Breaking Up With M: Cashless Limits Under Limited Commitment*

Tai-Wei Hu
University of Bristol

Guillaume Rocheteau
University of California, Irvine

This version: July 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 mitigated 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 Luis Araujo, Wei Cui, Cyril Monnet, Russell Wong, Randall Wright, Yidan Yan, Yu Zhu, Shengxing Zhang, and seminar participants at Carnegie Mellon University, FRB Philadelphia, Federal Reserve Board, Renmin University of China, SaMMF, University of Bristol, University College London, University of California at Davis, University of California at Irvine, University of California at Riverside, University of Essex, University of Hawaii at Manoa, University of Paris Pantheon-Assas for their comments.

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 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. 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 the 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. We distinguish between the aggregate relevance of cash for total welfare and its distributional relevance for the

[1] For some empirical evidence on the evolution toward cashlessness, see the Online Appendix E.

[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."

allocation of surplus between banks and consumers.

Throughout the paper, we abstract from two common motivations for eliminating cash—namely, tax evasion 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.

The distributional relevance of cash under complete information 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 #1. *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.*

In the Online Appendix C, we consider a variant of the environment in which

[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) adopt a similar approach to endogenize the fraction of credit trades. However, in their model there are heterogeneous 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.

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.

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. *Under costly enforcement, when banks are monopolistic and consumers are privately informed, the removal of cash at the cashless limit reduces aggregate welfare by generating financial exclusion and exacerbating 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. *Consider three situations ranked according to banks' market power. (i) If banks have no bargaining power and no information regarding consumers' liquidity needs, removing cash is neutral. (ii) If banks obtain full bargaining power* *but remain uninformed, cash removal reduces aggregate welfare. (iii) If banks have both 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. *Under costly record-keeping, provided inflation is not too high, the elimination of cash at the cashless limit 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 The rise of cashlessness and government responses

Many economies are transitioning rapidly toward cashless payments, including the United States. Measured by cash withdrawals, reliance on physical currency fell between 2013 and 2023 from 13 to 6 percent of GDP in advanced economies and from 31 to 15 percent in developing economies. The U.S. Diary of Consumer Payment Choice documents a similar shift in payment instruments (Bayeh et al., 2025): between 2015 and 2024, the share of payments made in cash fell from 33 to 14 percent, while the share made by credit card rose from 18 to 35 percent. In value terms, cash accounted for 10 percent of payments in 2015 and less than 4 percent in 2024. By this measure, the United States is already close to a cashless economy. More detailed evidence is provided in Online Appendix E.

Yet governments in several jurisdictions have taken steps to preserve cash as a viable means of payment. Sweden has strengthened institutional support for cash, the European Court of Justice has ruled that cash should, in principle, be accepted, and some U.S. jurisdictions, including Philadelphia, require retailers to accept cash.[5] A central motivation is financial inclusion: eliminating cash may exclude some households from participating in the payment system. Our analysis reinforces this concern by showing that eliminating cash can generate financial exclusion or credit rationing even when cash demand is negligible and all consumers initially have access to banking services. More broadly, we show that the welfare consequences of eliminating cash depend on the source of limited commitment, the information structure, and the degree of intermediary market power.

1.2 Literature

Our paper contributes to the New-Monetarist literature studying environments in which money and credit coexist, e.g., Gu, Mattesini, and Wright (2016) and references therein. Within this literature, Lagos and Zhang (2022) develop a framework to study cashless limits in an economy where banks act as payment intermediaries, connecting sellers to consumers using credit, with fiat money serving as an outside option that disciplines banks' market power.[6] Their analysis focuses on a holdup problem arising from irreversible production decisions. By contrast, we treat limited commitment as the fundamental friction generating the coexistence of money and credit and explicitly model the institutions that mitigate this friction. Whereas banks

[5] The "Cash Inquiry," initiated by the Ministry of Finance of Sweden, proposes requiring businesses to retain the ability to accept cash, either through staffed services or automated systems. See also the Economist's article, "Why Europe is rediscovering the virtues of cash," Jan 8th, 2026.

[6] The argument that central bank money provides an outside option that disciplines banks' market power also appears in the literatures on CBDCs (e.g., Andolfatto, 2021; Chiu et al., 2023) and bank lending (e.g., Rocheteau, Wright, and Zhang, 2018; Head et al., 2025; Dong et al., 2026).

in Lagos and Zhang provide sellers with access to consumers through an exogenously available credit technology, banks in our model make credit feasible by operating enforcement or monitoring technologies. Consequently, the availability and terms of credit are determined endogenously. The model endogenizes the intensive margin of credit, through equilibrium loan sizes or debt limits, and the extensive margin, through endogenous participation in banking relationships, allowing us to characterize how eliminating cash affects credit rationing and financial exclusion.[7]

To mitigate limited commitment, we consider two institutions studied in the literature. The first, in the spirit of Diamond (1984), is delegated monitoring, whereby banks incur costly monitoring and enforcement expenditures in order to transform consumers' illiquid IOUs into inside money. See also Freixas and Rochet (1997, Section 2.4). The second institution is public record keeping, as in Kehoe and Levine (1993). Applications in New Monetarist environments include Cavalcanti and Wallace (1999), Sanches and Williamson (2010), Gu et al. (2013), and Bethune, Hu, and Rocheteau (2018), among others.[8] Unlike this literature, we allow for heterogeneity in credit access and bank market power, with the perfectly competitive banking environment arising as a special case. We further show that the welfare consequences of removing cash can reverse relative to economies with externally enforceable credit.

Another key friction in our model is private information about consumers' borrowing needs, which generates informational rents and distorts credit provision. The version of our model with monopolistic banks solving a mechanism-design problem is related to Choi and Rocheteau (2023) and references therein, although we focus on credit rather than deposit markets and allow participation in banking relationships to be endogenous through a participation cost.[9] A central implication is that the welfare consequences of eliminating cash depend not only on the degree of banks' market power but also on its source—bargaining power or private information.

Finally, Menzio and Spinella (2026) also revisit the cashless-limit debate in a model where money serves as a store of value and financial markets are frictional

[7]The endogenous participation in our model is related to Freeman and Kydland (2000) and Li (2011), among others.

[8]Kocherlakota and Wallace (1998) study advances in payment technologies in a random-matching economy where money coexists with a public record of past actions that is updated with a lag.

[9]Ahnert et al. (2025) also study how payment instruments interact with private information in credit markets, but their focus is on merchant anonymity in cash transactions and payment-data design rather than on the cashless-limit implications of limited commitment.

and characterized by noisy search. They show that the cashless limit may or may not coincide with the nonmonetary equilibrium depending on whether the limiting economy is competitive. Our competitive-search specification in the Online Appendix C delivers a similar cashless limit through a different mechanism: as entry eliminates banks' market power, the economy converges to a competitive allocation in which cash becomes irrelevant.

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 1, in which a good is traded in a 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.

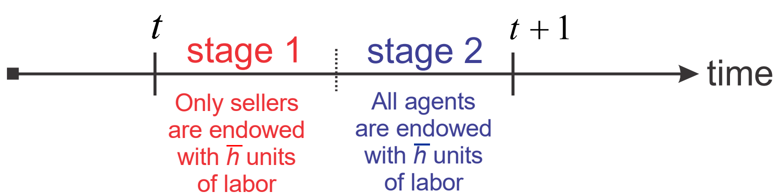

Figure 1: Description of time and endowments

Agents' lifetime preferences over consumption and work are represented by

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

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 (and stays constant forever) from a continuous distribution, $F(\varepsilon)$, with density $f(\varepsilon)$ and support $[0, \bar{\varepsilon}]$. The heterogeneity in ε captures differences in liquidity needs across consumers and will generate endogenous participation in banking relationships. 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 type → / 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 $\zeta^\varkappa_\tau \bar{h}$ units of labor in stage τ, where $\zeta^\varkappa_\tau \in \{0, 1\}$. Assumptions regarding $\zeta^\varkappa_\tau$ are summarized in Table 2. For buyers, $\zeta^1_1 = 0$ and $\zeta^1_2 = 1$. Thus, buyers can only work in stage 2. By contrast, for sellers, $\zeta^2_1 = \zeta^2_2 = 1$. So, sellers can work in both stages. We assume that $\bar{h}$ is sufficiently large so that the constraint, $h_{t,\tau} \le \zeta^\varkappa_\tau \bar{h}$ never binds in equilibrium whenever $\zeta^\varkappa_\tau = 1$.

agent's type → / endowment ↓	$\varkappa = 1$ (buyers)	$\varkappa = 2$ (sellers)
Stage 1: $\zeta^\varkappa_1$	=0	=1
Stage 2: $\zeta^\varkappa_2$	=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 liquid IOUs as liabilities, interpreted as deposits, and hold illiquid IOUs as assets, corresponding to loans to nonfinancial agents, the buyers. Figure 2 represents the trades between banks, buyers, and sellers in stages 1 (left panel) and 2 (right panel). Equivalently, buyers can directly issue IOUs to sellers, while banks act as guarantors of those obligations in exchange for an intermediation fee. Because the two interpretations are payoff-equivalent, we adopt the latter in what follows and restrict attention to a single type of debt instrument. We refer to this instrument as a *bond* in our analysis.

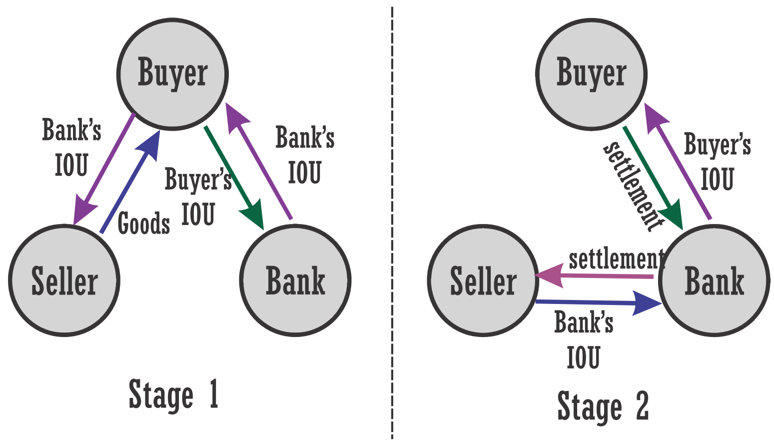

Figure 2: Bank credit

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. For tractability, we assume that relationships are permanent; allowing for stochastic destruction and rematching would not alter the main mechanisms. 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.[10] 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 $\sigma > 0$.[11] Without loss, we assume that sellers do not have access to banks and hence they cannot issue debt. 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.

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., $M_{t+1} = (1 + \pi)M_t$,

[10] 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.

[11] In principle, the enforcement cost could depend on all the dimensions of the contract, including 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 our results. For a specification with a strictly convex cost, see Bethune et al. (2022).

through lump-sum transfers (or taxes if $\pi < 0$) to buyers. We denote $1 + i \equiv (1+\rho)(1+\pi)$ and we assume $i > \sigma$. This assumption implies that supplying liquidity through bank credit is technologically less costly than self-insurance through money holdings.

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. 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. Credit arrangements are subject to limited commitment and therefore require monitoring or record-keeping technologies.

Under costly external enforcement, we interpret bank-issued payment instruments broadly as encompassing both credit- and debit-card transactions, since both are typically intermediated by banks and require monitoring technologies. In a debit-card interpretation, following Choi and Rocheteau (2023), buyers deposit funds with banks before entering stage 1. Deposits earn a gross real return R_d, implying an opportunity cost of liquidity equal to $\sigma \equiv (\beta^{-1} - R_d)/R_d$. This opportunity cost is isomorphic to the enforcement cost in the credit interpretation. Hence, the analysis in Sections 3–4 applies equally to credit- and deposit-based payment arrangements.

The re-interpretation is more delicate for the analysis in Section 5, where credit is sustained by public monitoring and self-enforcement rather than by costly external enforcement. In that environment, the bank-issued payment instrument is more naturally interpreted as credit-card debt rather than demand deposits. Under demand-deposit arrangements, there is no natural role for depositor default: consumers spend balances they already hold, whereas the self-enforcement mechanism in Section 5 relies on borrowers' incentives to repay in order to preserve future access to credit. The analysis in Section 5 therefore identifies a distinct channel through which cash can remain relevant at the cashless limit, with welfare implications opposite to those obtained under costly external enforcement. More broadly, the comparison highlights that the welfare consequences of eliminating cash depend critically on the mechanism through which limited commitment is mitigated by different payment instruments.

We introduce heterogeneity in liquidity needs by assuming that buyers differ in their stage-1 preferences. In Online Appendix E, we present 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.[12] 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. (2013). We introduce such a reputational mechanism on the banks' side in Online Appendix D. 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.

[12] 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 introducing 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:

$$m_\varepsilon(\underline{b}) \in \arg\max_{m' \geq 0} \left\{ -\frac{m'}{R_2^m} + \beta V^1(m', \varepsilon, \underline{b}) \right\}. \tag{5}$$

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,

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

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 $\partial V^1(m, \varepsilon, \underline{b})/\partial m = \varepsilon u'(m - \underline{b})$.

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

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

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 $-\underline{b} \geq 0$, hold 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

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

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 $-\underline{b} = d_\varepsilon$ 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 $\mathcal{G}_\varepsilon$, the 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,

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

The determination of the gain, $\mathcal{G}_\varepsilon(d_\varepsilon)$, is illustrated graphically in Figure 3. It is the difference between the net surplus from trade when buyers have and do not have access to a debt limit, d_ε, and real money balances are chosen optimally. From (15), if d_ε is lower than $m^c_\varepsilon \equiv m_\varepsilon(0)$ 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 $\mathcal{S}(d_\varepsilon, \varepsilon) - \mathcal{S}(m_\varepsilon(0), \varepsilon)$.

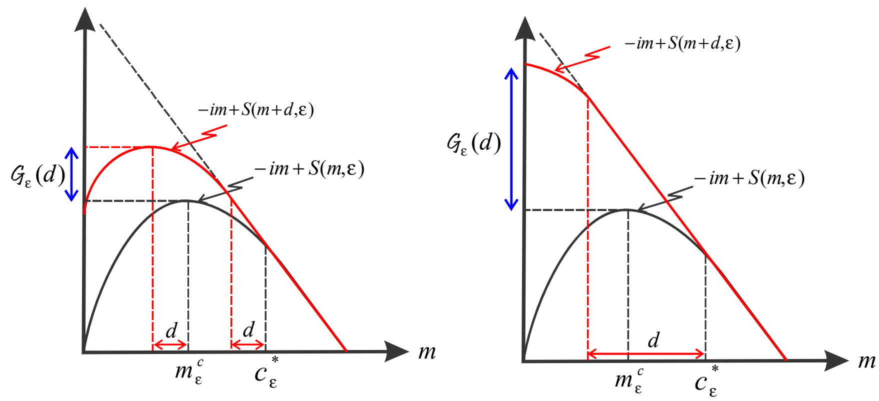

Figure 3: Value to the buyer of access to credit. Left panel: $d < m^c_\varepsilon$. Right panel: $d > m^c_\varepsilon$.

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.[13] The 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.,

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

[13] 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. Also, if there was a probability that the bank could renegotiate the terms of the contract in stage 1, then banked buyers would want to accumulate money holdings as in Rocheteau et al. (2018).

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, $\mathcal{G}_\varepsilon - \sigma d_\varepsilon - k$.

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

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

It is determined graphically in Figure 4. The threshold, ε_R, is decreasing in i and increasing in k. It approaches 0 as k tends to 0.

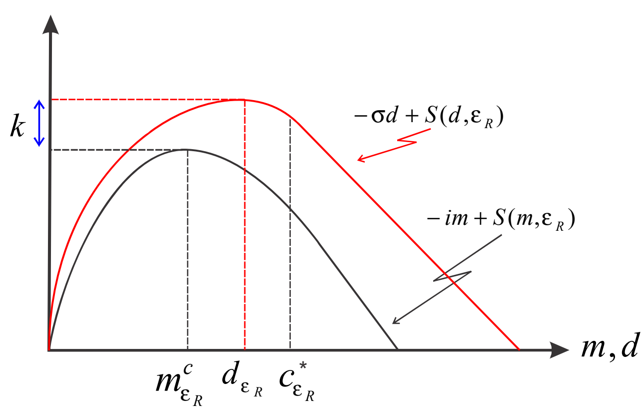

Figure 4: Determination of the exclusion threshold, ε_R

3.4 The value of money

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

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

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 individual real money holdings for unbanked buyers, $m_\varepsilon^c \equiv m_\varepsilon(0)$, debt limits, d_ε, aggregate real money balances, an exclusion threshold, ε_R, 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 $i > \sigma$, 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 exists 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 corresponding convergent sequence of monetary equilibria with aggregate real money balances $\{\text{M}_n\}_{n=1}^{+\infty}$, where each $\text{M}_n > 0$. 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. *Distributional relevance of cash. 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 $\theta > 0$, eliminating cash has no effect on aggregate welfare. Cash changes buyers' outside options and therefore alters the gains from forming a lending relationship. However, the outside option enters the joint surplus only as an additive term and hence does not affect the marginal condition determining the optimal debt limit. Consequently, the loan size d_ε, which maximizes the joint surplus, remains unchanged when cash is removed. Since consumption allocations in stage 1 are unaffected, aggregate welfare also remains unchanged.

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.

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' liquidity needs, and we consider that scenario in this section. It is this informational asymmetry that limits 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

$$\upsilon(\varepsilon) \equiv \varepsilon - \frac{1 - F(\varepsilon)}{f(\varepsilon)}. \tag{23}$$

It is strictly increasing, and $\upsilon(\varepsilon)/\varepsilon$ 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 $\mathcal{E}^b \subset [0, \bar{\varepsilon}]$.[14] 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

$$\int_{\mathcal{E}^b} (\phi_\varepsilon - \sigma d_\varepsilon - k) \, \mathrm{d}F(\varepsilon). \tag{24}$$

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 greater than or equal to the fee charged by the bank, ϕ_ε, that is,

$$\max_{m \geq 0} \left\{ \mathcal{S}(m + d_\varepsilon, \varepsilon) - im - \phi_\varepsilon \right\} \geq U^m(\varepsilon, i) \equiv \max_{m \geq 0} \left[-im + \mathcal{S}(m, \varepsilon) \right]. \tag{25}$$

Thus, (25) specifies that the utility of the buyer net of fee must be greater than the utility from being unbanked. As before, we use m_ε^c to denote the optimal m in $U^m(\varepsilon, i)$. The optimal menu must also satisfy incentive-compatibility (IC) constraints requiring that, for any $\varepsilon,\ \varepsilon' \in \mathcal{E}^b$,

$$\max_{m \geq 0} \left\{ \mathcal{S}(m + d_\varepsilon, \varepsilon) - im - \phi_\varepsilon \right\} \geq \max_{m \geq 0} \left\{ \mathcal{S}(m + d_{\varepsilon'}, \varepsilon) - im - \phi_{\varepsilon'} \right\}, \tag{26}$$

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

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

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 holding money, i, is larger than the monitoring/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 $d_\varepsilon \geq m_\varepsilon^c$

[14] 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.

for all $\varepsilon \in \mathcal{E}^b$. Using this observation and textbook 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

$$U^b(\varepsilon) \equiv \varepsilon u(d_\varepsilon) - d_\varepsilon - \phi_\varepsilon, \quad \forall \varepsilon \in \mathcal{E}^b, \tag{28}$$

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, (23), by substituting ϕ_ε from (28) into (24) 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*

$$\max_{\varepsilon_R, d_\varepsilon \geq m^c_\varepsilon} \int_{\varepsilon_R}^{\bar{\varepsilon}} \{\upsilon(\varepsilon)u(d_\varepsilon) - (1+\sigma)\, d_\varepsilon - k - U^m(\varepsilon_R, i)\}\, \mathrm{d}F(\varepsilon). \tag{29}$$

From (29), 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

$$[\upsilon(\varepsilon_R)u\left(d_{\varepsilon_R}\right) - (1+\sigma)\, d_{\varepsilon_R}] - \left[\upsilon(\varepsilon_R)u\left(m^c_{\varepsilon_R}\right) - (1+i)\, m^c_{\varepsilon_R}\right] = k. \tag{30}$$

This condition is identical to the one obtained under complete information, (19), except that the valuation, ε, has been replaced by the virtual valuation, $\upsilon(\varepsilon)$. This also implies that, as in Proposition 1, money and credit coexist if and only if $k < \mathcal{G}_{\bar{\varepsilon}}(d_{\bar{\varepsilon}}) - \sigma d_{\bar{\varepsilon}}$ since $\upsilon(\bar{\varepsilon}) = \bar{\varepsilon}$.

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

$$\upsilon(\varepsilon)u'(d_\varepsilon) = 1 + \sigma. \tag{31}$$

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 (31), 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.[15]

Proposition 3 ***(Optimal menu of lending contracts.)*** *Suppose that* $0 < k < \mathcal{G}_{\bar{\varepsilon}}(d_{\bar{\varepsilon}}) - \sigma d_{\bar{\varepsilon}}$ *so that serving the highest buyer type is profitable. Then there exists a unique monetary equilibrium in which a nonempty set of buyers is banked.*

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

$$\upsilon(\widetilde{\varepsilon}) = \left(\frac{1+\sigma}{1+i}\right)\widetilde{\varepsilon}, \tag{32}$$

the optimal menu of contracts offered by each bank is

$$\begin{aligned} (d_\varepsilon, \phi_\varepsilon) &= (m^c_\varepsilon, im^c_\varepsilon) & \forall \varepsilon \in [\varepsilon_R, \tilde{\varepsilon}] \quad (33) \\ &= \left(d^*_\varepsilon, id_{\tilde{\varepsilon}} + \int_{\tilde{\varepsilon}}^{\varepsilon} \frac{\partial \mathcal{S}[d(x), x]}{\partial d} d'(x) \mathtt{d}x\right) & \forall \varepsilon \in (\tilde{\varepsilon}, \bar{\varepsilon}] \quad (34) \end{aligned}$$

where ε_R *is the unique solution to (30) with* $d_{\varepsilon_R} = m^c_{\varepsilon_R}$*, i.e.,*

$$k = (i-\sigma)m^c_{\varepsilon_R}. \tag{35}$$

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 (30) with* $d_{\varepsilon_R} = d^*_{\varepsilon_R}$.

For k sufficiently small, there are banked buyers for whom the IR constraint binds, $\varepsilon_R < \tilde{\varepsilon}$. For all ε above $\tilde{\varepsilon}$, $d_\varepsilon \geq m^c_\varepsilon$ 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 that under complete information, thereby generating a welfare loss.

For all ε below the threshold, $\tilde{\varepsilon}$, the participation constraint, $d_\varepsilon \geq m^c_\varepsilon$, 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

[15] Note that in (34) 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.

participation constraint at equality while generating a positive profit, $(i-\sigma)d_\varepsilon$, for the bank. 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. When k is sufficiently high, all banked buyers face slack IR constraints, $\varepsilon_R > \tilde{\varepsilon}$.

Relative to the complete-information benchmark, private information distorts credit provision downward for all banked buyers. In particular, buyers with relatively high liquidity needs ($\varepsilon > \tilde{\varepsilon}$) do earn informational rents, but the bank reduces debt limits below the complete-information level in order to limit those rents. We interpret this distortion as credit rationing along the intensive margin. Along the extensive margin, ε_R is also higher under private information than under complete information.

4.2 On the relevance of cash under private information

The remainder of the section examines the relevance of cash in the cashless limit. To do so, we first consider the optimal contract in the no-cash economy. Since this is equivalent to having $i = +\infty$ (and hence $m^c_\varepsilon = 0$ for all ε), it follows from Lemma 3 that the optimal bank contract is given by $d_\varepsilon = d^*_\varepsilon > 0$ for all $\varepsilon > \varepsilon^0_R$, with the exclusion threshold, ε^0_R, as the unique solution to

$$\max_{d\geq 0} \left\{ \upsilon(\varepsilon^0_R)u\left(d\right) - (1+\sigma)d \right\} = k. \tag{36}$$

The following proposition evaluates the welfare cost from removing cash as the fixed cost of banking, k, approaches 0.

Proposition 4 ***(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, $\mathrm{M}_n \to 0$, 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, ε_R, 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.

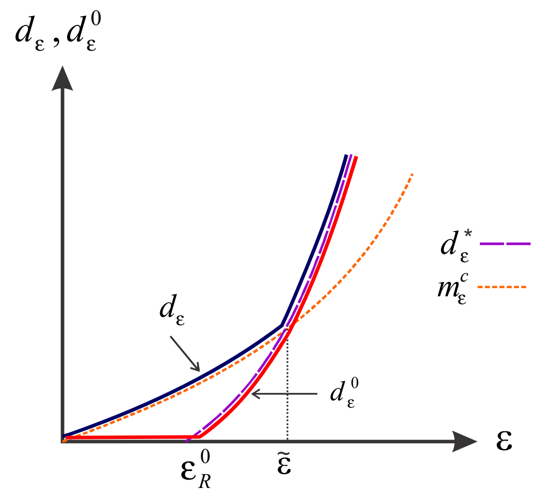

Figure 5: Debt limits at the cashless limit (d_ε) and in the no-cash economy (d_ε^0)

In Figure 5, we represent the debt limits at the cashless 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, either because their access to credit or debt limits deteriorate, or because the bank extracts a larger share of their surplus. Aggregate welfare falls as well.[16] This result holds not only at the limit where $k = 0$, but also for k strictly positive but sufficiently small.

[16] 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.

The key friction underlying the relevance of cash at the cashless limit is private information, which requires banks to leave informational rents to high-liquidity-needs buyers. To limit these rents, banks distort debt limits and exclude some lower types. Cash affects this screening problem because it determines buyers' outside options. Although cash is almost of no use 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' effective market power. These changes are illustrated in Figure 6.

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. 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 in Online Appendix C. These environments can be interpreted as instances of cash irrelevance in the sense of Woodford (1998).

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' effective market power by eliminating the informational friction so 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.

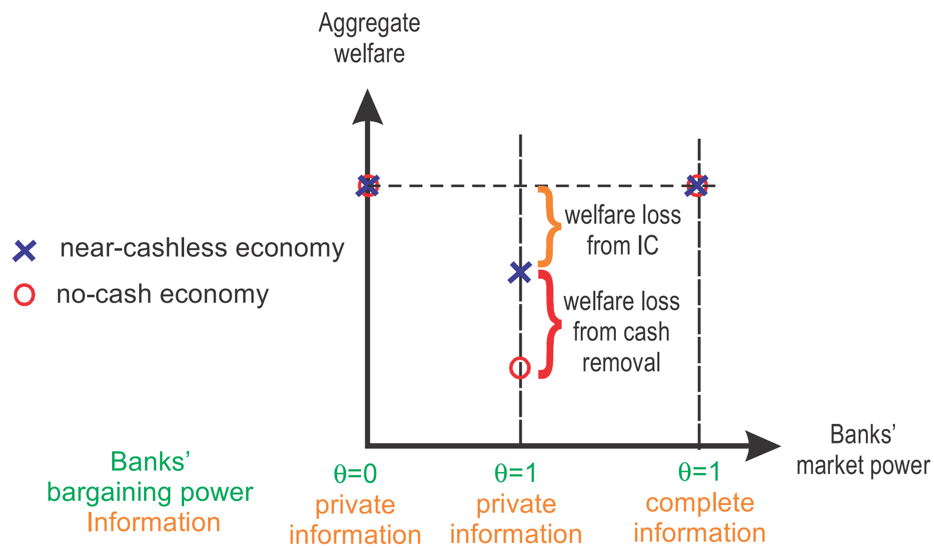

Figure 6: Welfare and market power in near-cashless and no-cash economies

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 raise aggregate welfare by reducing screening distortions and the welfare cost of cash removal, while also shifting surplus from consumers toward banks.

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 the bank's power to extract surplus arises solely from bargaining strength but banks remain imperfectly informed, eliminating cash may be socially harmful despite the gain, $\mathcal{L}(i)$. In this case, the additional distortions to credit supply induced by informational 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).[17] Banks lack the technology to directly enforce loan repayment but have access to buyers' credit histories. Repayment is therefore self-enforced through reputational incentives. 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.

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.[18] Upon formation of the match, in stage 2, the banking fee is paid prior to the extension of credit. Figure 7 illustrates the timing of events within the bank-buyer relationship.

Banks can fully commit to repay their IOUs and incur a fixed cost k whenever they issue IOUs, which may represent the cost of accessing and maintaining buyers' credit records or, more generally, the cost of operating the payment arrangement. The fixed cost is incurred at the time the loan is repaid, in stage 2.

[17] 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).

[18] 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.

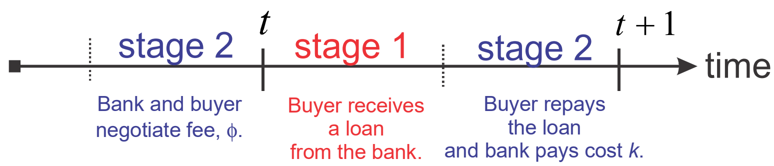

Figure 7: Timing of the bank-buyer lending relationship

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:

$$W^1(\omega,\varepsilon) = \max_{c,h,m'} \left\{ u(c) - h + \beta V^1(m',\varepsilon,\underline{b}_\varepsilon) \right\} \tag{37}$$
$$\text{s.t. } c + (1+\pi)\, m' + \phi_\varepsilon = h + \omega + T^1 \text{ and } h \leq \bar{h}.$$

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 is equal to*

$$-(1+\rho)\phi_\varepsilon + \mathcal{G}_\varepsilon(d_\varepsilon; i) \tag{38}$$

where

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

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

$$\hat{\phi}_\varepsilon = k + \theta \left[\mathcal{G}_\varepsilon(d_\varepsilon; i) - k\right]. \tag{39}$$

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 (39) into (37) and use the expression for $U^m(\varepsilon, i)$ from (25) to reexpress the value function, $W^1(0, \varepsilon)$, as

$$\begin{aligned} \rho W^1(0,\varepsilon) &= (1+\rho)\left[T^1 + u(c_1^*) - c_1^*\right] + U^m(\varepsilon, i) \\ &\quad + (1-\theta)\max\{\mathcal{G}_\varepsilon(d_\varepsilon; i) - k, 0\}. \end{aligned} \tag{40}$$

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

$$\rho \tilde{W}^1(0,\varepsilon) = (1+\rho)\left[T^1 + u(c_1^*) - c_1^*\right] + U^m(\varepsilon, i). \tag{41}$$

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 (40) and (41), d_ε solves

$$\rho d_\varepsilon = (1-\theta)\left[\mathcal{G}_\varepsilon(d_\varepsilon; i) - k\right]. \tag{42}$$

Equation (42) is represented graphically in Figure 8 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 (42). 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, (42) 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

$$\mathcal{E}^b \equiv \{\varepsilon \in [0, \bar{\varepsilon}] : \mathcal{G}_\varepsilon(d_\varepsilon; i) \geq k\}. \tag{43}$$

We show in the Appendix (see Lemma 5) that $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$ for some $\varepsilon_R > 0$.

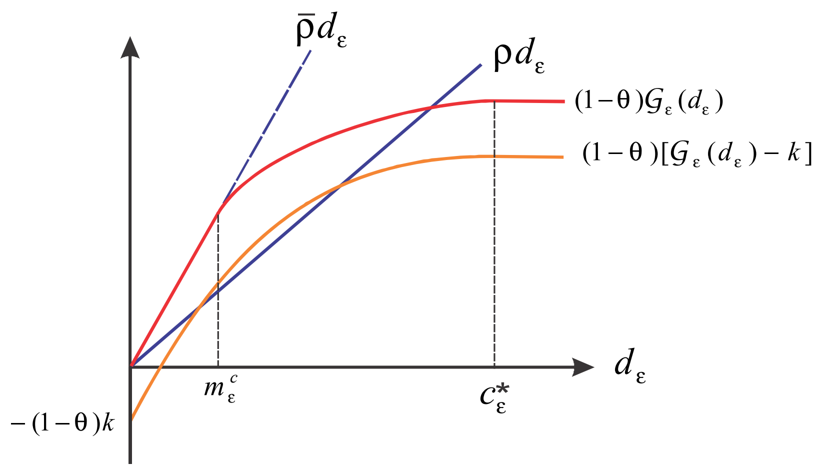

Figure 8: Determination of the debt limit

We define an equilibrium as a list of debt limits, (d_ε), the largest solutions to (42), a set of buyers with access to credit, $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$, solution to (43), and aggregate real money balances,

$$\mathrm{M} = \int_0^{\varepsilon_R} m_\varepsilon^c \mathrm{d}F(\varepsilon). \tag{44}$$

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 parameters according to $\pi = (i - \rho)/(1 + \rho)$.[19]

Proposition 5 ***(Monetary equilibria under limited commitment.)*** *Consider a sequence of economies characterized by $\theta < 1$ and $\{k_n\}_{n=1}^{+\infty}$. Define*

$$\overline{\rho}(i, \theta) = (1 - \theta)i. \tag{45}$$

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 is sufficiently small. Note that the condition $\rho < \overline{\rho}$ implicitly requires $\pi > 0$. The threshold, $\bar{\rho}$, is represented graphically in Figure 8 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, $d_\varepsilon = 0$.

[19] 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.

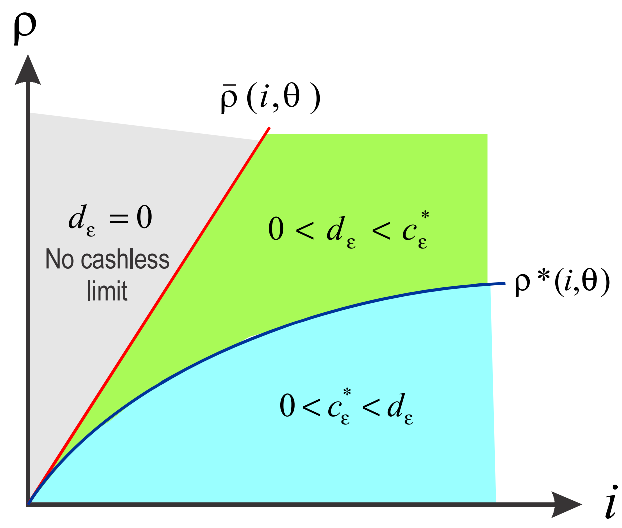

Figure 9: Thresholds for ρ below which debt limits are positive, $d_\varepsilon > 0$, and borrowing constraints are slack, $c_\varepsilon^* < d_\varepsilon$, when $k = 0$.

On the other hand, the threshold $\bar{\rho}$ is increasing in i, as illustrated in Figure 9. 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 6 ***(Relevance of cash at the cashless limit.)*** *Consider a sequence of economies, $\{k_n\}_{n=1}^{+\infty}$, featuring $\rho < \overline{\rho}(i,\theta)$ and $\lim_{n\to+\infty} k_n = 0$.*

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

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

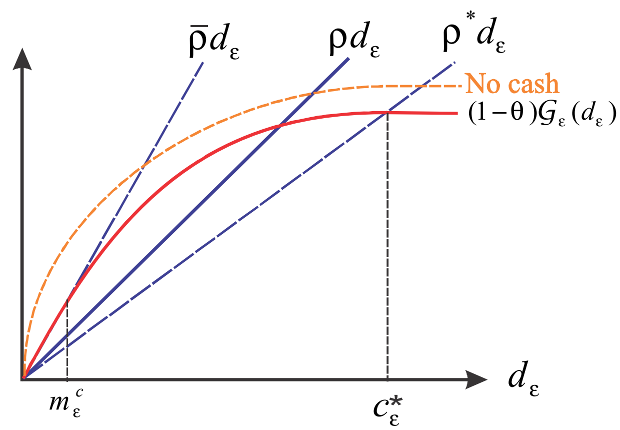

Figure 10: Effect of cash removal at the cashless limit on debt limits

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 10. In the case of CRRA preferences, $u(c) = c^{1-a}/(1-a)$ with $a \in (0,1)$,

$$\rho^* = (1-\theta)\left(\frac{a}{1-a}\right)\left[1 - \frac{1}{(1+i)^{\frac{1-a}{a}}}\right]. \tag{46}$$

We represent ρ^* as a function of i in the right panel of Figure 9. 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 10, then $d_\varepsilon < c_\varepsilon^*$ for some ε. When cash is removed, the stage-1 consumption rises for previously liquidity-constrained buyers, thereby increasing aggregate welfare. Graphically, the dashed orange curve representing the reputational loss from defaulting, $(1-\theta)\mathcal{G}_\varepsilon(d_\varepsilon; i)$, shifts upward.

An important implication of Proposition 6 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, cash is a fallback option when a bank refuses to lend. It therefore disciplines banks and mitigates screening distortions. When $\theta = 0$, this channel is inactive because fees are competitive and debt limits are efficiently determined by the enforcement technology. With self-enforcement, cash is instead a fallback option after default. Removing it raises the cost of exclusion, strengthens repayment incentives, and supports larger debt limits, thereby raising aggregate welfare. This mechanism operates even when banks have no market power.

When banks possess bargaining power, $\theta > 0$, the removal of cash raises aggregate welfare, yet buyers need not be better off. From (40), a buyer's stage-1 utility consists of two components: the status-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 fully commit to repay 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. (2013), 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 Online 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, removing 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. We showed that the welfare consequences of removing cash in economies near the cashless limit depend 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 both aggregate and consumer welfare provided that banks possess no bargaining power. If banks have market power, however, the distributional effects of cash removal remain. These findings indicate that money in 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. For example, as an increasing share of transactions takes place on online platforms requiring digital forms of payment, the government could 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 level of financial literacy in digital payments—if some groups still find it too cumbersome to use, financial exclusion will remain a significant concern, especially for the more vulnerable groups.

Appendix: Proofs of propositions

We prove all propositions here, relegating lemmas to Online Appendix A.

Proof of Proposition 1. The equilibrium has a recursive structure. From (12), we determine $m_\varepsilon(0)$ for all ε. Given $m_\varepsilon(0)$, we can compute $\mathcal{G}_\varepsilon$ from (15) for all ε. With d_ε determined by (17), $d_\varepsilon > m_\varepsilon(0)$ for all ε as $\sigma < i$. Moreover, 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 $\varepsilon_R > 0$ (which may be higher than $\overline{\varepsilon}$ and in that case no buyer is banked). We obtain aggregate real money balances, $\textsc{m}$, from (20) and the value of money, $\upsilon_{t,1} = \upsilon_{t,2}$, from (21). As a result, $\textsc{m} > 0$, and the equilibrium is monetary. This closes the determination of a steady-state monetary equilibrium, which is unique by construction. Finally, if $k < \mathcal{G}_{\overline{\varepsilon}}(d_{\overline{\varepsilon}}) - \sigma d_{\overline{\varepsilon}} = \max_{d \geq 0} [\mathcal{S}(d, \overline{\varepsilon}) - \sigma d] - \max_{m \geq 0} [\mathcal{S}(m, \overline{\varepsilon}) - im]$, then $\varepsilon_R < \overline{\varepsilon}$ and hence the equilibrium features some credit. ■

Proof of Proposition 2. <u>Part 1</u>. Note that d_ε is the same with or without cash according to (17). Suppose that $\theta = 0$. Then, from (18) and (22) we have that $\phi_\varepsilon = \phi_\varepsilon^0$ for all ε. Hence, both at the cashless limit and in the no-cash economy, 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. The stage-2 consumption of each agent is c_1^*. In stage-1, the consumption of a type-ε buyer is equal to d_ε, solution to (17) whether the economy is at the cashless limit or in the no-cash economy. So aggregate welfare is not affected by the presence of cash. Debt limits, consumption allocations, and monitoring costs are identical with and without cash, while fees are transfers between buyers and banks. Hence aggregate welfare is unchanged. However, buyers' welfare is lower, and banks' welfare is higher, in the no-cash economy since $\phi_\varepsilon^0 > \phi_\varepsilon$ for any $\theta > 0$ according to (18) and (22). ■

Proof of Proposition 3. From Lemma 3, the optimal contract solves (29). We solve it by two stages. First we take ε_R as given and solve for d_ε, and then we solve for the optimal ε_R. For a given ε_R, the Lagrangian for the problem is

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

The solution then satisfies the following FOC:

$$\upsilon(\varepsilon)u'(d_\varepsilon) - (1+\sigma) + \lambda(\varepsilon) = 0,\ \ \lambda(\varepsilon)(d_\varepsilon - m_\varepsilon^c) = 0,\ \ \lambda(\varepsilon) \geq 0. \tag{47}$$

These admit a unique solution:

$$\lambda(\varepsilon) = -\upsilon(\varepsilon)u'(m_\varepsilon^c) + (1+\sigma) \text{ and } d_\varepsilon = m_\varepsilon^c \text{ for all } \varepsilon < \tilde{\varepsilon}, \tag{48}$$

$$\lambda(\varepsilon) = 0 \text{ and } u'(d_\varepsilon) = \frac{1+\sigma}{\upsilon(\varepsilon)} \text{ for all } \varepsilon \geq \tilde{\varepsilon}, \tag{49}$$

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

$$u'(m_\varepsilon^c) = \frac{1+i}{\varepsilon} = \frac{1+\sigma}{\upsilon(\varepsilon)}. \tag{50}$$

Note that d_ε that solves (49) is the same as d_ε^* given in (31). 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 $\varepsilon < \tilde{\varepsilon}$. Note that both m_ε^c and d_ε^* are strictly increasing in ε and hence monotonicity is satisfied, and they cross exactly once at $\tilde{\varepsilon}$. Now, if $\varepsilon_R < \tilde{\varepsilon}$, then the solution is exactly the same as (48) and (49); otherwise, we have (49) only. Given the optimal d_ε, define

$$\Pi(\varepsilon_R) = \int_{\varepsilon_R}^{\bar{\varepsilon}} \left\{\upsilon(\varepsilon)u(d_\varepsilon) - (1+\sigma)\, d_\varepsilon - k - U^m(\varepsilon_R, i)\right\} \mathtt{d}F(\varepsilon).$$

Then,

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

For $\varepsilon_R \leq \tilde{\varepsilon}$, $d_{\varepsilon_R} = m^c_{\varepsilon_R}$ and hence

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

which is strictly increasing in ε_R. For $\varepsilon_R > \tilde{\varepsilon}$, the derivative is

$$\begin{aligned} & \frac{\mathtt{d}}{\mathtt{d}\varepsilon_R}\left\{\upsilon(\varepsilon_R)\left[u(d_{\varepsilon_R}) - u(m^c_{\varepsilon_R})\right] - \left[(1+\sigma)d_{\varepsilon_R} - (1+i)m^c_{\varepsilon_R}\right]\right\} \\ = \ & \upsilon'(\varepsilon_R)\left[u(d_{\varepsilon_R}) - u(m^c_{\varepsilon_R})\right] + \left(1 - \frac{\upsilon(\varepsilon_R)}{\varepsilon_R}\right)(1+i)\left(\frac{\mathtt{d}}{\mathtt{d}\varepsilon}m^c_{\varepsilon_R}\right) > 0 \end{aligned}$$

since $\frac{\upsilon(\varepsilon)}{\varepsilon} \leq 1$ for all ε, and hence it is also strictly increasing in ε_R. Note that

$$\left\{\upsilon(\varepsilon_R)\left[u(d^*_{\varepsilon_R}) - u(m^c_{\varepsilon_R})\right] - \left[(1+\sigma)d^*_{\varepsilon_R} - (1+i)m^c_{\varepsilon_R}\right]\right\} = \mathcal{G}_{\overline{\varepsilon}}(d_{\overline{\varepsilon}}) - \sigma d_{\overline{\varepsilon}}$$

at $\varepsilon_R = \overline{\varepsilon}$. Thus, if $k \in [\tilde{k}, \mathcal{G}_{\overline{\varepsilon}}(d_{\overline{\varepsilon}}) - \sigma d_{\overline{\varepsilon}})$, the solution to $\Pi'(\varepsilon_R) = 0$ has a unique solution that 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 proves Proposition **3.** ■

Proof of Proposition 4. Since $k_n \to 0$, we only consider the case where $k < \widetilde{k}$ and hence $\varepsilon_R < \tilde{\varepsilon}$. In this case, we can measure aggregate welfare as follows (note that for aggregate welfare, fees for banks do not appear as they cancel out with the labour needed by buyers):

$$\begin{aligned} \mathcal{W}(i) \ \equiv \ & 3\left[u(c^*_1) - c^*_1\right] + \int_0^{\varepsilon_R(i)} \left\{\varepsilon u[m^c_\varepsilon(i)] - m^c_\varepsilon(i)\right\} \mathtt{d}F(\varepsilon) \\ & + \int_{\varepsilon_R(i)}^{\tilde{\varepsilon}(i)} \left\{\varepsilon u[m^c_\varepsilon(i)] - (1+\sigma)m^c_\varepsilon(i) - k\right\} \mathtt{d}F(\varepsilon) \\ & + \int_{\tilde{\varepsilon}(i)}^{\overline{\varepsilon}} \left\{\varepsilon u\left(d^*_\varepsilon\right) - (1+\sigma)d^*_\varepsilon - k\right\} \mathtt{d}F(\varepsilon). \end{aligned} \tag{51}$$

The first term on the right side of (51) 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.

<u>Part 1.</u> Now, for k small, the optimal contract in Proposition 3 depends on k_n only through $\varepsilon_{R,n}$, which by (35) converges to zero as $k_n \to 0$. Hence, for the economy

indexed by n, the aggregate real balances are given by $\mathrm{M}_n = \int_0^{\varepsilon_{R,n}} m_\varepsilon^c \mathrm{d}F(\varepsilon) \to 0$ as $k_n \to 0$, and the allocations converge. Hence, a cashless limit exists.

<u>Part 2.</u> Now, as $k_n \to 0$, the exclusion threshold $\varepsilon_{R,n}^0$ in the no-cash economy that solves (36) with $k = k_n$ converges to $\varepsilon_{R,\infty}^0 > 0$ that solves $\upsilon(\varepsilon_{R,\infty}^0) = 0$. In contrast, as noted earlier, the exclusion threshold $\varepsilon_{R,n}$ that solves (35) in the monetary equilibrium converges to zero, and hence $\varepsilon_{R,\infty} = 0 < \varepsilon_{R,\infty}^0$. Note also that by (32), $\tilde{\varepsilon}(i) > \varepsilon_{R,\infty}^0$. 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 $k = k_n$,

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

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

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

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

$$u'(d_\varepsilon^*) = \frac{1+\sigma}{\upsilon(\varepsilon)} > \frac{1+i}{\varepsilon} = u'(m_\varepsilon^c),$$

where the inequality follows from $\varepsilon < \tilde{\varepsilon}$, and the facts that the inequality becomes an equality at $\varepsilon = \tilde{\varepsilon}$ by (32) and that $\upsilon(\varepsilon)/\varepsilon$ is strictly increasing. Moreover, if we let $\widehat{d}_\varepsilon$ solve $\varepsilon u'(\widehat{d}_\varepsilon) = 1+\sigma$, then $d_\varepsilon^* \leq m_\varepsilon^c < \widehat{d}_\varepsilon$ for all $\varepsilon < \tilde{\varepsilon}$. Since $[\varepsilon u\left(d\right) - (1+\sigma)d]$ 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

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

where $d(\varepsilon)$ denotes d_ε, the debt limit under the optimal contracts (note that $d_\varepsilon = 0$ for all $\varepsilon \leq \varepsilon_R^0$). 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 Proposition 5. We start with the following lemma.

Lemma 5 *Suppose that $\theta < 1$. For each $\varepsilon \in (0, \bar{\varepsilon})$, (42) admits a largest solution with $d_\varepsilon > m_\varepsilon^c > 0$ for all k sufficiently small if and only if $\rho < \overline{\rho} = (1-\theta)i$. This solution is increasing in ε. The set of banked buyers is $\mathcal{E}^b = [\varepsilon_R, \bar{\varepsilon}]$ for some $\varepsilon_R > 0$.*

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 $\widehat{d_{\bar{\varepsilon}}} > m_{\bar{\varepsilon}}^c$. 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 (42) under $\bar{\varepsilon}$ that is larger than $m_{\bar{\varepsilon}}^c$, 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 $m_{\bar{\varepsilon}}^c$. If $k_n < \overline{k}$, then the largest solution to (42) under $\bar{\varepsilon}$ satisfies $d_{\bar{\varepsilon}} > m_{\bar{\varepsilon}}^c$. Hence, the buyer of type $\bar{\varepsilon}$ enters the credit market if and only if $k_n < \overline{k}$. Moreover, if $k_n < \overline{k}$, there is a threshold $\varepsilon_{R,n} < \bar{\varepsilon}$ so that buyers of types below $\varepsilon_{R,n}$ 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 $\varepsilon > 0$,

$$\lim_{k_n \to 0} \mathcal{G}_\varepsilon(d_\varepsilon) > 0. \tag{52}$$

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 $\rho < \overline{\rho}$, $d_\varepsilon > m_\varepsilon^c$. This proves (52) and hence $\varepsilon_{R,n} \to 0$ as $k_n \to 0$. So, $\mathrm{M}_n \to 0$. ■

Proof of Proposition 6. Part 1. Note that $d_\varepsilon \geq c_\varepsilon^*$ at the limit $k = 0$ if

$$\rho \leq \rho_\varepsilon^*(i) \equiv \frac{(1-\theta)[\mathcal{S}(c_\varepsilon^*, \varepsilon) - U^m(\varepsilon, i)]}{c_\varepsilon^*}. \tag{53}$$

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

$$\rho^*(i) = \inf\left\{\rho_\varepsilon^*(i) : \varepsilon \in (0, \bar{\varepsilon}]\right\}. \tag{54}$$

Note that ρ_ε^* is continuous in ε. We now show that $\rho^*(i)$ defined by (54) is strictly smaller than $\bar{\rho}$. Now, fix some $\varepsilon > 0$. At $\rho = \bar{\rho}$, as illustrated by Figure 10, any d satisfying (42) has $d \leq m_\varepsilon^c$. 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 $\rho \in (\rho^*(i), \overline{\rho})$, 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

$$\mathcal{W}(i) \equiv 3\left[u(c_1^*) - c_1^*\right] + \int_0^{\bar{\varepsilon}} \mathcal{S}[d_\varepsilon(i), \varepsilon]\mathtt{d}F(\varepsilon),$$

where $d_\varepsilon(i)$ denotes the debt limits under the cashless limit, which solves (42). Note that the debt limits under the no-cash economy are 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 strictly increases the aggregate welfare.

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

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

$\rho_\varepsilon^*(i)$ 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.

Part 2. Let $\rho_0 = \rho^*(i)$. From (40), at the cashless limit $k = 0$, a type-ε buyer's life time utility is

$$\rho W^1(0, \varepsilon) = (1+\rho)\left[T^1 + u(c_1^*) - c_1^*\right] + U^m(\varepsilon, i) + (1-\theta)\mathcal{G}_\varepsilon(d_\varepsilon; i). \tag{55}$$

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

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

where $d_\varepsilon(i)$ is the debt limit under the cashless limit and $d_\varepsilon(\infty)$ is that under the no-cash economy. At $\rho = \rho_0$, the first term is zero but the second term is positive, since, as illustrated in Figure 10, $c^*_\varepsilon \leq d_\varepsilon(i) < d_\varepsilon(+\infty)$ and hence $\mathcal{S}(d_\varepsilon(i), \varepsilon) = \mathcal{S}(d_\varepsilon(+\infty), \varepsilon)$ for all ε. So removing cash decreases all buyers' utility. By continuity, for an interval of ρ's above ρ_0, removing cash decreases (at least some) buyers' surpluses. But it increases aggregate welfare as in Part 1. ■

References

[1] Ahnert, Toni, Peter Hoffmann, and Cyril Monnet (2025). Payments and privacy in the digital economy. Journal of Financial Economics, 169, 104050.

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

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

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

[5] Bayeh, Berhan, Isaiah Nardone, Shaun O'Brien, and Hailey Phelps (2025). "Findings from the diary of consumer payment choice." Federal Reserve Financial Services.

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

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

[12] Dong, Mei, Stella HuangFu, Timothy Kam, and Sam Ng (2026). “Banking concentration and loan-rate markups in a frictional loan market”, European Economic Review 188, 105393.

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

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

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

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

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

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

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

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

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

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

[23] Lagos, Ricardo (2025). “Onetary economics at 30: a reexamination of money in cashless limiting monetary economies,” NBER Working Paper 34155.

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

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

[26] Li, Yiting (2011). “Currency and checking deposits as means of payment,” Re-

view of Economic Dynamics 14, 403-417.

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

[28] Menzio, Guido and Saverio Spinella (2026). “More on onetary Economics.” New York University Working Paper.

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

[30] 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.

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

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

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

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

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

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

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

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