The Limits of *onetary Economics*: On Money as a Medium of Exchange in Near-Cashless Credit Economies*

Ricardo Lagos[†]
New York University

Shengxing Zhang[‡]
London School of Economics

May 2019

Abstract

We study the transmission of monetary policy in credit economies where money serves as a medium of exchange. We find that—in contrast to current conventional wisdom in policy-oriented research in monetary economics—the role of money in transactions can be a powerful conduit to asset prices and ultimately, aggregate consumption, investment, output, and welfare. Theoretically, we show that the cashless limit of the monetary equilibrium (as the cash-and-credit economy converges to a pure-credit economy) need not correspond to the equilibrium of the nonmonetary pure-credit economy. Quantitatively, we find that the magnitudes of the responses of prices and allocations to monetary policy in the monetary economy are sizeable—even in the cashless limit. Hence, as tools to assess the effects of monetary policy, monetary models without money are generically poor approximations—even to idealized highly developed credit economies that are able to accommodate a large volume of transactions with arbitrarily small aggregate real money balances.

Keywords: Asset prices; Cashless; Credit; Leverage; Liquidity; Margin; Monetary policy
JEL classification: D83, E52, G12

*This paper has been presented under the title *The Limits of onetary Economics* at Bank of Canada, Berkeley, Chicago, Chinese University of Hong Kong, Federal Reserve Banks of Atlanta, Minneapolis, Philadelphia, and St. Louis, MIT, NYU, Penn, Princeton, Queens College (CUNY), Rochester, Simon Fraser, Stanford, Tsinghua University (PBC School of Finance), UCLA, UNC-Chapel Hill, Universidad Di Tella, and Yale. We thank colleagues at these institutions for their valuable feedback.

[†]Lagos acknowledges support from the C.V. Starr Center for Applied Economics at NYU.

[‡]Zhang acknowledges support from the Centre for Macroeconomics at LSE, and the ESRC-NSFC Grant on *Building Debt Capital Markets in China*.

1 Introduction

A large body of work in macroeconomics rests on the premise that artificial economies without money are well suited to study the effects of 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. What underlies this moneyless approach to monetary economics is the received wisdom that the medium-of-exchange role of money is quantitatively irrelevant in the transmission of monetary policy.

The intuitive argument runs as follows: aggregate real money balances are a small fraction of aggregate real output in modern economies (e.g., inverse velocity of the monetary base tends to be relatively small), so policy-induced changes in real money balances are bound to have very small effects on output. Therefore, the argument goes, the traditional monetarist mechanism whereby changes in monetary policy are transmitted to the economy through changes in real money balances is basically irrelevant, and 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. There are two ways in which these reduced-form models have been used to support the view that medium-of-exchange considerations can be safely ignored.

First, the fact that the monetary equilibrium in these 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 an inverse velocity that is as small as in the data can be well approximated by an economy where real money balances are simply assumed to play no role in monetary transmission—an economy without money, for instance. Second, parametrized versions of these reduced-form models have been used to claim that, for realistic values of inverse velocity, the role of real money balances in monetary transmission is quantitatively insignificant.

In this paper we show that when the trading frictions that make money useful in exchange are modeled explicitly, the medium-of-exchange role of money is a significant and resilient channel for the transmission of monetary policy. Specifically, the two arguments that have been put forward to justify and encourage the use of models without money to study monetary policy are overturned when we replace the reduced-form formulations with more primitive and more general micro foundations. First, we show that along the cashless limit, the prices and allocations in the monetary equilibrium (as the cash-and-credit economy converges to a pure-credit economy) need not converge to the prices and allocations of the economy without money. So as a matter of pure theory, it would be incorrect to regard the economy with no money as an arbitrarily good approximation to an economy where credit has developed sufficiently to render equilibrium aggregate real balances negligible. Second, we show that this discontinuity is quantitatively significant: the effects of monetary policy in the explicit medium-of-exchange economy remain large even as aggregate real balances converge to zero along the cashless limit. The key insight is that along the cashless limit, even as the volume of transactions that involve cash converges to zero, the fact that individuals have the option to trade with cash implies that the vanishing volume of cash transactions feeds back into the prices negotiated in non-cash pure-credit transactions. We show that this logic applies as long as credit-market intermediaries have some market power.

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. Since in their model this outside option is an increasing function of the borrower's real balances, this mechanism offers a theory for the passthrough from the nominal policy interest rate to the real borrowing rate that the money holder has to pay for the loan. Agents in our model encounter some trading situations that exhibit a similar mechanism, but these particular trading situations are not the relevant ones for our main results. What prevents the medium-of-exchange transmission mechanism from dissipating in the near-cashless economies we study, is the fact that money affects the terms of trade in transactions that do not involve money. The reason is that the *option* to engage in monetary exchange improves the bargaining position of a trader when negotiating with a financial intermediary even though neither the trader nor the intermediary holds money nor wishes to hold money.

Another angle on our results, is that the option to engage in monetary trade disciplines the market power that financial intermediaries can exert over their clients—even in trades that do not involve money. Along a cashless limit, the aggregate volume of monetary trade vanishes, but the individual off-equilibrium option to engage in monetary trade remains available and credible. Therefore, it would be incorrect to infer that money cannot matter quantitatively simply based on the observation that it accounts for a small share of transactions. At a general level, our results are an application to monetary economics of the simple idea that credible outside options can drive outcomes even if they are not exercised in equilibrium. This idea is of course 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.[1]

The basic structure 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, 2019b), which in turn adopts some elements from Duffie et al. (2005). The major difference with Lagos and Zhang (2015, 2019b) is that here we allow investors to buy assets on margin. Aside from capturing an important aspect of trade in financial markets, credit is essential to study the role of monetary policy in near-cashless economies, which is the goal of this paper.

The rest of the paper is organized as follows. Section 2 describes the basic model. Equilibrium is characterized in Section 3. Section 4 presents the theoretical results for the cashless limit. Section 5 conducts quantitative theoretical exercises designed to gauge the magnitude and empirical relevance of the medium-of-exchange transmission mechanism. Section 6 places

[1] 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 mark-up 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.

our contribution in the context of the existing literature on monetary economics without money. Section 7 concludes. Appendix A develops two extensions of the theory. The first proposes an asset-pricing based channel through which the pure medium-of-exchange mechanism transmits monetary policy to aggregate consumption, investment, and output. The second explores the robustness of the main theoretical results to alternative credit arrangements. Appendix B studies efficiency and welfare. Appendix C presents additional theoretical results on the aggregate effects of monetary policy. Appendix D contains all proofs. Appendix E contains robustness checks for the quantitative results.

2 Model

2.1 Environment

Time is represented by a sequence of periods indexed by $t = 0, 1, ...$ Each time period is divided into two subperiods where different activities take place. There is a continuum of infinitely lived *investors*, each identified with a point in the set $\mathcal{I} = [0, N_I]$, with $N_I \in \mathbb{R}_+$, and a continuum of infinitely lived *brokers*, each identified with a point in the set $\mathcal{B} = [0, N_B]$, with $N_B \in \mathbb{R}_+$.

There is a continuum of production units with measure $A^s \in \mathbb{R}_{++}$ that are active every period. Every active unit yields an exogenous *dividend* $y_t \in \mathbb{R}_+$ of a perishable consumption good at the end of the first subperiod of period t. (Each active unit yields the same dividend as every other active unit, so $y_t A^s$ is the aggregate dividend.) At the beginning of every period, every active unit is subject to an independent idiosyncratic shock that renders it permanently unproductive with probability $1 - \eta \in [0, 1)$. If a production unit remains active, its dividend in period t is $y_t = \gamma_t y_{t-1}$ where γ_t is a nonnegative random variable with cumulative distribution function Γ, i.e., $\Pr(\gamma_t \leq \gamma) = \Gamma(\gamma)$, and mean $\bar{\gamma} \in (0, (\beta\eta)^{-1})$. The time-$t$ dividend becomes known to all agents at the beginning of period t, and at that time each failed production unit is replaced by a new unit that yields dividend y_t in the initial period and follows the same stochastic process as other active units thereafter (the dividend of the initial set of production units, $y_0 \in \mathbb{R}_{++}$, is given at $t = 0$). In the second subperiod of every period, every agent has access to a linear production technology that transforms the agent's effort into a perishable homogeneous consumption good (the *general good*).

For each active production unit there is a durable and perfectly divisible *equity share* that represents the bearer's ownership of the production unit and confers the right to collect dividends. At the beginning of every period $t \geq 1$, each investor receives an endowment of $(1-\eta)\, A^s$ equity shares corresponding to the new production units. (When a production unit fails, its equity share disappears.) There is a second financial instrument, money, that is intrinsically useless (it is not an argument of any utility or production function, and unlike equity, ownership of money does not constitute a right to collect any resources). The quantity of money at time t is denoted A_t^m. The initial quantity of money, $A_0^m \in \mathbb{R}_{++}$, is given and $A_{t+1}^m = \mu A_t^m$, with $\mu \in \mathbb{R}_{++}$. A monetary authority injects or withdraws money via lump-sum transfers or taxes to investors in the second subperiod of every period. At the beginning of period $t = 0$, each investor is endowed with a portfolio of equity shares and money.

As is standard in monetary theory, in order to preserve a meaningful role for money, we assume investors are anonymous and unable to commit. If in addition there were also complete lack of enforcement, investors would be unable to borrow, and would have no alternative but to fund first-subperiod purchases with money. In order to allow for credit in purchases of equity, we incorporate a limited form of enforcement by letting investors issue bonds that are collateralized by the equity shares they own. Specifically, some investors can issue bonds in the first subperiod of t, each representing a claim to one unit of the general good to be delivered in the second subperiod of t. We assume the bond issued in the first subperiod of t is collateralized in the sense that if the debtor defaults when the bond is due (at the beginning of the second subperiod of t), then the bond holder appropriates a fraction $\lambda \in [0, 1]$ of the equity shares that the debtor owns at the time of default.[2]

The market structure is as follows. In the second subperiod, all agents can trade the consumption good produced in that subperiod, equity shares, and money, in a spot Walrasian market. In the first subperiod, the trade of equity shares, money, and collateralized bonds is organized as follows. Two distinct Walrasian markets operate concurrently: a *bond market*, and an *equity market*. All bond brokers have access to the bond market, where they can trade collateralized bonds and money. All investors have access to the equity market, where they can trade equity and money. In addition, a fraction of investors also gains access to the bond market indirectly, by engaging in bilateral trades with brokers whom they meet at random. Specifically, with probability $1-\alpha \in [0, 1]$ an individual investor contacts a broker, in which case the investor is able to trade equity, money, and bonds. With probability α an investor does not contact a broker, in which case the investor only has access to the equity market where he

[2] This credit arrangement is similar to the one in Barro (1976) and Kiyotaki and Moore (1997, 2005).

can trade equity and money. This meeting process is independent and identical across investors and over time. Once a broker and an investor have made contact, the pair negotiates the quantities of bonds and money that the broker will trade in the corresponding first-subperiod competitive bond market on behalf of the investor, and an intermediation fee for the broker's services. We assume the terms of the trade between an investor and a broker are determined by Nash bargaining, where an investor has bargaining power $\theta \in [0,1]$.[3] The timing is that first-subperiod trading ends before production units yield dividends. Hence, equity is traded *cum dividend* in the first subperiod and *ex dividend* in the second subperiod.[4]

The preferences of an individual agent of type $j \in \{B, I\}$ (where "B" denotes "broker" and "I" denotes "investor") are represented by

$$\mathbb{E}_0^j \sum_{t=0}^{\infty} \beta^t \left(\varepsilon_t y_t \mathbb{I}_{\{j=I\}} + c_t - h_t \right),$$

where $\beta \in (0,1)$ is the discount factor, c_t is consumption of the homogeneous good that is produced, traded, and consumed in the second subperiod of period t, and h_t is the utility cost from exerting h_t units of effort to produce this good. The variable y_t is the quantity of the dividend good that an investor consumes at the end of the first subperiod of period t, and ε_t denotes the realization of a valuation shock that is distributed independently over time and across investors, with a differentiable cumulative distribution function G on the support $[\varepsilon_L, \varepsilon_H] \subseteq [0, \infty]$, and $\bar{\varepsilon} = \int \varepsilon dG(\varepsilon)$. Each investor learns the realization ε_t at the beginning of period t. The indicator function $\mathbb{I}_{\{j=I\}}$, which equals 1 if $j = I$ or 0 if $j = B$, reflects the assumption that brokers get no utility from the dividend good.[5] The expectation operator $\mathbb{E}_0^B$ is with respect to the probability measure induced by the dividend process and the broker's random trading process in the first subperiod. The expectation operator $\mathbb{E}_0^I$ is with respect to the probability measure induced by the dividend process, the investor's valuation shock, and the investor's random trading process in the first subperiod.

[3]The assumption that investors have direct access to the competitive equity market is made for analytical simplicity. It amounts to regarding the equity market as a conventional Walrasian market where investors bear zero transaction costs, which is a good approximation to organized exchanges such as the New York Stock Exchange. We have studied the role of search and bargaining frictions in equity markets in previous work (see, e.g., Lagos and Zhang (2015, 2019b)), so here we focus on the role that these frictions play in the credit market.

[4]As in conventional search models of financial over-the-counter markets (e.g., Duffie et al. (2005) and Lagos and Rocheteau (2009)), an investor must own the equity share in order to consume the dividend.

[5]This assumption implies that brokers have no direct motive for holding equity. It is easy to relax, but we adopt it because it is the standard benchmark in the search-based literature on over-the-counter markets (see Duffie et al. (2005), Lagos and Rocheteau (2009), and Lagos et al. (2011)).

2.2 Discussion of institutional background and modeling assumptions

In modern financial markets, brokers and dealers allow certain investors to purchase eligible securities *on margin*. "Margin" is an extension of credit from a broker-dealer to an investor using the investor's own securities as collateral. Funds borrowed on margin may be used for any purpose, including the purchase of securities. In particular, funds borrowed on margin are often used for the purchase of the same securities being pledged as collateral. Interest on the borrowed funds accrues over the period of time that the loan is outstanding. The use of margin is regulated by financial regulatory organizations, certain securities exchanges, and the broker-dealer holding the margin account. Buying on margin is standard practice among sophisticated investors. Hedge funds, for example, typically work with *prime brokers* that offer a package of services, including securities lending, global custody, and financing in the form of margin loans.[6] Retail brokerage companies offer small or less sophisticated investors cash brokerage accounts that require all securities purchases to be funded with cash. However, they often also offer investors who meet certain additional requirements, margin accounts that allow to borrow against eligible securities.[7]

The typical margin loan works as follows. A broker extends an investor a loan of L dollars in order to purchase A dollars worth of an asset, e.g., a stock. The investor's own capital or *equity* in this transaction is $E = A - L$. The stock is pledged as collateral to secure the loan, and E is also known as the *haircut* (or *down payment*). The investor's equity, E, expressed as a proportion of the value of the stock, A, is the *margin* on the loan, i.e., $\mathcal{M} = E/A$. In other words, the *margin* is the proportion of the value of the purchase of a security financed by the investor's own funds. The investor's *leverage* is the reciprocal of the margin, i.e., $\mathcal{L} = A/E$, and the *loan-to-value ratio* is $\mathcal{R} = L/A$.

Margin accounts are subject to the rules of the Federal Reserve Board, the Financial Industry Regulatory Authority (FINRA), and securities exchanges such as the New York Stock Exchange, as well as the brokerage firm's margin policies, which are sometimes more stringent than those of the regulators. The Federal Reserve Board Regulation T that regulates the extension of credit by securities brokers and dealers in the United States, specifies a minimum *initial margin requirement* of 50% for new, or initial, stock purchases. (The Federal Reserve has the authority

[6]Prime brokers are large investment banks or securities firms. The top list includes Bank of America Meryll Lynch, Barclays, Credit Suisse, Deutsche Bank, Goldman Sacks, J.P. Morgan, Morgan Stanley, and UBS.

[7]Examples of retail brokers that offer margin accounts to small investors include Charles Schwab Corporation, E-Trade, Fidelity Investments, Scottrade, TD Ameritrade, TradeStation, and USAA Brokerage.

to change this initial margin requirement, but it has been constant since 1974.) The rules of FINRA and the exchanges supplement the requirements of Regulation T by placing additional *maintenance margin requirements* on customer accounts. In particular, FINRA Rule 4210 requires that a customer maintains a minimum margin of 25% at all times. Under these rules, if the customer's margin falls below the minimum maintenance margin, which may be set by some brokers to be higher than 25%, the customer may be required to deposit more funds in the margin account to meet the minimum maintenance margin requirement. This is referred to as a *margin call*. Failure to meet the margin call may cause the broker to liquidate the securities in the customer's account in order to bring the account back up to the required maintenance margin level.

The loans that investors receive by issuing collateralized bonds in the first subperiod, are the theoretical counterparts of margin loans in actual financial markets. Suppose, as is the case in the model, that the investor can borrow up to a fraction λ of the value of the stock, and chooses to borrow that amount, i.e., $L = \lambda A$. In this case, the margin is $\mathcal{M} = 1 - \lambda$, leverage is $\mathcal{L} = (1 - \lambda)^{-1}$, and the loan-to-value ratio is $\mathcal{R} = \lambda$. In the model, the investor has no need to roll over the margin loan, so $\mathcal{M} = 1 - \lambda$ can be interpreted as the *initial margin requirement*, which is constant and typically similar across brokerage firms and eligible stocks.[8] In the model, as in actual markets, investors who want to buy stocks and are able to obtain margin loans from brokers can take on leverage, while those who cannot get margin loans must rely on their own funds.

We model margin loans as bilateral agreements with terms that are negotiated between the parties, as is the case for large investors in real-world markets. In the model, an investor may trade equity with no access to a margin loan (with probability α), or may trade equity with access to margin loans (with probability $1 - \alpha$). This is a simple way to model the multitude of circumstances investors may face in actual financial markets. For example, $\alpha > 0$ captures the fact that on a given day some investors may be unable to secure (mutually acceptable terms for) margin loans from brokers. The market structure in the first subperiod represents a prototypical over-the-counter (OTC) trading arrangement that involves finding a suitable counterparty and then negotiating the terms of the trade. This OTC structure is quite general in that it nests several perfectly competitive benchmarks as special cases. For example, $\alpha = 1$ corresponds to an

[8]Since in the theory the investor repays the loan at the end of the period when it was issued, there is no role for maintenance margin requirements or margin calls.

equity market that is competitive and frictionless (i.e., with no search or bargaining frictions), where all trades must be paid for with money. Alternatively, $1-\alpha=\theta=1$ corresponds to competitive, frictionless, and fully integrated equity and loan markets.

2.3 Bargaining and portfolio problems

Investors may find themselves in one of two trading situations in the OTC round of trade. With probability α, an investor is only able to trade equity and money competitively in the equity market. With probability $1-\alpha$, an investor simultaneously trades money and equity in the equity market, and bargains with a broker over the quantities of bonds and money that the broker trades in the bond market on behalf of the investor, as well as over the broker's intermediation fee. In this case, the outcome of the negotiation with the broker is determined by Nash bargaining with investor bargaining power θ. The broker's fee is expressed in terms of the general good and paid by the investor in the second subperiod.[9] To simplify the exposition, we assume brokers are mere matchmakers and cannot hold assets (equity, money, or bonds) for their own account.[10] An individual investor's portfolio at the beginning of period t is represented by a vector $\boldsymbol{a}_t=(a_t^m,a_t^s)\in\mathbb{R}_+^2$, i.e., it consists of $a_t^m\in\mathbb{R}_+$ units of money and $a_t^s\in\mathbb{R}_+$ equity shares.[11]

Let $W_t(\boldsymbol{a}_t,a_t^b,k_t)$ denote the maximum expected discounted payoff at the beginning of the second subperiod of period t of an investor who is holding portfolio $(\boldsymbol{a}_t,a_t^b)$ and has to pay a fee k_t. Consider an investor who enters period t with (pre-trade) portfolio $\boldsymbol{a}_t=(a_t^m,a_t^s)$ and valuation ε. With probability α, the investor is only able to trade in the equity market and his post-trade portfolio is $\hat{\boldsymbol{a}}_t(\boldsymbol{a}_t,\varepsilon)=(\hat{a}_t^m(\boldsymbol{a}_t,\varepsilon),\hat{a}_t^s(\boldsymbol{a}_t,\varepsilon),\hat{a}_t^b(\boldsymbol{a}_t,\varepsilon))$, with $\hat{a}_t^b(\boldsymbol{a}_t,\varepsilon)=0$ and

$$(\hat{a}_t^m(\boldsymbol{a}_t,\varepsilon),\hat{a}_t^s(\boldsymbol{a}_t,\varepsilon))=\arg\max_{(\hat{a}_t^m,\hat{a}_t^s)\in\mathbb{R}_+^2}\varepsilon y_t\hat{a}_t^s+W_t\left(\hat{a}_t^m,\hat{a}_t^s,\mathbf{0}\right) \tag{1}$$
$$\text{s.t. } \hat{a}_t^m+p_t\hat{a}_t^s\leq a_t^m+p_ta_t^s,$$

where $\mathbf{0}$ represents $(0,0)$. In this case the investor's bond holding is fixed at its beginning-of-period value of 0 since he has no access to the bond market.

[9] In related work (Lagos and Zhang (2015)), we instead assume that the investor must pay the intermediation fee on the spot, i.e., with money or equity. The formulation we use here simplfies the analysis, and the economic mechanisms of interest are essentially unchanged.

[10] The assumption that brokers cannot take long or short positions in the bond, or carry money or equity overnight, is without loss. In the working paper version (Lagos and Zhang (2019a)) we show that bond brokers choose not to trade for their own account when allowed to do so.

[11] Since bonds issued in the OTC round of period $t-1$ are settled in the second subperiod of $t-1$, there are no bonds outstanding at the beginning of period t.

With probability $1-\alpha$, the investor can simultaneously trade in the equity market and bargain over a bond-market trade with a broker. In this case, the bargaining outcome consists of an intermediation fee to the broker, $k_t(\boldsymbol{a}_t, \varepsilon)$, and a post-trade portfolio, $\overline{\boldsymbol{a}}_t(\boldsymbol{a}_t, \varepsilon) = (\overline{a}_t^m(\boldsymbol{a}_t, \varepsilon), \overline{a}_t^s(\boldsymbol{a}_t, \varepsilon), \overline{a}_t^b(\boldsymbol{a}_t, \varepsilon))$, that solves

$$\max_{(\overline{\boldsymbol{a}}_t, k_t)\in\mathbb{R}_+^2\times\mathbb{R}\times\mathbb{R}_+} \left[\varepsilon y_t \overline{a}_t^s + W_t(\overline{\boldsymbol{a}}_t, k_t) - \varepsilon y_t \hat{a}_t^s(\boldsymbol{a}_t, \varepsilon) - W_t\left(\hat{a}_t^m(\boldsymbol{a}_t, \varepsilon), \hat{a}_t^s(\boldsymbol{a}_t, \varepsilon), \mathbf{0}\right)\right]^\theta k_t^{1-\theta} \tag{2}$$
$$\text{s.t. } \overline{a}_t^m + p_t \overline{a}_t^s + q_t \overline{a}_t^b \leq a_t^m + p_t a_t^s$$
$$\varepsilon y_t \hat{a}_t^s(\boldsymbol{a}_t, \varepsilon) + W_t\left(\hat{a}_t^m(\boldsymbol{a}_t, \varepsilon), \hat{a}_t^s(\boldsymbol{a}_t, \varepsilon), \mathbf{0}\right) \leq \varepsilon y_t \overline{a}_t^s + W_t(\overline{\boldsymbol{a}}_t, k_t)$$
$$W_t\left(\overline{a}_t^m, (1-\lambda)\,\overline{a}_t^s, 0, k_t\right) \leq W_t(\overline{\boldsymbol{a}}_t, k_t),$$

where $\overline{\boldsymbol{a}}_t = \left(\overline{a}_t^m, \overline{a}_t^s, \overline{a}_t^b\right)$. Notice that if the investor and the broker were unable to reach an agreement, the investor can still trade equity and cash in the equity market at the terms specified by (1). Hence, the outcome (1) is the investor's outside option in his bargaining problem with the broker. The investor's gain from trade corresponding to an outcome $(\overline{\boldsymbol{a}}_t, k_t)$ consists of the continuation payoff $\varepsilon y_t \overline{a}_t^s + W_t(\overline{\boldsymbol{a}}_t, k_t)$, minus the investor's outside option, $\varepsilon y_t \hat{a}_t^s(\boldsymbol{a}_t, \varepsilon) + W_t\left(\hat{a}_t^m(\boldsymbol{a}_t, \varepsilon), \hat{a}_t^s(\boldsymbol{a}_t, \varepsilon), \mathbf{0}\right)$, namely the payoff the investor achieves in (1). The first constraint on (2) is the budget constraint the investor faces in the OTC round when he is able to trade simultaneously in the equity and the bond market. The second constraint ensures the trade is incentive compatible for the investor, and the restriction $k_t \in \mathbb{R}_+$ ensures the trade is incentive compatible for the broker. The third constraint ensures the investor will prefer to repay in the following subperiod any debt he may have issued in the previous OTC round, rather than default and forfeit a fraction λ of his post-trade equity holding.

Let $V_t(\boldsymbol{a}_t, \varepsilon)$ denote the maximum expected discounted payoff of an investor with valuation ε and portfolio $\boldsymbol{a}_t \equiv (a_t^m, a_t^s)$ at the beginning of the OTC round of period t. Define $\boldsymbol{\phi}_t \equiv (\phi_t^m, \phi_t^s)$, where ϕ_t^m denotes the real price of money, and ϕ_t^s the real *ex dividend* price of equity in the second subperiod of period t (both expressed in terms of the general good). Then,

$$W_t(\boldsymbol{a}_t, a_t^b, k_t) = \max_{(c_t, h_t, \tilde{\boldsymbol{a}}_{t+1})\in\mathbb{R}_+^4} \left[c_t - h_t + \beta\mathbb{E}_t \int V_{t+1}\left(\boldsymbol{a}_{t+1}, \varepsilon\right) dG(\varepsilon)\right] \tag{3}$$
$$\text{s.t. } c_t + \boldsymbol{\phi}_t \tilde{\boldsymbol{a}}_{t+1} \leq h_t + \boldsymbol{\phi}_t \boldsymbol{a}_t + a_t^b - k_t + T_t,$$

where $\mathbb{E}_t$ is the conditional expectation over the next-period realization of the dividend, $\tilde{\boldsymbol{a}}_{t+1} \equiv (\tilde{a}_{t+1}^m, \tilde{a}_{t+1}^s)$, $\boldsymbol{a}_{t+1} = (\tilde{a}_{t+1}^m, \eta\tilde{a}_{t+1}^s + (1-\eta)\,A^s)$, $\boldsymbol{\phi}_t \boldsymbol{a}_t$ denotes the dot product of $\boldsymbol{\phi}_t$ and $\boldsymbol{a}_t$, and $T_t \in \mathbb{R}$ is the real value of the time-t lump-sum monetary transfer. Similarly, let $W_t^B(k_t)$ denote the maximum expected discounted payoff at the beginning of the second subperiod of period t of a broker who has earned fee k_t. Then,

$$W_t^B(k_t) = k_t + \beta \mathbb{E}_t V_{t+1}^B, \tag{4}$$

where V_{t+1}^B denotes the maximum expected discounted payoff of a broker at the beginning of the OTC round of period $t+1$.

The value function of an investor who enters the OTC round of period t with portfolio $\boldsymbol{a}_t$ and valuation ε is

$$\begin{aligned} V_t(\boldsymbol{a}_t, \varepsilon) = & \ \alpha \left[\varepsilon y_t \hat{a}_t^s(\boldsymbol{a}_t, \varepsilon) + W_t\left(\hat{\boldsymbol{a}}_t(\boldsymbol{a}_t, \varepsilon), 0\right)\right] \\ & + (1-\alpha)\left[\varepsilon y_t \overline{a}_t^s(\boldsymbol{a}_t, \varepsilon) + W_t\left(\overline{\boldsymbol{a}}_t(\boldsymbol{a}_t, \varepsilon), k_t(\boldsymbol{a}_t, \varepsilon)\right)\right]. \end{aligned} \tag{5}$$

The value function of a broker who enters the OTC round of period t is

$$V_t^B = \alpha^B \int W_t^B\left[k_t(\tilde{\boldsymbol{a}}_t, \varepsilon)\right] dH_t(\tilde{\boldsymbol{a}}_t, \varepsilon) + \left(1-\alpha^B\right) W_t^B(0), \tag{6}$$

where $\alpha^B \in (0,1)$ is the probability a broker contacts an investor, and H_t is the joint cumulative distribution function over the portfolios and valuations of the investors that a broker may contact in the OTC round of trade of period t.

3 Equilibrium

Let A_{It}^m and A_{It}^s denote the quantities of money and equity shares, respectively, held by investors at the beginning of the OTC round of period t (after depreciated production units have been replaced). That is, $A_{It}^m = N_I \int a_t^m dF_t(\boldsymbol{a}_t)$ and $A_{It}^s = N_I \int a_t^s dF_t(\boldsymbol{a}_t)$, where F_t is the cumulative distribution function over portfolios $\boldsymbol{a}_t = (a_t^m, a_t^s)$ held by investors at the beginning of the OTC round of period t. Let $\tilde{A}_{It+1}^m$ and $\tilde{A}_{It+1}^s$ denote the total quantities of money and shares held by all investors at the end of period t, i.e., $\tilde{A}_{It+1}^k = \int_{\mathcal{I}} \tilde{a}_{it+1}^k di$ for $k \in \{m, s\}$, where $\tilde{a}_{it+1}^k$ denotes the quantity of asset k held at the end of period t by the individual investor identified with the point $i \in \mathcal{I}$. Thus, $A_{It+1}^m = \tilde{A}_{It+1}^m$ and $A_{It+1}^s = \eta \tilde{A}_{It+1}^s + (1-\eta) A^s$. For asset $k \in \{b, m, s\}$, let $\bar{A}_{It}^k = (1-\alpha) N_I \int \overline{a}_t^k(\boldsymbol{a}_t, \varepsilon) dH_t(\boldsymbol{a}_t, \varepsilon)$, and $\hat{A}_{It}^k = \alpha N_I \int \hat{a}_t^k(\boldsymbol{a}_t, \varepsilon) dH_t(\boldsymbol{a}_t, \varepsilon)$, with $\hat{a}_t^b(\boldsymbol{a}_t, \varepsilon) = 0$. We are now ready to define equilibrium.

Definition 1 *An equilibrium is a sequence of prices, $\{p_t, q_t, \phi_t^m, \phi_t^s\}_{t=0}^\infty$, portfolio allocations and fees in the OTC market, $\{\hat{\boldsymbol{a}}_t(\cdot), \bar{\boldsymbol{a}}_t(\cdot), k_t(\cdot)\}_{t=0}^\infty$, and end-of-day portfolios, $\{\tilde{\boldsymbol{a}}_{t+1}\}_{t=0}^\infty$, such* *that for all t: (i) the portfolios and fees in the OTC market solve (1) and (2); (ii) taking prices and the bargaining protocol as given, the end-of period portfolios solve (3); and (iii) prices are such that all Walrasian markets clear, i.e.,* $\tilde{A}^s_{It+1} = A^s$ *(the end-of-period t Walrasian market for equity clears),* $\tilde{A}^m_{It+1} = A^m_{t+1}$ *(the end-of-period t Walrasian market for money clears),* $\bar{A}^b_{It} = 0$ *(the period t OTC market for bonds clears),* $\hat{A}^s_{It} + \bar{A}^s_{It} = A^s$ *(the period t OTC market for equity clears), and* $(\hat{A}^m_{It} + \bar{A}^m_{It} - A^m_t)\mathbb{I}_{\{\phi^m_t>0\}} = 0$ *(the money market clears in the OTC round of of period t). An equilibrium is "monetary" if* $\phi^m_t > 0$ *for all t.*

The first step toward characterizing equilibrium is to find the bargaining outcomes. For any $(y, z) \in \mathbb{R}^2$, it is convenient to define the "indicator correspondence" $\chi : \mathbb{R}^2 \to [0, 1]$ by

$$\chi(y,z) \begin{cases} = 1 & \text{if } y < z \\ \in [0,1] & \text{if } y = z \\ = 0 & \text{if } z < y. \end{cases} \tag{7}$$

The following lemma characterizes post-trade portfolios in the OTC market for an economy with no money.

Lemma 1 *Consider the economy with no money, and let*

$$\varepsilon^n_t \equiv \frac{\bar{\phi}^s_t - \phi^s_t}{y_t}, \tag{8}$$

where $\bar{\phi}^s_t$ *denotes the price of an equity share in terms of bonds. Consider an investor who enters the OTC round of period t with equity holding* a^s_t *and valuation* ε*. Then:*

(i) If the investor does not contact a broker, the post-trade equity holding is $\hat{a}^s_t(a^s_t) = a^s_t$*.*

(ii) If the investor contacts a broker, the bargaining problem has a solution only if

$$\lambda < \frac{\bar{\phi}^s_t}{\phi^s_t}, \tag{9}$$

the post-trade portfolio is

$$\overline{a}^s_t(a^s_t, \varepsilon) = \chi(\varepsilon^n_t, \varepsilon)\frac{\bar{\phi}^s_t}{\bar{\phi}^s_t - \lambda\phi^s_t}a^s_t \tag{10}$$

$$\overline{a}^b_t(a^s_t, \varepsilon) = \bar{\phi}^s_t\left[1 - \chi(\varepsilon^n_t, \varepsilon)\frac{\bar{\phi}^s_t}{\bar{\phi}^s_t - \lambda\phi^s_t}\right]a^s_t, \tag{11}$$

and the intermediation fee for the broker is

$$k_t(a^s_t, \varepsilon) = (1-\theta)(\varepsilon - \varepsilon^n_t)\,y_t\left[\chi(\varepsilon^n_t, \varepsilon)\frac{\bar{\phi}^s_t}{\bar{\phi}^s_t - \lambda\phi^s_t} - 1\right]a^s_t. \tag{12}$$

Part (i) of Lemma 1 states that in a nonmonetary economy, investors who do not contact a broker are unable to trade in the OTC round.[12] Part (ii) states that an investor who contacts a broker can buy or sell equity and take a long or short position in bonds. From (10) and (11), if $\varepsilon < \varepsilon_t^n$, the investor sells all his equity for bonds. Conversely, if $\varepsilon_t^n < \varepsilon$, the investor shorts the bond in order to take a long position in equity. Condition (12) indicates the broker earns a fee on these transactions as long as $\varepsilon \neq \varepsilon_t^n$.

Notice the nonmonetary benchmark is capable of supporting trade in the OTC round, but only among investors with access to credit. The reason is that although only equity can be traded in the equity market, and only bonds can be traded in the bond market, investors with access to both markets can trade in both markets *simultaneously*, which implies these investors are actually able to exchange these securities at a price of $\bar{\phi}_t^s$ bonds per equity share.[13] Since each bond is a claim to 1 unit of the second-supberiod consumption good, we can think of $\bar{\phi}_t^s$ as the real cum dividend price in the OTC round of an equity share, expressed in terms of the second-subperiod consumption good. In the nonmonetary economy there is an implied real interest rate on bonds (expressed in terms of equity shares), denoted i_t^n. In the OTC round, an investor with access to credit can use 1 unit of equity to purchase $\bar{\phi}_t^s$ bonds. These bonds deliver $\bar{\phi}_t^s$ general goods in the following subperiod, when the relative price of general goods in terms of equity shares is $1/\phi_t^s$. Thus, $i_t^n \equiv \bar{\phi}_t^s/\phi_t^s - 1$, or using (8),

$$i_t^n = \frac{\varepsilon_t^n y_t}{\phi_t^s}. \tag{13}$$

Notice that with (13), (9) can be written as $\lambda < 1 + i_t^n$, so (9) could only be violated if the net

[12]In a nonmonetary economy, investors with no access to the credit market cannot buy equity because they have no way to pay sellers, and cannot sell equity because they have no way to collect payment from buyers.

[13]Our market structure implies that investors with access to the equity market can trade equity *and* money, while investors with access to the bond market can trade bonds *and* money. This market structure is different from a conventional cash-in-advance formulation, e.g., in the spirit of Lucas (1980), where agents with access to the equity market would be restricted to trading equity *for* money, and agents with access to the bond market would be restricted to trading bonds *for* money. The fact that the nonmonetary equilibrium supports some trade in equity and bonds makes it clear that our formulation does not require agents to hold cash (in advance or otherwise) to be able to trade equity or bonds. In contrast, absent money, trade in equity or bonds would be impossible in a conventional cash-in-advance formulation, which by assumption would require every equity and bond purchase to be paid in cash. Neither does the payment structure in our model fit the predetermined cash-good/credit-good taxonomy assumed in Lucas and Stokey (1983). In our model, at a given point in time, equity may resemble a "cash good" to investors who do not contact a credit broker, but a "credit good" to investors who are simultaneously in contact with the equity market and a credit broker. Similarly, over time, for a given investor, equity may resemble a "cash good" at times when the investor only contacts the equity market, or a "credit good" at times when the investor also contacts a credit broker simultaneously. Another way in which our model is different from a standard cash-in-advance formulation is that, as shown below, transaction velocity is endogenous (it changes with the monetary policy stance), and can exceed 1 in a model period.

real interest were negative, which will not be the case in equilibrium.[14]

The following lemma characterizes equilibrium post-trade portfolios in the OTC market for an economy with money.

Lemma 2 *Consider the economy with money, and let*

$$\varepsilon_t^* \equiv (p_t\phi_t^m - \phi_t^s)\frac{1}{y_t} \tag{14}$$

$$\varepsilon_t^{**} \equiv \varepsilon_t^* + (1 - q_t\phi_t^m)\left[\mathbb{I}_{\{q_t\phi_t^m<1\}}\frac{p_t}{q_t} + \mathbb{I}_{\{1<q_t\phi_t^m\}}\lambda\phi_t^s\right]\frac{1}{y_t}. \tag{15}$$

Consider an investor who enters the OTC round of period t with portfolio $\boldsymbol{a}_t$ and valuation ε. Then:

(i) If the investor does not contact a broker, the post-trade portfolio is

$$\hat{a}_t^m(\boldsymbol{a}_t,\varepsilon) = [1 - \chi(\varepsilon_t^*,\varepsilon)](a_t^m + p_t a_t^s) \tag{16}$$

$$\hat{a}_t^s(\boldsymbol{a}_t,\varepsilon) = \chi(\varepsilon_t^*,\varepsilon)\frac{1}{p_t}(a_t^m + p_t a_t^s). \tag{17}$$

(ii) If the investor contacts a broker, the bargaining problem has a solution only if

$$\lambda < \frac{p_t}{q_t\phi_t^s}, \tag{18}$$

and in that case the post-trade portfolio is

$$\begin{aligned}\overline{a}_t^m(\boldsymbol{a}_t,\varepsilon) = {} & \left\{\mathbb{I}_{\{1<q_t\phi_t^m\}}[1-\chi(\varepsilon_t^{**},\varepsilon)] + \mathbb{I}_{\{q_t\phi_t^m=1\}}\mathbb{I}_{\{\varepsilon<\varepsilon_t^{**}\}}[1-\chi(q_t\phi_t^m,1)]\right\}(a_t^m + p_t a_t^s) \\ & + \mathbb{I}_{\{q_t\phi_t^m=1\}}\mathbb{I}_{\{\varepsilon=\varepsilon_t^{**}\}}\tilde{a}_t^m\end{aligned} \tag{19}$$

$$\begin{aligned}\overline{a}_t^s(\boldsymbol{a}_t,\varepsilon) = {} & \left\{\mathbb{I}_{\{q_t\phi_t^m=1\}}\mathbb{I}_{\{\varepsilon_t^{**}<\varepsilon\}} + \left[1-\mathbb{I}_{\{q_t\phi_t^m=1\}}\right]\chi(\varepsilon_t^{**},\varepsilon)\right\}\frac{a_t^m + p_t a_t^s}{p_t - \lambda q_t\phi_t^s} \\ & + \mathbb{I}_{\{q_t\phi_t^m=1\}}\mathbb{I}_{\{\varepsilon=\varepsilon_t^{**}\}}\tilde{a}_t^s\end{aligned} \tag{20}$$

$$\overline{a}_t^b(\boldsymbol{a}_t,\varepsilon) = -\frac{1}{q_t}\left\{[\overline{a}_t^m(\boldsymbol{a}_t,\varepsilon) - a_t^m] + p_t[\overline{a}_t^s(\boldsymbol{a}_t,\varepsilon) - a_t^s]\right\}, \tag{21}$$

where

$$(\tilde{a}_t^m, \tilde{a}_t^s) \in \left\{\mathbb{R}_+^2 : \tilde{a}_t^m + (p_t - q_t\lambda\phi_t^s)\tilde{a}_t^s = a_t^m + p_t a_t^s\right\},$$

[14] To see why (9) is necessary for the bargaining outcome to be well defined, consider the budget constraint and the collateral constraint of an investor who contacts a broker, namely $\bar{\phi}_t^s \overline{a}_t^s + \overline{a}_t^b = \bar{\phi}_t^s a_t^s$, and $-\lambda\phi_t^s\overline{a}_t^s \leq \overline{a}_t^b$. These two conditions imply the borrowing constraint $-\lambda\phi_t^s a_t^s \leq (1 - \lambda\phi_t^s/\bar{\phi}_t^s)\,\overline{a}_t^b$. This constraint would be slack for all $\overline{a}_t^b < 0$ if (9) were violated, meaning that an investor with $\varepsilon > \varepsilon_t^n$ would be able (and willing) to take an infinitely long position in the stock. Intuitively, if (9) were violated, then an investor who starts with no wealth can sell b collateralized bonds to purchase $b/\bar{\phi}_t^s$ equity shares and this leveraged purchase would leave the investor's borrowing constraint slack, since $b < \lambda\phi_t^s b/\bar{\phi}_t^s$.

and the intermediation fee is

$$k_t(\boldsymbol{a}_t, \varepsilon) = (1-\theta)\Big\{(\varepsilon y_t + \phi_t^s)\left[\overline{a}_t^s(\boldsymbol{a}_t,\varepsilon) - \hat{a}_t^s(\boldsymbol{a}_t,\varepsilon)\right] + \phi_t^m\left[\overline{a}_t^m(\boldsymbol{a}_t,\varepsilon) - \hat{a}_t^m(\boldsymbol{a}_t,\varepsilon)\right] + \overline{a}_t^b(\boldsymbol{a}_t,\varepsilon)\Big\}. \tag{22}$$

Part (i) of Lemma 2 states that the individual investor with valuation ε who does not contact a credit broker in the OTC round of trade, uses all his money balances to buy equity if $\varepsilon_t^* < \varepsilon$, and sells all his equity for money if $\varepsilon < \varepsilon_t^*$. Part (ii) describes the bargaining outcome and post-trade portfolio of an investor who contacts a credit broker, and can therefore simultaneously trade bonds and equity. To offer a simple interpretation of the post-trade allocation in this case, suppose $q_t\phi_t^m < 1$ (which will be the case if the equilibrium interest rate on the bond is positive). In this case, if $\varepsilon_t^{**} < \varepsilon$, the investor short sells as much of the bond as allowed by the collateral constraint, and uses the proceeds from the short sale along with all his pre-trade money balances to buy equity. Conversely, if $\varepsilon < \varepsilon_t^{**}$, the investor sells all his pre-trade equity holding and uses the proceeds from the sale, and all his pre-trade money balances, to buy bonds. The broker extracts a fee whenever the investor has positive gain from trade.

For what follows, it is useful to think in terms of the interest rate implied by the inside bond. First, notice that with 1 unit of money an investor can buy $\frac{1}{q_t}$ bonds, which in total yield $\frac{1}{q_t}$ general goods in the following subperiod, and this is equivalent to $\frac{1}{q_t\phi_t^m}$ dollars. Thus, the competitive nominal rate on a collateralized loan is

$$i_t^m \equiv \frac{1}{q_t\phi_t^m} - 1. \tag{23}$$

Since the loan is repaid within the period, this is also a notion of real interest rate on these loans, with loan and repayment expressed in terms of the general good.[15] Another notion of interest rate, one that corresponds to the interest rate we defined for the economy with no money, is the real interest rate on bonds expressed in terms of equity shares. If an investor uses 1 unit of money to purchase bonds, this is equivalent to exchanging $\frac{1}{p_t}$ worth of equity shares for $\frac{1}{q_t}$ bonds that deliver $\frac{1}{q_t}$ general goods in the following subperiod, and this is equivalent to $\frac{1}{q_t\phi_t^s}$ equity shares. Hence, the real rate on a bond (in terms of equity shares) is

$$i_t^s \equiv \frac{p_t}{q_t\phi_t^s} - 1. \tag{24}$$

[15] A bond investment of $\frac{1}{\phi_t^m}$ dollars is equivalent to a bond investment of 1 unit of the general good. The $\frac{1}{\phi_t^m}$ dollars allow to buy $\frac{1}{q_t\phi_t^m}$ bonds, which in total yield $\frac{1}{q_t\phi_t^m}$ general goods. So the gross real interest in terms of general goods is also $\frac{1}{q_t\phi_t^m}$.

With (24), (18) can be written as $\lambda < 1 + i_t^s$, so (18) could only be violated if the net real interest were negative, which will not be the case in equilibrium.[16]

Notice i_t^m is the interest rate an investor would face if he had direct access to the competitive bond market. However, access to the bond market is intermediated by a broker who charges a fee $k_t(\boldsymbol{a}_t, \varepsilon)$ to an investor with pre-trade portfolio $\boldsymbol{a}_t$ and valuation ε who wishes to trade $\overline{a}_t^b(\boldsymbol{a}_t, \varepsilon)$ bonds. Thus, the fee the investor pays per unit of bond traded is $\frac{k_t(\boldsymbol{a}_t,\varepsilon)}{|\overline{a}_t^b(\boldsymbol{a}_t,\varepsilon)|}$ and the effective interest rate at which the investor can borrow (if $\varepsilon_t^{**} < \varepsilon$) or lend (if $\varepsilon < \varepsilon_t^{**}$) is $i_t^m - \frac{k_t(\boldsymbol{a}_t,\varepsilon)}{\overline{a}_t^b(\boldsymbol{a}_t,\varepsilon)}$, which using Lemma 2 can be written as

$$i_t^l(\varepsilon) = i_t^m + (1-\theta)\frac{(\varepsilon - \varepsilon_t^{**})\, y_t}{\varepsilon_t^{**} y_t + \phi_t^s}. \tag{25}$$

We focus the analysis on equilibria with an active bond market, which implies $q_t\phi_t^m \leq 1$ or equivalently, that the net nominal interest rate on bonds, i_t^m, is nonnegative.[17] In this case, (15) implies $\varepsilon_t^{**} y_t = (p_t/q_t - \phi_t^s)$ and notice that $(\varepsilon_t^{**} - \varepsilon_t^*)\, y_t = (1/q_t - \phi_t^m)\, p_t \geq 0$.

Hereafter, we consider recursive (stationary) equilibria, i.e., equilibria in which real equity prices (and real money balances, if the equilibrium is monetary) are time-invariant linear functions of the aggregate dividend.

Definition 2 *A recursive nonmonetary equilibrium (RNE) is a nonmonetary equilibrium in which real equity prices (general goods per equity share) are time-invariant linear functions of the aggregate dividend, i.e.,* $\phi_t^s = \phi^s y_t$, *and* $\bar{\phi}_t^s = \bar{\phi}^s y_t$ *for some* $\phi^s, \bar{\phi}^s \in \mathbb{R}_+$.

In a RNE, $\varepsilon_t^n = \left(\bar{\phi}_t^s - \phi_t^s\right)\frac{1}{y_t} = \bar{\phi}^s - \phi^s \equiv \varepsilon^n$, and the real interest rate on the bond, i.e., $i_t^n \equiv \bar{\phi}_t^s/\phi_t^s - 1$ as defined in (13), is

$$i^n = \frac{\varepsilon^n}{\phi^s}. \tag{26}$$

[16] To see why condition (18) is necessary for the bargaining outcome of an investor with access to credit to be well defined, consider the budget constraint and the collateral constraint of an investor who contacts a broker, namely $\overline{a}_t^m + p_t\overline{a}_t^s + q_t\overline{a}_t^b = a_t^m + p_t a_t^s$ and $-\lambda\phi_t^s\overline{a}_t^s \leq \overline{a}_t^b$. These two conditions imply the borrowing constraint $-\lambda\phi_t^s\left[a_t^s + (a_t^m - \overline{a}_t^m)/p_t\right] \leq (1 - \lambda q_t\phi_t^s/p_t)\,\overline{a}_t^b$. This constraint would be slack for all $\overline{a}_t^b < 0$ if (18) were violated, meaning that an investor with $\varepsilon > \varepsilon_t^{**}$ would be able (and willing) to take an infinitely long position in the stock. Intuitively, notice that if (18) is violated, then an investor with access to credit who starts with no wealth, can sell b collateralized bonds for money and use the monetary proceeds to purchase bq_t/p_t equity shares; this leveraged purchase would leave the investor's borrowing constraint slack, since $b < \lambda\phi_t^s bq_t/p_t$.

[17] Lemma 10 in Appendix D establishes that $q_t\phi_t^m \leq 1$ is necessary for bonds to trade in equilibrium. We focus on equilibria with an active bond market because if the bond is not traded, money is the only means of payment and the equilibrium conditions specialize to those in Lagos and Zhang (2019b).

Definition 3 *A recursive monetary equilibrium (RME) is a monetary equilibrium in which: (i) real equity prices (general goods per equity share) are time-invariant linear functions of the aggregate dividend, i.e.,* $\phi_t^s = \phi^s y_t$, $p_t \phi_t^m \equiv \bar{\phi}_{mt}^s = \bar{\phi}_m^s y_t$, *and* $p_t/q_t \equiv \bar{\phi}_{bt}^s = \bar{\phi}_b^s y_t$ *for some* $\phi^s, \bar{\phi}_m^s, \bar{\phi}_b^s \in \mathbb{R}_+$*; and (ii) real money balances are a constant proportion of output, i.e.,* $\phi_t^m A_t^m = Z A^s y_t$ *for some* $Z \in \mathbb{R}_{++}$.

In a RME, $\varepsilon_t^* = (p_t \phi_t^m - \phi_t^s) \frac{1}{y_t} = \bar{\phi}_m^s - \phi^s \equiv \varepsilon^*$, $\varepsilon_t^{**} = (p_t/q_t - \phi_t^s) \frac{1}{y_t} = \bar{\phi}_b^s - \phi^s \equiv \varepsilon^{**}$, $p_t = \frac{(\varepsilon^* + \phi^s) A_t^m}{Z A^s}$, $\phi_t^m = \frac{Z A^s y_t}{A_t^m}$, and

$$q_t = \frac{(\varepsilon^* + \phi^s)\, A_t^m}{(\varepsilon^{**} + \phi^s)\, Z A^s y_t}. \tag{27}$$

Thus, $\phi_{t+1}^s/\phi_t^s = \bar{\phi}_{mt+1}^s/\bar{\phi}_{mt}^s = \bar{\phi}_{bt+1}^s/\bar{\phi}_{bt}^s = \gamma_{t+1}$, $p_{t+1}/p_t = \mu$, and $\phi_t^m/\phi_{t+1}^m = q_{t+1}/q_t = \mu/\gamma_{t+1}$. In a RME, the interest i_t^m defined in (23) becomes

$$i^m = \frac{\varepsilon^{**} - \varepsilon^*}{\varepsilon^* + \phi^s} \tag{28}$$

and the interest rate i_t^s defined in (24) becomes

$$i^s = \frac{\varepsilon^{**}}{\phi^s}. \tag{29}$$

Notice that in a RME, (18) becomes $\lambda < 1 + i^s$, which is satisfied since $\lambda \in [0,1]$. Also, in a RME the interest rate $i_t^l(\varepsilon)$ defined in (25) becomes

$$i^l(\varepsilon) = i^m + (1-\theta)\frac{\varepsilon - \varepsilon^{**}}{\varepsilon^{**} + \phi^s}. \tag{30}$$

Let $q_{t,k}^B$ denote the nominal price in the second subperiod of period t of an N-period risk-free pure discount nominal bond that matures in period $t+k$, for $k = 0, 1, 2, ..., N$ (so k is the number of periods until the bond matures). Imagine the bond is illiquid in the sense that it cannot be traded in the OTC market. Then, in a stationary monetary equilibrium, $q_{t,k}^B = (\bar{\beta}/\mu)^k$, and

$$i^p = \frac{\mu - \bar{\beta}}{\bar{\beta}} \tag{31}$$

is the time-t nominal yield to maturity of the bond with k periods until maturity. Throughout the analysis we let $\bar{\beta} \equiv \beta\bar{\gamma}$ and maintain the assumption $\mu > \bar{\beta}$ (but we consider the limiting case $\mu \to \bar{\beta}$). Since there is a one-to-one mapping between the policy variable μ and the interest rate i^p, we can regard i^p as the nominal *policy rate* chosen by the monetary authority.

Hereafter, we consider a formulation of the model where the length of the time period is arbitrarily short, which allows us to deliver sharp theoretical results. This can be interpreted as an approximation to a continuous-time version of our discrete-time economy. To this end, we first generalize the discrete-time model by allowing the period length to be an arbitrary constant, and then take the limit as this constant becomes arbitrarily small. Let Δ denote the length of the model period, and define the discount rate, r, the expected dividend growth rate, g, the depreciation rate, δ, and the money growth rate, π, as $\beta \equiv (1+r\Delta)^{-1}$, $\bar{\gamma} \equiv 1+g\Delta$, $\eta \equiv 1-\delta\Delta$, $\mu \equiv 1+\pi\Delta$. Over a time period of length Δ, the dividend is $y_t\Delta$, and utility from consumption of the dividend good is $\varepsilon y_t \Delta$. We focus on recursive equilibria where, as $\Delta \to 0$, real asset prices are time-invariant linear functions of the *dividend rate*, y_t. Specifically, let $\Phi_t^s(\Delta)$ and $\Phi_t^m(\Delta) A_t^m$ denote the real equity price and the real aggregate money balance, respectively, in the discrete-time economy with time periods of length Δ. We look for recursive equilibria of this discrete-time economy such that $\Phi_t^s(\Delta) = \Phi^s(\Delta) y_t \Delta$ and $\Phi_t^m(\Delta) A_t^m = \mathrm{Z}(\Delta) A^s y_t \Delta$, where $\Phi^s(\Delta)$ and $\mathrm{Z}(\Delta)$ are time-invariant functions with the property that $\lim_{\Delta\to 0} \Phi^s(\Delta)\Delta = \phi^s$ and $\lim_{\Delta\to 0} \mathrm{Z}(\Delta)\Delta = Z$, with $\phi^s, Z \in \mathbb{R}$. Hence (26), (28), (29), (30), and (31) generalize to $i^n = \frac{\varepsilon^n}{\Phi^s(\Delta)}$, $i^m = \frac{\varepsilon^{**}-\varepsilon^*}{\varepsilon^*+\Phi^s(\Delta)}$, $i^s = \frac{\varepsilon^{**}}{\Phi^s(\Delta)}$, $i^l(\varepsilon) = i^m + (1-\theta)\frac{\varepsilon-\varepsilon^{**}}{\varepsilon^{**}+\Phi^s(\Delta)}$, and $i^p = \frac{(r+\pi-g+r\pi\Delta)\Delta}{1+g\Delta}$, respectively.[18]

Define

$$\rho^n \equiv \lim_{\Delta\mapsto 0} \frac{i^n}{\Delta} = \frac{\varepsilon^n}{\phi^s}$$
$$\rho^m \equiv \lim_{\Delta\mapsto 0} \frac{i^m}{\Delta} = \frac{\varepsilon^{**}-\varepsilon^*}{\phi^s}$$
$$\rho^s \equiv \lim_{\Delta\mapsto 0} \frac{i^s}{\Delta} = \frac{\varepsilon^{**}}{\phi^s}$$
$$\rho^l(\varepsilon) \equiv \lim_{\Delta\mapsto 0} \frac{i^l(\varepsilon)}{\Delta} = \rho^m + (1-\theta)\frac{\varepsilon-\varepsilon^{**}}{\phi^s}$$
$$\rho^p \equiv \lim_{\Delta\mapsto 0} \frac{i^p}{\Delta} = r+\pi-g.$$

Intuitively, ρ^n is the competitive real interest rate on inside bonds in a nonmonetary economy, ρ^m and ρ^s are the competitive nominal and real interest rates on inside bonds in a monetary economy, $\rho^l(\varepsilon)$ is the intermediated nominal interest rate available to an investor with valuation ε who has access to credit in a monetary economy, and ρ^p is the policy nominal interest rate

[18]The discrete-time formulation we laid out previously, corresponds to a special case of this formulation with $\Delta = 1$, $\Phi_t^s(1) \equiv \phi_t^s$, $\Phi_t^m(1) \equiv \phi_t^m$, $\Phi^s(1) \equiv \phi^s$, and $\mathrm{Z}(1) \equiv Z$.

that can be controlled by the monetary authority (e.g., by changing π). Notice that $\rho^p = r + \bar{\pi}$ is a Fisher equation that equates the nominal rate interest rate (on an illiquid outside bond) to the real risk-free interest rate, r, plus an expected inflation rate, $\bar{\pi} \equiv \pi - g$ (measured with the price of the general consumption good).[19] For what follows, it is useful to let $\mathcal{Z} \equiv \rho Z$, $\iota \equiv \rho^p/\rho$, and $\varphi \equiv \rho\phi^s$ in a nonmonetary, or $\varphi^n \equiv \rho\phi^s$ in a nonmonetary economy, where

$$\rho \equiv \lim_{\Delta \mapsto 0} \frac{1}{\Delta} \frac{1 - \bar{\beta}\eta}{\bar{\beta}\eta} = r + \delta - g.$$

The factor ρ can be interpreted as a *capitalization rate* for equity holdings, so φ (or φ^n) and $\mathcal{Z}$ are the "flow values" of an equity share and a unit of real balances, respectively, and ι is the policy rate up to a convenient normalization. In the remainder, we focus on the limiting economy that obtains as $\Delta \to 0$.

Proposition 1 *There exists a unique recursive nonmonetary equilibrium,* $(\varepsilon^n, \varphi^n)$*. Moreover,*

$$\varphi^n = \bar{\varepsilon} + (1 - \alpha)\,\theta \left[\int_{\varepsilon_L}^{\varepsilon^n} (\varepsilon^n - \varepsilon)\, dG(\varepsilon) + \frac{\lambda}{1 - \lambda} \int_{\varepsilon^n}^{\varepsilon_H} (\varepsilon - \varepsilon^n)\, dG(\varepsilon) \right] \tag{32}$$

and $\varepsilon^n \in [\varepsilon_L, \varepsilon_H]$ *is the unique solution to*

$$G(\varepsilon^n) = \lambda. \tag{33}$$

The asset price in the nonmonetary equilibrium can be decomposed into three components. The first term in (32) is the expected value of the dividend flow. The second term in (32) is the expected gain from exercising the option of reselling the asset in the OTC market. The third term in (32) reflects the expected marginal value of the asset when it is pledged as collateral for shorting the bond. We will label these components the *(value of the) resale option*, and the *(value of the) pledge option*, respectively. From (33), it is clear the marginal investor valuation that clears the OTC market, ε^n, is increasing in λ, reflecting the fact that as λ increases, the collateral constraint is relaxed, and higher valuation investors are able to absorb a higher

[19] The (gross) inflation rate in terms of general goods is $\phi_t^m/\phi_{t+1}^m = \mu \frac{y_t}{y_{t+1}} \equiv 1 + \tilde{\pi}_{t+1}$, which is stochastic. A measure of this average (expected) inflation is

$$\left[\mathbb{E}_t \frac{1}{1 + \tilde{\pi}_{t+1}} \right]^{-1} = \frac{\mu}{\bar{\gamma}} = \frac{1 + \pi\Delta}{1 + g\Delta} \equiv 1 + \bar{\pi}\Delta,$$

so as $\Delta \to 0$, we get $\bar{\pi} = \pi - g$ as a measure of the expected inflation rate in the nominal price of general goods. Since $p_{t+1}/p_t - 1 = \mu - 1 = \pi\Delta$, the inflation rate in the nominal price of equity shares is π.

proportion of the asset holdings. In particular, $\varepsilon^n \to \varepsilon_H$ as $\lambda \to 1$. For a given ε^n, the asset price is increasing in λ, but in the general equilibrium, a larger λ implies a larger ε^n, which in turn implies an investor is less likely to receive a valuation shock large enough to want to use the asset as collateral, and this force tends to make the price decreasing in λ. It is possible to show, however, that the former effect always dominates, and φ^n is increasing in λ even after taking into account the general equilibrium effect (see Lemma 14 in Appendix D).

Let

$$\bar{\iota}(\lambda) \equiv \frac{[\alpha + (1-\alpha)(1-\theta)](\bar{\varepsilon} - \varepsilon_L) + (1-\alpha)\theta\left[\varepsilon^n - \varepsilon_L + \frac{1}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon - \varepsilon^n)\,dG(\varepsilon)\right]}{\bar{\varepsilon} + (1-\alpha)\theta\left[\int_{\varepsilon_L}^{\varepsilon^n}(\varepsilon^n - \varepsilon)\,dG(\varepsilon) + \frac{\lambda}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon - \varepsilon^n)\,dG(\varepsilon)\right]} \tag{34}$$

$$\hat{\iota}(\lambda) \equiv \frac{\left[\alpha + (1-\alpha)\left(1 + \theta\frac{\lambda}{1-\lambda}\right)\right]\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon - \varepsilon^n)\,dG(\varepsilon)}{\bar{\varepsilon} + \int_{\varepsilon_L}^{\varepsilon^n}(\varepsilon^n - \varepsilon)\,dG(\varepsilon) + (1-\alpha)\theta\frac{\lambda}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon - \varepsilon^n)\,dG(\varepsilon)}. \tag{35}$$

The following proposition offers a complete characterization of the set of RME.

Proposition 2 *(i) If $\hat{\iota}(\lambda) < \iota < \bar{\iota}(\lambda)$ then there exists a unique recursive monetary equilibrium, $(\varepsilon^*, \varepsilon^{**}, \varphi, \mathcal{Z})$. The asset prices are*

$$\varphi = \varphi^n + [\alpha + (1-\alpha)(1-\theta)]\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^* - \varepsilon)\,dG(\varepsilon) \tag{36}$$

$$\mathcal{Z} = \frac{\alpha G(\varepsilon^*)}{[1 - G(\varepsilon^*)]\alpha + 1 - \alpha}\varphi. \tag{37}$$

The marginal valuations are $\varepsilon^{**} = \varepsilon^n$ and the unique $\varepsilon^* \in (\varepsilon_L, \varepsilon^n)$ that satisfies

$$\frac{[\alpha + (1-\alpha)(1-\theta)]\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon - \varepsilon^*)\,dG(\varepsilon) + (1-\alpha)\theta\left[\varepsilon^n - \varepsilon^* + \frac{1}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon - \varepsilon^n)\,dG(\varepsilon)\right]}{\varphi^n + [\alpha + (1-\alpha)(1-\theta)]\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^* - \varepsilon)\,dG(\varepsilon)} = \iota.$$

(ii) If $0 < \iota \leq \hat{\iota}(\lambda)$ then there exists a unique recursive monetary equilibrium, $(\varepsilon^*, \chi, \varphi, \mathcal{Z})$. The asset prices are

$$\varphi = \bar{\varepsilon} + \int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^* - \varepsilon)\,dG(\varepsilon) + (1-\alpha)\theta\frac{\lambda}{1-\lambda}\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon - \varepsilon^*)\,dG(\varepsilon)$$

$$\mathcal{Z} = \frac{\alpha G(\varepsilon^*) + (1-\alpha)\frac{1}{1-\lambda}[G(\varepsilon^*) - \lambda]}{[1 - G(\varepsilon^*)]\left[\alpha + (1-\alpha)\frac{1}{1-\lambda}\right]}\varphi.$$

The marginal valuations are $\varepsilon^{**} = \varepsilon^*$, where $\varepsilon^* \in [\varepsilon^n, \varepsilon_H)$ (with $\varepsilon^* = \varepsilon^n$ only if $\iota = \hat{\iota}(\lambda)$) is the unique solution to

$$\frac{\left[\alpha + (1-\alpha)\left(1 + \theta\frac{\lambda}{1-\lambda}\right)\right]\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon - \varepsilon^*)\,dG(\varepsilon)}{\bar{\varepsilon} + \int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^* - \varepsilon)\,dG(\varepsilon) + (1-\alpha)\,\theta\frac{\lambda}{1-\lambda}\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon - \varepsilon^*)\,dG(\varepsilon)} = \iota$$

and

$$\chi = \frac{\lambda}{1-\lambda}\frac{1 - G(\varepsilon^*)}{G(\varepsilon^*)}$$

is the proportion of the financial wealth that investors with access to credit with valuation lower than ε^* hold in the form of bonds (they hold the remaining $1-\chi$ fraction in cash).

Figure 1 illustrates the existence regions in the space of parameters ι (vertical axis) and λ (horizontal axis). The boundaries $\iota = \bar{\iota}(\lambda)$ and $\iota = \hat{\iota}(\lambda)$ define three regions.[20] As shown in Proposition 1, a nonmonetary equilibrium exists for any parametrization, and therefore for every configuration of parameters shown in Figure 1. For each $\lambda \in [0,1]$, $\bar{\iota}(\lambda)$ is the largest nominal policy rate consistent with existence of monetary equilibrium. No monetary equilibrium exists for $\iota \geq \bar{\iota}(\lambda)$, so in that case the nonmonetary equilibrium is the unique equilibrium. For every $\iota < \bar{\iota}(\lambda)$, the nonmonetary equilibrium coexists with a monetary equilibrium. For each $\lambda \in (0,1]$, the monetary equilibrium is qualitatively different depending on whether the policy rate is higher or lower than $\hat{\iota}(\lambda)$.

If $\hat{\iota}(\lambda) < \iota < \bar{\iota}(\lambda)$, then credit is scarce in the sense that the nominal interest rate on margin loans, ρ^m, is positive. In this case, investors without access to credit who have valuations higher than ε^* use all their money to take a long position in equity, and investors with access to credit who have valuations higher than ε^{**} use all their money and short the bond up to the collateral constraint to purchase equity. Investors without access to credit who have valuations lower than ε^* sell all their equity holdings for money, and investors with access to credit who have valuations lower than ε^{**} sell all their equity holdings and use all the proceeds along with their pre-trade money holdings to take a long position in bonds. Notice that in this parameter range, $\varepsilon_L < \varepsilon^* < \varepsilon^{**} = \varepsilon^n$, so the marginal equity buyer without access to credit has a lower valuation

[20]It is easy to prove that $\hat{\iota}(\lambda) \leq \bar{\iota}(\lambda)$ for all $\lambda \in [0,1]$ (with "=" only if $\lambda = 0$), and that $\hat{\iota}(0) = \bar{\iota}(0) = \frac{\bar{\varepsilon} - \varepsilon_L}{\bar{\varepsilon}}$, $\hat{\iota}(1) = 0$, and

$$\bar{\iota}(1) = \frac{(1-\alpha)\,\theta\,(\varepsilon_H - \varepsilon_L) + [\alpha + (1-\alpha)(1-\theta)](\bar{\varepsilon} - \varepsilon_L)}{\bar{\varepsilon} + (1-\alpha)\,\theta\,(\varepsilon_H - \bar{\varepsilon})}.$$

Hence, $\hat{\iota}(1) < \bar{\iota}(0) = \hat{\iota}(0) \leq \bar{\iota}(1)$. Also, $\bar{\iota}(1) - \bar{\iota}(0) = \frac{(1-\alpha)\theta(\varepsilon_H - \bar{\varepsilon})\varepsilon_L}{[\bar{\varepsilon} + (1-\alpha)\theta(\varepsilon_H - \bar{\varepsilon})]\bar{\varepsilon}}$, so $0 \leq \bar{\iota}(1) - \bar{\iota}(0)$ holds with "=" only if $(1-\alpha)\,\theta = 0$. In particular, notice $\theta = 0$ (or $\alpha = 1$) implies $\bar{\iota}(\lambda) = \bar{\iota}(0)$ for all λ.

than the marginal equity buyer with access to credit, and the latter has the same valuation he would have in the nonmonetary equilibrium. The fact that $\varepsilon^* < \varepsilon^{**}$ reflects that investors with access to credit have the option of investing money in the interest-yielding inside bond, while this option is not available to investors without access to credit. Notice that the equity price in this region of the parameter space is larger than the equity price in the nonmonetary equilibrium. The first reason (captured by the term multiplied by α in (36)) is that in a monetary equilibrium there are equity-for-money trades in the equity market so the investor without access to credit can resell equity if her valuation is relatively low, while this is not possible in the nonmonetary equilibrium. The second reason (captured by the term multiplied by $(1-\alpha)(1-\theta)$ in (36)) is that the option of being able to sell equity for money in the equity market improves the outside option of an investor with access to credit when he bargains for the terms of the trade with the broker who helps the investor take a long position in the bond.

Conversely, if $0 < \iota \leq \hat{\iota}(\lambda)$, then real balances are abundant and credit demand is relatively weak, so $\rho^m = 0$. In this case money is not dominated by inside bonds in terms of rate of return. Investors without access to credit who have valuations larger than ε^* use all their money to buy equity while investors without access to credit who have valuations lower than ε^* sell all their equity for money. Investors with access to credit who have valuations higher than ε^* use all their money and short the bond up to the collateral constraint to purchase equity. Investors with access to credit who have valuations lower than ε^* sell all their equity holdings, and are indifferent between holding the proceeds (and any pre-trade money holdings) in the form of money or bonds. Notice that in this parameter range, $\varepsilon_L < \varepsilon^n \leq \varepsilon^* = \varepsilon^{**} < \varepsilon_H$, so the marginal equity buyer without access to credit has the same valuation as the marginal equity buyer with access to credit, and this marginal valuation is higher than the marginal valuation an investor with access to credit would have in the nonmonetary equilibrium. The fact that $\varepsilon^* = \varepsilon^{**}$ reflects that the nominal rate on inside bonds is equal to zero, so investors with access to credit and investors without access to credit have the same valuation for money in the OTC round, even though only the former can lend it in the bond market.

Interestingly, the monetary equilibrium becomes a more robust trading arrangement when the loan-to-value ratio λ increases. To see this, notice that

$$\frac{\partial \bar{\iota}(\lambda)}{\partial \lambda} = \frac{(1-\alpha)\theta\varepsilon_L \int_{\varepsilon^n}^{\varepsilon_H} (\varepsilon - \varepsilon^n)\, dG(\varepsilon)}{\left\{\{1-\lambda[1-(1-\alpha)\theta]\}\bar{\varepsilon} + (1-\alpha)\theta\left[\int_{\varepsilon_L}^{\varepsilon^n} (\varepsilon^n - \varepsilon)\, dG(\varepsilon) - \lambda\hat{\varepsilon}\right]\right\}^2} \geq 0$$

with "=" only if $(1-\alpha)\theta = 0$. In other words, the maximum rate of inflation consistent with monetary equilibrium is increasing with λ. So as long as $0 < (1-\alpha)\,\theta$, money and credit (in the form of leveraged purchases of equity) behave as complements, rather than substitutes, in the general equilibrium. The reason money has value in the monetary equilibrium is that high-valuation investors use it to finance a long position in the equity. So for larger λ, this value of money in exchange is enhanced because the equity shares that an investor with access to credit buys with money can themselves be used as collateral to short the bond and take an even larger long position in the equity.

The following result summarizes the implications of monetary exchange and monetary policy for asset prices.

Proposition 3 *(i) The real asset price in the monetary equilibrium is higher than in the non-monetary equilibrium, i.e.,* $\varphi^n \leq \varphi$ *for all* $\iota \in [0, \bar{\iota}(\lambda)]$, *with "=" only if* $\iota = \bar{\iota}(\lambda)$. *Moreover,* $\varphi \leq \varepsilon_H$, *with "=" only if* $\iota = 0$.

(ii) In a monetary equilibrium, φ *is decreasing in* ι. *If* $\hat{\iota}(\lambda) < \iota < \bar{\iota}(\lambda)$, *then* ε^* *(the valuation of the marginal investor with no access to credit) is decreasing in* ι. *If* $0 < \iota \leq \hat{\iota}(\lambda)$, *then* ε^* *(the valuation of the marginal investor) is decreasing in* ι.

4 Cashless limits

In this section we study the properties of the monetary equilibrium in situations where agents can economize in the use of cash, and in settings where aggregate real cash balances are small relative to the aggregate real value of the stock of financial assets being traded. In other words, we study economies that can be construed as approximations to *pure-credit* or *cashless economies*.[21] To this end, we characterize the limiting equilibrium as either: (*i*) $\lambda \to 1$, (*ii*) $\iota \to \bar{\iota}(\lambda)$, or (*iii*) $\alpha \to 0$. The first limit, studied in Proposition 4, approximates an economy where investors with access to credit can buy assets with zero margin (infinite leverage). Thus, in this case a measure $1-\alpha$ of investors can greatly economize on cash use along the intensive margin of borrowing. The second limit, studied in Proposition 5, operates directly on the opportunity cost of holding money, and approximates an economy where this cost is so high that aggregate real balances become negligible. The third limit, studied in Proposition 6,

[21] We follow the tradition in monetary theory to use "cash" as a synonym of *unbacked money*. Hence, by *cashless economy*, we mean a *moneyless economy* akin to the "pure credit system" envisioned by Wicksell (1936, p. 62-70) that Woodford (2003) uses to motivate his moneyless approach to monetary economics. In lay terms, in contrast, "cash" often refers to currency, so with this alternative terminology, it would be possible for a payment system to be cashless but not moneyless.

approximates an economy where—regardless of the equilibrium value of money—nobody *needs* money to trade in the sense that it is budget feasible for every individual investor to long equity by shorting the bond without ever using money as means of payment. In each of these limits we characterize aggregate real balances, transaction velocity, and the role of monetary policy on trading activity and asset prices.

In a stationary equilibrium, we can define *transaction velocity* of money, $\mathcal{V}$, as the ratio of the nominal value of all transactions to the money supply, i.e.,

$$\mathcal{V} \equiv \left[\alpha G(\varepsilon^*) + (1-\alpha)\, G(\varepsilon^{**})\right] \frac{\varphi}{\mathcal{Z}} \tag{38}$$
$$= \begin{cases} \frac{\{\alpha[1-G(\varepsilon^*)]+1-\alpha\}[\alpha G(\varepsilon^*)+(1-\alpha)\lambda]}{\alpha G(\varepsilon^*)} & \text{if } \hat{\iota}\,(\lambda) < \iota < \bar{\iota}\,(\lambda) \\ \frac{\left(\alpha+\frac{1-\alpha}{1-\lambda}\right)G(\varepsilon^*)[1-G(\varepsilon^*)]}{\alpha G(\varepsilon^*)+\frac{1-\alpha}{1-\lambda}[G(\varepsilon^*)-\lambda]} & \text{if } 0 < \iota \leq \hat{\iota}\,(\lambda)\,. \end{cases}$$

Transaction velocity is a useful measure of the efficiency of cash use, i.e., a measure of the average number of transactions that can be supported per dollar in circulation.[22]

Proposition 4 *Consider an economy with* $\alpha \in (0,1)$. *Let*

$$\varphi^n_{\lambda=1} \equiv \lim_{\lambda\to 1} \varphi^n = \bar{\varepsilon} + (1-\alpha)\,\theta\,(\varepsilon_H - \bar{\varepsilon})\,.$$

As $\lambda \to 1$, $\varepsilon^{**} = \varepsilon^n \to \varepsilon_H$,

$$\frac{\mathcal{Z}}{\varphi} \to \frac{\alpha G(\varepsilon^*)}{[1-G(\varepsilon^*)]\,\alpha + 1 - \alpha} \tag{39}$$

$$\mathcal{V} \to \frac{\{\alpha\,[1-G(\varepsilon^*)] + 1 - \alpha\}\,[\alpha G(\varepsilon^*) + 1 - \alpha]}{\alpha G(\varepsilon^*)} \tag{40}$$

$$\varphi \to \varphi^n_{\lambda=1} + [\alpha + (1-\alpha)(1-\theta)] \int_{\varepsilon_L}^{\varepsilon^*} (\varepsilon^* - \varepsilon)\, dG(\varepsilon)\,, \tag{41}$$

where $\varepsilon^* \in [\varepsilon_L, \varepsilon_H]$ *is the unique solution to*

$$\frac{(1-\alpha)\,\theta\,(\varepsilon_H - \varepsilon^*) + [\alpha + (1-\alpha)(1-\theta)] \int_{\varepsilon^*}^{\varepsilon_H} (\varepsilon - \varepsilon^*)\, dG(\varepsilon)}{\bar{\varepsilon} + [\alpha + (1-\alpha)(1-\theta)] \int_{\varepsilon_L}^{\varepsilon^*} (\varepsilon^* - \varepsilon)\, dG(\varepsilon) + (1-\alpha)\,\theta\,(\varepsilon_H - \bar{\varepsilon})} = \iota. \tag{42}$$

[22]Money in our model may be traded many times in an OTC round. Consider a dollar initially spent by an equity buyer. If the dollar ends up in the hands of an equity seller who has no access to a credit broker, the dollar is not further circulated. But if the dollar ends up in the hands of an equity seller with access to a credit broker, it flows from the seller through the margin loans market to an equity buyer with access to credit who borrows. The dollar, now in the hands of the borrower, will travel further, to a seller and possibly further to another margin-loan borrower, until it eventually ends up, unspent, in the post-OTC portfolio of an equity seller who carries it into the following subperiod (if $q_t\phi^m_t < 1$, this seller is necessarily someone without access to credit, or possibly a seller with access to credit, if $q_t\phi^m_t = 1$). In this sense, we can think of margin loans as a liquidity saving mechanism that helps sustain monetary equilibrium.

Proposition 4 shows that, as $\lambda \to 1$, real balances remain positive and velocity remains bounded as long as $\alpha > 0$ and $\iota < \bar{\iota}(1)$. This is not surprising since money demand in the OTC round is supported by low-valuation investors without access to credit, and money demand in the second subperiod is supported by the probability of being a high-valuation investor without access to credit in the following period. The portfolio problem of an investor with access to credit remains well defined in the OTC round of trade even as λ approaches 1 because the equilibrium real interest rate in that limit becomes very high, i.e., $\rho^s \to \frac{\varepsilon_H}{\varphi}\rho$, which in equilibrium limits investors' desire to short the bond.

Proposition 5 *Consider an economy with $\alpha \in (0,1)$. As $\iota \to \bar{\iota}(\lambda)$, $\varepsilon^{**} = \varepsilon^n$, $\varepsilon^* \to \varepsilon_L$,*

$$\frac{\mathcal{Z}}{\varphi} \to 0 \tag{43}$$

$$\mathcal{V} \to \infty \tag{44}$$

$$\varphi \to \varphi^n. \tag{45}$$

Proposition 5 is a familiar "nonmonetary limit" in micro-founded monetary models: as the opportunity cost of holding money becomes very large, real balances approach zero. In this case, since the real value of the equity purchases of the $1-\alpha$ investors with access to credit remains positive in the limit, the transaction velocity of money diverges to infinity. The transaction velocity corresponding only to investors without access to credit, however, remains bounded. To see this, notice we can decompose $\mathcal{V}$ into two components, i.e., $\mathcal{V} = \hat{\mathcal{V}} + \bar{\mathcal{V}}$, where $\hat{\mathcal{V}} \equiv \alpha G(\varepsilon^*)\frac{\varphi}{\mathcal{Z}}$ and $\bar{\mathcal{V}} \equiv (1-\alpha)\,G(\varepsilon^{**})\frac{\varphi}{\mathcal{Z}}$ are the contributions to velocity of transactions by investors without access to credit, and investors with access to credit, respectively. Explicitly,

$$\hat{\mathcal{V}} = \begin{cases} [1-G(\varepsilon^*)]\,\alpha + 1 - \alpha & \text{if } \hat{\iota}(\lambda) < \iota < \bar{\iota}(\lambda) \\ \frac{\alpha G(\varepsilon^*)[1-G(\varepsilon^*)]\left(\alpha + \frac{1-\alpha}{1-\lambda}\right)}{\alpha G(\varepsilon^*) + \frac{1-\alpha}{1-\lambda}[G(\varepsilon^*)-\lambda]} & \text{if } 0 < \iota \leq \hat{\iota}(\lambda) \end{cases}$$

$$\bar{\mathcal{V}} = \begin{cases} \frac{(1-\alpha)\lambda\{[1-G(\varepsilon^*)]\alpha + 1 - \alpha\}}{\alpha G(\varepsilon^*)} & \text{if } \hat{\iota}(\lambda) < \iota < \bar{\iota}(\lambda) \\ \frac{(1-\alpha)G(\varepsilon^*)[1-G(\varepsilon^*)]\left(\alpha + \frac{1-\alpha}{1-\lambda}\right)}{\alpha G(\varepsilon^*) + \frac{1-\alpha}{1-\lambda}[G(\varepsilon^*)-\lambda]} & \text{if } 0 < \iota \leq \hat{\iota}(\lambda). \end{cases}$$

Hence, as $\iota \to \bar{\iota}(\lambda)$, $\bar{\mathcal{V}} \to \infty$, but $\hat{\mathcal{V}} \to 1$.

For the following result it is useful to let $\bar{\varsigma}(\alpha)$ and $\hat{\varsigma}(\alpha)$ denote the bounds defined in (34) and (35), but regarded as functions of α, i.e.,

$$\bar{\varsigma}(\alpha) \equiv \frac{[\alpha+(1-\alpha)(1-\theta)](\bar{\varepsilon}-\varepsilon_L)+(1-\alpha)\theta\left[\varepsilon^n-\varepsilon_L+\frac{1}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)\right]}{\bar{\varepsilon}+(1-\alpha)\theta\left[\int_{\varepsilon_L}^{\varepsilon^n}(\varepsilon^n-\varepsilon)\,dG(\varepsilon)+\frac{\lambda}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)\right]}$$

$$\hat{\varsigma}(\alpha) \equiv \frac{\left\{\alpha+(1-\alpha)\left[(1-\theta)+\theta\frac{1}{1-\lambda}\right]\right\}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)}{\bar{\varepsilon}+\int_{\varepsilon_L}^{\varepsilon^n}(\varepsilon^n-\varepsilon)\,dG(\varepsilon)+(1-\alpha)\theta\frac{\lambda}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)}.$$

Figure 2 is analogous to Figure 1, but illustrates the existence regions using the boundaries $\bar{\varsigma}(\alpha)$ and $\hat{\varsigma}(\alpha)$ in the space of parameters ι (vertical axis) and $1-\alpha$ (horizontal axis).

Proposition 6 *Consider an economy with $\alpha \in [0,1]$ and $\lambda \in (0,1]$. Let*

$$\tilde{\varphi}^n \equiv \lim_{\alpha\to 0}\varphi^n = \bar{\varepsilon}+\theta\left[\int_{\varepsilon_L}^{\varepsilon^n}(\varepsilon^n-\varepsilon)\,dG(\varepsilon)+\frac{\lambda}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)\right]. \tag{46}$$

As $\alpha \to 0$,

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

$$\frac{\mathcal{Z}}{\varphi} \to 0 \tag{47}$$

$$\mathcal{V} \to \infty \tag{48}$$

$$\varphi \to \tilde{\varphi} \equiv \tilde{\varphi}^n + (1-\theta)\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^*-\varepsilon)\,dG(\varepsilon), \tag{49}$$

where $\varepsilon^* \in (\varepsilon_L, \varepsilon^n)$ is the unique solution to

$$\frac{(1-\theta)\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon-\varepsilon^*)\,dG(\varepsilon)+\theta\left[\varepsilon^n-\varepsilon^*+\frac{1}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)\right]}{\bar{\varepsilon}+(1-\theta)\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^*-\varepsilon)\,dG(\varepsilon)+\theta\left[\int_{\varepsilon_L}^{\varepsilon^n}(\varepsilon^n-\varepsilon)\,dG(\varepsilon)+\frac{\lambda}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)\right]} = \iota. \tag{50}$$

(ii) If $0 < \iota \le \hat{\varsigma}(0)$, then

$$\frac{\mathcal{Z}}{\varphi} \to \frac{G(\varepsilon^*)-\lambda}{1-G(\varepsilon^*)} \tag{51}$$

$$\mathcal{V} \to \frac{G(\varepsilon^*)\left[1-G(\varepsilon^*)\right]}{G(\varepsilon^*)-\lambda} \tag{52}$$

$$\varphi \to \tilde{\varphi} \equiv \bar{\varepsilon}+\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^*-\varepsilon)\,dG(\varepsilon)+\theta\frac{\lambda}{1-\lambda}\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon-\varepsilon^*)\,dG(\varepsilon), \tag{53}$$

where $\varepsilon^* \in [\varepsilon^n, \varepsilon_H)$ is the unique solution to

$$\frac{\left(1-\theta+\theta\frac{1}{1-\lambda}\right)\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon-\varepsilon^*)\,dG(\varepsilon)}{\bar{\varepsilon}+\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^*-\varepsilon)\,dG(\varepsilon)+\theta\frac{\lambda}{1-\lambda}\int_{\varepsilon^*}^{\varepsilon_H}(\varepsilon-\varepsilon^*)\,dG(\varepsilon)} = \iota. \tag{54}$$

Proposition 6 considers the limiting economy as the fraction of investors who do not have access to credit, vanishes. This limiting economy is one where it is budget feasible for virtually every investor to finance equity purchases by shorting the bond, i.e., investors can purchase equity shares even if they come into the period carrying no cash. As explained in the context of Proposition 2, in a monetary equilibrium, investors with no access to credit who have relatively low valuation for equity are the ones who always demand money in the OTC trading round, while investors with access to credit who have relatively low valuation demand money in the OTC round only if $0 < \iota \leq \hat{\varsigma}(0)$. Therefore, the first-order implication of letting α approach zero is that the extensive margin of money demand from investors with no access to credit, i.e., the number of investors with no access to credit who wish to hold money, approaches zero.[23]

If the nominal policy rate is relatively low, i.e., if $0 < \iota \leq \hat{\varsigma}(0)$ as in part (*ii*) of Proposition 6, then the aggregate money demand from investors with no access to credit vanishes in the limit, but the aggregate money demand from investors with access to credit who have low valuation remains positive in the limit. The reason why money can have value in this limiting economy can be understood at least from two perspectives. First, money demand can be positive overnight because the collateral constraint is expected to be binding for investors with access to credit, so money can allow them to take a long position in equity that is larger than the one they would be able to take relying only on the margin loan.[24] Second, low-valuation investors with access to credit are willing to hold cash at the end of the OTC round because the nominal rate on the bond (i.e., the opportunity cost of holding cash in the OTC round) is zero when $0 < \iota \leq \hat{\varsigma}(0)$; this makes them indifferent between holding their wealth in money or bonds, and they hold some of each. As a result, in economies that satisfy $0 < \iota \leq \hat{\varsigma}(0)$, i.e., economies with relatively low inflation and relatively low ability to take on leverage, real balances converge to a positive limit (i.e., (51)), and velocity converges (i.e., (52)) as $\alpha \to 0$.

If the nominal policy rate is relatively high, i.e., if $\hat{\varsigma}(0) < \iota < \bar{\varsigma}(0)$ as in part (*i*) of

[23]The restriction that $\lambda \neq 0$ in the statement of Proposition 6 is because, as shown in Figure 1, $\lambda = 0$ implies $\hat{\varsigma}(\alpha) = \bar{\varsigma}(\alpha)$ for all α, and therefore in that case monetary equilibrium exists only for $\iota = \hat{\varsigma}(\alpha) = \bar{\varsigma}(\alpha)$. From Figure 1 we can also see that $\lambda = 1$ implies $\hat{\varsigma}(\alpha) = 0$ for all α, so low-valuation investors who have access to credit hold money in the OTC round only if $\iota = \hat{\varsigma}(\alpha) = 0$. Thus, if $\lambda = 1$, the range in part (*ii*) of Proposition 6 collapses to a single point, and only the parameter region in part (*i*) is generic in that case.

[24]Notice that in the range $0 < \iota \leq \hat{\varsigma}(0)$, this argument requires $\lambda < 1$. The reason is that from Figure 1 we can see that $\lambda = 1$ implies $\hat{\varsigma}(\alpha) = 0$ for all α, so low-valuation investors with access to credit hold money in the OTC round only if $\iota = \hat{\varsigma}(\alpha) = 0$. In other words, if $\lambda = 1$ then the range in part (*ii*) of Proposition 6 collapses to a single point (money is only held exactly at the Friedman rule), and only the parameter region in part (*i*) is generic in that case.

Proposition 6, then real balances converge to zero (i.e., (47)), and transaction velocity diverges to infinity (i.e., (48)) as $\alpha \to 0$. The immediate reason for this result is that given this high policy rate, aggregate money demand in the OTC round vanishes as $\alpha \to 0$, for two reasons. First, low-valuation investors without access to credit would be willing to hold money, but the aggregate number of investors without access to credit vanishes as $\alpha \to 0$. Second, low-valuation investors with access to credit are unwilling to hold money because money is dominated in rate of return by the collateralized inside bond. Since virtually nobody wishes to hold money in the OTC round, money has no value in the limiting economy as $\alpha \to 0$ with $\iota \in (\hat{\varsigma}(0), \bar{\varsigma}(0))$. The key result is that

$$\lim_{\alpha \to 0} \frac{\mathcal{Z}}{\varphi} = \lim_{\alpha \to 0} \frac{1}{\mathcal{V}} = 0 < \lim_{\alpha \to 0} (\varphi - \varphi^n) = (1 - \theta) \int_{\varepsilon_L}^{\varepsilon^*} (\varepsilon^* - \varepsilon) \, dG(\varepsilon) . \tag{55}$$

Notably, even though real balances and velocity converge to their nonmonetary equilibrium levels as $\alpha \to 0$, the real equity price in this cashless limit of the monetary economy exceeds the (corresponding limit of the) nonmonetary-equilibrium price by the value of a resale-option term in (55). Since ε^* is a function of ι (from (50)), it follows that in the cashless limit the asset price is still responsive to monetary policy, and that the magnitude of this response remains bounded away from zero even though the quantity of real balances converges to zero. This result stands in contrast with the more conventional result that the monetary equilibrium prices and allocations converge to their nonmonetary equilibrium counterparts when real balances vanish, e.g., as it happens with the cashless limit in Proposition 5.

Why does the cashless limit in Proposition 6 not correspond to the nonmonetary equilibrium? The answer is in (55) and, at least algebraically, it is simple: because the following two conditions are met: (a) the limit, as $\alpha \to 0$, of the marginal valuation of investors without access to credit is strictly larger than ε_L (i.e., the value of ε^* that solves (50) satisfies $\varepsilon_L < \varepsilon^*$, given that $\hat{\varsigma}(0) < \iota < \bar{\varsigma}(0)$), and (b) $\theta < 1$. Condition (a) is a result of the fact that the asset demand from leveraged high-valuation investors with access to credit sustains an equilibrium asset price that in the limit is high enough to encourage investors without access to credit who have low enough valuations, to sell the asset for cash at that price. Condition (b) is a parametric condition on the market power of financial intermediaries. The case with $\theta = 1$ corresponds to a market structure where bond brokers have no market power to extract intermediation fees, and in this case the equity price in the cashless limit corresponds to the asset price in the nonmonetary equilibrium. To understand this result at a conceptual level, it is useful to review the components of the limiting equity price

$$\lim_{\alpha\to 0}\varphi=\bar{\varepsilon}+\theta\frac{\lambda}{1-\lambda}\int_{\varepsilon^n}^{\varepsilon_H}(\varepsilon-\varepsilon^n)\,dG(\varepsilon)$$
$$+\theta\int_{\varepsilon_L}^{\varepsilon^n}(\varepsilon^n-\varepsilon)\,dG(\varepsilon)+(1-\theta)\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^*-\varepsilon)\,dG(\varepsilon)\,.$$

The first term is the expected value of the dividend flow. The second term is the investor's share θ of the expected value of the pledge option on the asset (i.e., the marginal value to the investor of the asset when it is pledged as collateral for shorting the bond). The third term is the investor's share θ of the expected value of the resale option on the asset. The sum of the first three terms equals $\tilde{\varphi}^n$ (the nonmonetary equity price in the economy with $\alpha=0$). The fourth term is the key term since it is the component of the asset price that is present in the cashless limit but not in the nonmonetary equilibrium. This term captures the improvement in the investor's bargaining position with the bond broker due to the fact that the investor's outside option is not just autarky, but rather trading in the equity market without access to bonds. To see this, consider an investor with access to credit who draws $\varepsilon\in[\varepsilon_L,\varepsilon^{**}]$. He wishes to *sell* all his equity and simultaneously bargains with a broker to take a long position in the bond. The investor's outside option in this bargain is the payoff of trading in the equity market without access to the bond market. The particular kind of trade that this option would entail, depends on the investor's valuation. If $\varepsilon\in[\varepsilon_L,\varepsilon^*]$, the investor's outside option amounts to acting like an investor without access to credit, i.e., *selling* equity for cash. In this case the expected capital gain from reselling the asset is $\int_{\varepsilon_L}^{\varepsilon^*}(\varepsilon^*-\varepsilon)\,dG(\varepsilon)$, and the bargaining outcome assigns a fraction $1-\theta$ of this gain from trade to the investor. This is precisely the right side of (55), i.e., the additional term in the equity price at the cashless limit.[25]

Intuitively, the reason for (55) is that an investor with access to credit who has valuation $\varepsilon\in[\varepsilon_L,\varepsilon^*]$ has the option to sell equity for cash in the equity market; this option improves his bargaining position with bond brokers, and allows the investor to earn a larger share of the gain from his portfolio reallocation (from equity to bonds). The key result is that the value of this option to the individual investor remains strictly positive in the limit as $\alpha\to 0$, even as aggregate real money balances are converging to zero along with the volume of cash

[25]If $\varepsilon\in[\varepsilon^*,\varepsilon^{**}]$, then the investor's outside option again amounts to acting like an investor without access to credit, which in this case entails *buying* equity with cash. Hence, since the equity would not be sold for this range of valuations, it would not deliver a gain associated with reselling, and therefore the investor's outside option of trading in the equity market with no access to bonds does not appear in the expected resale value component of the asset. Algebraically, this is why the upper bound of the integral in (55) is ε^* rather than ε^n.

transactions in the equity market. This may seem counterintuitive, in light of three natural questions. First, given that aggregate real money balances vanish in the cashless limit, why doesn't the value of the investor's option to engage in monetary exchange also vanish in the cashless limit? Second, if an investor with access to credit who has valuation $\varepsilon \in [\varepsilon_L, \varepsilon^*]$ were to exercise his option to sell equity for cash in the equity market along the trajectory toward the cashless limit (he actually does not exercise this option in the equilibrium), who would be on the other side of this transaction, buying equity with cash? Third, is it credible for an investor with access to credit who has valuation $\varepsilon \in [\varepsilon_L, \varepsilon^*]$ to threaten the broker to sell equity in exchange for money along the cashless limit with vanishing aggregate real money balances? The answer to the first question is that an individual investor is atomistic, so he can always execute his desired trade no matter how small trade volume in the competitive equity market has become along the trajectory toward the cashless limit. The answer to the second question is that the counterparties of an investor with access to credit who sells equity for cash in the equity market belong to the following two groups of investors (both of whom are holding strictly positive real money balances along the cashless trajectory): (a) the $(1-\alpha)[1-G(\varepsilon^{**})]$ investors with access to credit with valuations in $[\varepsilon^{**}, \varepsilon_H]$, and ($b$) the $\alpha[1-G(\varepsilon^*)]$ investors without access to credit with valuations in $[\varepsilon^*, \varepsilon_H]$ that are still around along the trajectory toward the cashless limit. Finally, the threat to sell equity for cash in the equity market and hold on to the cash (instead of dealing with the credit broker and selling equity to long the bond) is credible because, although aggregate money demand and therefore aggregate real balances are converging to zero along the cashless limit (as shown by (47)), each investor without access to credit (or each investor with access to credit who threatens to mimic the behavior of an investor without access to credit) is willing to demand cash in exchange for equity. Formally, although the limit of aggregate real money balances is zero, the limit of aggregate real money balances *per investor without access to credit* is strictly positive, i.e.,

$$\lim_{\alpha \to 0} \frac{\mathcal{Z}/\varphi}{\alpha} = \lim_{\alpha \to 0} \frac{G(\varepsilon^*)}{1 - G(\varepsilon^*)\,\alpha} = \lim_{\alpha \to 0} G(\varepsilon^*) > 0.$$

This is a key difference with the conventional inflationary cashless limit of Proposition 5. In that case, $\lim_{\iota \to \bar{\iota}(\lambda)} \varepsilon^* = \varepsilon_L$, and therefore

$$\lim_{\iota \to \bar{\iota}(\lambda)} \frac{\mathcal{Z}/\varphi}{\alpha} = \lim_{\iota \to \bar{\iota}(\lambda)} \frac{G(\varepsilon^*)}{[1 - G(\varepsilon^*)]\,\alpha + 1 - \alpha} = 0,$$

which is why the value of the option to sell equity for cash in the equity market (for a low-valuation investor with access to credit) converges to zero in the cashless limit of Proposition

5, and consequently the equity price in the cashless limit is equal to the equity price in the nonmonetary equilibrium.

The equilibrium in the cashless limit as $\alpha \to 0$ behaves as follows. The resale value option in (55) is positive because even in the limit, an individual investor with access to credit can credibly threaten to sell his equity for cash in the equity market, and this off-equilibrium threat improves his bargaining position with the broker, allowing the investor to secure a larger share of the trade surplus. Since the resale value option is positive in the limit, the asset price remains relatively high, and in particular, higher than the price in the nonmonetary equilibrium (where, with no cash valued in equilibrium, the investor cannot threaten to sell equity for cash in the equity market to improve his bargaining position). In this sense, there is a sort of discontinuity in the equilibrium value of the investor's outside option. On the one hand, this option is absent in the nonmonetary equilibrium. On the other hand, in the limit as $\alpha \to 0$, aggregate real balances converge to 0 but the limit of the value of the option to sell equity for cash is positive for any investor with access to credit. Thus, the resale value option of an investor with access to credit in (55) appears discontinuous when taking the limit with respect to α and comparing it to its counterpart in a nonmonetary equilibrium, and the asset price inherits this "discontinuity." In turn, the relatively high equity price in the cashless limit rationalizes the fact that not all investors without access to credit prefer buying equity rather than holding cash in the OTC round of trade, which manifests itself as $\varepsilon^* - \varepsilon_L > 0$ and $\lim_{\alpha \to 0} \frac{\mathcal{Z}/\varphi}{\alpha} > 0$ in the cashless limit.

To summarize, the result (55) depends on two fundamental features of the environment. First, bond brokers must have at least some degree of market power, i.e., $\theta < 1$. Second, investors must have at least some ability to take on leverage, i.e., $\lambda > 0$.[26]

5 Quantitative analysis

We regard a unit of time as corresponding to one year. The number of outstanding shares is normalized to 1, i.e., $A^s = 1$. The dividend growth rate is independently lognormally distributed over time, with mean .04 and standard deviation .12 per annum (e.g., as documented in Lettau and Ludvigson (2005), Table 1). That is, $y_{t+1} = e^{x_{t+1}} y_t$, with $x_{t+1} \sim \mathcal{N}\left(g, \Sigma^2\right)$, where $g = \mathbb{E}\left(\log y_{t+1} - \log y_t\right) = .04$ and $\Sigma = SD\left(\log y_{t+1} - \log y_t\right) = .12$. Over the sample period 1994-

[26] This is so because $\hat{\varsigma}(0) \leq \bar{\varsigma}(0)$ for all $\lambda \in [0,1]$, but with "=" only if $\lambda = 0$, so the parameter region in part (i) of Proposition 6 collapses to a single point if $\lambda = 0$. But for any positive λ, no matter how small, the region of discontinuity has positive measure in parameter space.

2007, the average nominal policy rate was 4.47% per annum, and the average inflation rate was 2.69% per annum.[27] Thus, we set $\rho^p = .0447$ and $\bar{\pi} \equiv \pi - g = .0269$, implying a real rate of $r = \rho^p - \bar{\pi} = .0178$ per annum.[28] The parameter δ can be taken as a proxy of the riskiness of stocks; we choose $\delta = .075$, i.e., a productive unit has a 92.5 percent probability of remaining productive each year. We set $\lambda = .75$, which implies a margin requirement of 25 percent.[29]

The distribution of idiosyncratic valuations, G, is assumed to be lognormal, and we normalize $\bar{\varepsilon} = 1$, i.e., $ln(\varepsilon) \sim \mathcal{N}(-\frac{1}{2}\Sigma_\varepsilon^2, \Sigma_\varepsilon^2)$. The parameters α, θ, and Σ_ε are calibrated so that, given the rest of the parametrization, the model is consistent with the following three facts: (a) the real asset price falls by about 11 basis points in response to a 1 basis point increase in the nominal policy rate, as in the high-frequency empirical estimates in Lagos and Zhang (2019b); (b) transaction velocity of money is 25 per day, which is the average daily number of times a dollar turns over in CHIPS (Clearing House Interbank Payments System); and (c) the median spread on margin loans is about 2.3%, i.e., $\rho^l\left(G_h^{-1}(.5)\right) - \rho^m = .023$ with $G_h(\varepsilon) \equiv \frac{G(\varepsilon)-G(\varepsilon^{**})}{1-G(\varepsilon^{**})}$, which is the current spread (over the fed funds rate) that a typical prime broker charges a large investor.[30] This procedure delivers $\alpha = .0406$, $\theta = .1612$, and $\Sigma_\varepsilon = 2.0784$. In Appendix E we assess the robustness of the quantitative results to alternative calibration strategies.

Our quantitative exercises consist of reporting asset price responses to changes in ρ^p for all $(\alpha, \lambda, \theta) \in [0,1]^3$. Specifically, we focus on the semi-elasticity of the asset price, ϕ^s, with respect to ρ^p, the policy rate.[31] Since the value is always negative, we report $\mathcal{S} = \left|\frac{d\phi^s/\phi^s}{d\rho^p}\right|$.

Figure 4 reports $\mathcal{S}$ for economies indexed by $(\alpha, \lambda) \in [0,1] \times \{.50, .75, .90, .99\}$. Our baseline calibration ensures that $\mathcal{S} \approx 11$ for $\lambda = .75$ and $\alpha = .0406$. The main finding is that the

[27]For the policy rate we use the 3-month Eurodollar futures rate (series IEDCS00 produced by the CME Group available via Datastream). The annual average inflation rate is imputed as $[CPI(January\_2008)/CPI(January\_1994)]^{1/14} - 1$, where $CPI(Month\_Year)$ is monthly CPI index available from FRED at https://fred.stlouisfed.org/series/CPIAUCSL.

[28]To streamline the presentation, here we assume r is constant and therefore associate changes in the policy rate ρ^p with changes in π. In Lagos and Zhang (2019b), which corresponds to a special case of this model with no credit, i.e., if either $\lambda = 0$ or $\alpha = 1$, we allow for the possibility that when the policy rate changes by $\Delta\rho^p$, the real rate changes by $\Delta r = w\Delta\rho^p$ and the inflation rate changes by $\Delta\pi = (1-w)\Delta\rho^p$, where $w \in [0,1]$ indexes the degree of passthrough from nominal rates to real rates.

[29]As mentioned in Section 2.2, FINRA Rule 4210 requires that a customer maintains a minimum margin of 25% at all times.

[30]For example, in June 2017, when the effective fed funds rate was about 1.04% per annum, Morgan Stanley's *Margin Interest Rate Schedule* specified an (annualized) interest rate on margin loans of 3.375% for debit balances of $50,000,000 or more. See https://www.morganstanley.com/wealth-disclosures/pdf/Margin_Interest_Rate.pdf.

[31]We know from our work in Lagos and Zhang (2019b) that in this type of model the response of the equity price to a one-time unanticipated change in the nominal policy rate is a good approximation to the response to the same change in a nominal policy rate that follows a stochastic (Markov) process that is as persistent as in the data, e.g., as in the quantitative implementation of the stochastic model in Lagos and Zhang (2019b).

response of the asset price to nominal rate shocks remains significant for a wide range of values of λ, and even in the pure-credit limiting economy that obtains as $\alpha \to 0$. For example, the semi-elasticity is even larger than 11 as $\alpha \to 0$ if investors are able to leverage up to 100 times their wealth. Figure 5 reports $\mathcal{S}$ for economies indexed by $(\alpha, \theta) \in [0,1] \times \{.16, .25, .70, .99\}$. Again, the response of the asset price to nominal rate shocks remains significant in the limiting economy as $\alpha \to 0$. For example, as $\alpha \to 0$, the semi-elasticity remains above 4 even if investors are able to capture 70% of the gains in trades intermediated by bond brokers. Figure 6 reports $\mathcal{S}$ for economies indexed by $(\alpha, \rho^p) \in [0,1] \times \{.03, .04, .0447, .05\}$. This exercise shows that for every level of α, the asset price response tends to be larger in environments with a lower background nominal policy rate.

Figures 7, 8, and 9 offer a comprehensive summary of the magnitude of the effects of monetary policy in limiting economies with $\alpha \to 0$. For a wide range of economies indexed by a pair ρ^p and λ, Figure 7 reports the value of $\mathcal{S}$ in the pure-credit limit that obtains as $\alpha \to 0$. The level sets in the right panel show it is not easy to find reasonable parametrizations that imply a value of $\mathcal{S}$ below 5. Figures 8 and 9 tell a similar story. Figure 8, for example, shows that, as predicted by the theory, $\mathcal{S} = 0$ in the pure-credit cashless limit of economies with no credit-market frictions or markups, i.e., economies with $\lambda = \theta = 1$. In contrast, $\mathcal{S}$ is positive and sizable for economies in which is $\theta < 1$, even if $1 - \theta$ is relatively small.

6 Discussion

6.1 On the moneyless approach to monetary economics

Our results on the medium-of-exchange role of money in the transmission of monetary policy run counter to a large body of work that follows a moneyless approach to monetary economics. This moneyless approach was advocated by Woodford (1998) and, based on the treatments in Woodford (2003) and Galí (2008), is now considered by many “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 sweeping 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 these results 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) effectively makes money and bonds the same asset (with 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 in the Euler equations of the limiting economy, this approximation result is used to justify neglecting monetary variables 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 is equivalent to simply assuming a period utility function of the form $U(c,m) = u(c) + Av(m)$ of consumption, c, and real money balances, m, for given functions $u(\cdot)$ and $v(\cdot)$, 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 in the limiting case, which is independent of money demand for transactions, can be used to approximate the monetary equilibrium in any case in which medium-of-exchange frictions exist but are small.

The textbook treatments of monetary policy in Woodford (2003) and Galí (2008) assign a very limited role to money. 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 $U(c,m) = u(c) + Av(m)$. This separable specification is justified by showing that, in the context of a competitive model with no credit frictions, if $U(c,m)$ 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, $M_t/(p_tY_t)$, where M_t/p_t denotes aggregate real money balances, and Y_t denotes GDP. Woodford (2003, p. 113) and Galí (2008, p. 31) argue that since $M_t/(p_tY_t)$ 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 $U(c,m) = u(c) + Av(m)$, 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 (e.g., Proposition 6) is in the spirit of Wicksell's "pure credit economy" and in line with the motivation for Woodford's cashless limit. Generically, however, our results stand in contrast with Woodford's: we find that in general the medium-of-exchange role of money is important for monetary transmission, and remains a significant conduit for monetary policy even in the cashless limit. As $\alpha \to 0$, real balances converge to zero, transaction velocity goes to infinity, and the monetary economy converges to a limit where monetary policy still has significant effects on welfare, asset prices, consumption, investment, and output. 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's. It is the case where financial intermediaries have no market power, i.e., $\theta = 1$. So, in order to argue that monetary frictions are irrelevant in cashless limiting economies or almost irrelevant near-cashless economies, it is necessary to also adopt the view that investors are always able to reap the entire share of the gains from trade when interacting with intermediaries in financial markets.[32] Our theoretical point here is that $\theta = 1$ is nongeneric. Whether the perfectly competitive case with $\theta = 1$ is the relevant case for applied work, is likely to be ultimately an empirical issue that deserves further study. We are aware of no evidence that $\theta = 1$ is the norm empirically, even in the financially advanced economies with low levels inverse velocity of the monetary base that Woodford (2003) and Galí (2008) argue are well approximated by the moneyless approach to monetary policy.[33]

For $\theta < 1$, our theory provides counterexamples to the typical claims (based on the reduced-form arguments outlined above) commonly used to endorse the moneyless approach. For example, Woodford (2003, p. 32) claims that the basic model in his book "abstracts from monetary frictions, in order to focus attention on more essential aspects of the monetary transmission mechanism...". Galí (2008, p. 10) claims that "...there is generally no need to specify a money demand function, unless monetary policy itself is specified in terms of a monetary aggregate, in which case a simple log-linear money demand schedule is postulated." The results we presented above (in particular those in Section C and Section 5) and in related work (Lagos and Zhang (2019b)) indicate that traditional medium-of-exchange considerations are in fact an essential aspect of the monetary transmission mechanism—even in the cashless limit or in near-cashless economies in which liquidity-saving mechanisms have developed sufficiently to make the inverse velocity of the monetary base very small. Any attempt to assess the macroeconomic effects of monetary policy without such considerations is necessarily incomplete.

[32] Our formulation also assumes individual investors are subject to borrowing limits; a collateral constraint in the model of Section 2, and a given borrowing limit in the model of Section A.2). This can be interpreted as a second departure from what may be described as "frictionless financial markets." However, our cashless results hold even if $\lambda \to 1$ in the economy of Section 2, and if $\Lambda \to \infty$ in the economy of Section A.2.

[33] Notice that for a given $\theta < 1$, our cashless limiting economy is different depending on the underlying monetary policy, ι, and credit conditions, as captured by λ in the baseline model. If the policy rate and leverage are relatively large as in part (i) of Proposition 6 then money is dominated in rate of return, real balances converge to zero, and asset prices respond to ι even in the cashless limit. If the policy rate and leverage are relatively low as in part (ii) of Proposition 6, then as $\alpha \to 0$, real balances are not dominated in rate of return, do not converge to zero, and monetary policy remains effective in the limit. In both cases our conclusions regarding the effects of ι on prices and allocations in the pure-credit limiting economy is clearly at odds with the limiting irrelevance result in Woodford (2003).

6.2 On reduced-form models of money demand

There are well known critiques of reduced-form models of money. Kareken and Wallace (1980) for example, state two. The first, is that assuming money is an argument of a utility (or a production) function is an instance of "implicit theorizing," by which they mean that while there may be stories that can be told to justify the approach (e.g., that money provides unmodelled transaction services), the assumptions implicit in these stories cannot be regarded as primitives, and unless the underlying environment is made explicit, the internal consistency of the theory cannot be assessed. The second criticism is that reduced-form specifications beg too many questions; explain too little. What is the thing called "money"? Is it a private liability? A government liability? A commodity? If it is a government liability, which one? If there are many countries, does the liability issued by the government of country A enter the utility function of a citizen of country B? Compelling as they may be, these two criticisms are ignored in most of applied monetary economics. The reason, we suspect, is that these criticisms may not seem too serious in practice. Suppose one wants to study the effects of a monetary policy shift in an advanced economy like the United States. The common practitioner's view would be that in this context, there are some reasonable choices for the assets that play the role of money, and that the unmodelled medium-of-exchange frictions subject to the Kareken-Wallace implicit theorizing critique are likely to be small anyway.[34]

The near-cashless results in Woodford (2003) and Galí (2008) differ from ours because they rely on a money-in-the-utility (MIU) formulation that can sometimes fail to capture important aspects of the role money plays in transactions. For this reason it may be useful to show the limitations of their reduced-form approach in the context of our economic environment. We want to stress that these limitations are relevant even after we have agreed on which assets should be included in the utility function, and remain relevant even for practical, routine questions in monetary policy in the context of advanced economies with highly developed liquidity-saving mechanisms and credit-based payment arrangements resulting in low aggregate real balances.

The equilibrium conditions for our model can be obtained from the following reduced form:

$$\max_{\{c_t, h_t, \tilde{\boldsymbol{a}}_{t+1}\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left[\mathcal{U}\left(c_{1t}, c_{2t}; \boldsymbol{s}_t\right) + c_t - h_t \right] \tag{56}$$

[34]The 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)).

$$\text{s.t. } c_t + \phi_t^s \tilde{a}_{t+1}^s + \phi_t^m \tilde{a}_{t+1}^m = h_t + (\bar{\varepsilon} y_t + \phi_t^s)\, a_t^s + \phi_t^m a_t^m + T_t$$
$$c_{1t} = \frac{a_t^m}{p_t} y_t$$
$$c_{2t} = a_t^s y_t$$
$$\boldsymbol{a}_{t+1} = \left(\tilde{a}_{t+1}^m, \eta \tilde{a}_{t+1}^s + (1-\eta)\, A^s\right),$$

with

$$\mathcal{U}\,(c_1, c_2; \boldsymbol{s}_t) \equiv \overline{\mathcal{U}}_{1t} c_1 + \overline{\mathcal{U}}_{2t} c_2,$$

where $\boldsymbol{s}_t \equiv (y_t, p_t, q_t, \boldsymbol{\phi}_t, \varepsilon_t^*, \varepsilon_t^{**})$, $\overline{\mathcal{U}}_{1t} \equiv \frac{p_t}{y_t}\,(\bar{v}_{It}^m - \phi_t^m)$, $\overline{\mathcal{U}}_{2t} \equiv \frac{1}{y_t}\,[\bar{v}_{It}^s - (\bar{\varepsilon} y_t + \phi_t^s)]$, and $\bar{v}_{It}^k \equiv \int v_{It}^k\,(\varepsilon)\, dG(\varepsilon)$ for $k \in \{m, s\}$ is defined in Lemma 5.

If we focus on recursive equilibrium, then $\mathcal{U}\,(c_1, c_2; \boldsymbol{s}_t) = U\,(c_1, c_2; \boldsymbol{x}_t)$, where

$$U\,(c_1, c_2; \boldsymbol{x}) \equiv u^z c_1 + u^s c_2, \tag{57}$$

where $\boldsymbol{x} \equiv (\phi^s, \varepsilon^*, \varepsilon^{**})$, and u^z and u^s are given by

$$u^z = [\alpha + (1-\alpha)(1-\theta)] \int_{\varepsilon^*}^{\varepsilon_H} (\varepsilon - \varepsilon^*)\, dG(\varepsilon) + (1-\alpha)\,\theta \left[\varepsilon^{**} - \varepsilon^* + \frac{1}{1-\lambda} \int_{\varepsilon^{**}}^{\varepsilon_H} (\varepsilon - \varepsilon^{**})\, dG(\varepsilon)\right] \tag{58}$$

$$u^s = [\alpha + (1-\alpha)(1-\theta)] \int_{\varepsilon_L}^{\varepsilon^*} (\varepsilon^* - \varepsilon)\, dG\,(\varepsilon) + (1-\alpha)\,\theta \left[\int_{\varepsilon_L}^{\varepsilon^{**}} (\varepsilon^{**} - \varepsilon)\, dG\,(\varepsilon) + \frac{\lambda}{1-\lambda} \int_{\varepsilon^{**}}^{\varepsilon_H} (\varepsilon - \varepsilon^{**})\, dG\,(\varepsilon)\right]. \tag{59}$$

Notice that, u^z and u^s are implicit functions of $\boldsymbol{x} \equiv (\phi^s, \varepsilon^*, \varepsilon^{**})$.[35] The first-order conditions for (56) are

$$\phi_t^m \geq \beta \mathbb{E}_t \left[\frac{\partial U\,(c_{1t+1}, c_{2t+1}; \boldsymbol{x}_{t+1})}{\partial c_{1t+1}} \frac{y_{t+1}}{p_{t+1}} + \phi_{t+1}^m\right], \text{ with “ = ” if } 0 < \tilde{a}_{t+1}^m$$
$$\phi_t^s = \beta \eta \mathbb{E}_t \left[\left(\bar{\varepsilon} + \frac{\partial U\,(c_{1t+1}, c_{2t+1}; \boldsymbol{x}_{t+1})}{\partial c_{2t+1}}\right) y_{t+1} + \phi_{t+1}^s\right].$$

If we focus on a recursive equilibrium of the discrete-time economy with period length Δ, and let $\Delta \to 0$, these first-order conditions become

$$\iota \geq \frac{u^z}{\varphi}, \text{ with “ = ” if } 0 < \mathcal{Z} \tag{60}$$
$$\varphi = \bar{\varepsilon} + u^s. \tag{61}$$

[35]Instead of using $\boldsymbol{x}$ to index U, we could use the set of all parameters, $\Pi = (\beta, \eta, \bar{\gamma}, G, \theta, \alpha, \lambda, \mu)$ since the equilibrium objects $(\phi^s, \varepsilon^*, \varepsilon^{**})$ are themselves constant functions of Π in a recursive equilibrium.

If money is held, (60) and (61) imply

$$\frac{u^z}{\bar{\varepsilon} + u^s} = \iota.$$

The right side of this condition is the opportunity cost of carrying a dollar between periods, and the left side is the marginal rate of substitution between money and equity.

If we substitute (58) and (59), (60) and (61) become identical to the last two conditions in Lemma 13, so formulation (56) is in this sense equivalent to our micro model of investor behavior. Hence, we can define recursive equilibrium as before: a vector, $(\varepsilon^*, \varepsilon^{**}, \varphi, \mathcal{Z})$, that satisfies the first two conditions in Lemma 13, together with (60) and (61) (taking into account that u^z and u^s are given by (58) and (59)).

Suppose instead, that we regard u^z and u^s as a pair of fixed parameters, as would be the case in a prototypical MIU formulation that ignores all the micro details of the OTC trading round, and instead proxies for them with a utility function U given by (57) that is taken parametrically and is meant to capture the "convenience yield" or "liquidity services" of certain assets. One obvious problem with this approach is that, instead of regarding U as an *indirect utility function* (i.e., and equilibrium object) it treats U as if it were a primitive, i.e., exogenous and invariant to changes in the policy and market-structure parameters. In contrast, in a structural micro-founded model such as the one we develop in Section 2, changes in the policy rate ι, credit conditions λ, and market-structure parameters α and θ, change the shape of the indirect utility function U through their effects on $\boldsymbol{x}$, or more explicitly, through their effects on u^z and u^s, that are functions of those parameters, as is clear from (58) and (59). An equilibrium for the MIU economy would be a pair, $(\varphi, \mathcal{Z})$, that satisfies (60) and (61) taking u^z and u^s parametrically. For this MIU economy, the equilibrium is simple: The equity price is given explicitly by (61); it consists of the expected dividend, $\bar{\varepsilon}$, and the "convenience yield" of equity, u^s. A robust feature of this formulation (exploited in the work that advocates studying monetary policy ignoring money-demand considerations) is that because U is separable in real balances, the equilibrium level of real balances plays no role in the determination of the rest of the equilibrium. In this case, for instance, the real asset price, φ, is independent of $\mathcal{Z}$ and therefore, independent of the policy rate, ι. Equilibrium real balances are given by (60), namely, monetary equilibrium exists only if $\iota = u^z/\varphi$, and in this case the equilibrium aggregate real balance is any $\mathcal{Z} \in [0, \infty)$.[36]

[36]The indeterminacy of $\mathcal{Z}$ in the monetary equilibrium in this example is due to the fact that U is not only separable, but also linear in real balances. If U where separable but not linear in real balances, then (61) would be unchanged and (60) (at equality) would determine a unique $\mathcal{Z}$.

The upshot is that the properties of the equilibrium prices $(\varphi, \mathcal{Z})$ that result from treating U as a primitive are in general very different from the properties of the equilibrium prices of the model where the role of money and equity in the exchange process is accounted for explicitly.

Obvious though this "Lucas critique" type of observation may be, it also turns out to be a critical shortcoming of the reduced-form approach, especially when used to draw conclusions on the importance of the medium-of-exchange function of money and its role in the transmission of monetary policy. To give a concrete example, based on the MIU formulation one would conclude that φ is independent of monetary policy. However, once the exchange process through which money yields liquidity services is spelled out, e.g., as in the model of Section 2, one learns that φ is decreasing in ι because u^s is really a decreasing function of ι. As another example, consider a cashless limit or a near-cashless economy like the ones often used to justify ignoring monetary frictions in the New Keynesian textbooks. The way such a cashless limit (e.g., resulting from a sequence of improvements in liquidity-saving mechanisms or credit-based payment arrangements) can be captured in the reduced-form model is by driving the marginal value of real balances to zero in the assumed utility function. For example, in our case with U given by (57), suppose we start with a policy rate consistent with monetary equilibrium, i.e., $\iota = u^z/\varphi$. Then as the parameter u^z is reduced, real balances go to zero but this has no effect on the rest of the equilibrium; in particular the real equity price (61) is unchanged because u^s is taken to be a parameter in the MIU formulation. The irrelevance of the medium-of-exchange role of money is as strong as it can be in this MIU formulation: the rest of the equilibrium (i.e., asset prices, and in the model with capital accumulation, also consumption, investment, and output) are invariant to $\mathcal{Z}$ in the cashless limit, but also away from it. In the micro-founded model, in contrast, u^s is a function of the credit conditions λ, the market-structure parameters α and θ, and in particular a function of the policy rate ι—even in the cashless limit, e.g., as $\mathcal{Z} \to 0$ because $\alpha \to 0$ as in part (i) of Proposition 6.

In sum, the widespread New Keynesian view that medium-of-exchange monetary frictions are unimportant for monetary transmission relies on two irrelevance results that are based on: (i) a reduced-form specification of cash and credit transactions that fails to capture the effects of monetary policy on prices and allocations that remain significant even in near-cashless economies, and (ii) a presumption that the financial markets implicit in the reduced-form specification are frictionless (and in particular that investors are always able to reap the entire share of the gains from trade when trading with intermediaries in those unmodelled cash and credit markets).

We have shown that if the role of money as a medium of exchange is modeled explicitly, and financial markets exhibit realistic frictions (e.g., credit takes the form of collateralized loans intermediated by brokers who have at least some degree of market power), then monetary policy conducted by means of changing a nominal interest rate is transmitted to the real allocations—even in the cashless limit, i.e., even if access to credit is so generalized that real balances are negligible. Along this cashless limit, the path of the monetary aggregate becomes irrelevant but the monetary frictions that give money value as a medium of exchange remain important for the transmission of monetary policy. These monetary frictions are resilient in that they remain operative even when real balances are negligible as is the case in the cashless limit. To summarize, these results are different from those obtained from reduced-form models because: (*i*) the true indirect utility function for money and consumption is changing along the cashless limit, so MIU formulations that assume the utility function is stable to policy or changes in the market structure are not be able represent the limit of the monetary equilibrium of a model such as ours, where money and intermediated credit help agents overcome commitment and double-coincidence-of-wants frictions; and (*ii*) the financial market in our model is not frictionless: credit is limited and involves intermediaries who have some degree of market power.

6.3 Price level determination in the cashless limit

The price level determination in limiting cashless economies is a recurrent theme in Woodford (2003), so for comparison purposes, it may be useful to explain the behavior of the price level along the cashless limit of our economy. In a RME, the price level (measured by the nominal price of equity shares) is $p_t = \frac{\varepsilon^* + \phi^s}{ZA^s} A_t^m$.[37]

In the discrete-time formulation with period length is Δ, $p_t = \left[\varepsilon^* + \Phi^s(\Delta)\right] A_t^m / \left[\mathrm{Z}(\Delta) A^s\right]$, so as $\Delta \to 0$, we get

$$p_t = \frac{\varphi}{\mathcal{Z} A^s} A_t^m = \begin{cases} \frac{[1-G(\varepsilon^*)]\alpha + 1 - \alpha}{\alpha G(\varepsilon^*)} \frac{A_t^m}{A^s} & \text{if } \hat{\imath}(\lambda) < \iota < \bar{\imath}(\lambda) \\ \frac{[1-G(\varepsilon^*)]\left(\alpha + \frac{1-\alpha}{1-\lambda}\right)}{\alpha G(\varepsilon^*) + \frac{1-\alpha}{1-\lambda}[G(\varepsilon^*) - \lambda]} \frac{A_t^m}{A^s} & \text{if } 0 < \iota \leq \hat{\imath}(\lambda) \end{cases}$$

and

$$\lim_{\alpha \to 0} p_t = \begin{cases} \frac{1}{A^s G(\varepsilon^*)} \lim_{\alpha \to 0} \frac{A_t^m}{\alpha} & \text{if } \hat{\imath}(\lambda) < \iota < \bar{\imath}(\lambda) \\ \frac{1 - G(\varepsilon^*)}{G(\varepsilon^*) - \lambda} \frac{A_t^m}{A^s} & \text{if } 0 < \iota \leq \hat{\imath}(\lambda), \end{cases}$$

[37]The price level measured by the nominal price of general goods is just $\frac{1}{\phi_t^m} = \frac{p_t}{(\varepsilon^* + \phi^s) y_t}$, so we focus the analysis on p_t.

where ε^* is characterized by (50) if $\hat{\iota}(\lambda) < \iota < \bar{\iota}(\lambda)$, and by (54) if $0 < \iota \leq \hat{\iota}(\lambda)$. If $0 < \iota \leq \hat{\iota}(\lambda)$, then $\lim_{\alpha \to 0} p_t$ is necessarily finite for any path $\{A_t^m\}_{t=0}^{\infty}$. If the monetary authority wishes to implement a certain price path for a recursive monetary equilibrium of the cashless limiting economy with $\iota \in (\hat{\iota}(\lambda), \bar{\iota}(\lambda))$, then it can simply choose a money supply process $\{A_t^m\}_{t=0}^{\infty}$ given by $A_t^m = \alpha M_t$ with $\dot{M}_t = \pi M_t$, which implements a price level in the cashless limit that is equal to

$$\lim_{\alpha \to 0} p_t = \frac{M_t}{A^s G(\varepsilon^*)}.$$

By choosing the level of M_0, the monetary authority can implement any price level in the RME of the cashless limiting economy. Intuitively, the monetary authority can always implement a price level that remains well defined (i.e., finite) even in the cashless limit, simply by keeping the money supply *per investor without access to credit* stable along the cashless limit.

7 Conclusion

We conclude by mentioning what we think are three promising avenues for future work. First, the model we have presented is tailored to the transmission of monetary policy that operates through financial markets. We have chosen this route guided by the empirical and quantitative results in Lagos and Zhang (2019b). Here we have shown how, through its effect on asset prices, monetary policy ultimately influences aggregate consumption, investment, and output. With some minor changes, our theory could be reinterpreted as a model of trade in an input that is reallocated across production units with different productivities. This alternative formulation would imply a more direct transmission of money shocks to the real economy. Second, our model offers a leverage-based theory of transaction velocity, a variable that is empirically relevant yet difficult to model satisfactorily in more conventional monetary models. Third, the model implications for how asset prices and their responses to monetary policy shocks depend on credit conditions and leverage could lead to fruitful empirical work.

References

AFONSO, G. AND R. LAGOS (2015): "Trade Dynamics in the Market for Federal Funds," *Econometrica*, 83, 263–313.

BARRO, R. J. (1976): "The Loan Market, Collateral, and Rates of Interest," *Journal of Money, Credit and Banking*, 8, 439–456.

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

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

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.

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.

KIYOTAKI, N. AND J. MOORE (1997): "Credit Cycles," *Journal of Political Economy*, 105, 211–248.

——— (2005): "Liquidity and Asset Prices," *International Economic Review*, 46, 317–349.

KRISHNAMURTHY, A. AND A. VISSING-JORGENSEN (2012): "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 120, 233–267.

LAGOS, R. AND G. ROCHETEAU (2009): "Liquidity in Asset Markets With Search Frictions," *Econometrica*, 77, 403–426.

LAGOS, R., G. ROCHETEAU, AND P.-O. WEILL (2011): "Crises and Liquidity in Over-the-Counter Markets," *Journal of Economic Theory*, 146, 2169–2205.

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.

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

——— (2019b): "Turnover Liquidity and the Transmission of Monetary Policy," *American Economic Review*, forthcoming.

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.

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

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.

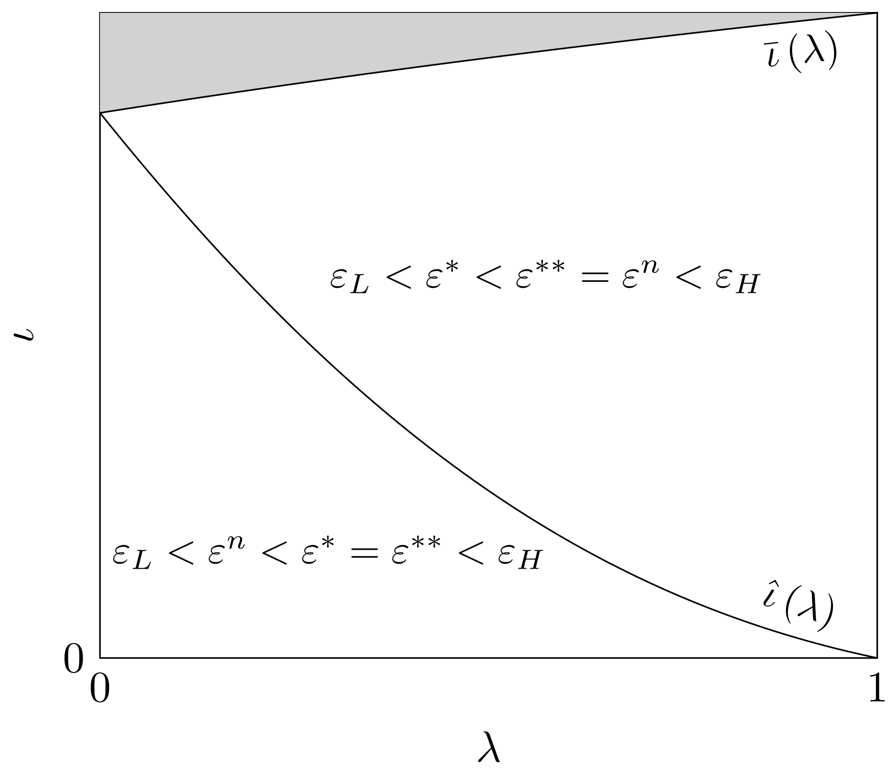

Figure 1: Existence regions for monetary equilibrium. There is no monetary equilibrium in the shaded region. Nonmonetary equilibrium exists everywhere.

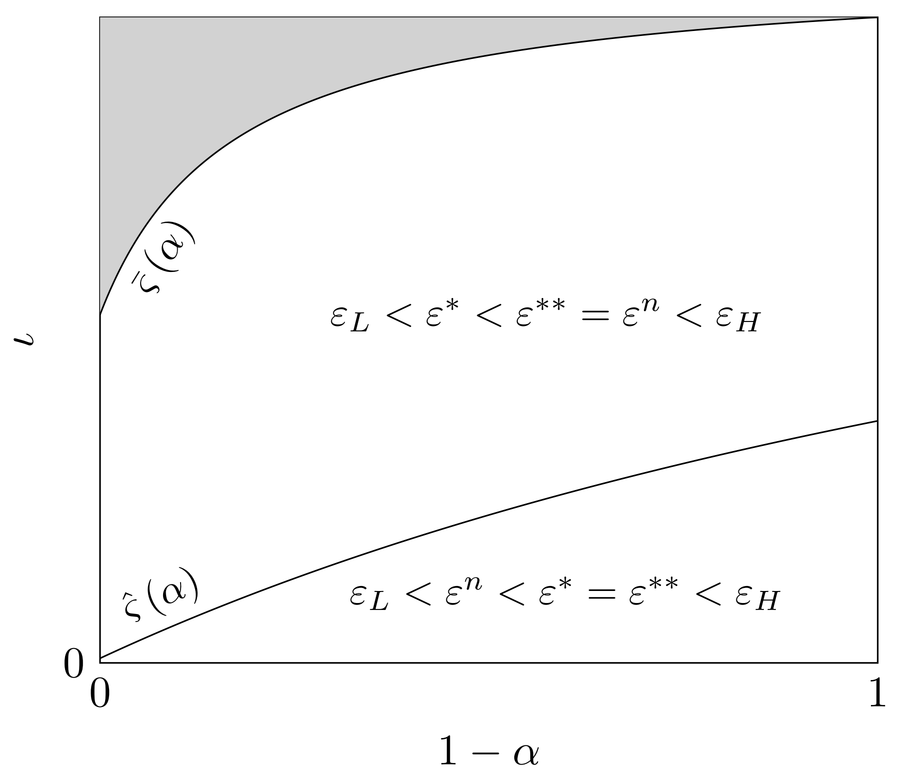

Figure 2: Existence regions for monetary equilibrium. There is no monetary equilibrium in the shaded region. Nonmonetary equilibrium exists everywhere.

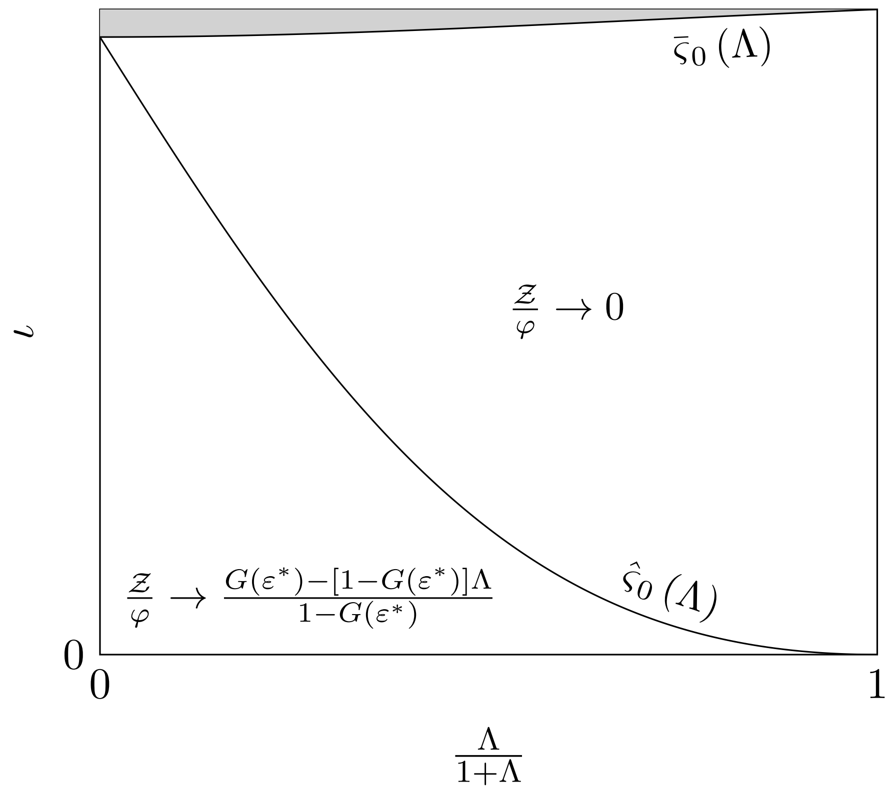

Figure 3: Existence regions for monetary equilibrium in the limit as $\alpha \to 0$ for the economy with unsecured credit. There is no monetary equilibrium in the shaded region. Nonmonetary equilibrium exists everywhere. Along a monetary equilibrium, real balances $\mathcal{Z}$ converge to 0 if $\hat{\varsigma}_0(\Lambda) < \iota < \bar{\varsigma}_0(\Lambda)$ and to a positive level if $0 < \iota \leq \hat{\varsigma}_0(\Lambda)$.

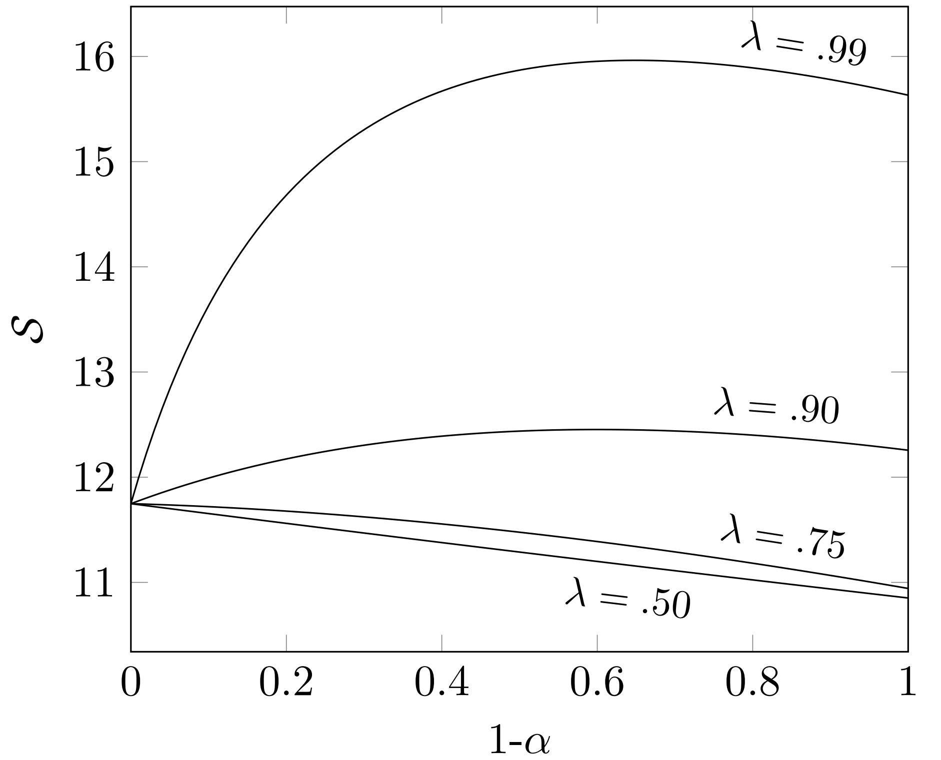

Figure 4: Semi-elasticity of the asset price with respect to the nominal policy rate for economies with different levels of leverage, λ, and access to credit, $1 - \alpha$.

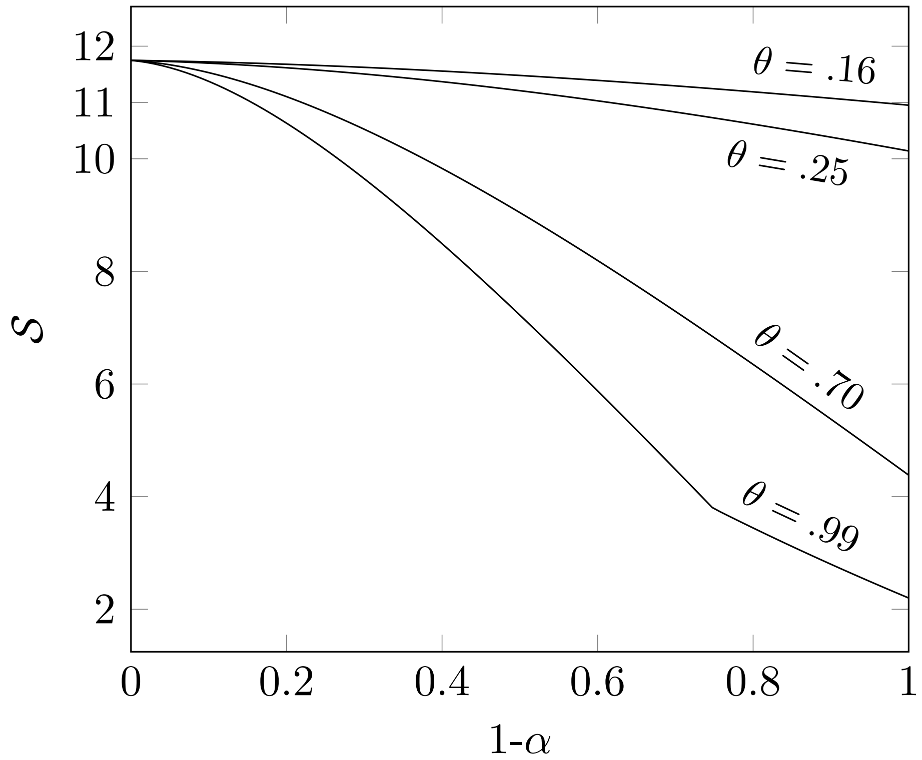

Figure 5: Semi-elasticity of the asset price with respect to the nominal policy rate for economies with different market power of brokers, $1-\theta$, and access to credit, $1-\alpha$.

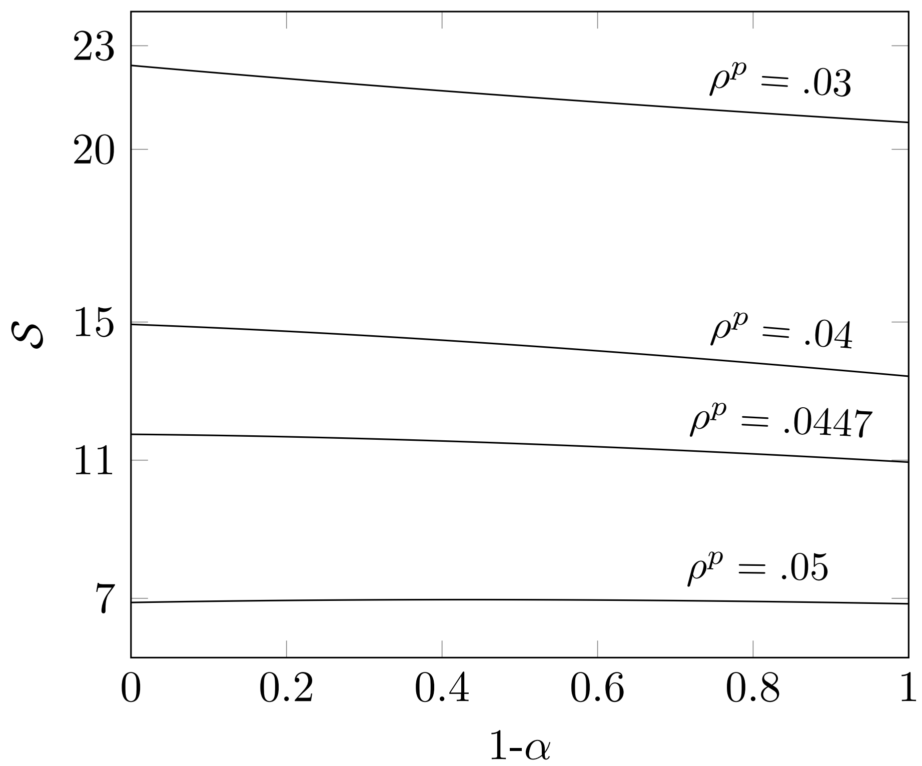

Figure 6: Semi-elasticity of the asset price with respect to the nominal policy rate for economies with different monetary regimes, ρ^p, and access to credit, $1 - \alpha$.

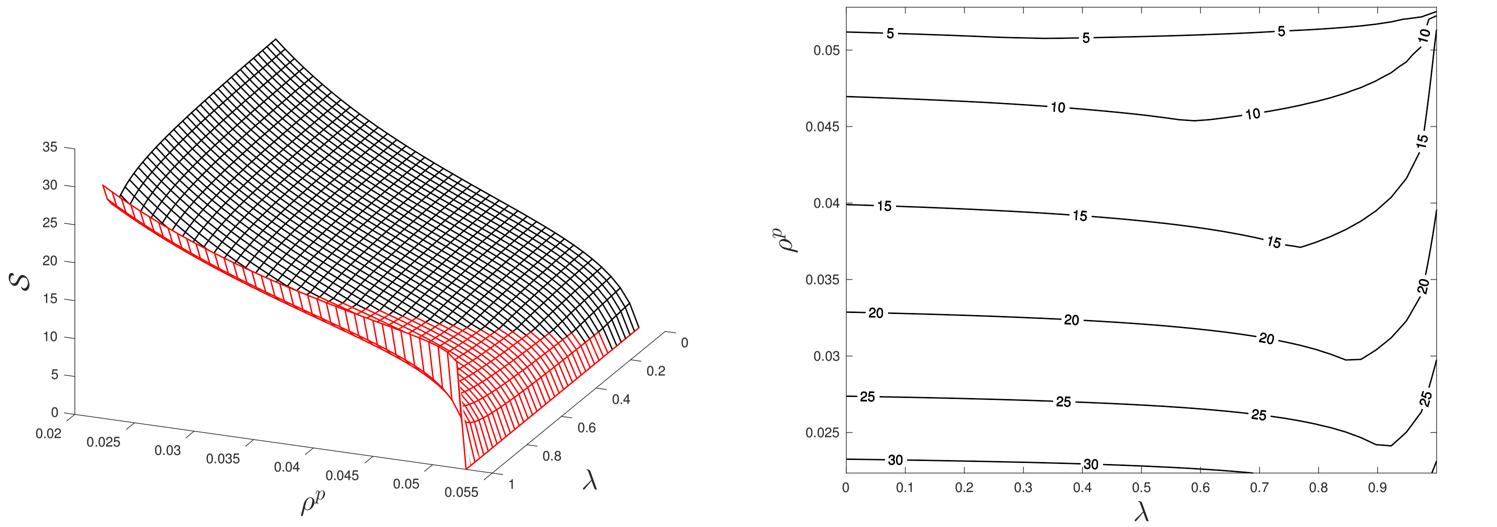

Figure 7: Semi-elasticity of the asset price with respect to the nominal policy rate as functions of λ and ρ^p in limiting economies with $\alpha \to 0$. The right panel shows the level sets for $\mathcal{S}$ corresponding to the left panel.

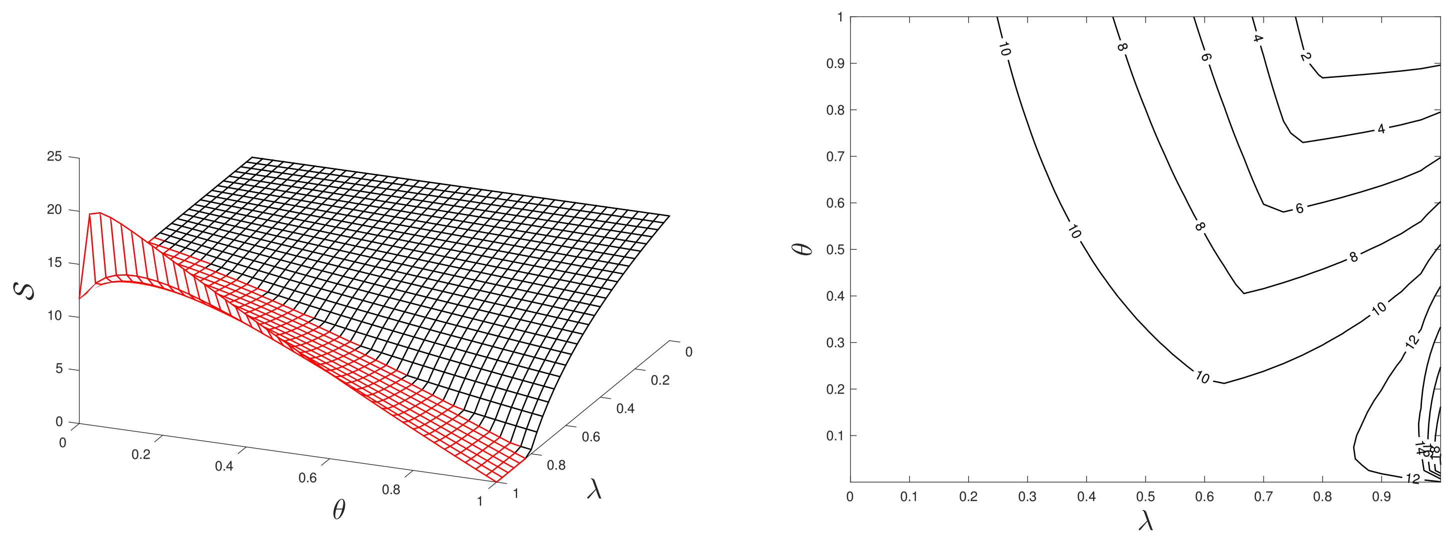

Figure 8: Semi-elasticity of the asset price with respect to the nominal policy rate as functions of λ and θ in limiting economies with $\alpha \to 0$. The right panel shows the level sets for $\mathcal{S}$ corresponding to the left panel.

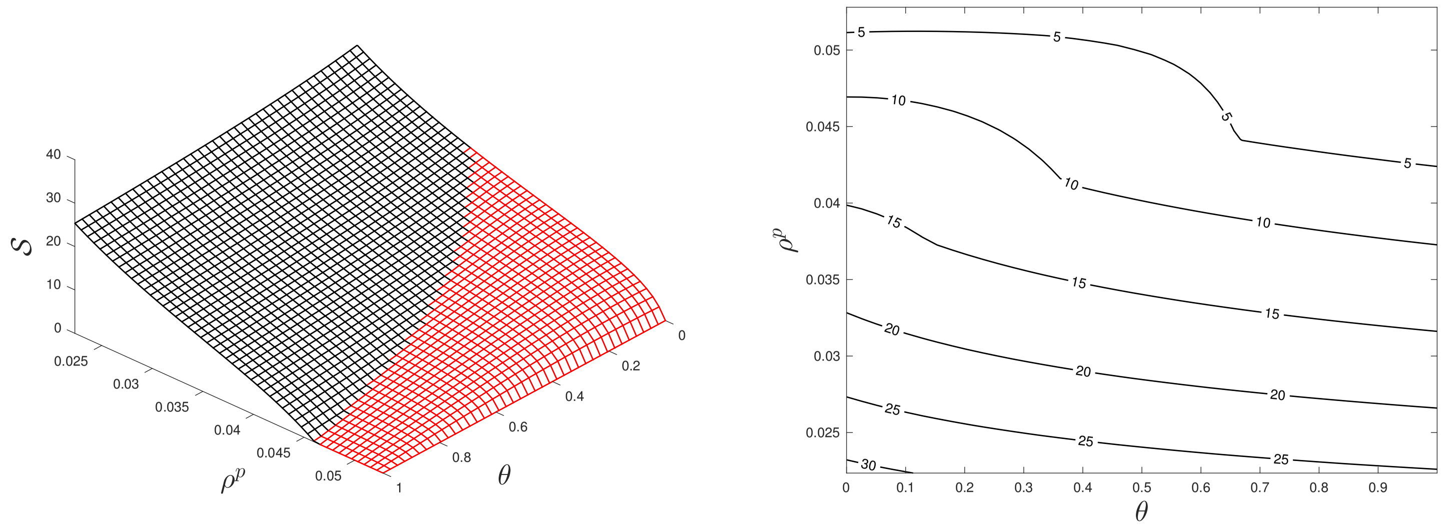

Figure 9: Semi-elasticity of the asset price with respect to the nominal policy rate as functions of θ and ρ^p in limiting economies with $\alpha \to 0$. The right panel shows the level sets for $\mathcal{S}$ corresponding to the left panel.