The Limits of *onetary Economics*: On Money as a Constraint on Market Power

Ricardo Lagos*
New York University

Shengxing Zhang†
London School of Economics

October 2021

Abstract

We formulate a generalization of the traditional medium-of-exchange function of money in contexts where there is imperfect competition in the intermediation of credit, settlement, or payment services used to conduct transactions. We find that the option to settle transactions with money strengthens the stance of sellers of goods and services vis-à-vis intermediaries, and show this mechanism is operative even for sellers who never exercise the option to sell for money. These *latent* money demand considerations imply that in general, in contrast to current conventional wisdom in policy-oriented research in monetary economics, monetary policy can remain effective through medium-of-exchange transmission channels—even in highly developed credit economies where the share of monetary transactions is negligible.

Keywords: Cashless, credit, liquidity, money, monetary policy.
JEL classification: D83, E52, G12.

*Lagos acknowledges support from the C.V. Starr Center for Applied Economics at NYU.
†Zhang acknowledges support from the Centre for Macroeconomics at LSE.

1 Introduction

A large body of work in macroeconomics rests on the premise that artificial economies without money are well suited to study monetary policy. In fact, most of the work in modern monetary economics that caters to policymakers abstracts from the usefulness of money altogether: there is typically no money in the models, or if there is money, it is merely held as a redundant asset. This moneyless approach to monetary economics relies on the received wisdom that medium-of-exchange and money-demand considerations are irrelevant for the transmission of monetary policy in the context of advanced economies whose credit-based settlement mechanisms have developed sufficiently to make the velocity of some monetary aggregates very high.[1]

The intuitive argument runs as follows: aggregate real money balances are a small fraction of aggregate real output in modern economies (e.g., velocity of the monetary base tends to be relatively high), so policy-induced changes in real money balances are bound to have small effects on output. Therefore, the argument goes, there is no significant loss in basing monetary policy advise on models where real money balances do not interact with the real allocation—or are simply assumed to be equal to zero. This intuition has been formalized in the context of economies where the role of money in exchange is not modeled explicitly, but rather, is proxied by either assuming money is an argument of a utility function, or by imposing that certain purchases be paid for with cash acquired in advance.

The view that medium-of-exchange considerations are inconsequential rests on two results. First, the fact that the monetary equilibrium in some reduced-form models is continuous under a certain "cashless limit" (e.g., obtained by taking to zero either the marginal utility of real balances in a model where money enters the utility function, or the fraction of "cash goods" in a cash-credit goods version of a cash-in-advance model where money becomes a redundant asset) has been used to conclude that a monetary economy with velocity that is as high as in the data can be well approximated by an economy without money or medium-of-exchange considerations. Second, parametrized versions of these reduced-form models have been used to claim that, for realistic values of velocity, money and medium-of-exchange considerations are quantitatively insignificant.

In this paper we show that the two results used to justify the moneyless approach can be

[1]The moneyless approach was pioneered by Woodford (1998). It is advocated by the textbook treatments of sticky-price models (e.g., Woodford (2003) and Galí (2008)), as well as by textbook treatments of the "Fiscal Theory of the Price Level" (e.g., Cochrane (2021)).

overturned when we replace the reduced-form formulations with a more explicit and general microstructure of payments. Specifically, we find that in the cashless limit, i.e., as credit-based settlement mechanisms evolve to make the velocity of money very high, the magnitude of the effect of monetary policy on consumption and welfare is given by a single sufficient statistic: the product of the *deposit spread* that bankers with market power impose on lenders, and the *price elasticity of demand* for the set of goods purchased with cash or credit. According to the theory, this statistic encodes all the empirical information needed to quantify the welfare cost of the purely monetary distortions associated with the opportunity cost of holding money in near-cashless economies. This result indicates that the folk wisdom that monetary policy analysis can ignore medium-of-exchange and money-demand considerations without loss is only correct for a much more limited set of monetary environments than previously recognized.

Our theory formalizes a generalization of the traditional medium-of-exchange foundation for money demand in contexts where credit, settlement, or payment services involve financial intermediaries with some degree of market power (e.g., banks, broker-dealers, credit card companies). The threat to settle transactions directly for money—a *latent money demand*—strengthens the stance of sellers of goods or services vis-à-vis intermediaries. The role of money as a discipline device for imperfectly competitive financial intermediaries opens a novel conduit for monetary transmission that operates through the effects of changes in the opportunity cost of holding money on money demand (actual and latent), and ultimately on relative prices and allocations. This latency, or off-equilibrium role of money, is distinct from the traditional medium-of-exchange function that money performs when it is actively exchanged to overcome other trading frictions, such as double-coincidence-of-wants problems, lack of commitment, and lack of enforcement. Unlike the conventional medium-of-exchange role that emphasizes buyer-side incentives to carry money, the role of monetary exchange as safeguard against intermediary market power remains relevant even in cashless limiting economies where highly developed credit and settlement arrangements make transaction velocity of money arbitrarily high. When money serves as a *latent* medium of exchange, changes in the value of money influence the terms of trade of *all sellers*—even those who never use money to trade. Thus, along the cashless limit, even though the aggregate volume of monetary trade vanishes, changes in the incentives to demand money have nonnegligible macroeconomic impact because the individual off-equilibrium option to engage in monetary trade still operates as a discipline device against intermediary market power. As a result, it is not possible to conclude that money cannot matter quantitatively solely on the basis that it accounts for a small share of transactions.[2]

The rest of the paper is organized as follows. Section 2 streamlines the main idea of the paper. Section 3 describes the economic environment, presents the solution to the social planner's problem, formulates the individual optimization and bilateral bargaining problems, and defines equilibrium. Section 4 characterizes the equilibrium of the nonmonetary economy. Section 5 characterizes the stationary monetary equilibrium. Section 6 derives prices and allocations in the cashless pure-credit limit, and discusses welfare and the issue of price-level determination in limiting cashless economies. Section 7 draws connections with related work (Section 7.1), discusses the limitations of the moneyless approach to monetary economics (Section 7.2), and explains the shortcomings of reduced-form models of money demand (Section 7.3). Appendices A through E contain all proofs, as well as characterizations of dynamic and sunspot monetary equilibria and their respective cashless pure-credit limits.

2 On money as a constraint on market power—a preview

A measure 1 of consumers wish to buy consumption (*good 1*) from a measure 1 of producers in a competitive market. Consumers get utility $u(c)$ from c units of good 1, and producers have a constant marginal cost of production, κ. Consumers can settle their purchases in two ways: on the spot with money, or with a credit arrangement whereby they deliver a numéraire (*good 2*) at a later time. The marginal utility from consuming, and the marginal cost of producing good 2 are equal to 1. After having produced good 1, a fraction $1-\alpha \in [0,1]$ of producers receive a shock that prevents them from accessing the credit-based payment arrangement; they can only settle their sales for money. The remaining α fraction of producers can settle their sales using the credit arrangement and/or money. For simplicity, assume any unsold inventory of good 1 depreciates fully after the trading round. Credit is intermediated by banks who act

[2] The elementary notion that credible outside options can drive outcomes even if they are not exercised in equilibrium is ubiquitous in economics. In macroeconomics, it is a key equilibrium driving force in models with private information. In international economics and industrial organization, the notion that the option to engage in trade—even if no trade actually occurs—can be a key determinant of equilibrium outcomes and welfare goes back many years. For example, this is the key idea behind the breakdown of the equivalence of tariffs and quotas under imperfect competition in Bhagwati (1965), and the key idea underlying the notion of *contestable markets* in Baumol (1982). Another well known example in trade theory is Markusen (1981), who considers two identical countries with a monopolist producer in each. Under autarky, the equilibrium has a monopoly markup in each country, but if trade between the countries is possible, competition turns the two monopolists into Cournot duopolists, which reduces markups and increases welfare in both countries even though no trade actually occurs since the countries are identical.

competitively on the lending side (with consumers), but have market power on the deposit side (with producers). Let $1-\theta$ denote the share of a depositor's gain from lending that is appropriated by a bank. The parameter $\theta \in [0,1]$ indexes market power in the deposit market: $\theta = 0$ is the case in which banks have monopoly power, while $\theta = 1$ is the perfectly competitive case in which banks cannot extract rents from producers.

Begin by considering a *nonmonetary* economy that abstracts from the existence of money. Let φ^n denote the effective (*cum interest*) relative price of good 1 in terms of good 2 that consumers face when they purchase good 1 with credit in the nonmonetary economy. The optimization problems of producers and consumers in the market for good 1 imply

$$\kappa = \alpha\theta\varphi^n \leq \varphi^n = u'(c), \tag{1}$$

with "<" unless $\alpha\theta = 1$. The first equality in (1) is the optimality condition for producers, which equates their marginal cost, κ, to their expected marginal revenue, which equals a share θ of the effective price φ^n if the producer has access to a bank to settle the transaction (with probability α), and 0 otherwise. The second equality in (1) is the optimality condition for consumers, which equates their marginal utility to the *effective* relative price of good 1, φ^n. If we let y^n denote the level of consumption implemented by the nonmonetary equilibrium, then (1) implies $u'(y^n) = \vartheta^n\kappa$, where $\vartheta^n \equiv \frac{1}{\alpha\theta} \geq 1$, with ">" unless $\alpha\theta = 1$. From the equality on the left side of (1), we know the equilibrium relative price of good 1 in the nonmonetary economy is $\varphi^n = \frac{\kappa}{\alpha\theta}$.

Next, consider a monetary economy. Let M_t denote the money supply, and p_{it} be the nominal price of good $i \in \{1,2\}$ at time t. An equilibrium is *monetary* if $1/p_{it} > 0$ for $i \in \{1,2\}$ and all t. Let φ^m denote the relative price of good 1 in terms of good 2 that consumers face if they choose to purchase good 1 with money, i.e., $\varphi^m \equiv p_{1t}/p_{2t}$. Let ρ denote the real interest rate in the competitive market where consumers borrow from banks. Finally, let $\iota > 0$ denote a nominal interest rate (determined by a monetary authority) that represents the opportunity cost to the consumer of holding money between consecutive time periods.[3]

In any monetary equilibrium, consumers use credit and all the money they bring into the period to finance their purchases of good 1.[4] Since consumers use both means of payment,

[3]Think of ι as the interest rate on a one-period government-issued risk-free nominal bond that is illiquid in the sense that it cannot be used to settle purchases of good 1. The case with $\iota = 0$ can be interpreted as an implementation of the Friedman rule (simpler to analyze by first assuming $\iota > 0$, and then letting $\iota \to 0$).

[4]We know that consumers spend all their money on good 1 because the opportunity cost of holding money

they must be indifferent between them, which together with the fact that in any monetary equilibrium they carry money between periods, implies $\rho = \iota$.[5] The optimization problems of producers and consumers in the market for good 1 imply

$$\kappa = (1 + \alpha\theta\iota)\,\varphi^m \leq (1 + \iota)\,\varphi^m = u'(c)\,, \tag{2}$$

with "$<$" unless $\alpha\theta = 1$ (or $\iota = 0$). The first equality in (2) is the optimality condition for producers that equates their marginal cost, κ, to their expected marginal revenue, which consists of the producer's outside option from selling for money (φ^m), plus their share of the profit from lending the marginal revenue to consumers ($\theta\iota\varphi^m$), weighted by the probability of having access to the bank (α). The second equality in (2) is the optimality condition for consumers, which equates their marginal utility to the *effective* relative price of good 1, which equals $(1 + \iota)\,\varphi^m$ (i.e., the *cum interest* relative price of good 1 in terms of good 2). If we let y^m denote the level of consumption implemented by the monetary equilibrium, then (2) implies $u'(y^m) = \vartheta^m\kappa$, where $\vartheta^m \equiv \frac{1+\iota}{1+\alpha\theta\iota} \geq 1$, with "$>$" unless $\alpha\theta = 1$ (or $\iota = 0$). The first equality in (2) implies the equilibrium relative price of good 1 is $\varphi^m = \frac{\kappa}{1+\alpha\theta\iota}$.

On the nominal side of the economy, recall that consumers spend all their money to purchase good 1. Also, since $\iota > 0$, money is dominated in rate of return by bank deposits, so only the $1-\alpha$ producers with no access to banks choose to hold their sales revenue in the form of money (the remaining α producers hold them as interest-earning bank deposits). Thus, aggregate money demand equals $(1-\alpha)\,p_{1t}y^m$, so the money-market clearing condition implies the price level is $p_{1t} = \frac{M_t}{(1-\alpha)y^m}$, the real quantity of money is $Z \equiv \frac{M_t}{p_{1t}} = (1-\alpha)\,y^m$, and the transactions velocity of money is $\mathcal{V} \equiv \frac{p_{1t}y^m}{M_t} = \frac{1}{1-\alpha}$.

In a monetary economy with $\alpha < 1$, the $1-\alpha$ producers who wish to sell their inventory of good 1 but lack access to the credit-based payment arrangement have no alternative but to accept money. The limiting economy that results as $\alpha \to 1$ can be thought of as a *cashless economy* in which the credit, payment, or settlement arrangements are developed to a degree where all transactions can be settled without using money as a medium of exchange. Notice

between time periods is positive ($\iota > 0$), so they would not carry money that they do not plan to spend. We know that consumers use some credit alongside money to finance purchases of good 1 because if they did not borrow, the market for loans would have to clear at zero trade, which would imply $\rho = 0$. But $\rho = 0$ is inconsistent with an equilibrium in which no consumers borrow and all purchases of good 1 are financed with money, since $\rho = 0 < \iota$, which means the financing cost of borrowing to buy good 1 (i.e., ρ) is lower than the financing cost of paying for good 1 with money (i.e., ι).

[5]Intuitively, the interest-rate cost of financing good 1 with credit must equal the opportunity cost (in terms of foregone interest on an illiquid intertemporal nominal bond) of carrying money between time periods.

that

$$\lim_{\alpha\to 1}\frac{1}{p_{1t}}=\lim_{\alpha\to 1}\frac{1}{\mathcal{V}}=\lim_{\alpha\to 1}Z=0\leq\lim_{\alpha\to 1}\left(\vartheta^{n}-\vartheta^{m}\right)=\frac{1-\theta}{\theta\left(1+\theta\iota\right)}. \tag{3}$$

The first three equalities in (3) say that the price level and velocity are diverging, while aggregate real money balances are vanishing; i.e., the limit of the monetary equilibrium as $\alpha\to 1$ is indeed *cashless.* The inequality is equivalent to $\lim_{\alpha\to 1}\left(y^{m}-y^{n}\right)\geq 0$, and is strict provided $\theta<1$. This finding may seem puzzling because the only difference between credit, payment, or settlement possibilities in the economy without money relative to the cashless limit of the monetary economy is that the latter has a sliver of monetary transactions with real value converging to 0. So how is it possible that the difference $y^{m}-y^{n}$ remains strictly positive as $\alpha\to 1$, or equivalently, as the share of payments in which money is used as a medium of exchange becomes negligible?

The answer is that if $\theta<1$, the producers' *option* to engage in monetary trade—a sort of *latent money demand*—serves as a constraint on the market power of credit, payment, or settlement intermediaries. To see this, write the marginal expected revenue in the nonmonetary and monetary economies as $R^{n}\equiv\alpha\theta\varphi^{n}$ and $R^{m}\equiv\left(1+\alpha\theta\iota\right)\varphi^{m}$, respectively. In both economies the producers' expected marginal revenue is decreasing in the market power of the intermediaries, but R^{m} contains a term associated with an outside option (the term that survives when $\theta=0$) while R^{n} does not. To illustrate further, consider the monopoly limit, i.e., $\lim_{\theta\to 0}R^{n}=0<\lim_{\theta\to 0}R^{m}=\varphi^{m}$. In the nonmonetary economy nothing prevents monopoly intermediaries from extracting the whole revenue generated by producers. In contrast, in the monetary economy the maximum revenue that monopoly intermediaries can extract from producers is constrained by the revenue that producers can guarantee for themselves by exercising their option of selling for money—an option they never exercise in equilibrium, but that is not available in an economy without money. In the rest of the paper we formulate a fully specified monetary model to formalize the role of money as a constraint on market power in payments, and use the theory to study the implications for monetary economics.

3 Model

3.1 Environment

Time is represented by a sequence of periods indexed by $t\in\mathbb{T}\equiv\{0,1,...\}$, each divided into two subperiods. There are three types of infinitely lived agents: *bankers, consumers*, and *producers*, denoted B, C, and P, respectively. An agent of type $i \in \{B, C, P\}$ is represented with a point in the set $\mathcal{I}_i = [0, 1]$. In the first subperiod, each producer can supply labor that can be used as input in a linear technology to produce good 1, which is only consumed by consumers. Production of good 1 takes place at the beginning of the first subperiod, before agents engage in any trading activity. In the second subperiod, all agent types can supply labor that can be used as input in a linear production technology to produce good 2, which is consumed by all agents. Good 1 and good 2 cannot be stored across periods, but there is within-period storage: producers can transform every unit of unsold inventory of beginning-of-period good 1 into $\underline{\kappa} \in \mathbb{R}_+$ units of end-of-period good 2.

A monetary authority issues *money*—a financial security that is durable and intrinsically useless (i.e., it is not an argument of any utility or production function, and it is not a formal claim to goods or services). The quantity of money outstanding at the beginning of period t is denoted M_t, with $M_0 \in \mathbb{R}_{++}$ given and distributed uniformly among consumers. In the second subperiod of every period, the monetary authority injects or withdraws money via lump-sum transfers or taxes to consumers, so that the money supply evolves according to $M_{t+1} = \mu M_t$, with $\mu \in \mathbb{R}_{++}$.

In order to have a meaningful role for money as a medium of exchange, we assume that consumers are unable to commit, and producers cannot enforce consumers' promises (neither individually nor via collective punishment schemes). In order to have a role for financial intermediation, we assume bankers are endowed with the ability to enforce and commit. In particular, a banker can enforce a future payment promised by a consumer, and can commit to make a future payment to a seller. This special ability to trust consumers and be trusted by producers makes bankers well suited to act as financial intermediaries between consumers and producers. Specifically, consumers will issue bonds through bankers in the first subperiod of t, with each bond representing a claim to one unit of good 2 to be delivered to bond holders through bankers in the second subperiod of t.[6]

In the second subperiod, all agents can trade good 2 and money in a spot Walrasian market. In the first subperiod, trade is organized as follows. Two spot Walrasian markets operate

[6] Absent bankers, there would be complete lack of enforcement: consumers would be unable to borrow, and would have no alternative but to fund first-subperiod consumption of good 1 with money. The equations in the following sections also admit an equivalent interpretation. Instead of assuming that bankers have the special power to enforce and commit, one could assume consumers can themselves commit to repay, but that bond trade must be intermediated by bankers for reasons other than limited enforcement of contracts and limited commitment to honor them.

contemporaneously: a *goods market* where good 1 and money are traded, and a *bond market* where bonds and money are traded. All bankers have access to the bond market where they can trade bonds and money competitively. All consumers have access to the goods market where they can trade good 1 and money competitively, and to the bond market through a randomly chosen banker to whom they make a take-it-or-leave-it offer involving bonds and money. All producers have access to the goods market where they can trade good 1 and money competitively, but only a subset of producers have access to the bond market where they can trade bonds and money by engaging in bilateral trade with bankers whom they contact at random. Specifically, let $\alpha \in [0,1]$ denote the probability that a producer contacts a banker in any given period. Once the producer and the banker have made contact, the pair negotiates the quantities of bonds and money that the banker will buy or sell in the competitive bond market on behalf of the producer, and an intermediation fee for the banker's service. The banker's fee is expressed in terms good 2 and paid in the second subperiod. The terms of this bilateral trade are determined by Nash bargaining, where the producer has bargaining power $\theta \in [0,1]$.

The individual preferences of an agent of type $i \in \{B, C, P\}$ are represented by

$$\mathbb{E}_0^i \sum_{t=0}^{\infty} \beta^t \left[u\left(y_{it}\right) \mathbb{I}_{\{i=C\}} - \kappa y_{it} \mathbb{I}_{\{i=P\}} + v(x_t) - h_t \right],$$

where the expectation operator, $\mathbb{E}_0^i$, is with respect to the probability measure induced by the random trading process in the first subperiod, $\beta \in (0,1)$ is the discount factor, $u : \mathbb{R}_+ \to \mathbb{R}$ is the consumer's utility function for good 1, $\mathbb{I}_{\{\cdot\}}$ is an indicator function that equals 1 if the condition in the subscript is satisfied, and 0 otherwise, $\kappa \in \mathbb{R}_{++}$ is the producer's marginal (disutility) cost of producing good 1, y_{it} is the agent's consumption (if $i = C$) or production (if $i = P$) of good 1 in period t, x_t is the agent's consumption of good 2 in period t, and h_t is the agent's disutility of supplying labor h_t in the second subperiod of period t. We assume $u'' < u(0) = 0 < u'$, $v'' \leq v(0) = 0 < v'$, $\underline{\kappa} < \kappa$, and that there exist $x^\star, y^\star \in \mathbb{R}_{++}$ such that $v'\left(x^\star\right) = 1$ and $u'\left(y^\star\right) = \kappa$. For any $\varphi \in \mathbb{R}_+$, let $\mathrm{D}(\varphi) \equiv u'^{-1}\left(\varphi\right)$.

Let $\left\{y_{Ct}^\star, y_{Pt}^\star, (x_{it}^\star, h_{it}^\star)_{i \in \{B,C,P\}}\right\}_{t=0}^{\infty}$ denote efficient allocation that solves the problem of a social planner who maximizes the equally weighted sum of all agents' expected discounted utilities, where $y_{Ct}^\star$ is the individual consumption of good 1 in period t, $y_{Pt}^\star$ is the individual production of good 1, $x_{it}^\star$ is the individual consumption of good 2 of an agent of type i, and $h_{it}^\star$ is the individual production of good 2 of an agent of type i.

Proposition 1 *The efficient allocation is* $y^\star_{Ct} = y^\star_{Pt} = y^\star$ *and* $x^\star_{it} = h^\star_{it} = x^\star$ *for all* $i \in \{B, C, P\}$ *and all* t.

3.2 Individual optimization, bargaining, and definition of equilibrium

We begin by describing the individual optimization problems in the second subperiod of a typical period. Let $W^i_t(a^m_t, a^g_t)$ denote the maximum expected discounted payoff, at the beginning of the second subperiod of period t, of an agent of type $i \in \{B, C, P\}$ who enters the subperiod with $a^m_t \in \mathbb{R}_+$ units of money and a claim to $a^g_t \in \mathbb{R}$ units of good 2. Let $V^i_t(a^m_t)$ denote the maximum expected discounted payoff of an agent of type $i \in \{B, C, P\}$ with money holding a^m_t at the beginning of the first subperiod of period t. Then

$$W^i_t(a^m_t, a^g_t) = \max_{(x_t, h_t, a^m_{t+1}) \in \mathbb{R}^3_+} \left[v(x_t) - h_t + \beta V^i_{t+1}(a^m_{t+1}) \right], \tag{4}$$

$$\text{s.t. } p_{2t} x_t + a^m_{t+1} \leq p_{2t}(h_t + a^g_t) + a^m_t + T^m_t \mathbb{I}_{\{i=C\}},$$

where p_{2t} is the nominal price of good 2, and $T^m_t \in \mathbb{R}$ is the time t lump-sum monetary injection to an individual consumer. Next, consider the three individual optimization problems that each agent type faces in the first subperiod of a typical period t.

First, consider the portfolio problem of a banker at the end of the first subperiod of period t, i.e., after the round of bond-market trades with consumers and producers. Formally, let $\hat{W}^B_t(a^m_t, a^g_t)$ denote the maximum expected discounted payoff of a banker who has money holding a^m_t and a claim to a^g_t units of good 2, as he reallocates his portfolio of money and bonds in the bond market at the end of the first subperiod of period t (i.e., possibly after having executed a trade on behalf of a client).[7] Then

$$\hat{W}^B_t(a^m_t, a^g_t) = \max_{\bar{\boldsymbol{a}}_t \in \mathbb{R}_+ \times \mathbb{R}} W^B_t(\bar{a}^m_t, \bar{a}^g_t) \tag{5}$$

$$\text{s.t. } \bar{a}^m_t + q_t \bar{a}^b_t \leq a^m_t,$$

where $\bar{\boldsymbol{a}}_t = (\bar{a}^m_t, \bar{a}^b_t)$, $\bar{a}^g_t = a^g_t + \bar{a}^b_t$, and q_t is the nominal price of a bond in the bond market of time t. Let $\bar{\boldsymbol{a}}_{Bt}(a^m_t) = (\bar{a}^m_{Bt}(a^m_t), \bar{a}^b_{Bt}(a^m_t))$ denote the solution to the maximization in (5).

[7] In principle, the banker may be holding a nonzero bond position when reallocating his own portfolio at the end of the first subperiod. However, as will become clear when we formulate the relevant bargaining problem, it is without loss of generality to assume that the banker's portfolio after having provided intermediation services is the same as the banker's beginning-of-period portfolio, which has zero bonds.

Second, in the first subperiod of period t, a consumer with beginning-of-period money holding a_t^m solves

$$\max_{(\bar{y}_t,\bar{\boldsymbol{a}}_t)\in\mathbb{R}_+^2\times\mathbb{R}} \left[u(\bar{y}_t)+W_t^C(\bar{\boldsymbol{a}}_t)\right] \quad (6)$$
$$\text{s.t. } \bar{a}_t^m+p_{1t}\bar{y}_t+q_t\bar{a}_t^b\leq a_t^m,$$

where $\bar{\boldsymbol{a}}_t=(\bar{a}_t^m,\bar{a}_t^b)$, and p_{1t} is the nominal price of good 1. Let $(\bar{y}_{Ct}(a_t^m),\bar{\boldsymbol{a}}_{Ct}(a_t^m))$, with $\bar{\boldsymbol{a}}_{Ct}(a_t^m)=(\bar{a}_{Ct}^m(a_t^m),\bar{a}_{Ct}^b(a_t^m))$, denote the solution to the maximization in (6).

Third, consider a producer who entered period t with money holding a_t^m, produced inventory y_t of good 1 at the beginning of the period, and then does not contact a banker. This producer's individual decision problem in the first subperiod of t is

$$\max_{(\tilde{y}_t,\tilde{a}_t^m)\in\mathbb{R}_+^2} W_t^P(\tilde{a}_t^m,\tilde{a}_t^g) \quad (7)$$
$$\begin{aligned}\text{s.t. } \tilde{a}_t^m &\leq a_t^m+p_{1t}\tilde{y}_t\\ \tilde{y}_t &\leq y_t\\ \tilde{a}_t^g &= (y_t-\tilde{y}_t)\underline{\kappa}.\end{aligned}$$

The first constraint is the budget constraint. The second constraint indicates the producer can at most sell the inventory of good 1 produced at the beginning of the period. The last constraint reflects the producer's ability to transform each unit of unsold inventory of good 1 from the first subperiod into $\underline{\kappa}$ units of good 2 in the second subperiod. Let $(\tilde{y}_{Pt}(y_t,a_t^m),\tilde{a}_{Pt}^m(y_t,a_t^m))$ denote the quantity of the inventory of good 1 sold in the goods market, and the post-trade money holding that solve (7).

Fourth, consider a producer who entered period t with money holding a_t^m, produced inventory y_t of good 1 at the beginning of the period, and then contacts a banker. This producer simultaneously chooses the quantity of good 1 to sell in the goods market, $\bar{y}_{Pt}(y_t,a_t^m)$, and bargains over the post-trade portfolio of bonds and money, $\bar{\boldsymbol{a}}_{Pt}(y_t,a_t^m)=(\bar{a}_{Pt}^m(y_t,a_t^m),\bar{a}_{Pt}^b(y_t,a_t^m))$, as well as the banker's fee, $k_{Pt}(y_t,a_t^m)$. The outcome, $(\bar{y}_{Pt}(y_t,a_t^m),\bar{\boldsymbol{a}}_{Pt}(y_t,a_t^m),k_{Pt}(y_t,a_t^m))$, is the solution to

$$\max_{(\bar{y}_t,\bar{\boldsymbol{a}}_t,k_t)\in\mathbb{R}_+^2\times\mathbb{R}\times\mathbb{R}_+} \left[W_t^P(\bar{a}_t^m,\bar{a}_t^g)-W_t^P(\tilde{a}_t^m,\tilde{a}_t^g)\right]^\theta k_t^{1-\theta} \quad (8)$$
$$\begin{aligned}\text{s.t. } \bar{a}_t^m+q_t\bar{a}_t^b &\leq a_t^m+p_{1t}\bar{y}_t\\ \bar{y}_t &\leq y_t\\ W_t^P(\tilde{a}_t^m,\tilde{a}_t^g) &\leq W_t^P(\bar{a}_t^m,\bar{a}_t^g),\end{aligned}$$

where $\bar{\boldsymbol{a}}_t = (\bar{a}_t^m, \bar{a}_t^b)$, $\bar{a}_t^g = \bar{a}_t^b + (y_t - \bar{y}_t)\underline{\kappa} - k_t$, $\tilde{a}_t^m = \tilde{a}_{Pt}^m(y_t, a_t^m)$, and $\tilde{a}_t^g = (y_t - \tilde{y}_{Pt}(y_t, a_t^m))\underline{\kappa}$. The first constraint is the budget constraint the producer faces in the first subperiod when able to trade simultaneously in the goods market and the bond market. The second constraint states that the producer can at most sell the inventory of good 1 produced at the beginning of the period. The third constraint ensures the trade is incentive compatible for the producer. The restriction $k_t \in \mathbb{R}_+$ ensures the trade is also incentive compatible for the banker.[8] Notice that if the producer and the banker were unable to reach an agreement, the producer can still trade in the goods market. Hence, the outcome (7), which determines the gain from selling of a cash-only producer, acts as the cash-and-credit producer's outside option in his bargaining problem with the banker.

Let $V_t^i(a_t^m)$ denote maximum expected discounted payoff of an agent of type $i \in \{B, C, P\}$ who enters the first subperiod of period t with money holding a_t^m. For a banker,

$$\begin{aligned} V_t^B(a_t^m) &= \alpha \int W_t^B[\bar{a}_{Bt}^m(a_t^m), \bar{a}_{Bt}^b(a_t^m) + k_{Pt}(\tilde{a}_t^m)] dH_t(\tilde{a}_t^m) \\ &\quad + (1-\alpha)\, W_t^B[\bar{\boldsymbol{a}}_{Bt}(a_t^m)], \end{aligned} \tag{9}$$

where H_t is the beginning-of-period t cumulative distribution function of money holdings across producers. For a consumer,

$$V_t^C(a_t^m) = u[\bar{y}_{Ct}(a_t^m)] + W_t^C[\bar{a}_{Ct}^m(a_t^m), \bar{a}_{Ct}^b(a_t^m)]. \tag{10}$$

For a producer,

$$\begin{aligned} V_t^P(a_t^m) &= \max_{y_t \in \mathbb{R}_+} \Big\{ -\kappa y_t + \alpha W_t^P[\bar{a}_{Pt}^m(y_t, a_t^m), \bar{a}_{Pt}^g(y_t, a_t^m)] \\ &\quad + (1-\alpha)\, W_t^P[\tilde{a}_{Pt}^m(y_t, a_t^m), \tilde{a}_{Pt}^g(y_t, a_t^m)] \Big\}, \end{aligned} \tag{11}$$

where $\bar{a}_{Pt}^g(y_t, a_t^m) = \bar{a}_{Pt}^b(y_t, a_t^m) + [y_t - \bar{y}_{Pt}(y_t, a_t^m)]\underline{\kappa} - k_{Pt}(y_t, a_t^m)$, and $\tilde{a}_{Pt}^g(y_t, a_t^m) = [y_t - \tilde{y}_{Pt}(y_t, a_t^m)]\underline{\kappa}$. Let $y_{Pt}(a_t^m)$ denote the solution to the maximization in (11).

Let $A_{it}^m = \int a_t^m dF_{it}(a_t^m)$, where F_{it} is the cumulative distribution function of money holdings for agents of type $i \in \{B, C, P\}$ at the beginning of period t. For asset type $k \in \{m, b\}$, let $\bar{A}_{Bt}^k = \int \bar{a}_{Bt}^k(a_t^m)\, dF_{Bt}(a_t^m)$, $\bar{A}_{Ct}^k = \int \bar{a}_{Ct}^k(a_t^m)\, dF_{Ct}(a_t^m)$, $\bar{A}_{Pt}^k = \alpha \int \bar{a}_{Pt}^k(y_{Pt}(a_t^m), a_t^m)\, dF_{Pt}(a_t^m)$, and $\tilde{A}_{Pt}^m = (1-\alpha)\int \tilde{a}_{Pt}^m(y_{Pt}(a_t^m), a_t^m)\, dF_{Pt}(a_t^m)$. Also, let $\bar{Y}_{Ct} = \int \bar{y}_{Ct}(a_t^m)\, dF_{Ct}(a_t^m)$, $\bar{Y}_{Pt} =$

[8] If a banker who enters the bilateral negotiation with money holding $x \in \mathbb{R}_+$ and a claim to $y \in \mathbb{R}$ units of good 2 negotiates a fee $k_t \in \mathbb{R}$ with the producer, the banker's gain from trade is $\hat{W}_t^B(x, y + k_t) - \hat{W}_t^B(x, y)$. The formulation (8) uses the fact that (4) and (5) imply $\hat{W}_t^B(x, y + k_t) - \hat{W}_t^B(x, y) = k_t$.

$\alpha \int \bar{y}_{Pt}\left(y_{Pt}\left(a_t^m\right), a_t^m\right) dF_{Pt}\left(a_t^m\right)$, $\tilde{Y}_{Pt} = (1-\alpha) \int \tilde{y}_{Pt}\left(y_{Pt}\left(a_t^m\right), a_t^m\right) dF_{Pt}\left(a_t^m\right)$, and $\boldsymbol{p}_t \equiv (p_{1t}, p_{2t})$. We are now ready to define equilibrium.

Definition 1 *An equilibrium is a sequence of prices, $\{\boldsymbol{p}_t, q_t\}_{t\in\mathbb{T}}$, end-of-period money holdings, $\left\{a_{it+1}^m\right\}_{i\in\{B,C,P\},t\in\mathbb{T}}$, and production, supply, consumption, portfolios, and fees in the first subperiod, $\{y_{Pt}(\cdot), \tilde{y}_{Pt}(\cdot), \bar{y}_{it}(\cdot), \tilde{a}_{Pt}^m(\cdot), \bar{\boldsymbol{a}}_{it}(\cdot), \bar{\boldsymbol{a}}_{Bt}(\cdot), k_{it}(\cdot)\}_{i\in\{P,C\},k\in\{m,b\},t\in\mathbb{T}}$, such that for all $t \in \mathbb{T}$: (i) taking prices and the bargaining protocol as given, the end-of-period money holdings solve (4) for $i \in \{B, C, P\}$; (ii) the asset holdings and fees in the first subperiod solve (5), (6), (7), (8); (iii) beginning-of-period production $y_{Pt}(\cdot)$ satisfies (11); and (iv) prices are such that all Walrasian markets clear, i.e., $\left[\sum_{i\in\{B,C,P\}} A_{it+1}^m - M_{t+1}\right] \mathbb{I}_{\{\max \boldsymbol{p}_t<\infty\}} = 0$ (the end-of-period t Walrasian market for money clears), $\sum_{i\in\{B,C,P\}} \bar{A}_{it}^b = 0$ (the period t market for bonds clears), $\bar{Y}_{Ct} = \tilde{Y}_{Pt} + \bar{Y}_{Pt}$ (the market for good 1 clears), and $\left[\tilde{A}_{Pt}^m + \sum_{i\in\{B,C,P\}} \bar{A}_{it}^m - M_t\right] \mathbb{I}_{\{\max \boldsymbol{p}_t<\infty\}} = 0$ (the first-subperiod money market clears). An equilibrium is "monetary" if $\max \boldsymbol{p}_t < \infty$ for each $t \in \mathbb{T}$ and "nonmonetary" otherwise.*

4 Nonmonetary economy

We begin by characterizing equilibrium in an economy without money. In this context, $M_t = 0$ for all t, so only good 1 and the bond are traded in the first subperiod. Let φ_t^n denote the relative price of good 1 in terms of the bond in the first subperiod of period t. The following result characterizes equilibrium in a nonmonetary economy.

Proposition 2 *Assume $\varphi^n < u'(0)$, where*

$$\varphi^n = \kappa + \frac{1-\alpha\theta}{\alpha\theta}(\kappa - \underline{\kappa}). \tag{12}$$

There exists a unique equilibrium of the nonmonetary economy, and $\varphi_t^n = \varphi^n$ for all t. Consumption of good 1, $\bar{y}_C^n$, satisfies

$$u'\left(\bar{y}_C^n\right) = \varphi^n, \tag{13}$$

and production of good 1 is $y_P^n = \bar{y}_C^n/\alpha$.

The equilibrium described in Proposition 2 works as follows. Consumers demand $\bar{y}_C^n$ units of good 1 and pay by issuing the bond. The proportion α of producers who have access to a banker sell all their their inventory y_P^n of good 1 and accept the bond as payment, while the proportion $1-\alpha$ of producers without access to the bond market store their inventory y_P^n until the following subperiod.[9]

Consider the problem of a producer who is deciding the production of good 1 given a relative price $x \in \mathbb{R}_+$. The producer's expected unit profit is $\Pi^n(x) \equiv R^n(x) - \kappa$, where $R^n(x) \equiv (1-\alpha)\underline{\kappa} + \alpha[\underline{\kappa} + \theta(x - \underline{\kappa})]$ is the expected unit revenue.[10] Given the constant-returns production technology, a producer must expect to break even in an equilibrium with production of good 1. Thus, the equilibrium price, φ^n, must satisfy $\Pi^n(\varphi^n) = 0$, which is equivalent to (12). Since the consumer faces no financing constraints of any kind, she simply chooses her demand by equating her marginal utility to the market price, as in (13).

In the nonmonetary economy, the relative price of good 1 is higher than the marginal cost of production (i.e., $\kappa \leq \varphi^n$, with "=" only if $\alpha\theta = 1$), so consumption of good 1 is inefficiently low (i.e., $\bar{y}_C^n \leq y^\star$). This price markup has two sources. The first, is that producers have imperfect access to bankers (i.e., $\alpha < 1$). The second, is that bankers have market power when transacting with producers (i.e., $\theta < 1$). Each of these sources implies the producer's expected unit revenue is lower than the market price, i.e., $R^n(\varphi^n) < \varphi^n$. Together with the optimality conditions for the consumption and production decisions, this wedge between marginal revenue and price implies a wedge between marginal utility and the marginal cost of production, i.e., $\kappa = R^n(\varphi^n) < \varphi^n = u'(\bar{y}_C^n)$. In sum, the credit/payment/settlement frictions (i.e., $\alpha\theta < 1$) induce a markup in good 1 even though individual producers are price takers in that market.[11]

[9] Given the prices and allocations in Proposition 2, the rest of the equilibrium is immediate from Lemma 1 in the appendix.

[10] The first term in $R^n(x)$ reflects the fact that if the producer cannot trade with a banker (an event that occurs with probability $1-\alpha$) then he is unable to execute the bond trade needed to settle the sale of good 1, and therefore only earns the storage return, $\underline{\kappa}$, on the ex-ante investment, κ. The second term in $R^n(x)$ reflects the fact that if the producer can trade with a banker (an event that occurs with probability α) then he is able to settle the sale of good 1, but only gets revenue equal to his outside option (i.e., the storage return $\underline{\kappa}$) plus a share θ of the gain from selling a unit of good 1 out of the inventory rather than storing it until the following subperiod (i.e., $x - \underline{\kappa}$).

[11] This explanation takes as given our maintained assumption $\underline{\kappa} < \kappa$, which means that producing good 2 in the second subperiod is cheaper (in terms of labor input) than obtaining it by producing good 1 in the first subperiod and storing it until the second subperiod. If $\underline{\kappa} = \kappa$, then κ would be both, the producer's outside option in the negotation with a banker, and his expected unit revenue if he did not access a banker, and therefore $\varphi^n = \kappa$ would be the only relative price consistent with expected unit profit equal to zero. Intuitively, $\underline{\kappa} = \kappa$ makes financial access and banker market power irrelevant since it eliminates the post-production hold-up problem faced by a producer whose marginal cost of production, κ, exceeds the unit value of unsold inventory, $\underline{\kappa}$.

5 Monetary equilibrium

In a monetary economy it is useful to think of the *nominal policy rate* chosen by the monetary authority, as

$$\iota \equiv \frac{\mu - \beta}{\beta}. \tag{14}$$

(Throughout we assume $\beta < \mu$, but will consider the limiting case $\mu \to \beta$.) Also, define the relative price of good 1 in terms of good 2 implied by the nominal prices, i.e., $\varphi_t^m \equiv p_{1t}/p_{2t}$, real money balances as $Z_{it} \equiv M_t/p_{it}$ for $i \in \{1, 2\}$, and

$$\rho_t \equiv \frac{p_{2t}}{q_t} - 1, \tag{15}$$

which is the equilibrium interest rate on the inside bond.[12] Let $\mathcal{V}_t \equiv p_{1t}\bar{Y}_{Ct}/M_t$ denote the *velocity of money*, defined as the ratio of nominal purchases of good 1 to the stock of money.[13]The following result characterizes the stationary monetary equilibrium.[14]

Proposition 3 *Assume $\varphi^n < u'(0)$, and let*

$$\bar{\iota} \equiv \frac{1}{\alpha\theta} \frac{\kappa - \underline{\kappa}}{\underline{\kappa}}. \tag{16}$$

There exists a unique stationary monetary equilibrium provided $0 \le \iota < \bar{\iota}$. In the stationary monetary equilibrium, $\rho_t = \rho$, $\varphi_t^m = \varphi^m$, $Z_{it} = Z_i$ for $i \in \{1, 2\}$, $\mathcal{V}_t = \mathcal{V}$ for all t, $p_{it} = \frac{M_t}{Z_{it}}$ for $i \in \{1, 2\}$, and $q_t = \frac{p_{2t}}{1+\rho_t}$. In addition,

[12] An agent can use 1 unit of money to buy $\frac{1}{q_t}$ bonds, which in total yield $\frac{1}{q_t}$ units of good 2 in the following subperiod, which are in turn equivalent to $\frac{p_{2t}}{q_t}$ dollars. Since the bond is repaid within the period, ρ_t can be interpreted as a real interest rate on the bond (with loan and repayment measured in terms of the good 2). To see this, notice that an agent can use p_{2t} dollars to buy $\frac{p_{2t}}{q_t}$ bonds, which in total yield $\frac{p_{2t}}{q_t}$ units of good 2. Since investing p_{2t} dollars in the bond is equivalent to investing 1 unit of good 2, the gross real interest on the bond expressed in terms of good 2 is also equal to $\frac{p_{2t}}{q_t}$. Throughout we specialize the analysis to the case $0 \le \rho_t$ because $\rho_t < 0$ entails an arbitrage opportunity inconsistent with equilibrium.

[13] The main results are essentially unchanged if we defined velocity using consumption of good 2 or any combination of consumption of good 1 and good 2.

[14] Throughout we focus on equilibria in which good 1 is produced. The full set of dynamic equilibrium conditions is reported in Appendix C (Lemma 5). In Appendix C we also characterize dynamic monetary equilibria (Section C.2), and sunspot monetary equilibria (Section C.3).

(i) If $0 < \iota < \bar{\iota}$*, then*

$$\rho = \iota \tag{17}$$

$$\varphi^m = \frac{1}{1+\alpha\theta\iota}\kappa \tag{18}$$

$$Z_1 = \frac{1}{\varphi^m} Z_2 = (1-\alpha)\, y^m \tag{19}$$

$$\mathcal{V} = \frac{1}{1-\alpha}, \tag{20}$$

where $y^m \equiv \mathrm{D}((1+\iota)\,\varphi^m)$ *is the consumption (and production) of good 1.*

(ii) As $\iota \to 0$*,* $\varphi^m \to \kappa$*, and* $y^m \to y^\star$*, and any* $Z_1 \in [(1-\alpha)\,\kappa, \infty)$ *is consistent with equilibrium.*

(iii) As $\iota \to \bar{\iota}$*,* $\varphi^m \to \underline{\kappa}$*, and* $y^m \to y^n$*.*

The equilibrium described in Proposition 3 works as follows. Either consumers or bankers carry money balances from a period into the following one. Their decision to hold money overnight is determined by an Euler equation that equates the marginal opportunity cost of holding money, i.e., the policy rate ι, to the marginal return, which equals the market interest rate on the inside bond, ρ. Consumers demand y^m units of good 1 and pay with money and/or by issuing the bond. The proportion α of producers who have access to a banker sell all their their inventory, y^m, of good 1 and accept the bond as payment, while the proportion $1-\alpha$ of producers without access to the bond market sell all their inventory, y^m, for money. If $\iota > 0$, then money demand in the first subperiod is entirely accounted for by the unbanked producers.

Consider the problem of a producer who is deciding the production of good 1 given a relative price $\varphi^m \in \mathbb{R}_+$ and an interest rate $\rho \in \mathbb{R}_+$. The producer's expected unit profit is $\Pi^m(\varphi^m) \equiv R^m(\varphi^m) - \kappa$, where $R^m(\varphi^m) \equiv (1-\alpha)\,\varphi^m + \alpha\,(1+\theta\rho)\,\varphi^m$ is the expected *effective unit revenue*.[15] Given the constant-returns production technology, a producer must expect to break even in an equilibrium with production of good 1. Thus, the equilibrium price, φ^m, must satisfy $\Pi^m(\varphi^m) = 0$, which after substituting the Euler equation (17), is equivalent to (18).

[15]The first term in $R^m(\varphi^m)$ reflects the fact that if the producer cannot trade with a banker (an event that occurs with probability $1-\alpha$) then he is unable to earn the bond return on the proceeds from the monetary sale of good 1, and therefore only earns the relative price of selling for money, φ^m, on the ex-ante investment, κ. The second term in $R^m(\varphi^m)$ reflects the fact that if the producer can trade with a banker (an event that occurs with probability α) then he earns his outside option in the negotiation with the banker (i.e., φ^m, the unit revenue of a producer with no access to a banker) plus a share θ of the bond return on the unit revenue of monetary trade, i.e., $\rho\varphi^m$.

The consumer chooses her demand by equating her marginal utility to the *effective price* of good 1, i.e., $u'(y^m) = (1+\rho)\,\varphi^m$.[16]

In the monetary equilibrium, the effective price of good 1 that determines the quantity demanded and produced is higher than the marginal cost of production (i.e., $\kappa \leq (1+\iota)\,\varphi^m$, with "=" only if $\alpha\theta = 1$ or $\iota = 0$), so consumption of good 1 is inefficiently low (i.e., $y^m \leq y^\star$). To fix ideas, assume $\iota > 0$. The wedge between the marginal cost of production and the consumer's effective price has two sources. The first, is that producers have imperfect access to bankers (i.e., $\alpha < 1$). The second, is that bankers have market power when transacting with producers (i.e., $\theta < 1$). Each of these sources implies the producer's expected effective unit revenue is lower than the effective price to the consumer, i.e., $R^m(\varphi^m) < (1+\rho)\,\varphi^m$. Together with the optimality conditions for the consumption and production decisions, this wedge between the expected effective marginal revenue and the effective price implies a wedge between marginal utility and the marginal cost of production, i.e., $\kappa = R^m(\varphi^m) < (1+\rho)\,\varphi^m = u'(y^m)$. In sum, the credit/payment/settlement frictions (i.e., $\alpha\theta < 1$) induce an effective markup in good 1 even though individual producers are price takers in that market.[17]

The monetary aspects of the equilibrium are simple. The price level implied by (19) is $p_{1t} = \frac{M_t}{(1-\alpha)y^m}$ in terms of good 1, (or $p_{2t} = \frac{M_t}{(1-\alpha)\varphi^m y^m}$ in terms of good 2), i.e., p_{1t} is equal to the nominal quantity of money in circulation per unit of good 1 that is purchased with money. The velocity of money in (20) is immediate from the definition $\mathcal{V} \equiv p_{1t}y^m/M_t$, and corresponds to the intuitive idea that since all consumers can borrow, and the fraction α of producers with access to bankers can instantaneously lend their nominal revenue back to consumers, on average, each unit of money is spent $\mathcal{V} = \frac{1}{1-\alpha}$ times.[18]

---

[16]The effective price that determines the consumer's demand consists of the relative price of the monetary trade, φ^m, plus the financing cost per unit of good 1, i.e., $\rho\varphi^m$.

[17]This explanation assumes $\iota > 0$. If $\iota = 0$, then the rent a financial intermediary can extract from a seller is nill, and consequently so is the wedge between the expected effective marginal revenue to the seller and the effective price borne by the consumer.

[18]To see this in a different way, we could decompose the instantaneous trading activity in subperiod 1 into a countable number of notional trading rounds. At the beginning of the period, the whole money supply is in the hands of consumers who initially spend it all in what we regard as the *first spending round*. A fraction $1-\alpha$ of the money spent in the first round is paid to producers with no access to bankers, and is therefore not spent again in the same subperiod. But a fraction α of the first-round nominal spending is paid to producers with access to bankers, who instantaneously lend it out to bankers, who instantaneously lend it out to consumers. Consumers use this borrowed money to purchase good 1 in a second round of spending. A fraction $1-\alpha$ of the money used for second-round purchases is not spent again, but a fraction α is spent one more time, and then a fraction α of that spending is spent one more time, and so on. This iterative process implies that $(1-\alpha)\,\alpha^{k-1}$ is the probability that a given unit of money is spent exactly $k \in \{1, ..., \infty\}$ times, and therefore each unit of money is spent $\sum_{k=1}^{\infty} k\,(1-\alpha)\,\alpha^{k-1} = \frac{1}{1-\alpha}$ times on average.

The consumption allocation implemented by the stationary monetary equilibrium converges to the nonmonetary allocation as $\iota \to \bar{\iota}$, and to the efficient allocation as either $\iota \to 0$ or $\alpha\theta \to 1$. In terms of comparative statics, as long as $\alpha\theta < 1$, we have $\frac{\partial \varphi^m}{\partial \iota} < 0 < \frac{\partial (1+\iota)\varphi^m}{\partial \iota}$, so $\frac{\partial y^m}{\partial \iota} < 0$.[19]

6 Cashless limit

In this section we consider the limiting economy as $\alpha \to 1$, i.e., as the fraction of producers without access to bankers vanishes. The following corollary of Proposition 2 characterizes the limit as $\alpha \to 1$ of the equilibrium of the nonmonetary economy.

Corollary 1 *Assume $\varphi^{n*} < u'(0)$, where*

$$\varphi^{n*} \equiv \lim_{\alpha \to 1} \varphi^n = \kappa + \frac{1-\theta}{\theta}(\kappa - \underline{\kappa}). \tag{21}$$

Then,

$$\lim_{\alpha \to 1} \bar{y}_C^n = \lim_{\alpha \to 1} \bar{y}_P^n = \mathrm{D}(\varphi^{n*}). \tag{22}$$

The following corollary of Proposition 3 characterizes the limit as $\alpha \to 1$ of the stationary monetary equilibrium.[20]

Corollary 2 *Consider the monetary equilibrium characterized in Proposition 3. Assume $\varphi^{n*} < u'(0)$, and let $\bar{\iota}^* \equiv \frac{1}{\theta}\frac{\kappa - \underline{\kappa}}{\underline{\kappa}}$. For any $\iota \in [0, \bar{\iota}^*]$,*

$$\lim_{\alpha \to 1} Z_i = \lim_{\alpha \to 1} \frac{1}{p_{it}} = \lim_{\alpha \to 1} \frac{1}{\mathcal{V}} = 0 \text{ for } i \in \{1, 2\} \tag{23}$$

$$\lim_{\alpha \to 1} \varphi^m = \frac{1}{1+\theta\iota}\kappa \tag{24}$$

$$\lim_{\alpha \to 1} y^m = \mathrm{D}\left(\frac{1+\iota}{1+\theta\iota}\kappa\right). \tag{25}$$

From (23), as $\alpha \to 1$, real money balances converge to 0, and nominal prices and velocity of money diverge to infinity. That is, the monetary economy approaches a *cashless limit* as the proportion of producers who can settle sales of good 1 through the credit intermediaries approaches 1. Condition (24) establishes that, in the cashless limit, the relative price of good

[19]Intuitively, whenever $\alpha\theta < 1$ and $\rho > 0$, the effective price of good 1 faced by the consumer is a markup over marginal cost, i.e., $(1+\rho)\varphi^m - \kappa = \frac{1-\alpha\theta}{1+\alpha\theta\rho}\rho\kappa$. This markup is increasing the equilibrium bond rate, ρ, which in equilibrium is equal to the policy rate, ι. Therefore, consumption is decreasing in the nominal rate, ι.

[20]The analogous limits for dynamic and sunspot monetary equilibria can be found in Appendix C (Sections C.2.1 and C.3.1, respectively).

1 in terms of good 2 implied by the nominal prices converges to $\varphi^{m*} \equiv \frac{1}{1+\theta\iota}\kappa$, and therefore the effective relative price faced by consumers converges to $(1+\iota)\,\varphi^{m*}$. Condition (25) then establishes that consumption of good 1 converges to $y^{m*} \equiv \mathrm{D}((1+\iota)\,\varphi^{m*})$.

Notice that as long as $\iota > 0$ and $\theta < 1$, real consumption in the cashless limit, i.e., y^{m*}, responds to changes in the nominal policy rate ι, which is also the opportunity cost of holding money. Specifically, $\frac{\partial \varphi^{m*}}{\partial \iota} < 0 < \frac{\partial[(1+\iota)\varphi^{m*}]}{\partial \iota}$, so consumption and output are decreasing in the opportunity cost of holding money—even if real money balances are negligible, or equivalently, even if the real value of all purchases of good 1 that are carried out with money is virtually nil. How is this possible? More precisely, why is the opportunity cost of holding money, ι, still a relevant determinant of the allocation when real balances are virtually not being used in transactions?

The answer is based on two observations. First, although real money balances are becoming negligible, they are being held along the cashless limit. Thus, the Euler equation (17) that ties the policy rate, ι, to the equilibrium real interest rate that consumers internalize when they purchase consumption via credit, ρ, holds everywhere along the cashless limit. Hence even in the cashless limit, ι remains relevant for consumers through its effect on ρ. Second, everywhere along the cashless limit, as long as $\theta < 1$, the effective relative price of good 1 that consumers face is a markup over marginal cost that depends on the equilibrium real interest rate, ρ. In sum, even if real money balances are negligible (as they are far along the cashless limit), an increase in the nominal policy rate, ι, increases the equilibrium real interest rate, ρ, which in turn increases the effective markup that consumers pay for good 1 (and the wedge between the marginal cost of production and the consumer's marginal utility of consumption), which induces consumers to decrease their demand of good 1. The aggregate real quantity of money plays no role in this monetary transmission channel—only the opportunity cost of holding money does. As mentioned above, this mechanism is operative as long as $\iota > 0$ and $\theta < 1$. As is clear from (25), if either $\iota = 0$ or $\theta = 1$, then there is no wedge between the producer's marginal revenue and the effective relative price that consumers face (i.e., no markup for good 1), and therefore along the cashless limit, the allocation implemented by the stationary monetary equilibrium converges to the efficient allocation characterized in Proposition 1.

Together, Corollary 1 and Corollary 2 imply that, generically (in terms of the market-power parameter θ), the equilibrium of the pure-credit economy with no money is not a good approximation to the pure-credit (cashless) limit of the monetary economy. Formally, (21) and

(24) imply

$$\varphi^{n*} - (1+\iota)\,\varphi^{m*} = \frac{1-\theta}{\theta}\left(\frac{1}{1+\theta\iota}\kappa - \underline{\kappa}\right). \tag{26}$$

Notice that $\varphi^{n*} \geq (1+\iota)\,\varphi^{m*}$ for all $\theta \in [0,1]$ and all $\iota \in [0,\bar{\iota}^{*})$, with "=" only if $\theta = 1$, so $\mathrm{D}((1+\iota)\,\varphi^{m*}) \geq \mathrm{D}(\varphi^{n*})$, i.e., consumption is higher in the cashless limit of the monetary economy than in the nonmonetary economy. In other words, the allocation implemented by the cashless limit of the monetary equilibrium coincides with the allocation implemented by the equilibrium of the nonmonetary economy only if bankers have no market power over producers (i.e., $\theta = 1$). Generically, however, $\mathrm{D}((1+\iota)\,\varphi^{m*}) - \mathrm{D}(\varphi^{n*}) > 0$, and this difference is decreasing in θ. A monetary policy that makes money more valuable (e.g., by reducing the opportunity cost of holding it, ι) makes $\varphi^{n*} - (1+\iota)\,\varphi^{m*}$ larger, since it improves the producer's outside option of trading good 1 for money, which reduces the banker's effective market power, and ultimately reduces the markup in the market for good 1.[21] For any $\theta \in [0,1)$, the difference between the equilibrium allocation implemented by the pure-credit (cashless) limit of the monetary economy, and the allocation implemented by the pure-credit economy with no money, $\mathrm{D}((1+\iota)\,\varphi^{m*}) - \mathrm{D}(\varphi^{n*})$, is decreasing in ι.[22]

In order to understand why the option to sell for money is an effective and credible outside option for sellers in their negotiations with bankers even as aggregate real money balances become very small, notice that for all $\iota \in [0,\bar{\iota}^{*})$,

$$\lim_{\alpha\to 1} Z_1 = 0 < \lim_{\alpha\to 1}\frac{Z_1}{1-\alpha} = \mathrm{D}((1+\iota)\,\varphi^{m*}).$$

That is, aggregate demand for money in the first subperiod converges to zero, but the *individual demand for money* does not, in the sense that any individual producer who belongs to the vanishing population of producers without access to a banker is willing to accept money in exchange for good 1. Hence, when trading with a banker, the producer's threat to sell for cash is credible everywhere along the cashless limit.

[21] In contrast, $\lim_{\iota\to\bar{\iota}^{*}}[\varphi^{n*} - (1+\iota)\,\varphi^{m*}] = 0$ even if $\theta < 1$. That is, $\mathrm{D}((1+\iota)\,\varphi^{m*}) - \mathrm{D}(\varphi^{n*})$ can be made arbitrarily small by choosing a background monetary policy rate ι high enough to make the value of money sufficiently small. Intuitively, if expected inflation is very high, monetary exchange ceases to be an effective outside option for producers in their negotiations with banks.

[22] As we discuss in Section 7.2, in previous treatments of the cashless limit, e.g., Woodford (1998), the continuity argument relies on a background monetary policy akin to the Friedman Rule that makes the opportunity cost of holding cash equal to zero. In contrast, in our theory, the discrepancy between the pure-credit limit of the nonmonetary economy and the cashless limit of the monetary economy is *largest* at $\iota = 0$.

6.1 Price-level determination in cashless limiting economies

How can the nominal price level be determined in cashless economies? This seemingly paradoxical question is central to the literature that studies monetary policy in models without money, and much effort has been devoted to answering it (see, e.g., Woodford (1998, 2003), Cochrane (2005, 2021)). In this section we offer an alternative view of the determination of the price level in the cashless limit. In our approach, the price level in the cashless limiting economy is determined by equating demand and supply of money, much as in every textbook monetary model with money. Specifically, if the monetary authority wishes to implement a certain price path for a stationary monetary equilibrium of the cashless limiting economy with some $\iota \in (0, \bar{\iota})$, then it can simply choose a money supply process $\{M_t\}_{t=0}^{\infty}$ given by $M_t = (1-\alpha)\,\bar{M}_t$, with $\bar{M}_{t+1} = \mu \bar{M}_t$ for some $\mu \in \mathbb{R}_{++}$. By Corollary 2, we know this monetary policy implements a price level, p_{1t}, in the cashless limit of the stationary monetary equilibrium that is equal to

$$\lim_{\alpha \to 1} p_{1t} = \frac{\bar{M}_t}{\mathrm{D}\left(\frac{1+\iota}{1+\theta\iota}\kappa\right)}.$$

By choosing the level of $\bar{M}_0$, the monetary authority can implement any desired price level in the stationary monetary equilibrium of the cashless limiting economy. Intuitively, we can always implement a price level that remains well defined (i.e., finite) even in the cashless limit, simply by ensuring that the money supply *per producer without access to a banker* remains stable along the cashless limit.[23]

6.2 Welfare

In this section we compare the welfare associated with the nonmonetary and monetary equilibrium paths. To this end, we consider a measure of welfare based on the (equally weighted) sum of all agents' expected discounted utilities at the beginning of a period. We use $\mathcal{W}^{\star}$, $\mathcal{W}^{n}$, and $\mathcal{W}^{m}$ to denote the levels of welfare achieved by the planner's solution, the nonmonetary equilibrium, and the stationary monetary equilibrium, respectively. To streamline the exposition, we assume $v(x) = x$ (without meaningful loss of generality).

According to Proposition 1,

$$(1-\beta)\,\mathcal{W}^{\star} = u(\mathrm{D}(\kappa)) - \kappa \mathrm{D}(\kappa)\,. \tag{27}$$

[23]To streamline the exposition, here we have focused on price level determination in the stationary monetary equilibrium. The same logic can be used to determine the price level in the cashless limit of a dynamic equilibrium characterized in Proposition 6.

In the appendix (Lemma 6), we show that

$$(1-\beta)\,\mathcal{W}^n = u(\mathrm{D}(\varphi^n)) - \left[\kappa + \frac{1-\alpha}{\alpha\theta}(\kappa - \underline{\kappa})\right]\mathrm{D}(\varphi^n) \quad (28)$$

$$(1-\beta)\,\mathcal{W}^m = u(\mathrm{D}((1+\iota)\,\varphi^m)) - \kappa\mathrm{D}((1+\iota)\,\varphi^m), \quad (29)$$

with φ^n and φ^m as defined in Proposition 2 and Proposition 3, respectively.

The following proposition establishes that the stationary monetary equilibrium achieves a higher level of welfare than the economy with no money, and achieves the first-best level of welfare if monetary policy implements $\iota = 0$ (i.e., the Friedman rule). Welfare in the cashless limit of the stationary monetary equilibrium is strictly higher than in the nonmonetary economy provided $\theta < 1$ and $\iota < \lim_{\alpha\to 1}\bar{\iota}$. The condition $\theta < 1$ says that bankers have some degree of market power vis-à-vis producers. This condition is necessary for the producer's option to sell for cash to affect the terms of trade in pure-credit transactions intermediated by bankers. The condition $\iota < \lim_{\alpha\to 1}\bar{\iota}$ ensures an individual producer is willing to demand money in exchange for good 1 so that the producer's threat to circumvent the banker by trading for money is indeed credible.

Proposition 4 *(i) If $\alpha < 1$, welfare in the stationary monetary equilibrium is decreasing in the nominal interest rate, i.e., $\partial\mathcal{W}^m/\partial\iota < 0$, and*

$$\mathcal{W}^n < \mathcal{W}^m \leq \mathcal{W}^\star,$$

where the second inequality is strict unless $\iota = 0$.

(ii) Welfare in the cashless limit of the stationary monetary equilibrium is decreasing in the nominal rate as long as the nominal rate is positive and bankers have some market power against producers, i.e., $\partial\left[\lim_{\alpha\to 1}\mathcal{W}^m\right]/\partial\iota < 0$ as long as $0 < (1-\theta)\,\iota$. Moreover,

$$\lim_{\alpha\to 1}\mathcal{W}^n \leq \lim_{\alpha\to 1}\mathcal{W}^m \leq \lim_{\alpha\to 1}\mathcal{W}^\star,$$

where the first inequality is strict unless $(1-\theta)\,(\iota - \lim_{\alpha\to 1}\bar{\iota}) = 0$, and the second inequality is strict unless $(1-\theta)\,\iota = 0$.

The following proposition shows that in the cashless limit, for small values of ι, the magnitude of the effects of monetary policy on consumption and welfare depends on a single *sufficient statistic*: the product of the *deposit spread* that bankers with market power impose on lenders, and the *price elasticity of demand* for good 1. In the model, the latter is $\epsilon \equiv \frac{d[\ln \mathrm{D}(\varphi)]}{d \ln \varphi}$, where $\varphi \equiv (1+\iota)\, \varphi^m$ is the effective price of good 1. To formalize the theoretical counterpart of the deposit spread, notice that ρ_t is the interest rate a lender would earn if he had direct access to the competitive bond market. However, access to the bond market is intermediated by a banker who charges a fee. From Lemma 2, we know that along a monetary equilibrium, a producer who produces y_t at the beginning of the period and contacts a banker, lends $q_t \bar{a}^b_{Pt}(y_t) = p_{1t} y_t$ units of money in the first subperiod, for which he receives $p_{2t}[\varphi_t y_t - k_{Pt}(y_t, 0)]$ units of money (net of the banker's intermediation fee) in the second subperiod. Hence, net of the fee, the lender earns a *deposit rate* $\rho^D_t \equiv \frac{p_{2t}[\varphi_t y_t - k_{Pt}(y_t,0)]}{p_{1t} y_t} - 1$. From part (*iii*) of Lemma 2 (see appendix), $k_{Pt}(y_t, 0) = (1-\theta)\, \rho_t \varphi^m_t y_t$, so along a stationary monetary equilibrium, $\rho^D_t = \rho^D$ for all t, and $\rho^D = \theta \rho$. In other words, the banker earns a *deposit spread* $(\rho - \rho^D)/\rho = 1 - \theta$.

Proposition 5 *Let $\tau(\iota)$ denote the compensating variation associated with a deviation in the nominal policy rate from* 0 *(the Friedman rule) to ι in the cashless limit of the stationary monetary economy. Then,*

$$\frac{d\tau(\iota)}{d\iota} = -\frac{d\left[\ln \mathrm{D}\left(\varphi^*\right)\right]}{d\iota} \approx -(1-\theta)\, \epsilon.$$

7 Discussion

7.1 Applications and related work

The basic design of our model builds on Lagos and Wright (2005). The particular market structure is similar to the one we have used in Lagos and Zhang (2015, 2019, 2020), which in turn adopts some elements from Duffie et al. (2005). In Lagos and Zhang (2019) we study the effects of monetary policy in the cashless limit of an economy where investors can settle equity trades using money or margin loans that are intermediated by brokers with market power. That model is calibrated to match the empirical estimates of the asset price responses to monetary policy shocks, and used to obtain quantitative theoretical estimates of these responses in the cashless limit. A key difference with Lagos and Zhang (2019) is that here, monetary exchange is between buyers and sellers of a consumption good as in canonical monetary models (e.g., Samuelson (1958), Lucas (1980), or Lagos and Wright (2005)). In contrast with canonical monetary models, which emphasize the usefulness of money for buyers with limited access to credit, the baseline formulation of the model we develop here has buyers with unlimited access to credit, and therefore highlights the usefulness of monetary exchange for sellers who need a means to collect payment from buyers they do not trust. In this context, money is essential only to sellers with no access to the intermediated credit-based settlement.

The key mechanism underlying our results for near-cashless economies is that the mere existence of *some* valued money influences the terms of trade in credit transactions that may not involve money. The fact that traders' asset holdings affect the terms of trade in a bilateral bargain is commonplace in models of decentralized exchange. The mechanism arises naturally in models of over-the-counter trade with unrestricted asset holdings such as Afonso and Lagos (2015). In the search-based monetary literature there are also environments where money confers a strategic bargaining advantage to the agent who holds it. In Zhu and Wallace (2007), for example, the mechanism is embedded in the bargaining protocol, according to which holding money is akin to having more bargaining power. A more recent example is Rocheteau et al. (2018), where holding money improves a borrower's outside option in the bilateral bargain for a loan. In these papers actually holding certain assets (e.g., money) confers a strategic advantage. In contrast, what prevents the medium-of-exchange transmission mechanism from dissipating in the near-cashless economies we study, is that money affects the terms of trade in transactions between counterparties that neither hold nor wish to hold money on the equilibrium path.

We have chosen to interpret our market structure as one where there is a credit market intermediated by bankers where sellers can buy bonds issued by consumers. However, our equations (or minor variants of our equations) admit other interpretations. Here we outline two. First, the model also corresponds to a trading arrangement in the first subperiod where each consumer holds a checking account with a bank, and funds the checking account by taking a loan from the bank. The loans mature in the following subperiod, are payable in good 2, and bear an interest equal to ρ_t. In the first subperiod consumers pay for good 1 with money brought in advance and by debiting their checking accounts at the bank. A proportion $1-\alpha$ of sellers lack the technology to verify the debit transaction with the consumer's bank, and therefore can only accept money as payment. The remaining α producers are able to accept money and checks (debits on the buyer's checking account). A banked seller deposits all her revenue from sales with a bank in the form of an interest bearing deposit account, which is callable in terms of good 2 in the following subperiod, and earns a interest equal to ρ_t^D (as defined in Section 6.2).

An alternative interpretation is to think of the bank as a credit-card company. All consumers and a fraction α of sellers have a contract with a credit-card company (the remaining $1-\alpha$ sellers do not). The credit-card company charges consumers a borrowing rate ρ_t and passes on an interest rate ρ_t^D to the fraction α of sellers who extend credit to consumers through the credit-card company. To illustrate, suppose $\theta = 0$, which may be interpreted as a situation where sellers face a monopolist credit-card company. In a world with no money, the credit-card company would charge sellers a service fee large enough to extract all the gains from their trade with buyers. In a world with money, the same credit card company would be unable to charge a fee higher than the one that delivers sellers the same gains from trade that they could earn by settling all their sales with money.

Although our focus in this paper has been on money and monetary policy, the reinterpretations of the model that we described above suggest that the basic mechanisms we have studied in the context of government-issued fiat money as a constraint on market power would be relevant in any setting that involves market power in the intermediation of credit, settlement, or payment activities. More generally, the asset that serves as a constraint on the market power of the intermediaries may be any security that serves as alternative means of payment, such as a privately issued money (e.g., a cryptocurrency) or a central bank digital currency (CBDC), as in Andolfatto (2020) or Chiu et al. (2020).

7.2 On the moneyless approach to monetary economics

The moneyless approach to monetary economics was pioneered by Woodford (1998) and, based on the treatments in Woodford (2003) and Galí (2008), is now "the textbook" approach to monetary theory and practice. The common justification for doing monetary economics without money is the view that the frictions associated with the medium-of-exchange role are irrelevant in the transmission of monetary policy. This view rests on two specific results. Both results rely on a model where the medium-of-exchange role of money is not explicit, but rather is proxied by either assuming money is an argument of a utility function, or by imposing that certain purchases be paid for with cash acquired in advance. The first result is theoretical, and can be found in Woodford (1998). The second result is quantitative, and can be found in Woodford (2003) and Galí (2008). We discuss each of them in turn.

Woodford (1998) considers a version of the cash-in-advance economy of Lucas (1980) with "cash goods" and "credit goods" as in Lucas and Stokey (1983), but where the set of cash goods is represented with a parameter $\alpha \in (0, \gamma]$ for some $\gamma \in (0, 1)$. The economy with

$\alpha \to \gamma$ corresponds to the formulation with no credit goods of Lucas (1980). When $\alpha \to 0$, the economy is interpreted to be approaching a "cashless limit" where there are no cash goods, i.e., a conventional perfectly competitive nonmonetary model without a cash-in-advance constraint. In this context, the first result in Woodford (1998) is that under the assumption that the money supply sequence $\{M_t\}_{t=0}^{\infty}$ satisfies $M_t \geq \underline{M}$ for some $\underline{M} > 0$ for all t, then there is no monetary equilibrium in the limiting case $\alpha = 0$ (in the sense that the nominal price of cash goods, $\{p_t\}_{t=0}^{\infty}$, diverges to infinity). The second result is that given p_t is finite for all $\alpha > 0$, one cannot find a solution for the limiting case $\alpha = 0$ as an approximation to the small-α case. Woodford interprets this result to mean that in this model "the use of money in transactions is intrinsic to the model's ability to determine an equilibrium price level." Woodford then augments the model by assuming the government adopts a fiscal-monetary regime that ensures money is valued and held by private agents even if it is merely a redundant asset. Specifically, the government is assumed to: (i) maintain a strictly positive level of *nominal* government liabilities (so that *cash* taxes must be levied on the private sector in order to service the nominal debt), and (ii) pay a nominal interest on money balances (equal to the nominal interest on the government debt), where the nominal interest rate follows an exogenous rule described by a function $g(\cdot)$ of p_t, assumed to be continuously differentiable in the neighborhood of some p^*, with $g(p^*)$ chosen to ensure that money is held in the equilibrium of the economy with $\alpha = 0$ (i.e., to ensure the Euler equation for money holds with equality, and the relevant transversality condition satisfied given $p_{t+1} = p_t = p^*$ for all t). Condition (ii) guarantees money and bonds have the same rate of return, which ensures private agents are willing to hold money even though it is not useful in transactions. Notice that since money plays no role as a medium of exchange, there is no demand of money for private transactions that can be equated to the money supply to determine p_t. Condition (i), however, amounts to assuming a private-sector demand for money needed to meet the nominal tax liabilities with the government; this tax-induced money demand allows the price level, p_t, to be determined using the government budget constraint. In the context of the cash-credit cash-in-advance model under the fiscal-monetary regime described by conditions (i) and (ii), Woodford shows the central approximation result of his paper, namely that the equilibrium is continuous in the parameter α, i.e., the equilibrium of the economy with $\alpha = 0$ can be well approximated by the equilibrium of an economy with positive but small enough value of α.

Hence, against the background of a cash-in-advance economy subject to assumptions (i)

and (ii), the cashless limit just described, i.e., the economy with $\alpha = 0$ where money is a redundant asset with no role in exchange, can be regarded as a good approximation to a monetary economy where money is needed to satisfy a cash-in-advance constraint but only for a very small set of goods, i.e., the economy with α is positive but very small. Since there are no monetary variables or liquidity factors associated with the medium-of-exchange role of money in the Euler equations of the limiting economy, this approximation result is used to justify disregarding such considerations in Euler equations more generally, alluding to economies with "highly developed financial institutions" (meaning economies with low α). In this context, for the Euler equations for other durable assets, ignoring monetary variables and liquidity factors is equivalent to simply assuming a period utility function of the form $U(c_t, z_t) = u(c_t) + A\ell(z_t)$ of consumption, c_t, and real money balances, z_t, for given functions $u(\cdot)$ and $\ell(\cdot)$, and a constant $A \in \mathbb{R}_+$. Thus, the Woodford cashless-limit approximation result is often used to justify this specific money-in-the-utility-function formulation, sometimes with $A \approx 0$. In sum, the takeaway of Woodford (1998) is that the cashless equilibrium that obtains in the limiting case can be used to approximate the monetary equilibrium whenever the share of transactions in which money serves as a medium-of-exchange is small.

Modern textbook treatments of monetary policy deliberately assign a very limited role to money (e.g., see Woodford (2003, p. 32), Galí (2008, p. 10), and Cochrane (2021, p. 1-4)). For the most part, the medium-of-exchange role is either ignored, or when it is acknowledged, it is incorporated implicitly by assuming real money balances as an argument of the agents' utility functions (or some equivalent cash-in-advance formulation). The preferred specification is $U(c_t, z_t) = u(c_t) + A\ell(z_t)$. This separable specification is justified by showing that, in the context of a competitive model with no credit frictions, if $U(c_t, z_t)$ is nonseparable, then the elasticity of output with respect to a monetary shock that raises the nominal interest rate by one percentage point is proportional to inverse velocity, $M_t/(p_t Y_t)$, where M_t/p_t denotes aggregate real money balances, and Y_t denotes GDP. Specifically, Woodford (2003, p. 113) and Galí (2008, p. 31) argue that since $M_t/(p_t Y_t)$ is small in the data (e.g., with M_t interpreted as the monetary base), the effect of monetary policy on output that is attributable to monetary frictions is quantitatively small so it can be ignored, e.g., by considering the simpler formulation $U(c_t, z_t) = u(c_t) + A\ell(z_t)$, often even assuming $A \approx 0$.

The literature mentions several reasons why it may be interesting to study monetary policy in limit cashless economies such as the one with $\alpha \to 0$ in Woodford (1998). The first, as argued by Woodford (1998, p. 174), is that the hypothetical cashless limit may one day become a reality as a result financial innovations that continually reduce the quantity of the monetary base that needs to be held on average to carry out a given volume of transactions: "The only natural limit to this process is an ideal state of frictionless financial markets in which there is no positive demand for the monetary base at all, if it is dominated by other financial assets, and no determinate demand for it if it is not." The second, is that the cashless limit may be a useful thought experiment, as argued by Wicksell (1936, p. 70) when considering his "pure credit economy," defined as:

> *"... a state of affairs in which money does not actually circulate at all, neither in the form of coin (except perhaps as small change) nor in the form of notes, but where all domestic payments are effected by means of the Giro system and bookkeeping transfers. A thorough analysis of this purely imaginary case seems to me to be worth while, for it provides a precise antithesis to the equally imaginary case of a pure cash system, in which credit plays no part whatsoever. The monetary systems actually employed in various countries can then be regarded as combinations of these two extreme types. If we can obtain a clear picture of the causes responsible for the value of money in both of these imaginary cases, we shall, I think, have found the right key to a solution of the complications which monetary phenomena exhibit in practice."*

The cashless limit we considered in Section 6 is a formalization of Wicksell's verbal description of a "pure credit economy." It is also akin to the cashless limit in Woodford (1998), but set in the context of a monetary model with a more explicit and general microstructure of payments. Our model demonstrates that the cashless limiting result in Woodford (1998) only holds for a much more limited set of monetary environments than previously recognized. This finding is relevant since the cashless limiting result in Woodford (1998) is the theoretical basis for the folk wisdom that monetary policy analysis in the context of high-velocity economies can ignore the role of money in the mechanism of exchange without loss. For example, if credit, payment, or settlement intermediaries have significant market power, the medium-of-exchange role of money will remain a significant conduit for monetary policy even in the cashless limit. Specifically, as $\alpha \to 1$, real balances converge to zero, transaction velocity goes to infinity, and the monetary economy converges to a limit where monetary policy can still have significant effects on consumption, output, and welfare. There is one special case of our theory that delivers irrelevance results for the medium-of-exchange role of money in the cashless limit that are similar to Woodford (1998). It is the case where financial intermediaries have no market power, i.e., $\theta = 1$. So in order to conclude that monetary frictions are irrelevant in cashless limiting economies or almost irrelevant near-cashless economies, it is necessary to also adopt the view that financial, payment, or settlement intermediaries do not have significant market power with their depositors, lenders, or payees. Whether the perfectly competitive case with $\theta = 1$ is the relevant one for applied work is ultimately an empirical issue that deserves further study.[24]

The main insight from our theory is that if the intermediaries who offer the credit-based settlement services have market power, then even sellers with access to credit who neither hold, wish to hold, or choose to hold money on the equilibrium path, benefit from having the option to use money to settle sales—even if they never exercise it. The value of this option is reflected in equilibrium prices and allocations even as the measure of sellers with no access to credit vanishes along the trajectory toward a cashless pure-credit economy. As a result, as aggregate real money balances become negligible and the transaction velocity of money becomes arbitrarily large along the cashless limit, the *latent medium-of-exchange channel* of monetary transmission that operates through the opportunity cost of holding monetary assets remains operative, and determines the relative price of "cash goods" and "credit goods"—even in the cashless limit. In Proposition 5 we have shown that in the cashless limit, the magnitude of the effects of monetary policy on consumption and welfare depends on a single *sufficient statistic*: $(1 - \theta)\epsilon$, i.e., the product of the *deposit spread* that bankers with market power impose on lenders, and the *price elasticity of demand* for the goods purchased with cash or credit. Thus, in general, it is not possible to conclude that the role of money in exchange cannot matter quantitatively solely on the basis that it accounts for a small share of transactions.

7.3 On reduced-form models of money demand

The existing cashless limiting results that we reviewed in Section 7.2 have been obtained in the context of economies where the role of money in exchange is not modeled explicitly, but rather, is proxied by either assuming money is an argument of a utility function, or by imposing that

[24]We are aware of no evidence that a value of θ close to 1 is the empirically relevant case in the financially advanced economies with high velocity of the monetary base that Woodford (2003) and Galí (2008) argue are well approximated by the moneyless approach to monetary policy. In the United States, for example, there is evidence of substantial market power in deposit markets, see, e.g., Berger and Hannan (1989), Hannan and Berger (1991), Neumark and Sharpe (1992), Degryse and Ongena (2008), and Drechsler et al. (2017).

certain purchases be paid for with cash acquired in advance. On the other hand, our results are based on a model with a more general microstructure of payments, and more explicit foundations for the role of money in the exchange process. The purpose of this section is to reconcile the methodological gap between our microfounded approach and the reduced-form approach. To this end, we engineer a reduced-form monetary model (with money in the utility function) that is "equivalent" (in a sense that we make precise below) to our microfounded model.

This exercise delivers three lessons. The first is that conventional money-in-the-utility-function formulations cannot capture the role of money as market-power constraint that we are emphasizing in this paper, no matter how general the utility function is allowed to be (e.g., in terms of the cross interactions between real balances and consumption and leisure). The second lesson is that the Kareken-Wallace-Lucas critique makes it particularly problematic to draw conclusions on the relevance of the medium-of-exchange role of money in the transmission of monetary policy on the basis of a reduced-form model of money as they are conventionally formulated.[25] The third lesson is that there is indeed a fairly simple reduced-form that can represent the equilibrium of our microfounded model, and that is immune to the Kareken-Lucas-Wallace critique through the lens of our theory, but it requires certain modeling choices that are rather unorthodox from the standpoint of the reduced-form literature.

Consider the following reduced-form monetary model. Time is discrete and the horizon infinite. There is a unit measure of identical infinitely lived agents who consume two types of nonstorable final goods (*good 1* and *good 2*) and supply two types of labor (*labor 1* and *labor 2*). Each agent has access to a linear technology that transforms one unit of labor of type 2 into one unit of the final good of type 2. There is a set of intermediate goods of different types, each denoted by $i \in [0,1]$, that serve as input in the production of good 1. There is a competitive firm with access to a production technology that transforms bundles of intermediate goods of different types, i.e., $[y_t(i)]_{i\in[0,1]}$ with $y_t(i) \in \mathbb{R}_+$ for $i \in [0,1]$, into the final good of type 1.

[25] There are well known critiques of reduced-form models of money. Kareken and Wallace (1980) for example, formulate two. The first, is that reduced-form specifications explain too little because they exogenously specify which object will play the role of "money" rather than treating "moneyness" as an equilibrium outcome. For example, arguments of the reduced-form utility functions used in practice typically include real money balances (e.g., as in Sidrauski (1967), Galí (2008), and Woodford (2003)), but also government bonds, equity shares, and other financial assets (e.g., as in Krishnamurthy and Vissing-Jorgensen (2012)). The second critique, which Kareken and Wallace call *implicit theorizing* comprises a *Lucas critique*: a utility function with money balances as an argument is unlikely to be invariant to the monetary policy changes that the model is being used to analyze. This second critique is what we refer to as the *Kareken-Wallace-Lucas critique*.

Specifically, with an input bundle $[y_t(i)]_{i\in[0,1]}$, this firm can produce a quantity

$$y_{1t} = G\left([y_t(i)]_{i\in[0,1]}\right) \equiv \left(\int_0^1 y_t(i)^{\frac{\varepsilon-1}{\varepsilon}}\, di\right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{30}$$

of final good 1, where $\varepsilon \in (1,\infty)$. Each type $i \in [0,1]$ of intermediate good is produced by a monopolistically competitive firm that has access to a linear technology to transform each unit of labor of type 1 into one unit of the intermediate good of type i.

The agent's problem is

$$\max_{\{\boldsymbol{c}_t,\boldsymbol{h}_t,m_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U\left(\boldsymbol{c}_t, \boldsymbol{h}_t, \frac{m_t}{P_{1t}}\right) \tag{31}$$

$$\text{s.t. } P_{1t}c_{1t} + P_{2t}c_{2t} + m_{t+1} = w_{1t}h_{1t} + P_{2t}h_{2t} + m_t + \Pi_{1t} + T_t$$

and $0 \leq m_{t+1}$. The notation is as follows: $\beta \in (0,1)$ is the discount factor; U is the utility function (specified below); c_{jt} is the period-t consumption of the final good of type $j \in \{1,2\}$, and $\boldsymbol{c}_t \equiv (c_{1t}, c_{2t})$; $h_t(i)$ is the period-t supply of labor of type 1 to the firm that produces the intermediate good of type $i \in [0,1]$, $h_{1t} \equiv \int_0^1 h_t(i)di$, h_{2t} is the period-t supply of labor of type 2, and $\boldsymbol{h}_t \equiv (h_{1t}, h_{2t})$; m_t is the agent's nominal money holding at the beginning of period t; P_{jt} is the nominal price of final good $j \in \{1,2\}$ in period t; w_{1t} is the nominal wage for labor of type 1; $\Pi_t(i)$ is the period-t nominal profit from the firm that produces the intermediate good of type $i \in [0,1]$, and $\Pi_{1t} \equiv \int_0^1 \Pi_t(i)di$. The money supply, $\{M_t\}_{t=0}^{\infty}$, follows the same process as in Section 3, implemented via lump-sum transfers $T_t = M_{t+1} - M_t$ to the agents. The problem of the firm that produces the final good of type 1 is

$$\max_{[y_t(i)]_{i\in[0,1]}} P_{1t}y_{1t} - \int_0^1 p_t(i)\, y_t(i)\, di \text{ s.t. } (30), \tag{32}$$

where $p_t(i)$ denotes the nominal price of the intermediate good $i \in [0,1]$ in period t. Let $[\mathrm{Y}_t(p_t(i))]_{i\in[0,1]}$ denote the maximizer of (32). The problem of the firm that produces intermediate good $i \in [0,1]$ is

$$\Pi_t(i) = \max_{p_t(i),h_t(i)} \left[p_t(i)\, \mathrm{Y}_t(p_t(i)) - w_{1t}h_t(i)\right] \text{ s.t. } \mathrm{Y}_t(p_t(i)) \leq h_t(i). \tag{33}$$

Notice that (30)-(33) describe a conventional representative-agent economy with money in the utility function and monopolistic competition.

A monetary equilibrium of the reduced-form model with money in the utility function (described by (30)-(33)) can be summarized by a path $\left\{(c_{jt}, h_{jt}, y_{jt}, \mathcal{Z}_{jt})_{j\in\{1,2\}}, \phi_t, \pi_t\right\}_{t=0}^{\infty}$, where

c_{jt} is the consumption of final good $j \in \{1,2\}$, h_{jt} is the labor supply of type $j \in \{1,2\}$, y_{jt} is the production of final good $j \in \{1,2\}$, $\phi_t \equiv \frac{P_{1t}}{P_{2t}}$ is the relative price of final good 1 in terms of final good 2 faced by consumers, $\pi_t \equiv \frac{\Pi_{1t}}{P_{2t}}$ is the real profit (in terms of the final good 2) from the set of firms that produce the intermediate goods, and $\mathcal{Z}_{jt} \equiv \frac{M_t}{P_{jt}}$ is the aggregate real money balance (in terms of final good $j \in \{1,2\}$). In a stationary monetary equilibrium, $c_{jt} = c_j$, $h_{jt} = h_j$, $y_{jt} = y_j$ and $\mathcal{Z}_{jt} = \mathcal{Z}_j$ for $j \in \{1,2\}$, $\phi_t = \phi$, and $\pi_t = \pi$ for all t.

Consider the following specification of preferences:

$$U\left(\boldsymbol{c}_t, \boldsymbol{h}_t, \frac{m_t}{P_{1t}}\right) \equiv u\left(c_{1t}\right) + v\left(c_{2t}\right) + A\ell\left(\frac{m_t}{P_{1t}}\right) - Bh_{1t} - h_{2t}, \tag{34}$$

where the functions u and v are as described in Section 3, $A, B \in \mathbb{R}_{++}$ are regarded as preference parameters, and $\ell : \mathbb{R}_+ \to \mathbb{R}_+$ is an exogenous function with $\ell'' \leq 0 \leq \ell'$ that represents the "convenience yield" the agent gets from holding real money balances. In the appendix (Lemma 7) we show that along the stationary monetary equilibrium of this economy, $\phi = \frac{\varepsilon}{\varepsilon-1}B$, $c_1 = h_1 = y_1 = \mathrm{D}(\phi)$, $c_2 = h_2 = y_2 = x^\star$, $\pi = \frac{1}{\varepsilon}\phi\mathrm{D}(\phi)$, $\mathcal{Z}_1$ satisfies $\iota = \frac{A}{\phi}\ell'\left(\mathcal{Z}_1\right)$, and $\mathcal{Z}_2 = \phi\mathcal{Z}_1$. The money-in-the-utility (MIU) formulation described by (30)-(34) is relevant for our purposes because if we parametrize it by setting

$$\ell\left(\frac{m_t}{P_{1t}}\right) = \log\left(\frac{m_t}{P_{1t}}\right) \tag{35}$$

$$A = \frac{\iota\left(1-\alpha\right)\left(1+\iota\right)}{1+\alpha\theta\iota}\kappa\mathrm{D}\left(\frac{1+\iota}{1+\alpha\theta\iota}\kappa\right) \tag{36}$$

$$B = \frac{1+\left[1-\alpha\left(1-\theta\right)\right]\iota}{1+\alpha\theta\iota}\kappa \tag{37}$$

$$\varepsilon = \frac{1+\iota}{\alpha\left(1-\theta\right)\iota}, \tag{38}$$

with the parameters α, θ, ι, κ, and $\mathrm{D}(\cdot)$ as defined in Section 3, then it implements as a stationary monetary equilibrium the same real allocation as the economy with more explicit microfoundations for money demand that we presented in Section 3.

If the parameters of the MIU formulation satisfy (35)-(38), then: (*i*) $\mathrm{D}(\phi) = \mathrm{D}((1+\iota)\,\varphi^m)$, i.e., consumption and output of good 1 are equal in both economies (and this is also the labor employed in the production of good 1 in both economies); (*ii*) consumption and output of good 2 equal $x^\star$ in both economies (and this is also the labor employed in the production of good 2 in both economies); (*iii*) $\frac{1}{\varepsilon}\phi\mathrm{D}(\phi) = \alpha\left(1-\theta\right)\iota\varphi^m\mathrm{D}((1+\iota)\,\varphi^m)$, i.e., the aggregate real profit (in terms of good 2) earned by the entities with market power is equal in both economies (this profit accrues to the collection of bankers who serve producers in the economy of Section 3, and to the collection of firms that produce intermediate goods in the MIU economy); and (*iv*) $\mathcal{Z}_1 = \ell'^{-1}\left(\frac{\iota\phi}{A}\right) = (1-\alpha)\mathrm{D}((1+\iota)\,\varphi^m) = Z_1$, i.e., aggregate real money balances expressed in terms of good 1 (i.e., the measure of real money balances that enters the reduced-form utility function (34)) are equal in both economies.[26]

This "equivalence" between the micro-founded and the reduced-form models would appear to confirm the view that, instead of modeling the micro details of monetary exchange, there is no significant loss in assuming a utility function U intended to capture the "convenience yield" or "liquidity services" of certain assets. The problem with this view, however, is that it presumes that the Kareken-Wallace-Lucas implicit theorizing critique can be ignored. Specifically, it regards A, B, and ε as *parameters*, i.e., as exogenous and invariant to changes in the policy rate, ι, and market-structure parameters, α and θ. In contrast, according to the underlying microfounded model of Section 3, when viewed through the lens of the MIU representation, changes in the policy rate and market-structure parameters change the shapes of utility function, U, and of the production function, G, through their effects on the convenience-yield factor A, the disutility of labor supply, B, and the elasticity of substitution between intermediate inputs, ε, as indicated by (36)-(38).

This evident Karaken-Wallace-Lucas critique turns out to be a critical shortcoming of the reduced-form approach, especially when used to draw conclusions on the importance of the

[26]When the reduced-form parameter A satisfies (36), aggregate real money balances expressed in terms of good 1 are equal in both economies, i.e., $\mathcal{Z}_1 = Z_1 = (1-\alpha)\mathrm{D}((1+\iota)\,\varphi^m)$, while aggregate real money balances expressed in terms of good 2 satisfy $\mathcal{Z}_2 = \phi\mathcal{Z}_1 = (1+\iota)\,\varphi^m Z_1 > \varphi^m Z_1 = Z_2$. In other words, $\mathcal{Z}_2 = (1+\iota)\,Z_2$. This difference between $\mathcal{Z}_2$ and Z_2 is immaterial for the real allocation (consumption, labor, and output). It is an accounting discrepancy that stems from the fact that in the economy of Section 3, we are calculating real balances in terms of good 1 using the *accounting relative price of good 1*, i.e., $\varphi_t^m \equiv p_{1t}/p_{2t}$, while the consumer makes consumption decisions internalizing the *effective relative price of good 1*, i.e., $(1+\rho_t)\,\varphi_t^m$. This nuance is missing in the MIU economy, where the *accounting* and the *effective* relative prices are conceptually the same because, for pricing purposes (from the consumers' perspective), the producers and the bankers of the richer micro-founded model are implicitly consolidated into a single supply-side entity composed of the firms with market power that produce intermediate goods, along with the downstream competitive firm that produces the final good 1. To see this clearly (and to eliminate the accounting discrepancy if desired), we can define the nominal price of good 1 in the economy of Section 3 as the *effective price* that the consumer uses to make consumption decisions, i.e., $\bar{p}_{1t} \equiv (1+\rho_t)\,p_{1t}$. The corresponding measure of aggregate real money balances expressed in terms of good 1 would then be $\bar{Z}_{1t} \equiv \frac{M_t}{\bar{p}_{1t}} = \frac{Z_{1t}}{1+\rho_t}$, where Z_{1t} denotes the aggregate real balances as defined in Section 5 (using the *acccounting price*, p_{1t}, i.e., $Z_{1t} \equiv \frac{M_t}{p_{1t}}$). The aggregate real money balances expressed in terms of good 2 are still as defined in Section 5, i.e., $Z_{2t} \equiv \frac{M_t}{p_{2t}} = (1+\rho_t)\,\varphi_t^m \bar{Z}_{1t}$. If we replace condition (36) with $A = \frac{\iota(1-\alpha)}{1+\alpha\theta\iota}\kappa\mathrm{D}\left(\frac{1+\iota}{1+\alpha\theta\iota}\kappa\right)$, then in the stationary monetary equilibrium of the MIU economy, we get $\mathcal{Z}_1 = \bar{Z}_1 = \frac{1}{1+\iota}(1-\alpha)\mathrm{D}((1+\iota)\,\varphi^m)$ and $\mathcal{Z}_2 = Z_2 = (1-\alpha)\,\varphi^m\mathrm{D}((1+\iota)\,\varphi^m)$.

medium-of-exchange function of money and its role in the transmission of monetary policy. For a concrete example, consider the limit of the MIU formulation (34) as the parameter A approaches 0. This is a version of a "cashless limit" that is often used to motivate ignoring medium-of-exchange considerations in the New Keynesian textbooks (see, e.g., Woodford (2003) and Galí (2008)). As the parameter A approaches 0, the equilibrium condition $\iota = \frac{A}{\phi}\ell'(\mathcal{Z}_1)$ implies aggregate real balances, $\mathcal{Z}_1$, converge to 0. However, the MIU formulation implies c_1 is determined by $u'(c_1) = \frac{\varepsilon}{\varepsilon-1}B$, and since ε and B are treated as fixed parameters, monetary considerations (i.e., real money balances or the policy rate) have no influence on the real variables (consumption, production, and labor supply). The irrelevance of the medium-of-exchange role of money is as strong as it can be in this particular MIU formulation since equilibrium real variables are invariant to the monetary variables in the cashless limit, but also away from it.[27] In contrast, once the role of money in exchange is explicitly spelled out as in the model of Section 3, one learns that $\phi = (1+\iota)\varphi^m$ is increasing in ι, and therefore c_1 is decreasing in ι, both in the cashless limit and away from it. From (36)-(38), we see that the cashless limit that results from letting α approach 1 in the microfounded model in fact does imply that A approaches 0 as assumed in the New Keynesian literature, for example. But what the reduced-form approach misses is that, through the lens of the equivalent MIU formulation, monetary policy also operates through the reduced-form parameter ε: an increase in the policy rate ι induces a reduction in the elasticity of substitution ε, which increases the markup on the consumption good 1.

To summarize, our cashless limiting results (Section 6) differ from those in Woodford (1998, 2003) or Galí (2008) (Section 7.2) because standard MIU formulations cannot capture the function of money as a constraint on market power. The reason is that the "equivalent" MIU formulation that *is* able to capture the microeconomic interactions in our model, requires making some modeling choices that are rather unorthodox from the standpoint of the reduced-form literature, such as specifying preferences (the disutility of labor supply, B) and production functions (the elasticity of substitution of inputs, ε) that are functions of policy variables (the nominal policy rate, ι) and of parameters that describe the marketstructure of credit-based settlement (the market power of the credit intermediaries, θ).

[27] This "dichotomy" between real variables and monetary considerations is a well known property of the MIU formulation where utility is assumed to be separable in real money balances, which makes it a popular reduced-form specification in the literature that advocates dissociating monetary policy from money.

8 Conclusion

Over the past twenty five years a folk wisdom has developed, according to which abstracting from the role of money in the mechanism of exchange entails no meaningful loss for informing monetary policy. This view hinges on two specific results based on economies where the role of money is not modeled explicitly, but rather, is proxied by assuming money is an argument of a utility function. We have revisited these results in a theory with a more explicit description of the role of money in the exchange process and a more general microstructure of credit and payments. We have found that the folk wisdom that monetary policy analysis can ignore medium-of-exchange considerations without loss in high-velocity economies is only valid for a much more limited set of economic environments than previously recognized: Unless demand is sufficiently inelastic, or the credit and payment microstructure involves intermediaries with no significant market power, changes in the nominal policy rate can have substantial effects on the real allocations through a modified version of the traditional money-demand channel.

Our theoretical finding raises the concern that, contrary to the current folk wisdom, models that abstract from money need not be good approximations to monetary economies, which would make the widespread practice of basing monetary-policy advise on such *onetary* models, incomplete. We hope our theory will stimulate the kind of empirical and quantitative work necessary to allay (or confirm) this concern. We think there is much to be learned by restoring money to policy-oriented research in monetary economics.

References

Afonso, G. and R. Lagos (2015): "Trade Dynamics in the Market for Federal Funds," *Econometrica*, 83, 263–313.

Andolfatto, D. (2020): "Assessing the Impact of Central Bank Digital Currency on Private Banks," *Economic Journal*, forthcoming.

Baumol, W. J. (1982): "Contestable Markets: An Uprising in the Theory of Industry Structure," *The American Economic Review*, 72, 1–15.

Berger, A. N. and T. H. Hannan (1989): "The Price-Concentration Relationship in Banking," *Review of Economics and Statistics*, 71, 291–299.

Bhagwati, J. (1965): "On the Equivalence of Tariffs and Quotas," *Trade, Growth and the Balance of Payments*, 53–67.

Chiu, J., M. Davoodalhosseini, J. Jiang, and Y. Zhu (2020): "Bank Market Power and Central Bank Digital Currency: Theory and Quantitative Assessment," Working Paper.

Cochrane, J. H. (2005): "Money as Stock," *Journal of Monetary Economics*, 52, 501–528.

——— (2021): *The Fiscal Theory of the Price Level*, unpublished manuscript.

Degryse, H. and S. Ongena (2008): "Competition and Regulation in the Banking Sector: A Review of the Empirical Evidence on the Sources of Bank Rents," *Handbook of Financial Intermediation and Banking*, 2008, 483–554.

Drechsler, I., A. Savov, and P. Schnabl (2017): "The Deposits Channel of Monetary Policy," *Quarterly Journal of Economics*, 132, 1819–1876.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005): "Over-the-Counter Markets," *Econometrica*, 73, 1815–1847.

Galí, J. (2008): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press.

Hannan, T. H. and A. N. Berger (1991): "The Rigidity of Prices: Evidence from the Banking Industry," *American Economic Review*, 81, 938–945.

Kareken, J. H. and N. Wallace (1980): *Models of Monetary Economies: proceedings and contributions from participants of a December 1978 conference sponsored by the Federal Reserve Bank of Minneapolis*, Federal reserve bank of Minneapolis.

Krishnamurthy, A. and A. Vissing-Jorgensen (2012): "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 120, 233–267.

Lagos, R. and R. Wright (2005): "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 113, 463–484.

Lagos, R. and S. Zhang (2015): "Monetary Exchange in Over-the-Counter Markets: A Theory of Speculative Bubbles, the Fed Model, and Self-Fulfilling Liquidity Crises," Working Paper 21528, NBER.

——— (2019): "On Money as a Medium of Exchange in Near-Cashless Credit Economies," Working Paper 25803, NBER.

——— (2020): "Turnover Liquidity and the Transmission of Monetary Policy," *American Economic Review*, 110, 1635–1672.

Lucas, R. E. (1980): "Equilibrium in a Pure Currency Economy," *Economic Inquiry*, 18, 203–220.

Lucas, R. E. and N. L. Stokey (1983): "Optimal Fiscal and Monetary Policy in an Economy Without Capital," *Journal of Monetary Economics*, 12, 55–93.

Markusen, J. R. (1981): "Trade and the Gains from Trade with Imperfect Competition," *Journal of International Economics*, 11, 531–551.

Neumark, D. and S. A. Sharpe (1992): "Market Structure and the Nature of Price Rigidity: Evidence from the Market for Consumer Deposits," *Quarterly Journal of Economics*, 107, 657–680.

Rocheteau, G., R. Wright, and C. Zhang (2018): "Corporate Finance and Monetary Policy," *American Economic Review*, 108, 1147–86.

Samuelson, P. A. (1958): "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," *Journal of Political Economy*, 66, 467–482.

Sidrauski, M. (1967): "Inflation and Economic Growth," *Journal of Political Economy*, 75, 796–810.

Wicksell, K. (1936): *Interest and Prices: A Study of the Causes Regulating the Value of Money*, Macmillan and Co., Limited.

Woodford, M. (1998): "Doing Without Money: Controlling Inflation in a Post-Monetary World," *Review of Economic Dynamics*, 1, 173–219.

——— (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press.

Zhu, T. and N. Wallace (2007): "Pairwise Trade and Coexistence of Money and Higher-Return Assets," *Journal of Economic Theory*, 133, 524–535.