d_{\varepsilon'} > 0$ because $\varepsilon' < \varepsilon_0$. Thus, the type-$\varepsilon$ buyer is better off under the contract $(d_{\varepsilon_0}, \phi_{\varepsilon_0})$ than under $(d_{\varepsilon'}, \phi_{\varepsilon'})$ if

$$-\phi_{\varepsilon_0} + \varepsilon u(m^c_{\varepsilon}) - m^c_{\varepsilon} \geq -\phi_{\varepsilon'} + \varepsilon u(m^c_{\varepsilon}) - m^c_{\varepsilon} - i[m^c_{\varepsilon} - d_{\varepsilon'}].$$

Using the expression for fee, (51), to substitute ϕ_{ε_0} and $\phi_{\varepsilon'}$ out, the above inequality becomes

$$\int_{\varepsilon'}^{\varepsilon_0} u(d_x)\mathtt{d}x \geq \{\varepsilon_0 u(m^c_{\varepsilon}) - (1+i)m^c_{\varepsilon}\} - \{\varepsilon' u(d_{\varepsilon'}) - (1+i)d_{\varepsilon'}\}. \tag{70}$$

Using integration by parts, and the fact that d_{ε} satisfying (58),

$$\begin{aligned}
\int_{\varepsilon'}^{\varepsilon_0} u(d_x)\mathtt{d}x &= \varepsilon_0 u(d_{\varepsilon_0}) - \varepsilon' u(d_{\varepsilon'}) - \int_{\varepsilon'}^{\varepsilon_0} x\frac{1+\sigma}{\upsilon(x)}\left(\frac{\mathrm{d}}{\mathrm{d}x}d_x\right)\mathtt{d}x \\
&> \varepsilon_0 u(d_{\varepsilon_0}) - \varepsilon' u(d_{\varepsilon'}) - \int_{\varepsilon'}^{\varepsilon_0} (1+i)\left(\frac{\mathrm{d}}{\mathrm{d}x}d_x\right)\mathtt{d}x \\
&= \varepsilon_0 u(d_{\varepsilon_0}) - \varepsilon' u(d_{\varepsilon'}) - (1+i)(d_{\varepsilon_0} - d_{\varepsilon'}).
\end{aligned}$$

where the inequality follows from the fact that $\varepsilon' > \widetilde{\varepsilon}$, and hence, for all $x > \varepsilon'$

$$x\frac{1+\sigma}{\upsilon(x)} < (1+i).$$

Thus, noting that $d_{\varepsilon_0} = m^c_{\varepsilon}$ by the definition of ε_0, this proves (70).■

Appendix C. Competitive search equilibrium in the credit market

We first give a lemma that characterizes the optimal contract posted by banks.

Lemma 10 ***(Optimal contracts.)*** *The set $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$ with $\varepsilon_R > 0$ as the unique solution to*

$$\max_{d \geq 0} [\mathcal{S}(d, \varepsilon_R) - \sigma d] - \max_{m \geq 0} [\mathcal{S}(m, \varepsilon_R) - im] = k, \tag{71}$$

and for all $\varepsilon > \varepsilon_R$, the optimal contract, $(\phi_\varepsilon, d_\varepsilon, q_\varepsilon)$, solves

$$\sigma = \mathcal{S}'(d_\varepsilon, \varepsilon) \tag{72}$$

$$k = \alpha'(q_\varepsilon) \{[\mathcal{S}(d_\varepsilon, \varepsilon) - \sigma d_\varepsilon] - U^m(\varepsilon, i)\} \tag{73}$$

$$\phi_\varepsilon = \sigma d_\varepsilon + \theta(q_\varepsilon) \{[\mathcal{S}(d_\varepsilon, \varepsilon) - \sigma d_\varepsilon] - U^m(\varepsilon, i)\}, \tag{74}$$

where $\theta(q) \equiv \alpha'(q)q/\alpha(q)$. Moreover, even if the types are private information, it is still optimal for type-ε buyer to enter the submarket $(\phi_\varepsilon, d_\varepsilon, q_\varepsilon)$ given above.

Proof. From (23), the optimal contract solves:

$$(q_\varepsilon, d_\varepsilon, \phi_\varepsilon) = \arg\max_{(q,d,\phi)} \{\alpha(q) \{[\mathcal{S}(d, \varepsilon) - \sigma d] - [\mathcal{S}(m^c_\varepsilon) - im^c_\varepsilon]\} - kq\} \tag{75}$$

$$\phi = \frac{kq}{\alpha(q)} + \sigma d. \tag{76}$$

The first-order condition for the loan size is (72). The first-order condition for credit market tightness is (73). Using that $\alpha'(0) = 1$, it follows from (73) that the threshold for ε above which it is optimal to have an active submarket is ε_R solution to (71). Combining (73) with the free entry condition, $k = \alpha(q_\varepsilon)(\phi_\varepsilon - \sigma d_\varepsilon)/q_\varepsilon$, we obtain the expression for the intermediation fee, (74).

Now we show that our results are robust to the case where banks positing contracts without complete information on buyers' types. Specifically, we assume that banks post contracts that are observed by buyers. A contract consists of (ϕ, d, q), where ϕ is the fee (in per-period terms), d is the loan size and q is the queue length. Buyers can direct their search toward a specific contract. Each active contract (ϕ, d, q) is a submarket and the set of all active submarkets is denoted Ω.We denote $U^b(\varepsilon)$ the surplus of a type-ε buyer from credit trade in equilibrium, which is given by

$$U^b(\varepsilon) = \max_{(\phi,d,q)\in\Omega} \alpha(q) \{[\mathcal{S}(d, \varepsilon) - \phi] - [\mathcal{S}(m^c_\varepsilon, \varepsilon) - im^c_\varepsilon]\}. \tag{77}$$

Now, taking the buyers' utility levels in equilibrium as given, the bank can post contracts to maximize their profits. Following the market-utility approach as in Wright et al. (2021), we assume that if a bank posts a contract and a market tightness that gives a type-ε buyer exact the same utility as $U^b(\varepsilon)$, then the bank attracts only type-ε buyers.

A bank who wants to attract type-ε buyers then solve

$$\begin{aligned} U^s(\varepsilon) &= \max_{(\phi,d,q)} \frac{\alpha(q)}{q}(\phi - \sigma d) \\ \text{s.t. } U^b(\varepsilon) &= \alpha(q)\left\{[\mathcal{S}(d,\varepsilon) - \phi] - [\mathcal{S}(m^c_\varepsilon, \varepsilon) - i m^c_\varepsilon]\right\}. \end{aligned}$$

Note that free entry implies that in equilibrium, $U^s(\varepsilon) = k$.

Now, by duality, the optimal contract solves

$$\max_{(\phi,d,q)\in\Omega} \alpha(q)\left\{[\mathcal{S}(d,\varepsilon) - \phi] - [\mathcal{S}(m^c_\varepsilon, \varepsilon) - i m^c_\varepsilon]\right\} \text{ s.t. } \frac{\alpha(q)}{q}(\phi - \sigma d) = k,$$

which is the same contract as above. Moreover, this implies that buyers are willing to sort into the submarket intended for them. ■

Given Lemma 10, a competitive search equilibrium can be described as a list of two functions, $m^c_\varepsilon : [0, \bar{\varepsilon}] \mapsto \mathbb{R}^+$ and $(\phi_\varepsilon, d_\varepsilon, q_\varepsilon) : [0, \bar{\varepsilon}] \mapsto \mathbb{R}^{3+}$, a threshold, ε_R, and a M solutions to (12), (72), (73), (71), (74), and (24). Now we give the proof of Proposition 3.

Proof of Proposition 3. From (73), as k_n goes to 0, $\alpha'(q_n)$ goes to 0, which implies $q_n \to +\infty$ and $\alpha(q_n) \to 1$. Since banked buyers do not hold real money balances, $\text{M}_n \to 0$, i.e., the economy becomes cashless. Moreover, it can be shown that $\alpha'(q_n)q_n/\alpha(q_n) \to 0$. To see this, note that

$$1 - \alpha(q) = \int_q^{+\infty} \alpha'(x)dx,$$

where we used that $\alpha(+\infty) = 1$. Using that $\alpha' > 0$, $\int_q^{+\infty} \alpha'(x)dx > \int_q^{2q} \alpha'(x)dx$.Using that $\alpha'' < 0$, $\int_q^{2q} \alpha'(x)dx > q\alpha'(2q)$.Using the change of variable $z = 2q$,

$$1 - \alpha\left(\frac{z}{2}\right) > \frac{z}{2}\alpha'(z).$$

Taking the limit as $z \to +\infty$, and using that $\lim_{z\to\infty} \alpha(z/2) = 1$, it follows that $\lim_{z\to+\infty} z\alpha'(z) = 0$. Hence, $\lim_{q\to+\infty} \alpha'(q)q/\alpha(q) = 0$. Finally, from (74), $\theta(q) = \alpha'(q)q/\alpha(q) \to 0$ implies $\phi_\varepsilon \to \sigma d_\varepsilon$. Since this result does not depend on the presence of cash, removing cash has no effect on aggregate welfare at the cashless limit. ■

Appendix D. Self-enforcement of banks

In the main text we assumed that banks can fully commit to repay their obligations. Here we relax this assumption and assume instead that banks also have to rely on reputations to motivate their repayment. For simplicity we assume that banks can enforcement buyers to repay with a fixed cost k and, to monitor a loan of size d, with a cost of σd. The bank commits to pay these costs when the contract is signed; for example, the bank needs to set aside resources for monitoring at the time of contracting. We also assume that the buyer type ε is observable to the bank.

The timing is the same as before and in period-0 a bank is matched with a buyer, and the pair negotiate a contract between the two. The debt limit d_ε is to be endogenously determined by an incentive compatibility constraint on the bank below. For a given debt limit d_ε for the type-ε buyer, following the same logic as before, we obtain the solution

$$\phi_\varepsilon = \theta\{\mathcal{G}_\varepsilon(d_\varepsilon) - \sigma d_\varepsilon - k\} + \sigma d_\varepsilon + k. \tag{78}$$

where θ is the bank's bargaining power. Of course, such contract is profitable to both sides only if the surplus, $\mathcal{G}_\varepsilon(d_\varepsilon) - \sigma d_\varepsilon - k$, is nonnegative.

Now we consider incentive compatibility for bank repayment and optimal debt limit. For a bank to repay its debt, it must be the case that its future profits outweigh its temptation to default. Thus, the debt limit maximizes the joint surplus, subject to incentive-compatibility constraint for repayment: d_ε solves

$$\Pi_\varepsilon = \max_{d_\varepsilon} \mathcal{S}(d_\varepsilon, \varepsilon) - \sigma d_\varepsilon - k - U^m(\varepsilon, i) \text{ subject to } d_\varepsilon \leq \frac{\phi_\varepsilon - \sigma d_\varepsilon - k}{\rho}. \tag{79}$$

Lemma 11 *Let $\theta > 0$. The equation (79) has a unique solution with $d_\varepsilon > m^c_\varepsilon$ for all ε and for all positive k small if and only if*

$$\theta(i - \sigma) > \rho. \tag{80}$$

Moreover, in this case, the solution is given by

$$\varepsilon u'(d_\varepsilon) = 1 + \sigma + \frac{\lambda_\varepsilon}{1 + \lambda_\varepsilon}\frac{\rho}{\theta}, \tag{81}$$

for some $\lambda_\varepsilon \geq 0$ with $\lambda_\varepsilon > 0$ only if

$$\rho d_\varepsilon = \theta\{\varepsilon u(d_\varepsilon) - (1 + \sigma)d_\varepsilon - k - U^m(\varepsilon, i)\}.$$

Moreover, Π_ε strictly increases with ε.

Thus, there is a unique monetary equilibrium in which the bank serves buyers of types $\varepsilon \geq \varepsilon_R$, where ε_R is determined by $\Pi_{\varepsilon_R} = 0$.

Proof. To show that a solution with $d_\varepsilon > m_\varepsilon^c$ exists, we only need to show that there exists such debt limit that satisfies

$$\rho d_\varepsilon \leq \theta\{\mathcal{G}_\varepsilon(d_\varepsilon) - \sigma d_\varepsilon - k\}. \tag{82}$$

Note that this condition is almost identical to that in the case where buyers have limited commitment, except that we replace $1-\theta$ by θ here and $\sigma > 0$ here. Now, if $\rho \geq \theta(i-\sigma)$, for any $k > 0$, (82) does not hold for any $d > 0$.

Conversely, suppose that $\rho < \theta(i-\sigma)$ holds. Then, $\theta\{\mathcal{G}_\varepsilon'(0) - \sigma\} = \theta(i-\sigma) > \rho$ and $\theta)[\mathcal{G}_\varepsilon(0) - k] = -\theta k < 0$. For k small, $\mathcal{G}_\varepsilon(m_\varepsilon^c) - m_\varepsilon^c > k$ and hence there exist two values of $d > 0$ such that (82) holds with equality, and the larger one has $d > m_\varepsilon^c$. Let $\hat{d}_\varepsilon$ satisfies $u'(\hat{d}_\varepsilon) = (1+\sigma)/\varepsilon$. If $\hat{d}_\varepsilon > m_\varepsilon^c$ satisfies (82), then $\lambda_\varepsilon = 0$ and it is the optimal solution. Otherwise, the larger solution that satisfies (82) is smaller than $\hat{d}_\varepsilon$ and is the optimal solution.

Now,

$$\frac{\partial}{\partial \varepsilon}\Pi_\varepsilon \quad = \quad u(d_\varepsilon) - u(m_\varepsilon) + \frac{\lambda_\varepsilon}{1+\lambda_\varepsilon}\frac{\rho}{\theta}\left(\frac{\partial}{\partial \varepsilon} d_\varepsilon\right) > 0, \tag{83}$$

where the inequality follows from the fact that

$$\frac{\partial}{\partial \varepsilon} d_\varepsilon = \frac{\theta[u(d_\varepsilon) - u(m_\varepsilon)]}{\rho\left(1 - \frac{\lambda_\varepsilon}{1+\lambda_\varepsilon}\right)} > 0.$$

■

We have then the following proposition about the cashless limit.

Proposition 8 *Let $\pi > \sigma$ be given. There is a threshold $\overline{\rho}$ such that (80) holds for all $\rho < \overline{\rho}$. Consider a sequence of economies characterized by $\{k_n\}_{n=0}^{+\infty}$ with $\lim_{n\to+\infty} k_n = 0$.*

1. *Suppose that $\rho < \overline{\rho}$. As $n \to +\infty$, the economy becomes cashless, $m_n \to 0$.*
2. *At the limit, there are two thresholds $\rho^* < \rho_1 \leq \overline{\rho}$ such that the following holds. For all $\rho \in (\rho^*, \overline{\rho})$, removing cash increases the aggregate welfare. However, for all $\rho \in (\rho^*, \rho_1)$, there is a positive measure of buyers who are worse off when cash is removed.*

Proof. The construction of $\overline{\rho}$ as a function of π follows exactly the same reasoning as in the main text. Also, ρ^* is constructed to be the highest threshold such that $\rho \leq \rho^*$ implies that $d_\varepsilon = \hat{d}_\varepsilon$ for all ε. Thus, for $\rho > \rho^*$, $d_\varepsilon < \hat{d}_\varepsilon$ for a subset of buyers. As in the main text, at ρ^*, all buyers are worse off if cash is removed because of higher ϕ_ε. Thus, for ρ slightly higher this remains true by continuity. ■

References

[1] Alvarez, Fernando, and Urban Jermann (2000). "Efficiency, equilibrium, and asset pricing with risk of default." Econometrica 68, 775–97.

[2] Andolfatto, David (2021). "Assessing the impact of Central Bank digital currency on private banks", The Economic Journal 131, 525–540.

[3] Banet, Jeremie, and Lucie Lebeau (2025). "Central bank digital currency: Financial inclusion vs disintermediation". Working Paper.

[4] Bayeh, Berhan, Isaiah Nardone, Shaun O'Brien, and Hailey Phelps (2025). "Findings from the diary of consumer payment choice." Federal Reserve Financial Services. https://www.frbservices.org/news/research/2025-findings-from-the-diary-of-consumer-payment-choice

[5] Bethune, Zach, Tai-Wei Hu, and Guillaume Rocheteau (2018a). "Indeterminacy in credit economies," Journal of Economic Theory, 175, 556-584.

[6] Bethune, Zach, Tai-Wei Hu, and Guillaume Rocheteau (2018b). "Optimal credit cycles," Review of Economic Dynamics 27, 231-245.

[7] Bethune, Zachary, Guillaume Rocheteau, Russell Wong, and Cathy Zhang (2022). "Lending relationships and optimal monetary policy." Review of Economic Studies 89, 1833-1872.

[8] Cavalcanti, Ricardo De O., and Neil Wallace, (1999) "Inside and outside money as alternative media of exchange," Journal of Money, Credit and Banking 31, 443-457.

[9] Chiu, Jonathan, Seyed Mohammadreza Davoodalhosseini, Janet Jiang, and Yu Zhu (2023). "Bank market power and Central Bank digital currency: Theory and quantitative assessment," Journal of Political Economy 131, 1213-1248.

[10] Choi, Michael, and Guillaume Rocheteau (2023). "A model of retail banking and the deposits channel of monetary policy," Journal of Monetary Economics 139, 127-147.

[11] den Haan, Wouter, Garey Ramey, and Joel Watson (2003). "Liquidity flows and fragility of business enterprises." Journal of Monetary Economics 50, 1215–1241.

[12] Di Iorio, Alberto, Anneke Kosse and Ismail Mustafi (2025). "And so we pay: more digital and faster, with cash still in play." Committee on Payments and Market Infrastructures (CPMI) Briefs #8, March.

[13] Diamond, Douglas (1984). “Financial intermediation and delegated monitoring.” The Review of Economic Studies 51, 393–414.

[14] Drechsler, Itamar, Alexi Savov, and Philipp Schnabl (2017). “The deposits channel of monetary policy.” The Quarterly Journal of Economics 132, 1819–1876.

[15] Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen (2005). “Over-the-counter markets.” Econometrica 73, 1815-1847.

[16] Freeman, Scott, and Finn E. Kydland (2000). “Monetary Aggregates and Output.” American Economic Review 90, 1125-1135.

[17] Freixas, Xavier, and Jean-Charles Rochet (1997). Microeconomics of Banking. The MIT Press, Cambridge.

[18] Gu, Chao, Fabrizio Mattesini, and Randall Wright (2016). “Money and credit redux.” Econometrica 84, 1–32.

[19] Gu, Chao, Fabrizio Mattesini, Cyril Monnet, and Randall Wright (2013a). "Banking: A New Monetarist approach," Review of Economic Studies 80, 636-662.

[20] Gu, Chao, Fabrizio Mattesini, Cyril Monnet, and Randall Wright (2013b). "Endogenous credit cycles," Journal of Political Economy 121, 940 - 965.

[21] Head, Allen, Timothy Kam, Sam Ng, Guangqian Pan (2025). “Money and imperfectly competitive credit”, Journal of Economic Theory 228, 106050.

[22] Holmstrom, Bengt, and Jean Tirole (1997). “Financial intermediation, loanable funds, and the real sector.” Quarterly Journal of Economics 112, 663–91.

[23] Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill (2025). The economics of Over-the-Counter markets: a toolkit for the analysis of decentralized exchange, Blackwell’s.

[24] Kehoe, Timothy and David Levine (1993). "Debt constrained asset markets," Review of Economic Studies 60, 865-88.

[25] Khiaonarong, Tanai, and David Humphrey (2023). “Measurement and use of cash by half the world’s population.” IMF Working Paper 23/62.

[26] Kiyotaki, Nobuhiro, and John Moore (2002). “Evil is the root of all money.” American Economic Review 92, 62–66.

[27] Kocherlakota, Narayana (1998). "Money is memory", Journal of Economic Theory 81, 232-251.

[28] Kocherlakota, Narayana, and Neil Wallace (1998). "Incomplete record-keeping and optimal payment arrangements", Journal of Economic Theory 81, 272-289.

[29] Lagos, Ricardo, "Onetary economics at 30: a reexamination of money in cashless limiting monetary economies," NBER Working Paper 34155.

[30] Lagos, Ricardo and Guillaume Rocheteau (2005). "Inflation, output, and welfare." International Economic Review 46, 495-522.

[31] Lagos, Ricardo, and Randall Wright (2005). "A unified framework for monetary theory and policy analysis." Journal of Political Economy 113, 463-484.

[32] Lagos, Ricardo, Guillaume Rocheteau, and Randall Wright (2017). "Liquidity: A New-Monetarist perspective," Journal of Economic Literature 55, 371–440.

[33] Lagos, Ricardo, and Shengxing Zhang (2020). "Turnover liquidity and the transmission of monetary policy." American Economic Review 110, 1635-72.

[34] Lagos, Ricardo, and Shengxing Zhang (2022). "The limits of onetary economics: On money as a constraint on market power." Econometrica 90, 1177–204.

[35] Li, Yiting (2011). "Currency and checking deposits as means of payment," Review of Economic Dynamics 14, 403-417.

[36] Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green (1995). Microeconomic Theory, Oxford University Press.

[37] Moen, Espen (1997). "Competitive Search Equilibrium." Journal of Political Economy 105, 385–411.

[38] Ohlhaver, Puja, E. Glen Weyl, and Vitalik Buterin (2022). "Decentralized Society: Finding Web3's Soul," Working paper.

[39] Ostroy, Joseph, and Ross Starr (1990). "The transactions role of money," Handbook of Monetary Economics, Edited by B.M. Friedman and F.H. Hahn, Elsevier Science Publishers, Chapter 1, Volume 1, 3-62.

[40] Petrosky-Nadeau, Nicolas and Etienne Wasmer (2017). Labor, Credit, and Goods Markets, MIT Press.

[41] Rocheteau, Guillaume and Ed Nosal (2017). Money, Payments, and Liquidity, MIT Press.

[42] Rocheteau, Guillaume, and Randall Wright (2005). "Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium," Econometrica 73, 175-202.

[43] Rocheteau, Guillaume, Randall Wright, and Cathy Zhang (2018). "Corporate finance and monetary policy." American Economic Review 108, 1147-86.

[44] Rogoff, Kenneth (2014). "Costs and benefits to phasing out paper currency." Dinner speech at NBER Macroeconomics Annual Conference, April 11, 2014.

[45] Rogoff, Kenneth (2016). The Curse of Cash: How Large-Denomination Bills Aid Crime and Tax Evasion and Constrain Monetary Policy. Princeton University Press.

[46] Sanches, Daniel and Stephen Williamson (2010). "Money and credit with limited commitment and theft," Journal of Economic Theory 145, 1525-1549.

[47] Scharfstein, David and Adi Sunderam (2017). "Market power in mortgage lending and the transmission of monetary policy". Manuscript, Harvard Business School.

[48] Townsend, Robert (1980). "Models of money with spatially separated agents," Kareken, N Wallace (Eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis, 265-303.

[49] Venkateswaran, Venky, and Randall Wright (2014). "Pledgability and liquidity: A New Monetarist model of financial and macroeconomic activity." NBER Macroeconomics Annual 28, 227–70.

[50] Wallace, Neil (1980) "The overlapping generations model of fiat money." J Kareken, N Wallace (Eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis, 49-82.

[51] Wang, Yifei, Toni Whited, Yufeng Wu, and Kairong Xiao (2022). "Bank market power and monetary policy Transmission: Evidence from a structural estimation". The Journal of Finance 77, 2093-2141.

[52] Wasmer, Etienne and Phillipe Weil (2004). "The macroeconomics of labor and credit market frictions." American Economic Review 94, 944–963.

[53] Williamson, Stephen. (2012). "Liquidity, monetary policy, and the financial crisis: A new monetarist approach." American Economic Review 102, 2570-2605.

[54] Woodford, Michael (1998). "Doing without Money: Controlling Inflation in a Post-Monetary World", Review of Economic Dynamics 1, 173-219