Queen's Economics Department Working Paper No. 1481

Money, Credit and Imperfect Competition Among Banks

Allen Head
Queen's University

Timothy Kam
Australian National University

Sam Ng
Australian National University

Isaac Pan
University of Sydney

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

2-2022

Money, Credit and Imperfect Competition Among Banks *

Allen **Head**[†] Timothy **Kam**[‡] Sam **Ng**[§] Isaac **Pan**[¶]

February 7, 2022

Abstract

Using micro-level data for the U.S., we provide new evidence—at national and state levels—of a positive (negative) relationship between the standard deviation (coefficient of variation) and the average in bank lending-rate markups. In a quantitative theory consistent with these empirical observations, banks' lending market power is determined in equilibrium and is a novel channel of monetary policy. At low inflation, banks tend to extract higher markups from existing loan customers rather than competing for additional loans. As a result, banking activity need not be welfare-improving if inflation is sufficiently low. This result speaks to concerns regarding market power in the banking sectors of low-inflation countries. Normatively, under a given inflation target, welfare gains arise if a central bank can use additional liquidity-provision (or tax-and-transfer) instruments to offset banks' market-power incentives.

JEL codes: E41; E44; E51; E63; G21
Keywords: Banking and Credit; Markups Dispersion; Market Power; Stabilization Policy; Liquidity.

*We thank seminar participants at UC Irvine, the ANU, Queen's University, Keio University, Sungkyunkwan University, UWA, the Virtual Australian Macroeconomic Seminars, the Sogang University Workshop on Search and Matching in Money and Finance, the 2020 Summer Workshop on Money, Banking, Payments and Finance in Gerzensee, the Canadian Macro Study Group, and the 2021 North American Summer and the European Winter Meetings of the Econometric Society. We also thank Nicolas de Roos for a very helpful discussion. Allen Head and Timothy Kam are grateful for funding by the Australian Research Council Discovery Project grant DP210101688. Kam also thanks to the Queen's University Weatherall Fellowship and ARC Discovery Project grant DP180103680 for initial financial support.

[†]Queen's University. E-mail: `heada@econ.queensu.ca`
[‡]Australian National University; Sungkyunkwan University. E-mail: `tcy.kam@gmail.com`
[§]Australian National University. E-mail: `samiengmanng@gmail.com`
[¶]University of Sydney. E-mail: `guangqian.pan@sydney.edu.au`

1 Introduction

Throughout the world, there is increased policy interest in the link between industry market power and monetary policy (see, e.g., Duval, Davide Furceri, Lee and Tavares, 2021). In particular, there are also concerns about market power in the banking sector (see, e.g., in Australia, Canada and the U.S., respectively, Sims, 2016; Productivity Commission, 2018; Wilkins, 2019; Executive Order 14036, 2021). For example, the U.S. President Biden (Executive Order 14036, 2021) has recently called for, *inter alia*, a review of banking-sector market power with the following motivation:

> [O]ver the last several decades, as industries have consolidated, competition has weakened in too many markets .... [F]ederal Government inaction has contributed to these problems, with workers, farmers, small businesses, and consumers paying the price. ...
>
> [I]n the financial-services sector, consumers pay steep and often hidden fees because of industry consolidation.

Likewise, in other countries such as Australia. Sims (2016), Chairperson of the *Australian Competition and Consumer Commission*, commented that:

> [I]t is the way market participants gain, maintain and use their market power that may lead to poor consumer outcomes. ...Reforms that alter *incentives of banks* [*sic*] ... [a]imed directly at bolstering consumer power in markets, and reforms to the governance of the financial system, should be the prime focus of policy action.

Consider the U.S. as an example. There are high profit margins in the U.S. banking sector (with markups of around 90%) and evidence of imperfect interest-rate pass-through (with a Rosse-Panzar H-statistic of 50%).[1]

Market power among banks matters for the design and effectiveness of monetary policy. We focus on the role of banks in insuring against liquidity risk when they have market power and ask if and how monetary policy should take this into account. Specifically, in

[1] See Corbae and D'Erasmo (2015, 2018). The Rosse-Panzar H-statistic measures the degree of competition in the banking market. It measures the elasticity of banks' revenues relative to input prices. Under perfect competition, the H-statistic equals one since an increase in input prices raise marginal cost and total revenue by the same amount. Under a monopoly, an increase in input prices raises marginal costs, lowers output and revenues, so that the H-statistic is less than or equal to zero. When the H-statistic is between zero and unity, one usually presumes a monopolistically competitive industry.

It is also well documented that there is substantial concentration in the banking industry in all developed countries. For example, post-2007, the market share of the top-three banks in Germany is about 78%; in the U.K., this is roughly 58%; in Japan, this is 44%; and, in the U.S., the corresponding share is 35%. These statistics are averages across annual 2007-2019 time-series data (available from Bankscope, 2020). However, we will not focus on addressing such measures of market power in this paper as the notions of market share and oligopolistic market structure are beyond the scope of our theory.

a setting where bank market power is itself an *equilibrium phenomenon* we ask: How does imperfect competition among lenders affect the *distribution of loan rates* and the pass-through of monetary policy to loan rates? And, in the presence of market power, is it always the case that banks—as insurers of liquidity risk—improve economic welfare? Finally, if market power in banking responds endogenously to policy, then should monetary policy attempt to temper the response of banks' markups in the presence of aggregate demand shocks?

Empirical evidence using bank branch pricing data. We provide new empirical evidence using publicly-available U.S. data and loan-rates data constructed from the level of bank branches. Our focus is on the relationship between loan rate markups and their dispersion. First, we measure dispersion in loan rates for identical consumer loan products, controlling for geography and other confounding factors.[2] We refer to the remainder or unexplained dispersion as *residual* or *orthogonalized dispersion* in loan rates. Second, at national and state levels, we find evidence of a positive (negative) relationship between the standard deviation (coefficient of variation) of bank lending rate markups and their averages.[3]

A theory of imperfectly competitive bank lending. The evidence motivates our theory; a general equilibrium monetary economy with market power in banking and *ex-post* heterogeneity (*i.e.*, dispersion) in loan rates both of which arise in equilibrium and respond to policy. There is imperfect pass-through of monetary policy to loan rates in equilibrium and an active stabilization monetary policy which commits to redistributing liquidity can be welfare improving.

Our baseline model inherits the following ingredients from Berentsen, Camera and Waller (2007) (BCW): In each period, there are two sequential markets. We will follow the New Monetarist tradition and label the first market as the *decentralized market* (DM). In the DM, agents are anonymous and lack any commitment to honoring private contracts. This contractual friction is the source of market incompleteness and renders fiat money valuable for sustaining private exchange.

The second market is a *centralized market* (CM) that has Walrasian features. The CM enables *ex-post* heterogeneous agents to re-balance their asset portfolios through labor supply and consumption demand choices. Perfectly competitive firms hire labor to produce and sell a homogeneous good to consumers in the CM. In the DM, sellers transform their own labor services into goods on the spot, taking fiat money as payment from buyers of their goods. The government implements monetary and tax policies in the DM and CM. Absent banking

[2] Martín-Oliver, Salas-Fumás and Saurina (2007) and Martín-Oliver, Salas-Fumás and Saurina (2009) also find price dispersion in loan rates for identical loan products in the case of Spanish banks.

[3] Although not the focus of this paper, we have also documented similar evidence in terms of mortgage loan rates and their markups.

institutions, the model is isomorphic to a DM price-taking variant in Rocheteau and Wright (2005).

Our focus on banks as intermediaries between heterogeneous liquidity needs is similar to BCW. Heterogeneity among depositors and borrowers arises because *ex-ante* identical households experience idiosyncratic shocks that determine whether an agent is an *ex-post* "active" or an "inactive" buyer. "Inactive" buyers end up holding idle money which they can deposit in the banking system to avoid the inflation tax. "Active" buyers consume using their own money, and may also borrow additional money from banks. Thus, banks are institutions that help to insure individual liquidity risks by accepting nominal deposits and extending loan contracts in the DM. Repayments of loans and returns on deposits occur in the CM. Banks can perfectly monitor and enforce the terms of these contracts.

We depart from the environment studied in BCW with regard to our modeling of the banking system. Whereas they study a perfectly competitive credit market, in our environment, active buyers must search among banks for potential lines of credit. The search process we consider is an adaptation of the noisy consumer search model of Burdett and Judd (1983). Banks post loan rates anticipating that their potential customers will observe only a random sample of posted rates. In equilibrium, active buyers who have made contact with at least one lender, optimally chooses whether (and how much) to borrow. This equilibrium mechanism implies *ex-post* heterogeneity among *active* buyers depending on whether and at what rate they have the opportunity to borrow.[4] The model nests BCW's equilibrium with perfect competition in banking as the limit when all active buyers have at least two borrowing opportunities and monopoly banking as the other extreme when all active buyers have at most one such opportunity.

Calibrated to match aggregate money demand and the average loan markup, the model is consistent with observed relationships between loan rate dispersion and average markups as described above. Moreover, it exhibits a new, non-monotonic, channel through which monetary policy affects banking market power. At empirically plausible low inflation rates, dispersion in loan rates diminishes and banks tend to exploit their markup (*intensive margin*) more than their ability to attract borrowers (*extensive margin*). In this way banks effectively extract surplus from consumers in goods trades. As agents' need for insuring liquidity risk is low when inflation is low, welfare in an economy with active banks can be lower than in an environment without banking. This result speaks to why policymakers in many low-inflation countries may be particularly concerned with market power in the banking sector. As inflation rises, insurance of liquidity risk becomes more important. Moreover, banks respond to increased loan demand as households economize on cash balances by giving up

[4]For simplicity, our model features loan-rate markup heterogeneity driven purely from the loan-pricing or loan-demand aspect. Banks' cost of funds are homogeneous in equilibrium. We will delve into a particular modelling assumption that yields the latter result.

some of their markups in return for attracting more loan customers. In such cases, banks improve welfare just as they do in BCW's environment with perfect competition. In addition, if demand is low in the sense that not many consumers seek loans, then competition among banks along the extensive margin can lower the markup sufficiently to render banking welfare improving even at low inflation.

Since our model contains Bertrand pricing as a parametric limit, we can replicate the competitive banking setting of BCW as a special case. We can thus decompose agents' money demand into several components. One part captures the liquidity insurance role of banks—this is identical to that arising in the perfect competition case of BCW. We can, however, also isolate new marginal benefit and cost terms that capture the effects of equilibrium market power and its attendant loan-interest risk on agents' decisions to accumulate money.

A novel insight arises from this decomposition of money demand and we use it to show how policies that redistribute liquidity can affect agents' money and loan demands. With perfectly competitive banks redistributive tax instruments do not directly affect individual agents' money demand in equilibrium. Indirectly, however, they are crucial for counteracting sub-optimal interest rate movements by raising the competitive deposit rate when aggregate demand is low (see Berentsen and Waller, 2011).

This prompts us to analyze optimal stabilization policies in a version of the model with shocks to aggregate demand. This is in the same spirit as Berentsen and Waller (2011) and Boel and Waller (2019). In our environment, however, the optimal stabilization policy exploits the endogeneity of market power in banking. We deliberately shut down the fluctuations in the deposit rate that are the focus of BCW. This allows us to isolate the welfare improvement arising from the effect of stabilization policy on market power in bank lending. As in BCW, the policy works by redistributing *ex-post* heterogeneous agents' individual liquidity. Here, this effectively tempers banks' intensive-versus-extensive margin trade-off that drives changes in both the dispersion and average level of markups. The optimal policy outcome is akin to the maintenance of an "elastic currency" as practiced by Central Banks, including the U.S. Fed, although banking sector market power may not historically have been an important motivation for such policies.[5]

Related literature. Several other recent papers also consider imperfect competition in banking. Drechsler, Savov and Schnabl (2017) show that banks' ability to mark down on deposits is empirically important. Choi and Rocheteau (2021) assume depositor private

[5]This harks back to the *Aldrich-Vreeland Act* of 1908. The Act was enacted to implement elastic or emergency currency in response to the Bankers' Panic or Knickerbocker Crisis of 1907. The Act also led to the creation of a decentralized Federal Reserve model under the *Federal Reserve Act* of 1913. In the official title of the *Federal Reserve Act*, one finds the phrase: "[A]n Act to provide for the establishment of Federal reserve banks, to furnish an elastic currency ... [*etc*] (*sic*)." (We thank Randy Wright for suggesting this interpretation.)

information in deposit-contracts bargaining in a setting where banks engage in second-degree price discrimination. The authors use this mechanism to rationalize the bank-deposit channel of monetary policy documented in Drechsler et al. (2017).

Alternatively, others have studied oligopoly in the banking industry. Corbae and D'Erasmo (2021) model a market structure where big banks interact with small fringe banks and other non-bank lenders. The theory can generate an empirically relevant bank-size distribution and shows how regulatory policies affect banking stability through the market structure. Altermatt and Wang (2021) show how an oligopolistic banking market structure affects both the monetary policy transmission mechanism and bank defaults. Dong, Huangfu, Sun and Zhou (2021), endogenize the number of banks. Chiu, Davoodalhosseini, Jiang and Zhu (2019) focus on oligopolistic competition on the deposit side to study the effects of central bank digital currency. Our approach complements these papers by accounting for imperfect competition on the lending side (see also, Allen, Clark and Houde, 2019; Clark, Houde and Kastl, 2021, for more evidence using mortgage data).

Also, our theory generates empirically relevant dispersion of loan rates and loan-rate markups. Most of these other models feature homogeneous bank lending rates. An exception is Corbae and D'Erasmo (2021) where banks face exogenous idiosyncratic and aggregate shocks. Banks in their real model help depositors economize on monitoring costs and diversify borrower idiosyncratic project risk. The authors focus on banking sector market power in a real model and study how market power impinges on financial stability. We consider a monetary model in order to connect monetary policy to market power and loan-rate markups dispersion. Dispersion arises in equilibrium loan pricing as a consequence of noisy search on the part of borrowers. Banks play a role of insuring against the costs of having idle liquidity—an issue that arises in monetary equilibria and pertains to monetary policy.[6]

In addition, a version of our economy *without* banks may attain higher welfare in a monetary equilibrium than our benchmark economy with banking. (This depends non-monotonically on long-run inflation.) Our approach thus allows us to study the welfare enhancing role of banking and how it may be diminished or even eliminated when the degree of market power in banking responds endogenously to monetary policy.

The remainder of the paper is organized as follows. In Section 2, we provide micro-

[6]Of course, there are other aspects of banking that help improve welfare. Chang and Li (2018) consider the same mechanism for banks but extend it to incorporate fractional reserves and liquidity buffers (see also, Kashyap, Rajan and Stein, 2002). This gives rise to a non-neutral liquidity channel of monetary policy in their model. Gu, Mattesini, Monnet and Wright (2013) consider a setting with limited commitment in exchange. In their model, banks improve welfare since limited commitment in private contractual obligations prevents more efficient allocative outcomes in the absence of banks. Also, bank liabilities can serve as payment instruments. He, Huang and Wright (2008) consider the safe-keeping role of banks when there is a risk of asset theft. These various reasons imply that banks can support a more efficient allocation in equilibrium. We eschew these factors in our model and focus solely on the role of banks as potential institutions for insuring private liquidity risk.

data evidence on the relationship between consumer loan-rate markups and the dispersion of these markups. In Section 3, we lay out the details of the model and its component decision problems of households, firms, government and banks. In Section 4, we describe the stationary monetary equilibrium of the model economy and discuss the new features that arise in our model. In Section 5, we study the insights from the model quantitatively by first calibrating it to U.S. data. Using numerical results, we illustrate the effects of equilibrium market power in the banking sector. Here we also provide an empirically testable prediction on loan-rate dispersion and markups. In Section 6, we study an optimal interest-rate (or inflation) policy alongside redistributive taxation as a function of banking-sector, loan-demand uncertainty. We conclude in Section 7.

2 Empirical Evidence

We begin by documenting new empirical relationships between measures of dispersion and the average level of markups in consumer loan rates in the U.S. data. We use data obtained from *RateWatch*.[7] In this data, we have information starting from a granular level; that of specific bank branches. We aggregate this to the national level in our main regression results, but find similar results at the state level.[8]

Empirical insights. Dispersion will be measured as the standard deviation or the coefficient of variation. The main empirical insights are as follows: First, there is a positive relationship between the standard deviation and the average level of markups at monthly frequency. Second, we find a negative relationship between the coefficient of variation of markups and their average at the national level. Our theoretical model will be able to rationalize these empirical facts as an additional, externally-valid results.

2.1 Data

Branch-level interest rate data. *RateWatch* provides monthly interest-rate data at the branch level for several types of consumer lending products. Our baseline analysis focuses on unsecured consumers lending within a particular class of loans. This strategy allows us to rule out issues of observed (and unobserved) heterogeneity across borrowers and loan products. Also, this measure is the most consistent with our theoretical model's setting where there is equilibrium rate dispersion for a single type of consumer loan product. Specifically, we choose the most commonly used product for personal loans: Personal Unsecured Loan for

[7] See https://www.rate-watch.com/.

[8] In theory, one could even do the empirical analysis at the bank-branch or county level. However, in practice, the information is too sparse at many branches or counties to be informative at such levels.

Tier 1 borrowers.[9] Our primary sample includes 496,942 branch-month observations from January 2003 to December 2017, involving 11,855 branches. To calculate each branch's markup against the Federal Funds Rate, we collect daily effective federal funds data from Federal Reserve H15 report.

Bank and county controls. We obtain commercial banks' information from their call reports. Specifically, we collect information on each commercial bank's reliance on deposit financing, leverage ratio, credit risk and bank size.

The Federal Deposit Insurance Corporation (FDIC) provides branch-level deposit holdings information, for all FDIC-insured institutions. This can be found in the Summary of Deposits (SOD) dataset. We use this data set to approximate each branch's local market competition and the impact of its commercial-bank-branch network. To control for potential local-market competition effects, we calculate each branch's deposit share in its county, the Herfindahl-Hirschman Index (HHI) in each county's deposit holdings and the number of branches in the county. To measure one branch's parent commercial bank's branch network, we calculate one branch's deposit share in its parent bank, the Herfindahl-Hirschman Index (HHI) of the commercial bank's deposit holdings and the number of branch counts in the commercial bank.

We also control for county-level socioeconomic information. This includes median income, poverty rate, population and average house price. This information is obtained from census data. We also have county-level unemployment and number of business establishments from the Bureau of Labor Statistics, county-level real GDP and GDP growth from the Bureau of Economic Analysis to control for local economic activities.

2.2 Loan Markups

We use two measures of lending-rate markups: (1) the raw markup of lending rates over the federal funds rate; and (2) a markup orthogonalized using a set of control variables.

Raw markup. We calculate each branch's markup relative to the federal fund rate. Specifically, the branch-level raw markup is calculated as

$$Markup_{b,i,c,s,t} = (Rate_{b,i,c,s,t} - FF_t)/(1 + FF_t). \tag{2.1}$$

In this definition, b stands for a bank branch, i for the parent bank to which the branch belongs, c for the county in which branch is located, s for the state and t for the date that

[9]As a robustness check, we also use mortgage rates as the alternative variable to calculate loan-rate markups. Specifically, we choose 30-Year Fixed Mortgage rate with an origination size of $175,000. Our key results still hold when we use mortgage rates. Our results continue to hold if we use rates on personal loans with different borrower qualities (*i.e.*, different borrower "tier" definition).

RateWatch reports the branch rate information.

Residual or orthogonalized markups. Why the use of residual markups? Branch-level loan pricing could simultaneously be explained by local socioeconomic factors, deposit market competition, bank-branch networks, characteristics of banks and other fixed effects. These factors could determine locally different demands for loans and costs of bank funds. However, these confounding features will not be captured in our simpler model structure. In our model, the distribution of loan-rate markups will result from the single feature of noisy consumer search in equilibrium. To have an empirical counterpart that is as consistent as possible with what our model can say, we need to focus on an empirical measure of "residual markups".

We orthogonalize the branch-level markup with respect to these potential factors to obtain a measure of our residual markup or orthogonalized markup. We use the following OLS regression to obtain the residual $\epsilon_{b,i,c,s,t}$:

$$Markup_{b,i,c,s,t} = a_0 + a_1 X_{b,i,c,s,t} + a_2 X_{i,t} + a_3 X_{c,s,t} + \epsilon_{b,i,c,s,t}. \tag{2.2}$$

Here, $X_{b,i,c,s,t}$ represents branch-specific control variables including local deposit market competition and bank branch networks, $X_{i,t}$ represents commercial bank control variables and $X_{c,s,t}$ represents county-level socio-economic control variables. We then re-scale $\epsilon_{b,i,c,s,t}$ to match the mean and standard deviation of raw markups in our full sample and use it as our alternative specification for the markup. A detailed summary of these factors can be found in our Online Appendix A.2 (Table A.2).

2.3 Markup dispersion and mean regression

We estimate OLS regressions of the dispersion of markups ($Dispersion_t$) on their monthly average ($\overline{Markup}_t$):

$$Dispersion_t = b_0 + b_1 \overline{Markup}_t + \epsilon_t. \tag{2.3}$$

In the Regression Model (2.3), b_1 is the coefficient of interest and standard errors are clustered by month. The average markup, $\overline{Markup}_t$, is defined for either case of the raw markup or the orthogonalized markup definition above. We consider two measures of markup dispersion: the monthly standard deviation of markups (SD_t) and monthly coefficient of variation (CV_t).

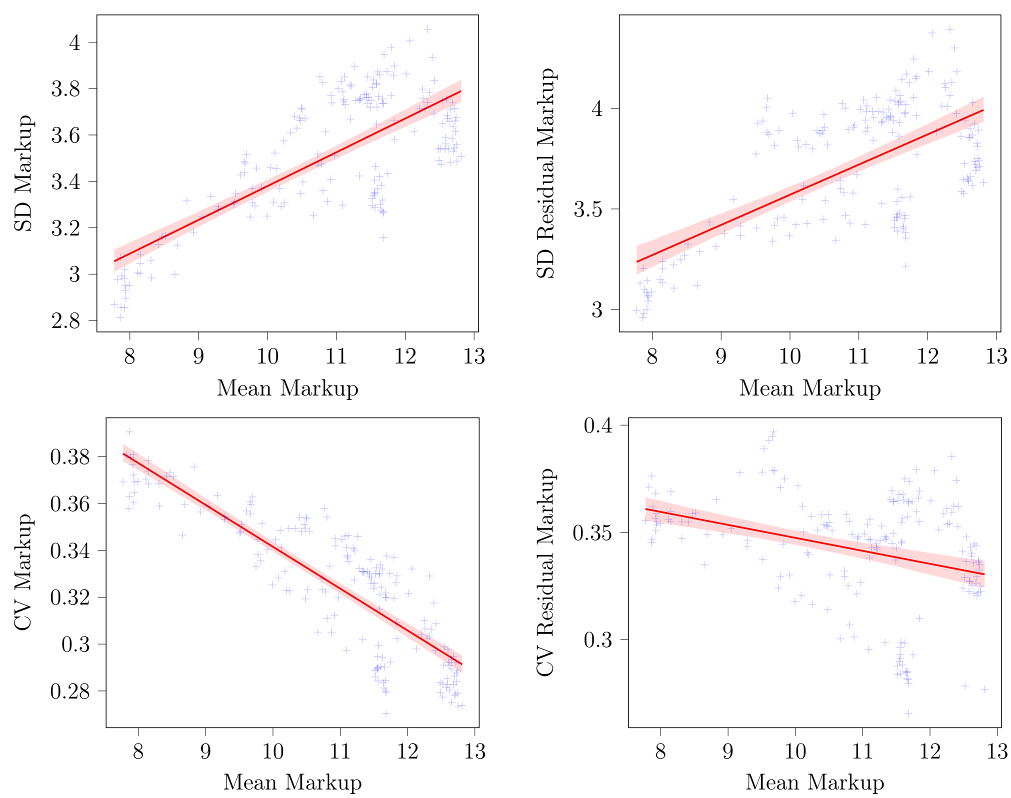
2.4 Results

First, we will illustrate our main empirical insights graphically using simple scatter plots. In Figure A.9, we have the correlations between our two measures of markup dispersion and the average markup.

Consider now the relationship between the standard deviation and mean of markups in the top row of Figure A.9. On the left is the relationship using the raw measure of markups. On the right is the same relationship for the orthogonalized version of markups—*i.e.*, our residual markup measure after controlling for various local, market, and social confounding factors. The standard deviations of both measures of markups are positively correlated with their averages. In particular, the correlation is 0.752.

The second row of Figure A.9, shows that the coefficients of variation of markups are negatively correlated with their averages. This holds for both raw (*left panel*) and orthogonalized (*right panel*) measures of markups. The correlation for the case of the raw markup is -0.857.

Figure 1: Markup dispersion measures and average at the national level (January 2003 to December 2017). Dispersion measures: SD (*standard deviation*) and CV (*coefficient of variation*). Data source: *RateWatch*, "Personal Unsecured Loans (Tier 1)." Least squares regression (line) with 95% (bootstrapped) confidence bands (shaded patches) superimposed.



Next, we reinforce the simple correlation insights from Figure A.9 by regressing our two measures of markup dispersion on markup averages, as described in Equation (2.3). The regression results are summarized in Table 1. From Columns (1) and (3) of Table 1, we can deduce a positive and statistically significant relationship between standard deviations and averages for the respective raw and orthogonalized markup measures. For the raw markup, the coefficient in Column (1) indicates that a one-percentage-point increase in the average markup is associated with a 0.146-percentage-point increase in its standard deviation. Alternatively, for the orthogonalized markup, Column (3) indicates that a one-percentage-point increase in the average markup is associated with a 0.192-percentage-point increase in the standard deviation. Similarly, from Columns (2) and (4), we can see negative and statistically significant relationships between the coefficients of variation and average markups, respectively, for the raw and orthogonalized measures of markups.

Table 1: Regression of markup dispersion on markup average (national-level data).

	Markup dispersion: $Dispersion_t$			
	Raw Markup		Orthogonalized markup	
	(1)	(2)	(3)	(4)
	SD_t	CV_t	SD_t	CV_t
$\overline{Markup}_t$	0.146***	−0.018***	0.192***	−0.014***
	(0.004)	(0.000)	(0.010)	(0.001)
Constant	1.924***	0.520***	1.621***	0.492***
	(0.039)	(0.004)	(0.111)	(0.009)
N	180	180	180	180
adj. R^2	0.554	0.733	0.333	0.259

Note: Standard errors in parentheses. $^{*}\, p < 0.1$; $^{**}\, p < 0.05$; $^{***}\, p < 0.01$

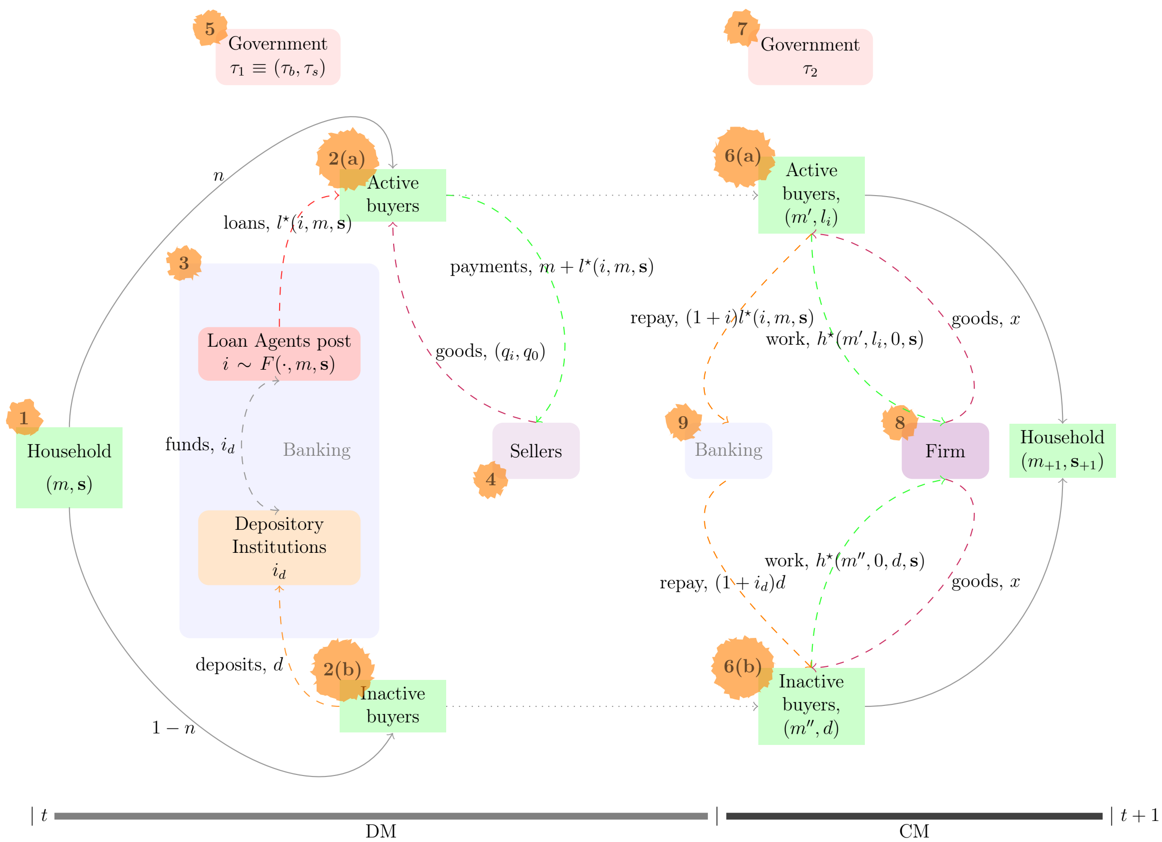
2.5 State-level analysis

We report the complementary, state-level results in our Online Appendix A.1. In particular, we consider markup dispersion at the state level by taking the standard deviation of branch markups from state s in month t. Consistent with the analysis at the national level, the standard deviations of markups are positively related to their average levels in the state-month panel data after controlling for state and time fixed effects. There is also a corresponding result in terms of the coefficient of variation that matches the insight at the national level. However, the estimates for this case are much noisier.

3 Model

Our model is motivated by the empirical evidence from Section 2 and encompasses the perfectly competitive banking model of Berentsen et al. (2007) (BCW) as a special case. In this environment, banks play the role of insuring individuals against liquidity risk: They take deposits from *ex-post* holders of idle money and make loans to those who require additional liquidity.[10] We generalize the loan market to one where there is noisy consumer search for loans (see, *e.g.*, Burdett and Judd, 1983). This new feature generates dispersion in loan rates in equilibrium and also results in an endogenous degree of banking market power that depends on government policy in equilibrium.

Figure 2: Agents, timing and actions in the sequential markets.



Overview. In the model, time is discrete and infinite. We use the following notational convention for date-dependent variables: $X \equiv X_t$ and $X_{+1} \equiv X_{t+1}$. As in Lagos and Wright (2005) or BCW, each date is divided into two sequential markets— respectively, a *Decentral-*

[10]Following BCW, we abstract from means of consumption smoothing other than banks and individually held money. In general, we could allow agents to own other assets (*e.g.*, claims to private equity or bonds). In order to rationalize equilibrium coexistence of fiat money alongside other asset claims, we could introduce costly asset liquidation in the frictional secondary asset market. This could be modelled, for example, as frictional over-the-counter trades as in Rocheteau and Rodriguez-Lopez (2014) and Duffie, Gârleanu and Pedersen (2005). This would render demand for multiple assets that have different liquidity premia in equilibrium. For the purposes of this paper, however, these are unnecessary features that would only obscure our main insights.

ized Market (DM) and a *Centralized Market* (CM). At the beginning of each date $t = 0, 1, ...,$ the aggregate stock of money in circulation is M. Let $\boldsymbol{\tau} = (\tau_b, \tau_s, \tau_2)$ denote a list of constant policy (tax or transfer) outcomes, where τ_b and τ_s are implemented in the DM and τ_2 is applied in the CM. We will denote the list of initial aggregate stock of money and policy outcomes by the vector notation $\mathbf{s} := (M, \boldsymbol{\tau})$. An individual with money balance m will have value and decision functions that depend on the vector $(m, \mathbf{s})$. The sequencing of events and actions are as follows (see also Figure 2):

1. A unit measure of households carry individual money balance, m, which is identical in equilibrium.

2. Each household observes the outcome of an individual shock:

   (a) With probability n, the household becomes an *active buyer*. That is, the household wishes to consume goods q early (in the DM). Since these agents are anonymous, they cannot trade with DM goods sellers on promises to repay in future states. Exchange will thus be supported by fiat money. They can top up their money balance by drawing from a line of credit with a *lending agent*, if they have succeeded in matching with such a lender.

   (b) With probability $1 - n$, the household becomes an *inactive buyer*. In this event, the household does not want to consume in the DM. However, the *ex-post* inactive buyer is left with idle money that is subject to inflation tax. If there is a banking sector, the inactive buyer can deposit their idle liquidity.[11]

3. The banking system is composed of perfectly-competitive *depository institutions* and imperfectly-competitive *lending agents*. Each of these groups is of measure one. The depository institutions take deposits from inactive buyers and commit to repay with nominal interest i_d. The depository institutions can also invest unused deposits in external markets at some nominal interest rate r. (In equilibrium, $r = i_d$.) They commit to supply funds to lending agents at the same rate and to return them to depositors in one period. Lending agents post their loan rate i, given their anticipation of being matched with active buyers and given the latter's optimal loan demands, $l^{\star}(i, m, \mathbf{s})$. Active buyers search for lines of credit from the lending agents. *Ex-post*, some may match with one, two or no lenders.

4. Sellers in the DM produce perishable goods on the spot and exchange these with active buyers who exhibit *ex-post* heterogeneity with regard to their access to loans at poten-

[11] Technically, this is a slight variation of the original setup of BCW. In BCW, there is an equivalent measure n of agents who become buyers in the DM. The remaining $1 - n$ become sellers in the DM. Here, we fix a unit measure of agents as always being sellers and re-label the $1 - n$ measure of agents as "inactive buyers". Substantively, this is still of the same form as BCW.

tially different rates. The buyers exchange some or all their liquidity for these goods given their optimal demands which vary with their borrowing costs.

5. In the DM, the government can tax from or transfer nominal amounts to households (*i.e.*, all buyer types) and sellers in the DM. These, respectively, are denoted by $\tau_b M$ and $\tau_s M$. (We will describe these policies in more detail later.)

6. In the CM, markets are Walrasian and perfectly competitive. The heterogeneous households supply different amounts of labor according to their optimal labor-supply rule $h^\star$ in a competitive labor market. This heterogeneity is a consequence of their different money-, loans- or deposit-balance outcomes by the end of the DM. These agents either repay their different loans (with interest) or collect their deposit returns. Households then combine their labor income and net wealth to purchase a general good x from a representative *firm* and to accumulate a money balance, m_{+1}, to carry into the following period.

7. In the CM, the goverment can also exact lump-sum taxes or transfers, $\tau_2 M$, on the various household types.

8. In the CM, a representative *firm* hires labor and produces the general good x using a linear technology.

9. Lending agents can enforce loan (plus interest) repayment and so the depository institutions have their funds returned. In turn, depository institutions pay back depositors (and external lenders) with interest at rate i_d.

This sequence of events repeats with households carrying m_{+1} at the start of date $t+1$. Following Lagos and Wright (2005) and BCW we assume households' utilities are quasi-linear so that all carry the same money balance, m_{+1} into period $t+1$. We now turn to detailed descriptions of the decision problems of each type of agent. In each case, we work backwards from the CM to the DM.

3.1 Households

Households' period utility is given by

$$\mathcal{U}(q,x,h) = u\left(q\right) + U\left(x\right) - h, \tag{3.1}$$

where $u(q)$ denotes the utility flow from consumption of the DM good q, $U\left(x\right)$ is the utility of consumption good x in the CM, and $-h$ is the disutility of labor. We assume that $u' > 0$, $u'' < 0$ and that u satisfies the usual Inada conditions. Likewise for U. For concreteness now, and anticipating the quantitative analyses later, we restrict our attention to the constant-relative-risk-aversion (CRRA) family of functions:

$$u(q) = \lim_{\hat{\sigma}\to\sigma} \frac{q^{1-\hat{\sigma}} - 1}{1-\hat{\sigma}}, \tag{3.2}$$

and, we will assume that $\sigma < 1$.[12]

3.1.1 Households in the Centralized Market

Consider a household at the beginning of the CM, with money, loan or deposit balances (m, l, d). Households discount payoffs between two time periods using subjective discount factor $\beta \in (0, 1)$. In the preceding DM, the agent may have been an active buyer (with $m \geq 0$, $l \geq 0$ and $d = 0$) or inactive buyer (with $l = 0$ and $m \geq d \geq 0$). Let V denote the value function of households at the beginning of a period (*viz.* the start of the next DM). Given her state, her maximal lifetime utility is given by

$$W(m, l, d, \mathbf{s}) = \max_{x,h,m_{+1}} \left[U(x) - h + \beta V(m_{+1}, \mathbf{s}_{+1})\right], \tag{3.3}$$

subject to

$$x + \phi m_{+1} = h + \phi m + \phi(1 + i_d)\, d - \phi(1 + i)\, l + \pi + T, \tag{3.4}$$

where ϕ is the date-t value of a unit of money in units of CM good x, i_d is the market interest rate on deposits d, i is some realized interest rate on the buyer's outstanding loan l, π is aggregate profit from bank ownership, and $T = \tau_2 M$ is any lump-sum tax or transfer from the government in the CM.

Using Equations (3.4) and (3.3), the problem may be rewritten as

$$W(m, l, d, \mathbf{s}) = \phi\left[m - (1 + i)\, l + (1 + i_d)\, d\right] + \pi + T + \max_{x,m_{+1}} \left\{U(x) - x - \phi m_{+1} + \beta V(m_{+1}, \mathbf{s}_{+1})\right\}. \tag{3.5}$$

The first-order conditions with respect to the choices of x and m_{+1}, respectively, are

$$U_x(x) = 1, \tag{3.6}$$

[12]This restriction is empirically motivated, as it is required to enable the model to fit long-run money-demand data well. We also consider the case of $\sigma > 1$ but we do not discuss it here for brevity. The knife-edge case of $\sigma = 1$ is not well-defined in terms of equilibrium characterization. The restriction with $\sigma < 1$ is also the case studied by Head, Liu, Menzio and Wright (2012).

and,

$$\beta V_m(m_{+1}, \mathbf{s}_{+1}) = \phi, \tag{3.7}$$

where $V_m(m_{+1}, \mathbf{s}_{+1})$ is the marginal value of an additional unit of money taken into period $t+1$. The envelope conditions are

$$W_m(m, l, d, \mathbf{s}) = \phi, \quad W_l(m, l, d, \mathbf{s}) = -\phi(1+i), \quad \text{and,} \quad W_d(m, l, d, \mathbf{s}) = \phi(1+i_d). \tag{3.8}$$

Note that $W(\cdot, \mathbf{s}_{+1})$ is linear in (m, l, d) and optimal decisions characterized by Equations (3.6) and (3.7) are independent of the agent's wealth.

3.1.2 Households in the Decentralized Market

Consider working backwards in the DM. We begin with the heterogeneous households' problem, after these agents have realized their status as active buyers (who have or have not matched with various lending agents) and inactive buyers.

***Ex-post* inactive buyers.** Conditional on being inactive in the DM (with probability $1-n$), a household with money holdings, m, can deposit d of this money with the depository institutions. She then has the *ex-post* value of continuing to the CM (in the same period), $W(m + \tau_b M - d, 0, d, \mathbf{s})$.

***Ex-post* active buyer with no line of credit.** First, consider an active household that has not succeeded in meeting a lending agent. In such an event, there is no possibility of borrowing additional money. Let p denote the money price of a DM good. *Ex post*, such a buyer has the following value:

$$B^0(m, \mathbf{s}) = \max_{0 \leq q_b \leq \frac{m+\tau_b M}{p}} \left\{ u(q_b) + W(m + \tau_b M - pq_b, 0, 0, \mathbf{s}) \right\}. \tag{3.9}$$

Using Equation (3.2), the agent's optimal demand for goods can be derived as

$$q_b^{0,\star}(m, \mathbf{s}) = \begin{cases} \frac{m+\tau_b M}{p} & \text{if } p < \hat{p} \\ (p\phi)^{-1/\sigma} & \text{if } p \geq \hat{p} \end{cases}, \tag{3.10}$$

where the cut-off price level, $\hat{p}$, will be determined in equilibrium and equals

$$\hat{p} \equiv \hat{p}(m, \mathbf{s}) = \phi^{\frac{1}{\sigma-1}} (m + \tau_b M)^{\frac{\sigma}{\sigma-1}}. \tag{3.11}$$

***Ex-post* active buyer with line(s) of credit.** Next, consider the post-match value of a buyer who has contacted at least one lending agent:

$$B(m,\mathbf{s}) = \max_{q_b \leq \frac{m+l+\tau_b M}{p}, l\in[0,\bar{l}]} \left\{u(q_b) + W(m+\tau_b M + l - pq_b, l, 0, \mathbf{s})\right\}. \tag{3.12}$$

In our environment, $\bar{l} = \infty$. This implies that loan contracts are perfectly enforceable, as in the baseline case of Berentsen et al. (2007).[13]

Using the Karush-Kuhn-Tucker conditions derived from Equation (3.12), we obtain the demands for special goods and loans. The demand for the special good is given by:

$$q_b^\star(m,\mathbf{s}) = \begin{cases} [p\phi(1+i)]^{-1/\sigma} & \text{if } 0 < p \leq \tilde{p}_i \text{ and } 0 \leq i \leq \hat{i} \\ \frac{m+\tau_b M}{p} & \text{if } \tilde{p}_i < p < \hat{p} \text{ and } i > \hat{i} \\ (p\phi)^{-1/\sigma} & \text{if } p \geq \hat{p} \text{ and } i > \hat{i} \end{cases}, \tag{3.13}$$

where

$$\hat{p} \equiv \hat{p}(m,\mathbf{s}) = \phi^{\frac{1}{\sigma-1}}(m+\tau_b M)^{\frac{\sigma}{\sigma-1}} \qquad \text{and} \qquad \tilde{p}_i = \hat{p}(1+i)^{\frac{1}{\sigma-1}}, \tag{3.14}$$

respectively, correspond to a maximal DM price at which the agent will use both his own liquidity and credit line from the bank, and, a maximal price at which the agent's purchase results in her being liquidity constrained. Since $\sigma < 1$, we have: $0 < \tilde{p}_i < \hat{p} < +\infty$.

The maximal interest rate at which a buyer is willing to borrow is given by

$$\hat{i} \equiv \hat{i}(m,\mathbf{s}) = (p\phi)^{\sigma-1}[\phi(m+\tau_b M)]^{-\sigma} - 1. \tag{3.15}$$

For any interest rate $i \in [0,\hat{i}]$, the buyer's loan demand is:

$$l^\star(i,m,\mathbf{s}) = \begin{cases} p^{\frac{\sigma-1}{\sigma}}[\phi(1+i)]^{-\frac{1}{\sigma}} - (m+\tau_b M) & p \in (0,\tilde{p}_i];\ i \in [0,\hat{i}] \\ 0 & p \in (\tilde{p}_i,\hat{p});\ i > \hat{i} \\ 0 & p \geq \hat{p};\ i > \hat{i}. \end{cases} \tag{3.16}$$

From the respective first cases of Equations (3.13) and (3.16), we can see that if the DM

[13] We can also consider the case with endogenous borrowing limits by solving a problem of limited commitment to repay loans. This was considered in an extended model in Berentsen et al. (2007). Unlike BCW, in our setting, bank-specific lending limits would have to be determined simultaneously with the equilibrium distribution of loan rates. In either case, there is no default in equililibrium. We could also consider exogenous default on loan repayments in the CM by parametrizing such events. However, since we considered loan products that have high-quality borrowers (*i.e.*, negligible default rates) in our empirical evidence, we avoid the from complication having to do with the possibility of default here.

good's relative price ($p\phi$) and interest on bank loans (i) are sufficiently low, the agent borrows to top up her own money balance and her goods and loan demands are decreasing in both i and $p\phi$. If, however, the special good's relative price and interest on borrowing are higher (*i.e.*, the intermediate case), the agent prefers not to borrow, but rather to spend all her money on the special good and be liquidity constrained. In this case the loan rate does not matter for demand. Finally, if $p\phi$ and i are sufficiently high, the buyer prefers not only not to borrow but also not to spend all her money balance on the special good. The cutoff prices $(\hat{p}, \tilde{p}_i, \hat{i})$ are functions of the state of the economy and public policy.

Households in the DM *ex-ante*. Now consider the beginning of period t when households are *ex-ante* homogeneous before the start of the DM. Given the money balance, m, such a household has maximal expected lifetime utility

$$V(m, \mathbf{s}) = n \left\{ \alpha_0 B^0(m, \mathbf{s}) + \alpha_1 \int_{[\underline{i}, \bar{i}]} B(i, m, \mathbf{s})\, \mathrm{d}F(i, m, \mathbf{s}) \right.$$
$$\left. + \alpha_2 \int_{\underline{i}(m,\mathbf{s})}^{\bar{i}(m,\mathbf{s})} B(i, m, \mathbf{s})\, \mathrm{d}\left[1 - (1 - F(i, m, \mathbf{s}))^2\right] \right\} + (1 - n) W(m + \tau_b M - d, 0, d, \mathbf{s}). \tag{3.17}$$

Conditional on being an active DM buyer with probability n, a household searches for a bank to obtain a line of credit, taking the distribution $F(\cdot, m, \mathbf{s})$ of banks' posted loan rates i as given. The function F will be an equilibrium object.[14] With probability $\alpha_0 \in (0, 1)$, the buyer fails to find a bank. Her value is then $B^0(m, \mathbf{s})$, which is given in Equation (3.9). With probability $\alpha_1 \in (0, 1 - \alpha_0)$, the buyer makes contact with one bank. Her ex-post value is then $B(i, m, \mathbf{s})$, from Equation (3.12), where i is a single draw from the distribution F. With probability $\alpha_2 = 1 - \alpha_0 - \alpha_1$, the buyer has two independent, randomly drawn meetings with two banks. Her ex-post value is then $B(i, m, \mathbf{s})$, again given by (3.12), but where i is the lower of two rates drawn from $F(\cdot, m, \mathbf{s})$, distributed as $1 - (1 - F(\cdot, m, \mathbf{s}))^2$.[15]

3.2 Competitive firms and sellers.

Firms in the Centralized Market. A unit measure of perfectly-competitive firms convert total labor supplied into the general good, x, and sell it to households in the CM. These firms

[14] We assume for now a compact support for $F(\cdot, m, \mathbf{s})$ as $\left[\underline{i}(m, \mathbf{s}), \bar{i}(m, \mathbf{s})\right]$. This will be a result proved in Lemma 13 in Online Appendix A.3.6.

[15] One can endogenize the maximal number of lenders that can be contacted by introducing a convex cost of search (see also, Burdett and Judd, 1983). Also, the contact probabilities α_k can be endogenized through a search-intensity decision margin (see, *e.g.*, Wang, 2016). Since these additional features do not really alter the insights, we maintain the simpler Burdett and Judd (1983) noisy-search setup here.

hire labor competitively in a spot labor market in the CM. Total CM labor supply is described in Section 4 below.

Sellers in the Decentralized Market. In the DM, a unit measure of sellers of goods behave much like households, except that sellers can produce the DM good on demand and do not value consuming it. DM sellers are analogous to Walrasian price-taking producers in Rocheteau and Wright (2005). Each DM seller has valuation

$$S(m, \mathbf{s}) = \max_{q_s} \{-c(q_s) + W(m + \tau_s M + pq_s, 0, 0, \mathbf{s})\} . \tag{3.18}$$

Here, $c(q)$ represents the cost of producing quantity q of goods, where $c(0) = 0$, $c_q(q) > 0$ and $c_{qq}(q) \geq 0$. The sellers' optimal production plan satisfies

$$c_q(q_s) = p\phi. \tag{3.19}$$

That is, the DM sellers produce to the point where the marginal cost of producing good q_s equals its relative price. It is straightforward to show that in equilibrium their valuation will be $S(0, \mathbf{s})$ at the start of each DM—*i.e.*, they optimally carry no money into the DM.

3.3 Banking

We discuss the two components of the banking sector: depository institutions and lending agents, in turn. The focus in this paper is on the nature of competition in lending, and so we will use the term *lending agents* and *banks* interchangeably from here on.

3.3.1 Depository institutions

First, consider the interaction between households and firms on the one hand and depository institutions on the other. Depository institutions take deposits and lend them to lending agents in the DM. In the CM, these depository institutions can perfectly enforce repayment of these loans and commit to return both deposits and contracted interest to individual depositors. In addition, these institutions can invest un-loaned deposits in a competitive external market. (As the lending side is imperfectly competitive with some lending agents meeting more, just enough, or too few borrowers, it may the case that not all deposits in the aggregate are lent by lending agents in equilibrium.) Such investments pay an exogenously specified nominal return, r.

Previewing the equilibrium we will consider, because these institutions behave competitively, they will be willing both to promise all depositors a gross return on deposits of $1 + i_d = r$, and to make advances to lending agents in exchange for repayment in the CM at gross rate r as well. Normally, only those households who are uninterested in consuming the special good this period will deposit. Other households will want to keep their money to make purchases in the DM (and possibly to borrow more, see below), and DM sellers will not carry money into the period at all. Note that the choice of r is exogenous, with this being enabled by our "small open economy" assumption. The value of r chosen corresponds to that arising in the baseline version of Berentsen et al. (2007), which features a perfectly competitive banking system. As such, households will be insured against holding idle balances here to the same extent that they are in that environment.[16]

3.3.2 Lending agents (banks)

Lending agents (*viz.* banks) contract with prospective borrowers before the latter trade with DM sellers and can enforce repayment (by households) of loans in the CM. These "banks" behave in a manner similar to that of sellers in the basic model of Burdett and Judd (1983). That is, they post a lending rate i, and commit to satisfying the demand for loans at that rate.

Households, who *ex-post* realize their active-buyer status, randomly pick a sample of posted loan rates. They are able to borrow the amount they desire at the lowest rate they observe.[17] We restrict attention to cases in which with probability α_k a prospective lender observes $k \in \{0, 1, 2\}$ quoted lending rate(s). On successfully making contact with a borrower, the lending agent is able to obtain funds from depository institutions at gross marginal cost $r = 1+i_d$. The details of the interest rate (price)-posting problem are described below. Again previewing equilibrium, lending agents will, on average, earn positive profits as all posted lending rates in equilibrium will exceed i_d. The lending agents can either retain these profits or return them lump-sum to households, firms or both in the CM.

Consider now the problem of a lending agent that takes the distribution of posted rates, F, as given and has marginal cost of funds i_d. If the bank posts a loan rate i, its expected profit is

$$\Pi(i, m, \mathbf{s}) = n\left[\alpha_1 + 2\alpha_2 \left(1 - F(i, m, \mathbf{s})\right) + \alpha_2 \zeta(i, m, \mathbf{s})\right] R(i, m, \mathbf{s}), \tag{3.20}$$

[16] We could also consider market power in deposit-rate setting, with dispersion in those rates, by replicating the noisy-search problem which we will describe on the lending side. To make tractable the two-sided search problem in such a setting, we can introduce a perfectly-competitive interbank market that intermediates between imperfectly-competitive depository agents and lending agents. This would expand our analysis but would not substantively change our results regarding market power and markup dispersion in lending.

[17] In principle DM sellers could also approach these lenders. However, they will have no need for loans in equilibrium. We therefore ignore this possibility directly in the model setup.

where

$$\zeta(i,m,\mathbf{s}) = \lim_{\varepsilon \searrow 0} \{F(i,m,\mathbf{s}) - F(i-\varepsilon,m,\mathbf{s})\}, \tag{3.21}$$

$$R(i,m,\mathbf{s}) = l^{\star}(i,m,\mathbf{s})[(1+i)-(1+i_d)], \tag{3.22}$$

$R(i,m,\mathbf{s})$ is the profit per customer served, $l^{\star}(i,m,\mathbf{s})$ is the demand for loans, and $n\alpha_2\zeta(i,m,\mathbf{s})$ is the measure of consumers that contact both this bank and another which has posted the same rate, i.[18]

With regard to banks' optimal choice of i, consider first a hypothetical bank serving borrowers who have contacted *only* this one bank. This bank's *realized* profit is

$$\Pi^m(i,m,\mathbf{s}) = n\alpha_1 R(i,m,\mathbf{s}), \tag{3.23}$$

where the superscript, m, denotes a bank serving only customers who observe a single rate—*i.e.*, a monopolist bank.

Second, consider a bank faced with customers who potentially observe more than one rate due to noisy search. The bank's realized profit is given by

$$\Pi^{\star}(m,\mathbf{s}) = \max_{i \in \text{supp}(F(\cdot,m,\mathbf{s}))} \Pi(i,m,\mathbf{s}) \tag{3.24}$$

subject to Equations (3.20), (3.21), (3.22) and (3.16).

As we restrict attention to linear pricing rules, we can prove that for any state and government policy, $\Pi^m(\cdot,m,\mathbf{s})$ is twice continuously differentiable, strictly concave and always positive-valued. Moreover, it can be shown that any bank facing more than one customer will also earn strictly positive profit. There is a maximal loan interest that is the smaller of either the monopolist's optimal loan rate $i^m(m,\mathbf{s})$ or the consumer's maximum willingness to pay $\hat{i}(m,\mathbf{s})$, where the latter depends on both the state and policy. The natural lower bound on loan rates is $i_d = (1+\tau)/\beta - 1$, and so the support of the distribution of posted loan rates is bounded. The support is also connected, as in Burdett and Judd (1983).[19]

From a bank's expected profit definition in Equation (3.20), it can be seen that each bank faces the following trade-off: It can raise its profit *per loan* by raising its loan rate (*i.e*, by increasing its *markup*). A bank can, however, raise the *measure* of borrowers it serves by lowering its posted rate. Since banks are *ex-ante* identical, we may think of the distribution $F(\cdot,m,\mathbf{s})$ as representing different pure-strategy choices or we may think of banks as mixing symmetrically over a range of interest rates that yield the same expected profit. In either

[18]We assume that in such cases prospective borrowers randomize between the two lenders. In equilibrium, the probability of a borrower observing two identical lending rates goes to zero.

[19]These results are derived formally in the Online Appendix in Sections A.3.1 to A.3.6.

interpretation, each borrower faces an individual distribution $F(\cdot, m, \mathbf{s})$ of random loan rates. We can derive a closed form for the distribution F as follows.

Lemma 1. *Suppose that the aggregate money stock grows by the factor $\gamma \equiv 1 + \tau > \beta$.*

1. *If $\alpha_1 \in (0,1)$, each borrower $(m, \mathbf{s})$ faces a unique non-degenerate, posted-loan-rate distribution $F(\cdot, m, \mathbf{s})$. This distribution is continuous with connected support:*

$$F(i, m, \mathbf{s}) = 1 - \frac{\alpha_1}{2\alpha_2}\left[\frac{R\left(\bar{i}, m, \mathbf{s}\right)}{R\left(i, m, \mathbf{s}\right)} - 1\right], \tag{3.25}$$

   where $\operatorname{supp}\left(F(\cdot, m, \mathbf{s})\right) = \left[\underline{i}(m, \mathbf{s}), \bar{i}(m, \mathbf{s})\right]$, $R\left(\underline{i}, m, \mathbf{s}\right) = \frac{\alpha_1}{\alpha_1 + 2\alpha_2} R\left(\bar{i}, m, \mathbf{s}\right)$ *and* $\bar{i}(m, \mathbf{s}) = \min\{\hat{i}(m, \mathbf{s}), i^m(m, \mathbf{s})\}$.

2. *If $\alpha_2 = 1$, then $F(\cdot, m, \mathbf{s})$ is degenerate at i_d:*

$$F(i, m, \mathbf{s}) = \begin{cases} 0 & \text{if } i < i_d \\ 1 & \text{if } i \geq i_d \end{cases}. \tag{3.26}$$

3. *If $\alpha_1 = 1$, $F(\cdot, m, \mathbf{s})$ is degenerate at the largest possible loan rate $\bar{i}$ such that*

$$F(i, m, \mathbf{s}) = \begin{cases} 0 & \text{if } i < \bar{i}(m, \mathbf{s}) \\ 1 & \text{if } i \geq \bar{i}(m, \mathbf{s}) \end{cases}. \tag{3.27}$$

We relegate the proof to Online Appendix A.3.7. This result is akin to the original notion of "firm equilibrium" in Burdett and Judd (1983, Lemma 2) and in the monetary version of Head and Kumar (2005, Proposition 3). For empirical relevance, we restrict attention to the first part of Lemma 1. That is, we focus on an equilibrium in which the distribution of rates is non-degenerate. With regard to the extent of market power, this case is sandwiched between the two familiar extremes: A Bertrand equilibrium and a monopoly-price equilibrium. These are described, respectively, in the second and third parts of Lemma 1.

3.4 Government

We maintain the notation for government policies from BCW. The government and monetary authority can make a lump-sum monetary injection or extraction in the CM, and can make targeted (positive or negative) transfers, to active and inactive households in the DM. These policy instruments are denoted as τ_1, and τ_2, in the DM and CM, respectively.

The total change to the money supply, $(\gamma - 1)M \equiv \tau M$, is split between DM and CM:

$$M_{+1} - M = (\gamma - 1)M = \tau_1 M + \tau_2 M, \tag{3.28}$$

where

$$\tau_1 := n\tau_b + (1-n)\tau_b + \tau_s \leq \tau. \tag{3.29}$$

The transfers $\tau_b M$ and $\tau_s M$ go to buyers and sellers in the DM, respectively.[20]

4 Stationary Monetary Equilibrium

We focus on a *stationary monetary equilibrium* (SME), in which the price level and money supply grow at the same constant rate: $\phi/\phi_{+1} = M_{+1}/M = \gamma \equiv 1 + \tau$. In this section, we characterize the components of a SME, with a particular focus on an equilibrium with money and credit. This is the equilibrium configuration that emerges in the calibrated model later.

As the price level $(1/\phi)$ is non-stationary, to obtain a well-defined stationary equilibrium we multiply nominal variables by ϕ. Let $z = \phi m$ and $Z = \phi M$ denote individual and aggregate real balances, respectively. Also, let $\rho = \phi p$ denote the relative price of DM to CM goods, and, $\xi = \phi l$ the real value of a loan. In a SME, DM sellers neither accumulate money in the CM nor borrow. Inactive DM households deposit all their money with depository institutions. Thus, we need only need consider the loan demand of active buyers. The stationary counterpart to the state-policy vector $(m, \mathbf{s})$ will now be $(z, \mathbf{z})$, where $\mathbf{z} = (Z, \boldsymbol{\tau})$.

4.1 The distribution of posted lending rates

Consider the case of $\alpha_1 \in (0, 1)$ as stated in Lemma 1. Rewriting the distribution of loan rates in terms of stationary variables, we have:

$$F(i, z, \mathbf{z}) = 1 - \frac{\alpha_1}{2\alpha_2}\left[\frac{R(\bar{i}, z, \mathbf{z})}{R(i, z, \mathbf{z})} - 1\right], \tag{4.1}$$

where $\text{supp}(F) = \left[\underline{i}(z, \mathbf{z}), \bar{i}(z, \mathbf{z})\right]$, $\underline{i}(z, \mathbf{z})$ solves

$$R(\underline{i}, z, \mathbf{z}) = \frac{\alpha_1}{\alpha_1 + 2\alpha_2} R(\bar{i}, z, \mathbf{z}), \qquad \bar{i}(z, \mathbf{z}) = \min\{i^m(z, \mathbf{z}), \hat{i}(z, \mathbf{z})\}, \tag{4.2}$$

[20] We restrict attention to identical transfers to all "active" and "inactive" buyers. This can be relaxed.

and,

$$R(i,z,\mathbf{z}) \equiv R(i,z,\mathbf{z}) = \left[\rho^{\frac{\sigma-1}{\sigma}}(1+i)^{-\frac{1}{\sigma}} - (z+\tau_b Z)\right](i-i_d) \tag{4.3}$$

is real bank profit per customer served.

We now have the following useful comparative static result regarding the relationship between household-level real balances and the distribution of posted lending rates:

Lemma 2. *Fix a long-run inflation rate $\tau > \beta - 1$, and let $\alpha_0, \alpha_1 \in (0,1)$. Consider any two real money balances z and z' such that $z < z'$. The induced loan-price distribution $F(\cdot, z, \mathbf{z})$ first-order stochastically dominates $F(\cdot, z', \mathbf{z})$.*

The proof can be found in Online Appendix A.6.1. In short, in a SME where households carry higher (lower) real balances into the DM, they are more (less) likely to draw lower loan-rate quotes, *ceteris paribus*. This reflects the fact that when potential borrowers carry low real balances into the period, demand for loans will be relatively high. All else equal, given strong loan demand, lending agents' optimal markups rise. Hence, the conclusion of Lemma 2.

4.2 Demand for money and bank credit

We now derive an equation describing CM agents' optimal money demand. A general expression for this is shown in Online Appendix A.4.[21] For clarity, we restrict attention to a stationary monetary equilibrium (SME) in which both *ex-ante* demand for money balances and *ex-post* demand for loans in the DM are positive. This will turn out to be the equilibrium configuration that emerges under our calibration when we consider a range of computational experiments later.[22]

Lemma 3. *Fix long-run inflation $\gamma \equiv 1 + \tau > \beta$ and let $\alpha_0, \alpha_1 \in (0,1)$. Assume that there is an SME in which real balances, $z^\star \in \left(0, \left(\frac{1}{1+\bar{i}(z^\star,\mathbf{z})}\right)^{\frac{1}{\sigma}}\right)$. Then,*

1. *the relative price of DM goods satisfies*

$$\rho = 1 < \tilde{\rho}_i(z^\star, \mathbf{z}) \equiv (z^\star)^{\frac{\sigma}{\sigma-1}}(1+i)^{\frac{1}{\sigma-1}}, \tag{4.4}$$

[21] This is done by taking the partial derivative of Equation (3.17) (i.e., marginal valuation of money) one period ahead, combining this with the first-order condition with respect to next-period money balance in Equation (3.7) and the optimal DM-good and loan demand functions in Equations (3.10) and (3.13).

[22] These equilibrium properties rely on sufficient conditions that are *per se* not dependent on model primitives. (See Proposition 5 further below for the details.) However, the sufficient condition is satisfied automatically in the computational experiments.

for any $i \in \text{supp}(F(\cdot; z^\star, \mathbf{z}))$*;* $\tilde{\rho}_i = \phi\tilde{p}_i$ *is the stationary transform of cut-off pricing function* $\tilde{p}_i$*, defined in Equation* (3.14)*;*

2. *loan demand is always positive; and,*

3. *money demand is given by the Euler equation*

$$\frac{\gamma - \beta}{\beta} = \underbrace{(1-n)i_d}_{[A]} + \underbrace{n\alpha_0\left(u'[q_b^0(z^\star, \mathbf{z})] - 1\right)}_{[B]} + \underbrace{n\int_{\underline{i}(z^\star,\mathbf{z})}^{\bar{i}(z^\star,\mathbf{z})} i\left[\alpha_1 + 2\alpha_2(1 - F(i, z^\star, \mathbf{z}))\right] dF\left(i, z^\star, \mathbf{z}\right)}_{[C]}. \tag{4.5}$$

The left-hand side of Equation (4.5) is the forgone nominal risk-free interest rate as a result of demanding additional money balance. The terms on the right-hand side of Equation (4.5) constitute the expected marginal benefit of holding money. The first term (A) is a benefit in terms of deposit interest earned by *ex-post* inactive buyers, just like that in Berentsen et al. (2007). The second term (B) is the liquidity premium on own money holding in delivering consumption value (in the case of no contact with a lender). The third term (C) comes about because an extra unit of money carried into the next period saves the agent an expected interest per unit of money not borrowed.

Since active DM buyers may fail to make contact with a lender with probability (α_0), and because it is costly to carry too much money into the next period ($\gamma > \beta$), agents may find it welfare improving *ex ante* to count on using bank credit to "top up" liquidity. This is in contrast to Berentsen et al. (2007) where the only potential gains from banking arise from the payment of interest to *inactive* buyers. Market power among banks manifests in loan-rate dispersion, *i.e.*, in a non-degenerate distribution of loan interest $F(\cdot, z, \mathbf{z})$. This tends to reduce the *ex-ante* value of real balances and be welfare reducing. As such, whether or not the presence of banking improves welfare is ambiguous.

To see this, after some algebra, we can rewrite Condition (4.5) as an asset pricing relation:

$$
\begin{aligned}
1 = &\underbrace{\frac{\alpha_0 \left(u'[q_b^0(z^\star, \mathbf{z})] - 1\right)}{i_d}}_{\text{Self-insurance: benefit-cost ratio}} \\
&+ \underbrace{\int_{\underline{i}(z^\star,\mathbf{z})}^{\bar{i}(z^\star,\mathbf{z})} \underbrace{\left[\alpha_1 + 2\alpha_2 \left(1 - F\left(i, z^\star, \mathbf{z}\right)\right)\right]}_{\text{Extensive margin}} \quad \underbrace{\left(\frac{i}{i_d}\right)}_{\text{Intensive margin}} \quad \mathrm{d}F(i, z^\star, \mathbf{z})}_{\text{Benefit-to-cost ratio from reduced borrowing} \equiv \text{Expected-transactions markup, } \hat{\mu}}.
\end{aligned} \tag{4.6}
$$

The left-hand side of this no-arbitrage condition in Equation (4.6) is the normalized, relative price of giving up CM consumption for real money balance today. On the right, we have the discounted expected real return from carrying money into the next-period DM. This return has two components inherited from the equivalent expression in Equation (4.5). These components—now measured relative to the opportunity cost of holding money (i_d)—are associated with the ability to consume without credit and with a reduced loan-interest burden, respectively. The latter term reflects the fact that households must consider both the cost of a particular loan (the *intensive* margin) and the likelihood of being able to choose among multiple lenders (the *extensive* margin).

Discussion: Two special cases. It is useful at this point to juxtapose our setting alongside two special cases. First, take Equation (4.5) to the special case of Berentsen et al. (2007) with perfectly competitive banks: If $\alpha_2 = 1$ and $\alpha_0 = 0$, then $F(\cdot, z, \mathbf{z})$ is degenerate on the singleton set $\{i = i_d\}$ (by Proposition 1). In this case, money demand is implicitly characterized by

$$
1 + \frac{\gamma - \beta}{\beta} = u'(q^{BCW}), \tag{4.7}
$$

where q^{BCW} will be a non-decreasing function of money balance in equilibrium. Note that our derivation of Equation (4.7) yields the same expression as Equation (22) in Berentsen et al. (2007) when we have linear production in the DM, *i.e.*, $c'(q_s) = 1$.

Second, consider a pure-currency economy without banks ($\alpha_0 = 1$). In this limit, agents must self-insure by carrying money. Money demand in this limit economy is implicitly described by

$$
1 + \frac{1}{n}\left(\frac{\gamma - \beta}{\beta}\right) = u'(q^{\text{no-bank}}). \tag{4.8}
$$

On the left-hand side of Equations (4.7) and (4.8) are the gross costs of carrying additional money into the DM, in the respective BCW and no-bank (or self-insurance) economies. On the right are their respective gross (utility) returns from consuming in the DM using their real money balance. Since perfectly competitive banks serve to insure agents who may end up with probability $(1-n)$ not requiring money, the liquidity premium on money $u'(q^{BCW})-1$ is exactly the competitive interest rate $i = i_d = (\gamma - \beta)/\beta$. However, in the case of a no-bank (or self-insurance) environment, the cost of carrying money is higher, as can be directly deduced by an additional term, $1/n > 1$, on the left-hand side of Equation (4.8). This raises the liquidity premium on money when agents cannot access banks to insure the risk of having idle liquidity. Thus, perfectly competitive banks would induce a higher consumption allocation than under self insurance if money has a return that is inferior to that on a risk-free asset. That is exactly why in Berentsen et al. (2007), banking always at least weakly improves the allocation and raises welfare.[23]

Comparing Equation (4.6) with (4.7), it is immediate that with noisy search, allocations will be dominated by the setting with perfectly competitive banks. However, things are a bit more subtle when we compare our setting of imperfectly competitive banks with that of a no-bank equilibrium. Under noisy consumer-loan search, holding less real money balance does not necessarily map into lower welfare since consumers can expand their consumption outcomes by borrowing from banks. However, more borrowing requirements entail more surplus being extracted by lenders with market power—*i.e.*, by holding more money balances agents are more likely to draw low rates (Lemma 2). To see these additional, opposing effects arising in an agent's *ex-ante* money balance decision, we can rewrite Equation (4.6) more compactly as:

$$1 + \frac{1}{\alpha_0}\left[1 - \hat{\mu}(\gamma)\right]\left(\frac{\gamma - \beta}{\beta}\right) = u'(q^{0,HKNP}), \tag{4.9}$$

where $q^{0,HKNP}$ is DM consumption of an *ex-post* active buyer who has no line of credit in our "HKNP" environment. (See Equation (4.12) later for its formula.) Notice that on the right-hand side of Equation (4.9), this is now merely the gross return to consuming with real balances in the event that the agent fails to make contact with any bank. The additional return from carrying money shows up in terms of loan-interest cost savings, now compactly tucked away in the *ex-ante* (or expected-transactions) markup term $\hat{\mu}$ on the left-hand side.

Now consider the left-hand side of Equation (4.9). Recall that since agents do not know *ex ante* the realization of their loan interest rate, their money balance decision must take into account the distribution of such rates. The latter is a consequence of equilibrium market power of banks when agents search noisily for credit. In Equation (4.9), this shows up as

[23] That is, comparing Equations (4.7) and (4.8), it is clear that $q^{BCW} > q^{\text{no-bank}}$ if the economy is always away from the Friedman rule (i.e., if $\gamma > \beta$). Since $n < 1$, the left-hand side of Equation (4.8) is a shift up of its counterpart on the left-hand side of Equation (4.7).

the distortion term $[1 - \hat{\mu}(\gamma)]/\alpha_0$ to the basic gross cost of money demand, $(\gamma - \beta)/\beta$. (The latter is the same term that appears on the left-hand side of Equation (4.7) in the perfect-competition case.)

On the one hand, inflation (γ) has a positive effect on the gross cost of holding money—*i.e.*, the term $(\gamma - \beta)/\beta$ on the left-hand side of Equation (4.9). We can visualize this effect as inflation shifting up the cost of holding money—*i.e.*, the graph of the left-hand side of Equation (4.9). On the other, since *ex-ante* markup $\hat{\mu}$ is strictly greater than one, the term $1 - \hat{\mu}(\gamma)$ is always negative. This tends to work in the opposite direction by shifting down the left-hand side of Equation (4.9). However, we can show that $\hat{\mu}(\gamma)$ is larger (smaller) for lower (higher) inflation (see Proposition 15 in Online Appendix A.7.) Hence, the net effect of long-run inflation, γ, is ambiguous. That is, there will be non-monotone effects of changes in the long-run inflation rate on relative money holdings and on welfare when comparing our noisy search economy to either the economy with perfectly-competitive banking or that with no banks. We will return to evaluate this new trade-off quantitatively in Section 5.

4.3 Goods market equilibrium in the DM and CM

The DM goods sellers are assumed to be Walrasian price takers. In equilibrium, DM sellers produce some level of q_s that attains marginal cost pricing: $\rho = c'(q_s)$. What is supplied, q_s, must equal aggregate demand in the DM for the good:

$$\begin{aligned} q_s(z, \mathbf{z}) \equiv c'^{-1}(\rho) = n\alpha_0 q_b^{0,\star}(z, \mathbf{z}) \\ + n\left[\int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} [\alpha_1 + 2\alpha_2 - 2\alpha_2 F(i, z, \mathbf{z})]\, q_b^{\star}(i, z, \mathbf{z})\, \mathrm{d}F(i, z, \mathbf{z})\right]. \end{aligned} \tag{4.10}$$

Given $x^\star = 1$, we can also verify that aggregate CM labor equals $x^\star$ due to the assumption of a linear CM production technology and market clearing in the labor and goods markets in the CM.

4.4 Equilibrium lending

In equilibrium lenders must earn non-negative profits. In aggregate, this requires that total interest collected on real loans ($\xi^\star(i, z, \mathbf{z})$) weakly exceeds that paid on total real deposits:

$$\begin{aligned} (1-n) i_d \delta^\star(z, \mathbf{z}) \equiv (1-n) i_d (z + \tau_b Z) \\ \leq n\left\{\int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} [\alpha_1 + 2\alpha_2 - 2\alpha_2 F(i, z, \mathbf{z})]\, i\, \xi^\star(i, z, \mathbf{z})\, \mathrm{d}F(i, z, \mathbf{z})\right\}. \end{aligned} \tag{4.11}$$

4.5 Summary: SME with money and credit

Definition 4. A *stationary monetary equilibrium with money and credit* is a steady-state allocation $(x^\star, z^\star, Z)$, allocation functions $\{q_b^{0,\star}(z^\star, \mathbf{z}), q_b^\star(\cdot, z^\star, \mathbf{z}), \xi^\star(\cdot, z^\star, \mathbf{z}\}$, and (relative) pricing functions $(\rho, F(\cdot; z^\star, \mathbf{z}))$ such that given government policy $\boldsymbol{\tau}$ satisfying Equation (3.29),

1. $x^\star = 1$;
2. $z^\star \equiv z^\star(\boldsymbol{\tau}) = Z$ solves (4.5);
3. given $z^\star$, $q_b^{0,\star}(z^\star, \mathbf{z})$ and $q_b^\star(\cdot, z^\star, \mathbf{z})$, respectively, satisfy
$$q_b^{0,\star}(z, \mathbf{z}) = \frac{z + \tau_b Z}{\rho}, \qquad \text{for } \rho < \hat{\rho}(z, \mathbf{z}), \tag{4.12}$$
and,
$$q_b^\star(i, z, \mathbf{z}) = [\rho(1+i)]^{-\frac{1}{\sigma}}, \qquad \text{for } 0 < \rho \leq \tilde{\rho}_i(z, \mathbf{z})(z, \mathbf{z}) \text{ and } 0 \leq i < \hat{i}(z, \mathbf{z}); \tag{4.13}$$
4. $\xi^\star(\cdot, z^\star, \mathbf{z})$ satisfies:
$$\xi^\star(i, z, \mathbf{z}) = \rho^{\frac{\sigma-1}{\sigma}}(1+i)^{-\frac{1}{\sigma}} - (z + \tau_b Z), \quad \text{for } \rho \in (0, \tilde{\rho}_i(z, \mathbf{z})], \quad i \in [0, \hat{i}(z, \mathbf{z})); \tag{4.14}$$
5. ρ solves (4.10);
6. $F(\cdot; z^\star, \mathbf{z})$ is determined by (4.1); and,
7. aggregate loans supplied is feasible according to (4.11).

Note that τ_s does not materially affect equilibrium determination, and so we can set $\tau_s = 0$ without changing our basic results. For the baseline calibration of the model, we also set $\tau_b = 0$, so that there is no redistributive tax or transfer policy in place. Later we will consider counterfactual analyses involving differential tax policies.

Under sufficient conditions, there exists a unique SME with money and credit:

Proposition 5. *Assume loan contracts are perfectly enforceable. If* $1+\tau \equiv \gamma > \beta$, $z^\star \in (0, \overline{z})$, *where* $\overline{z} = \left[1 + \bar{i}(z^\star, \mathbf{z})\right]^{-\frac{1}{\sigma}}$, *and there is a* $N(z^\star, \mathbf{z}) \in [0, 1]$ *such that* $n \geq N(z^\star, \mathbf{z})$, *then there exists a unique SME with both money and credit.*

For a proof, see Appendix A.6.4. Formal proofs of intermediate results can be found in Appendices A.6.1, A.6.2, and A.6.3. Here we sketch the basic idea. Fix $\gamma > \beta$. First we show that lending banks' posted loan-price distribution $F(\cdot, z, \mathbf{z})$ is decreasing (in the sense of first-order stochastic dominance) in households' real balance, z. The intuition is that as households carry more money into the DM, the marginal benefit of bank credit falls. See Lemma 2 for details. As such, households with higher z are more likely to observe a most-preferred interest rate that is not too high.

Second, with probability α_0, a household contacts no lending agent and so its marginal benefit from holding an extra dollar falls as real balances rise. Together, these factors establish that the right-hand side of (4.5) is a continuous and monotone decreasing function of z. Since the left-hand side of (4.5) is constant in z, there exists a unique real money balance $z^\star$ for a given $\gamma > \beta$.

The second condition ensures that $z^\star$ is bounded and that the maximal loan interest is not too high. This guarantees positive loan demand. The third condition requires that the measure of active DM buyers not be too small. Although neither of these conditions are determined solely by model primitives, but rather depend on equilibrium objects, we can easily verify them in our numerical calculations.

While the results obtain for $\gamma > \beta$, it is also of interest to consider the case of the Friedman rule ($\gamma = \beta$):

Proposition 6. *If $1 + \tau \equiv \gamma = \beta$, then there is no SME with loan interest rate dispersion. Moreover, if $\alpha_0 > 0$, the Friedman rule attains the first-best allocation $q^{\star,FB}$.*

The banking system is redundant at the Friedman rule, for the simple reason that it is costless to carry money across periods. As such households can insure themselves perfectly against the risk of not having trading opportunities. Therefore, there is no gain to redistributing liquidity across agents in an SME. From this point onward, we restrict attention to cases in which $\gamma > \beta$.

Remark. In a SME, we have that $z = z^\star(\boldsymbol{\tau}) = Z$, so we can collapse the state-policy vector $(z, \mathbf{z})$ into just $\mathbf{z}$ when we discuss results pertaining to an SME below.

5 Quantitative Analyses

In this section, we calibrate the baseline model to macro-level data and then use it to investigate the effects of various parameters and alternative policies. We also compare the model's predictions to micro-level empirical observations for external validity.

5.1 Baseline calibration

Our approach is to match the empirical money demand and the average lending markup (in percentage terms) in the macro data.[24] In our model, we measure the average (percentage) lending markup in a SME by

$$\mu(\gamma) = \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} \left[\frac{i}{i_d(\tau)} - 1 \right] \mathrm{d}F(i, \mathbf{z}), \tag{5.1}$$

where $\gamma = 1 + \tau$.

In terms of identification, the bank-contact probabilities (α_0, α_1) directly affect the lending-rate distribution, $F(\cdot, \mathbf{z})$, and thus banks' average markup relationship with their cost of funds i_d. The CM utility function, U, is assumed to be logarithmic. With quasi-linear preferences, real CM consumption is then given by $x^\star = \bar{U}_{CM}(U')^{-1}(A)$, where the scaling parameter, $\bar{U}_{CM}$, determines the relative importance of CM and DM consumption. The DM utility function is given by Equation (3.2). The DM production cost function is linear and has no parameter to be estimated. The parameters $(\bar{U}_{CM}, \sigma)$ are identified through the model-implied aggregate real money demand relationship with i_d.

We interpret a model period as a year and calibrate to annual data. The model has seven parameters: $(\tau, \beta, \sigma, \bar{U}_{\mathrm{CM}}, n, \alpha_0, \alpha_1)$. In the baseline setting, we assume that there are no redistributive policies so that the only policy in place is the long-term inflation target τ—*i.e.*, we set $\tau_b = \tau_s = 0$ and $\tau = \tau_2$.

External calibration. Some parameters can be determined directly by observable statistics. We use the Fisher relation to pin down the money growth rate (long-run inflation), τ, and discount factor, β. The share of inactive buyers (depositors) $\tilde{n} \equiv 1 - n$ is set to match the average share of household depositors with commercial banks per thousand adults in the United States.[25]

Internal calibration. We jointly choose the pairs $\left(\sigma, \bar{U}_{\mathrm{CM}}\right)$ and (α_0, α_1) to match, respectively, the aggregate relationships between nominal interest and money demand, and between nominal interest and average gross lending markup. These empirical relations are estimated by auxiliary fitted-spline functions. Intuitively, each pair of these parameters are identified by the shift (or position) and the overall shape of the respective spline approximants of the empirical relations.

[24]We use the $(\frac{\text{bank prime loan rate}}{\text{three-month U.S. Treasury Bill rate}}) - 1$ as a proxy for the average percentage markup in lending rates. We use the three-month T-bill rate to be consistent with the model and the empirical money demand data used in Lucas and Nicolini (2015). Alternatively, we could use the federal funds rate and this would not alter the general shape of the function.

[25]Source: FRED Series USAFCDODCHANUM, "Use of Financial Services—key indicators".

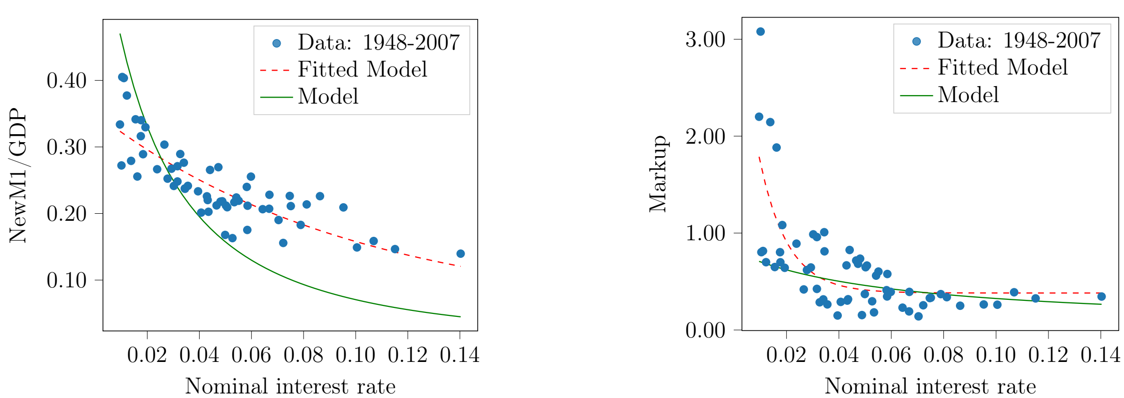
Our parameter values and targets are summarized in Table 2. We use annual data spanning from 1948 to 2007, so as to avoid both the Second World War and the Great Recession.[26] Figure 3 provides the respective scatterplots of the two empirical relationships (*blue circles*) just mentioned, the empirical spline models (*dashed-red lines*, "Fitted Model"), and our calibrated model's predictions (*solid-green lines*, "Model") for these relations.

Table 2: Calibration and targets.

Parameter	**Value**	**Empirical Targets**	**Description**
$1+\tau$	$(1+0.0382)$	Inflation rate[a]	Inflation rate
$1+i$	$(1+0.0481)$	3-month T-bill rate[a]	Nominal interest rate
β	0.9906	-	Discount factor, $\frac{(1+i)}{(1+\tau)}$
σ	0.525	Aux reg. $(i, M/PY)$[b]	CRRA (DM q)
$\bar{U}_{CM}$	0.79	Aux reg. $(i, M/PY)$[b]	CM preference scale
$\tilde{n}$	0.35	household depositors[c]	Proportion of inactive DM buyers
α_0, α_1	$0.04, 0.065$	Aux reg. (i,markup)[d]	Prob. $k = 0, 1$ bank contacts

[a] Annual nominal interest and inflation rates.
[b] Auxiliary statistics (data) via spline function fitted to the annual-data relation between the three-month T-bill rate (i) and Lucas and Nicolini (2015) New-M1-to-GDP ratio (M/PY).
[c] Household depositors with commercial banks per 1000 adults for the United States.
[d] Auxiliary statistics (data) via spline function fitted to the annual-data relation between the three-month T-bill rate (i) and average banking (gross) markup ratio, (bank loan prime rate)/i.

Figure 3: Aggregate money demand and average bank markup—model and data.



(a) Calibration and money demand data (b) Calibration and average (%) markup data

In Figure 3, it can be seen that our model's fit to aggregate money demand (i.e., the solid green graph in the *left panel*) is not absolutely perfect, especially at higher nominal interest rates. This is due to a tension between matching both real money demand and the average banking markup simultaneously. In the model, a higher nominal interest leads to both a reduction in real money demand and an increase in the cost of funds for banks (i_d).

[26] Note: data for the bank loan prime rate is only available from 1931 onward.

The latter effect reduces the average markup. In order to match the high average markup in data, DM buyers in the model would have to hold even lower real balances, given that real balances are inversely related to lending rates. Nonetheless, we view the model's fit under our benchmark calibration to be reasonable.

5.2 Comparative steady states

We now consider SMEs indexed by different steady-state rates of inflation, τ. We ask the following questions: First, what mechanisms are at work and how do they affect markups and the pass-through of monetary policy to loan rates? Second, what are the testable empirical predictions of these mechanisms? And finally, under what circumstances are agents *ex-ante* better off in an economy with banks than in one without them?

As noted above, in contrast to Berentsen et al. (2007), financial intermediation of the type studied here need not always be welfare improving. Moreover, in our theory the design of optimal cyclical liquidity depends directly on the long-run inflation target and indirectly on the effect of inflation on the state-dependent loan rate distribution, $F(\cdot, \mathbf{z})$.[27]

Banks' intensive-extensive profit margin trade-off. We begin with a discussion of the tradeoff faced by a lending bank in its rate setting decision. Figure 4 depicts realized profit per customer and posted loan rate densities for steady-state inflation rates at zero, five and ten percent. The *solid-red*, *dashed-green* and *dotted-blue* graphs are, respectively, associated with an annual inflation target of 0%, 5%, and 10%. (The width in the domains of these graphs indicate the shifts in the bounds of the distributions' supports in the respective experiments.)

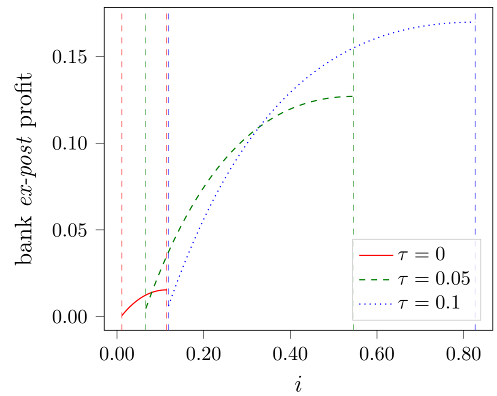
In Figure 4 we see that lenders trade off between profit per customer (the intensive margin) which is increasing in the posted loan rate, and, the number of customers that it successfully serves (the extensive margin) which is *decreasing* in the posted rate. As inflation, τ, rises, not only does the equilibrium support of F shift to the right, but the mass of the density also shifts rightward relative to the lower bound. We identify this latter effect as the extensive margin, as it implies that banks lose customers as they raise their lending rates.

Inflation and lending markups. We use the coefficient of variation of markups across banks as our measure of dispersion.

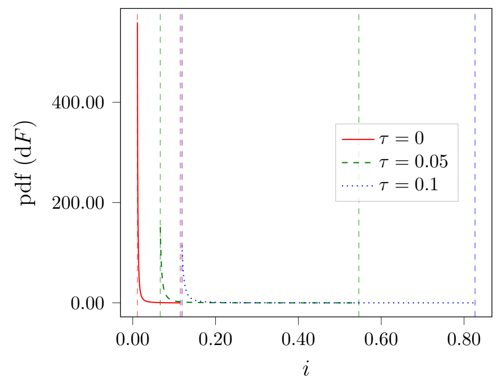
These two measures are depicted in the two upper panels of Figure 5. As an alternative, we also plot the standard deviation measure of dispersion in the lower panel of Figure 5. The relationship between average markup and these two markup dispersion measures agree with the empirical evidence in Section 2.

[27] We will discuss the optimal policy problem in Section 6.

Figure 4: Lenders' extensive-versus-intensive margin trade-off for various levels of long-run inflation, $\gamma = 1 + \tau$.

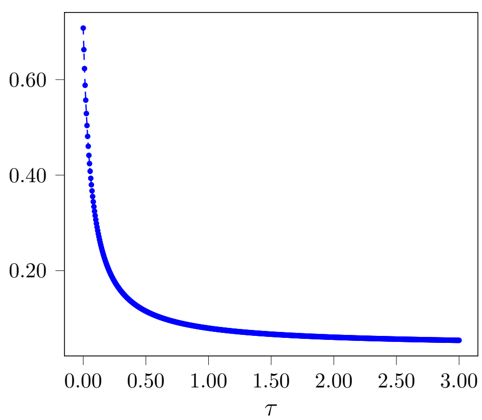


(a) intensive margin: profit per customer

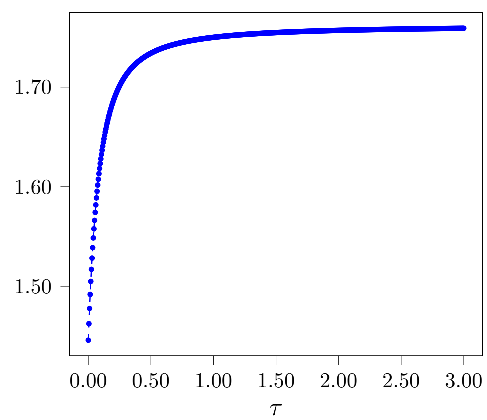


(b) extensive margin: loan-customer density

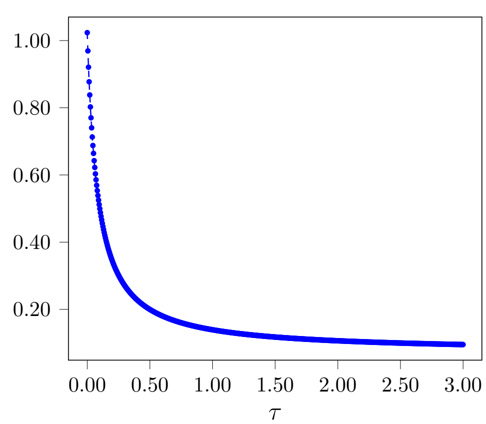
Figure 5: The effects of inflation on lending banks' market power for $\tau \in (\beta - 1, \bar{\tau}]$.



(a) loan (%) markups average



(b) loan (%) markups dispersion (CV)



(c) loan (%) markups dispersion (SD)

As trend inflation rises, market power measured by the average markup declines, and it declines especially sharply at low inflation. The average markup in Equation (5.1) is the ratio of two parts that are both going up with inflation. First, the numerator in Equation (5.1)—*i.e.*, the average loan rate—goes up with inflation. Second, the cost of funds for lending agents is just $i_d = (\gamma - \beta)/\beta$. Higher inflation ($\gamma$) clearly translates to higher i_d. For the result that average loan-rate markup falls with inflation, it must be that average loan rate itself is rising more slowly than the deposit rate. In our Online Appendix A.7, we identify sufficient conditions under which this can be proven.

On one hand, the average loan rate rises because higher inflation has a negative effect on real money balance (see Proposition 15 in Online Appendix A.7). Since the lending agents expect potential borrowers to have lower money balances when inflation is higher, the latter are likely to need to borrow more. Lenders thus have more incentive to charge higher loan rate and so average loan rates are higher with higher inflation. This reasoning also underlies the conclusion of Lemma 2.

On the other hand, borrowers demand fewer loans when loan rates they face are higher. Lenders must then "compete harder" for borrowers as inflation rises. Overall, the degree of competition measured in terms of average markups is falling with inflation.[28]

Non-monotone welfare consequences. The tension that we have discussed earlier in Section 4.2 leads to non-monotone contributions to welfare by bank lending. Recall that we contrasted Equation (4.5), or its equivalent in Equation (4.9), with the two extremes of a perfectly-competitive banking equilibrium and a no-bank equilibrium. The immediate (partial-equilibrium) suggestion from there was that the welfare-improving role of banks—away from the Friedman rule—may no longer be unambiguous when we have imperfect competition among banks.

The reason is that depending on long-run inflation ($\gamma = 1 + \tau$) there are offsetting forces. First, a lower γ will imply a higher z (in equilibrium). On the one hand, a higher z will increase consumption allocation overall, which should raise expected welfare. On the other hand, lower inflation (γ) will mean a higher markup (see Proposition 15 in Online Appendix A.7), with banks extracting more of the DM goods market surplus rent from active households

[28]We can see that the support of $F(\cdot, z, \mathbf{z})$ shifts right and becomes wider, reflecting an increase in dispersion. The scale-free, coefficient-of-variation measure in Figure 5 depicts exactly this point. Those banks posting the lowest rates, post closer to their marginal cost in an attempt to serve a large number of borrowers (many of whom have made contact with more than one bank). Meanwhile, those posting high rates (which for the most part serve only customers with no alternative) raise their rates by a large amount to take advantage of their borrowers' high marginal utility of consumption driven by their lower real balances. At higher inflation, the mass of $F(\cdot, z, \mathbf{z})$ spreads out from its lower bound so that posted interest rates become more dispersed. At the same time, that lower bound falls toward banks' marginal cost of funds. The average (percentage) markup is highest at low inflation (although it collapses to zero at the Friedman Rule) then converges back to zero as τ (and inflation) rises sufficiently.

that borrow and thus reducing welfare. Here, there is a new channel from monetary policy to welfare through the competitiveness of bank lending—whether it is welfare improving overall depends on the long-run inflation target, $\gamma = 1 + \tau$. Lower inflation exacerbates the welfare-eroding effects of bank lending market power while at the same time reducing the insurance benefits of banking.

Our welfare criterion is measured in terms of the *ex-ante* lifetime utility of homogeneous households, given a particular τ-policy-induced SME. In the baseline SME with noisy search for loans, the welfare function is

$$W^{HKNP}(\tau) = \frac{1}{1-\beta}\left[U(x^\star) - x^\star - c[q_s^\star(\mathbf{z})]\right] + \frac{n}{1-\beta}\left[\alpha_0 u[q_b^{0,\star}(\mathbf{z})] + \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} [\alpha_1 + 2\alpha_2 - 2\alpha_2 F(i,\mathbf{z})] u[q_b^\star(i,\mathbf{z})] \mathrm{d}F(i,\mathbf{z})\right], \tag{5.2}$$

where the functions $q_b^{0,\star}$, $q_b^\star$ and $q_s^\star$ are characterized by Equations (4.12), (4.13) and (4.10), respectively, and the equilibrium real money balance $z = z^\star = Z$ is dependent on policy τ. Also, recall that we have "switched off" any redistributive policies for now: $\tau_b = \tau_s = 0$ and $\tau = \tau_2$. For the no-bank equilibrium, the corresponding welfare function, $W^{\text{no-bank}}$, is equivalent to Equation (5.2) being set at the limit of $\alpha_0 = 1$ so that its last term on the right disappears. The BCW, perfect-competition case would have the welfare function

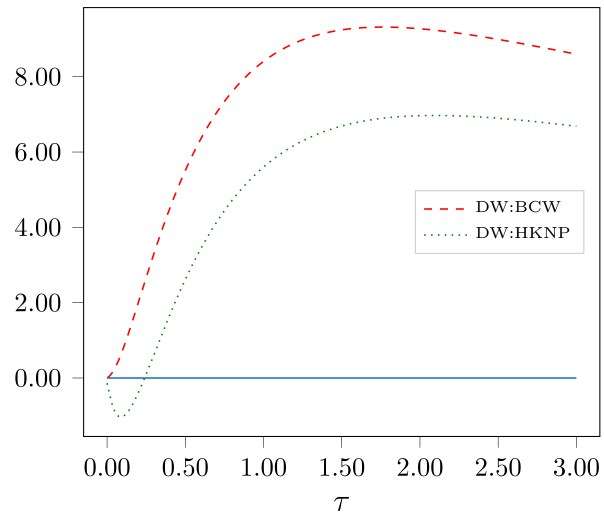
$$W^{BCW}(\tau) = \frac{1}{1-\beta}\left\{nu\left[q_b^\star(i_d, z^{BCW}, \mathbf{z})\right] - c\left[q_s^\star(z^{BCW}, \mathbf{z})\right] + U(x^\star) - x^\star\right\}, \tag{5.3}$$

where z^{BCW} is the equilibrium real money balance and in the BCW equilibrium, the competitive loan rate equals the deposit rate i_d.

To understand whether there is welfare gain from financial intermediation, we compare the household's lifetime utility in an economy with banks relative to an economy without banks. Following Berentsen et al. (2007), we define a difference in welfare value as $DW^e(\tau) = W^e(\tau) - W^{\text{no-bank}}(\tau)$, for a given τ-policy SME, for each economy $e \in \{HKNP, BCW\}$ under consideration. Figure 6 shows the welfare gain from banking in both our imperfectly competitive benchmark (*dotted-green graph*, HKNP) and the Berentsen et al. (2007) economy (*dash-dotted-red graph*, BCW) which our model nests by setting $\alpha_2 = 1$. In both the Berentsen et al. (2007) economy and our baseline the relationship between trend inflation and welfare gain from banking is non-monotonic.

In the BCW case, the welfare difference is increasing at low inflation and only begins to fall at very high inflation as the inflation tax eventually outweighs the gains from insurance. In our imperfectly competitive benchmark, however, the additional effect of imperfect com-

Figure 6: Welfare difference—HKNP or BCW banks versus no banks.



petition leads to a *negative* welfare effect of banking overall at low inflation. When trend inflation is low, the gains from insurance are also low and are easily outweighed by the high markups (recall Figure 5) which arise in equilibrium. As inflation rises, markups fall and the gains to insurance rise so that net effect of the banking system to be positive. Welfare gains are, however, always lower in the imperfectly competitive benchmark than in the BCW case. Note that the welfare gain from financial intermediation in both economies approaches to zero as $\tau \to \infty$. This is because the value of liquidity is of very little value at very high inflation.[29]

Insights. To sum up, there are two opposing welfare effects arising from anticipated inflation tax. On one hand, as in Berentsen et al. (2007), banks improve welfare by providing insurance against holding idle money in the DM as an inactive buyers. On the other, there is a negative, welfare-reducing effect emanating from banks' market power in the loan market. By raising the cost of additional funds, banks effectively extract surplus from buyers in DM trades, putting downward pressure on the value of real balances. Overall, whether banks raise or lower welfare in equilibrium depends on the *net* effect of these two opposing channels. As such, it depends on the extent of imperfect competition (measured by markups) in equilibrium. Imperfect competition in banking has the potential to offset the welfare-enhancing effects of financial intermediation. This finding suggests that policymakers—especially in low-inflation countries—may rightly be concerned with market power in the banking sector.

[29] In Online Appendix A.8, we consider an alternative welfare-comparison exercise. We replace the no-bank equilibrium here with an equilibrium where banks exist but agents are restricted to not be able to borrow or deposit excess liquidity. We label this as a *financially-autarkic* equilibrium. We can prove (and also numerically show) that welfare under any HKNP equilibrium always dominates its corresponding financially-autarkic equilibrium, if the long-run inflation target is away from the Friedman rule.

5.3 Imperfect pass-though: a testable empirical prediction

The previous analysis illustrates two implications of the theory: First, there is positive markup of lending rates on average over the cost of bank funds in equilibrium, implying the potential for imperfect pass-through of monetary policy to lending rates. Second, there is a positive correlation between equilibrium average markup and the standard deviation of markups in the loan market.[30] That is, as inflation rises, banks pass through the increase in costs *differentially* to their lending rates in a manner analogous to that described by Head, Kumar and Lapham (2010).

In our baseline calibration, the implied correlation between average markup and dispersion of markups matches the annualized version of the monthly data from Section 2. This provides an external validity check on the model's empirical relevance.

6 Optimal stabilization policy

In this section, we study an optimal stabilization policy in response to aggregate demand shocks by solving a version of the Ramsey problem.[31] Our policy exercise here both follows and contrasts with that considered Berentsen and Waller (2011). The difference in the policy implication here now is due to the nature of competition among banks. The mechanism in Berentsen and Waller (2011) depends on counteracting sub-optimal interest rate movements, raising the deposit rate when aggregate demand is low. Here optimal policy involves reducing banks' lending markups when aggregate demand is high. We isolate this channel (and eliminate that of BCW) by holding the deposit rate fixed as aggregate demand varies.

6.1 Aggregate demand shocks

Let n, the fraction of active DM buyers now fluctuate randomly. Changes in n will effectively function as aggregate demand shocks. Similarly, we could also consider a multiplicative shock ϵ to the utility of DM consumption. In what follows, we write the model with both types of shocks, although in our examples here we consider only shocks to n for brevity. Suppose now that n belongs to the set $[\underline{n}, \overline{n}]$, where $0 < \underline{n} < \overline{n} < 1$, and, ϵ is bounded in the set $[\underline{\epsilon}, \overline{\epsilon}]$, where, $0 < \underline{\epsilon} < \overline{\epsilon} < \infty$. Let $\omega = (n, \epsilon) \in \Omega$ denote the aggregate state vector, where $\Omega = [\underline{n}, \overline{n}] \times [\underline{\epsilon}, \overline{\epsilon}]$. There is a given probability density function ψ over Ω.

[30]We use the standard deviation and the coefficient of variation as alternative measures of dispersion in the model to be consistent with the empirical evidence in Section 2.

[31]Details of the problem setup can be found in Appendix A.9.

6.2 Monetary policy

The central bank commits to an overall long-run inflation target τ (or equivalently, a price path) and engages in state-contingent liquidity management. The timing and sequence of actions are as follows. First, a monetary injection, τM occurs at the beginning of the period (before shocks, *i.e.* ω, are realized). Second, the central bank injects a lump-sum amount of money to the DM agents up to $\tau_1(\omega)M = n\tau_b(\omega)+(1-n)\tau_b(\omega)+\tau_s(\omega)$ after all agents observe ω.[32] We assume the central bank can only tax in the CM but not the DM, so only transfers are permitted in the DM, i.e., $\tau_1(\omega) \geq 0$. Third, agents enter into banking arrangements and then exchange and production of DM goods take place. Lastly, agents trade goods, work and settle the financial claims (loans and deposits) in the subsequent CM. Finally, we assume that any state-contingent injection of liquidity received by the DM agents will be undone in the CM, *i.e.*, $\tau_2(\omega) = -\tau_1(\omega)$. This state-contingent policy can be thought of as a central-bank repo agreement where they sell money in the DM and promised to buy that back in the CM.[33]

Given the assumption that the DM state-contingent policy will be undone in the subsequent CM, the total change to the aggregate money stock is deterministic and given by

$$M_{+1} - M = (\gamma - 1)M = \tau M, \tag{6.1}$$

where $\gamma = 1 + \tau$ is the growth in money supply.

From here, we restrict attention to a stationary monetary equilibrium where end-of-period real money balances are both time and state invariant, *i.e.* $\phi M = \phi_{+1}M_{+1} = z$, for all $\omega \in \Omega$. In a stationary monetary equilibrium, money supply growth is

$$\frac{\phi}{\phi_{+1}} = \frac{M_{+1}}{M} = \frac{p_{+1}}{p} = \gamma = 1 + \tau.$$

Essentially, the central bank's hands are tied to price-level targeting via a given trajectory for the money stock, as in Berentsen and Waller (2011). What the central bank can choose

[32]Potentially, the central bank can treat the DM active buyers, the DM inactive buyers and the sellers differently with their state-contingent policy in response to ω shocks. Here, we shut down this channel. Moreover, since the sellers do not carry money in each DM, it is without loss of generality that we let $\tau_s(\omega) = 0$ in our numerical exercises.

[33]We also consider a redistributive transfer-and-tax scheme in response to aggregate demand shocks, for robustness checks. The result is qualitatively similar and hence we do not present it here. The key difference between our main exercise and the robustness checks is in terms of how policy actions are constrained. Here is an idea of this alternative policy setting. First, the government commits to an overall inflation target τ. Thus, the amount of money (τM) created by the government splits between the DM and the CM depending on the realization of the state. Second, the government may induce state-contingent, lump-sum transfer/tax in the form of $\tau_1(\omega)M$ in the DM. Third, state-contingent injections of liquidity to agents in the DM are followed by transfers/taxes in the CM to maintain the price-level target, i.e., $\tau_2(\omega)M = [\tau - \tau_1(\omega)]M$ for all ω. That is, if $\tau - \tau_1(\omega) < 0$, then the government taxes the households in the subsequent CM.

will be described next.

6.3 The central bank

Our environment is novel in that the market power of banks (distribution of lending rates, F) responds endogenously to both policy and the state of the economy. This creates an extra layer of policy trade-off relative to the perfectly competitive banking environment. To understand how the stabilization policy works here, we consider two policy regimes in response to aggregate demand fluctuations:

1. **An active central bank**: The central bank commits to an *ex-ante* optimal policy that maximizes social welfare in a steady-state equilibrium (SME):

$$\begin{aligned}\max_{\{q_b^0(\cdot,\omega),\ q_b(\cdot,\omega),\tau_b(\omega)\}_{\omega\in\Omega}} \quad & U(x) - x - c(q_s(\mathbf{z},\omega)) \\ & + \int_{\omega\in\Omega} n\alpha_0 \epsilon u\left[q_b^0(\mathbf{z},\omega)\right]\psi(\omega)\,\mathrm{d}\omega \\ & + \int_{\omega\in\Omega} n \int_{\underline{i}(\mathbf{z},\omega)}^{\bar{i}(\mathbf{z},\omega)} \left[\alpha_1 + 2\alpha_2\left(1 - F(i,\mathbf{z},\omega)\right)\right] \\ & \times \epsilon u\left[q_b(i,\mathbf{z},\omega)\right]\mathrm{d}F(i,\mathbf{z},\omega)\,\psi(\omega)\,\mathrm{d}\omega\end{aligned} \tag{6.2}$$

   subject to a stochastic version of the SME conditions and to the constraint on policy: $\frac{\gamma-\beta}{\beta} = \tau + \tau_1(\omega) + \tau_2(\omega)$, and, $\tau_1(\omega) = -\tau_2(\omega)$. The detail of the optimal policy problem can be found in our Online Appendix A.9. The policy plan prescribes ω-contingent liquidity injections. That is, $\tau_1(\omega) = \tau_b(\omega) \geq 0$.[34]

2. **A passive central bank**: In this regime, the policymaker is constrained by $\tau_1(\omega) = \tau_2(\omega) = 0$ for all $\omega \in \Omega$. In this case, equilibrium outcomes are similar to those of the deterministic baseline SME.

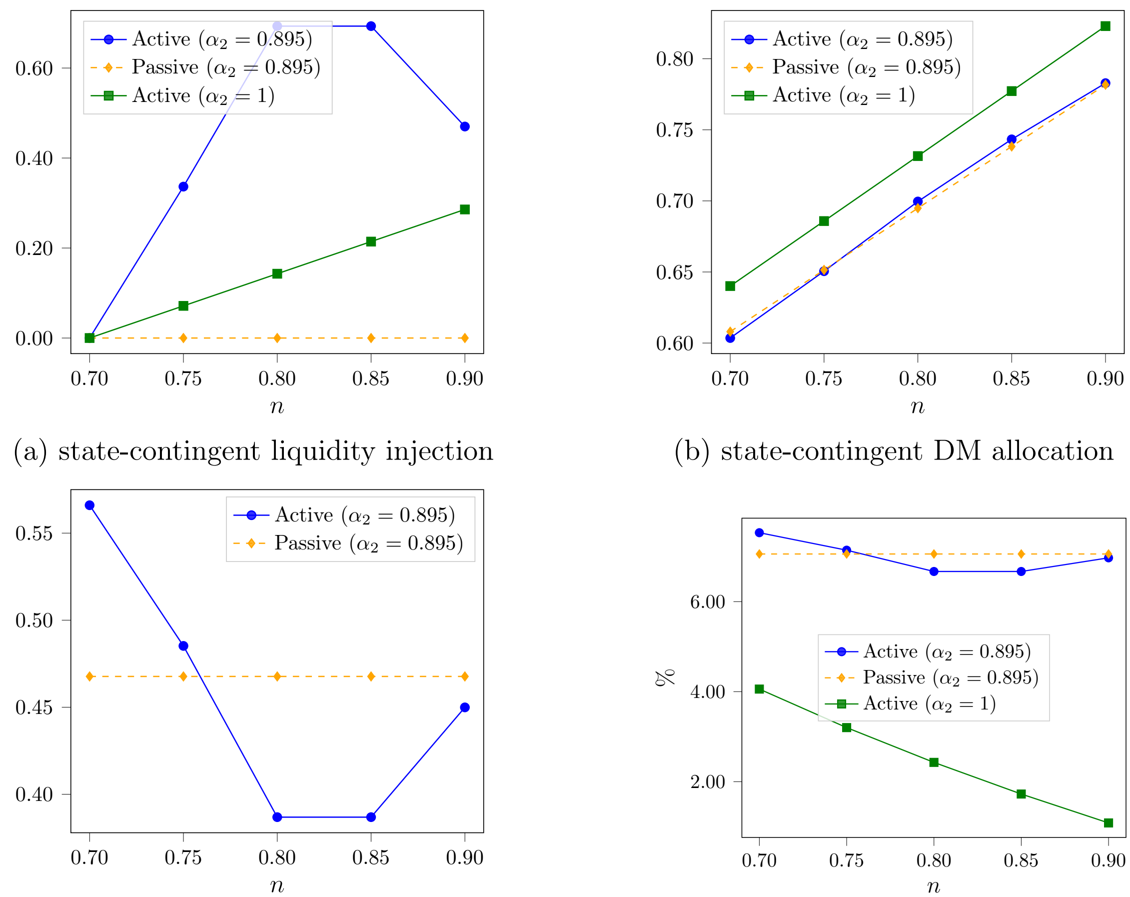
6.4 Discussion

For illustration, we consider only the case of shocks to the number of active DM buyers, n, and hold ϵ fixed at one. As such, we seek optimal policy solutions that only depend on n—*i.e.*, $\tau_b(n)$. We assume that n is distributed uniformly on $\{n_1, \ldots n_5\}$ where $n_i < n_{i+1}, i = 1, ..., 4$.[35]

[34] Since we are considering an optimal policy choice over SME allocations, we directly write these as functions of an SME state-policy vector augmented by ω—*i.e.*, $(\mathbf{z}, \omega)$.

[35] We have also considered taste shock ϵ for the DM goods to be a proxy for demand fluctuation. The results are similar but we do not discuss them for the sake of brevity. The key difference between the ϵ-shock case and the n-shock case is an extra moving part in the equilibrium loan-price distribution F. In particular, lending banks' trade-offs are changing with respect to both the ϵ-state-contingent policy and the ϵ-shock simultaneously. In the n-shock case, there is one less moving part in this respect.

Figure 7: The effects of state-contingent liquidity management on allocations, interest rates and bank market power.



(a) state-contingent liquidity injection

(b) state-contingent DM allocation

(c) state-contingent loan (%) markup

(d) state-contingent loan interest rate

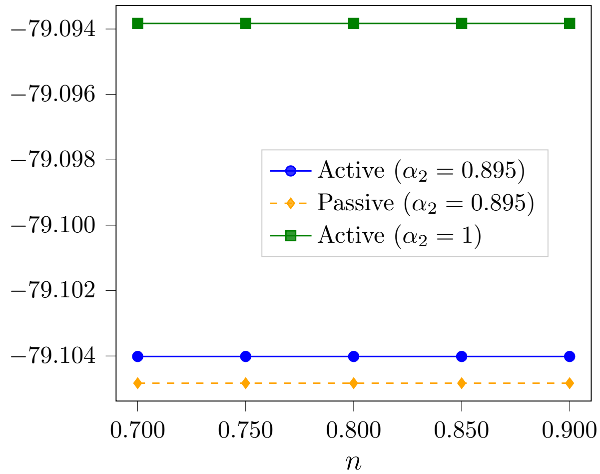
We fix the long-run inflation target at $\tau > \beta - 1$ (away from the Friedman rule). This has the interpretation of policy commitment to a long-run price path. Thus, banks' marginal cost of funds is also fixed at $i_d = \gamma/\beta - 1$.[36]

Figure 7 shows the effects of the optimal state-contingent liquidity management policy in both our imperfectly competitive benchmark (blue-circle and orange-diamond graphs) and in the BCW economy ($\alpha_2 = 1$) with the active central bank (green-square graphs). The blue-circle graphs correspond to our case with an active central bank. The orange-diamond ones are for our case with a passive central bank.

In our imperfectly competitive benchmark, active demand-side stabilization policy through liquidity provision results in higher *ex-ante* welfare for households than the case with passive policy as shown in Figure 8. However, imperfect competition among banks distorts some of

[36] If $\tau = \beta - 1$, then holding money is costless, so there is no need for banks or for stabilization policy. We set τ to equal the average inflation rate in the data.

Figure 8: Ex-ante welfare: active policy versus passive policy



the welfare gains from the active central bank's liquidity management policy.

To help with the economic intuition, we identify two opposing forces of state-contingent liquidity injections, $\tau_b(n)$, on both allocations and welfare.[37] First, higher $\tau_b(n)$ lowers the real money balance inducing higher markup dispersion (implied by first-order stochastic dominance). Second, higher $\tau_b(n)$ directly lowers both dispersion and the average markup by reducing the maximum (*i.e.* monopoly) loan rate. The net welfare consequence of stabilization policy thus depends on the magnitude of these two opposing forces across states.

We find that the active central bank commits to injecting more liquidity into the market in high-demand states relative to a passive policy regime. This liquidity injection induces relatively more (less) consumption in the state with *ex-post* more (fewer) active buyers than the passive policy regime. Similarly, the active policy induces a relative increase (reduction) in *ex-post* markups in low n states (high n states). This happens via the two opposing forces of state-contingent liquidity policy described above. Overall, efficiency gains from active policy here come from the ability of the central bank to reduce market power in the banking system when aggregate demand is high and many buyers are active and borrowing. This comes at the expense of higher markups when aggregate demand is low.

7 Conclusion

We construct and study a microfounded monetary economy where the market power of lenders (banks) is endogenous and responds to policy. We show that the model can account for incomplete pass-through of monetary policy to lending interest rates. The model implies a positive relationship between average markup in loan rates and its standard-deviation, and

[37] These can be deduced from Equations (A.9.7) and (A.9.8) in Appendix A.9.

a negative relationship between the average and the coefficient of variation. This is consistent with new evidence from micro-level data on U.S. consumer loans.

We also show that imperfect competition may render an otherwise useful banking system detrimental if inflation is sufficiently low. That is, an economy with no banks may achieve higher welfare than one in which they provide insurance against random liquidity needs. Our welfare analysis speaks to why policymakers in many low-inflation countries may, rightly, be concerned with market power in the banking sector. We also study an optimal liquidity provision policy in response to aggregate demand shocks under the constraint of a long-run inflation target. For a given inflation target, the optimal stabilization policy reduces market power in the banking system in states where there are more agents who want to consume and borrow (*i.e.*, when aggregate demand is high). However, the policy has to tolerate higher markups when aggregate demand is low.

References

Allen, Jason, Robert Clark, and Jean-François Houde, "Search Frictions and Market Power in Negotiated-Price Markets," *Journal of Political Economy*, 2019, *127* (4), 1550–1598. Cited on page(s): [5]

Altermatt, Lukas and Zijian Wang, "Oligopoly Banking, Risky Investment, and Monetary Policy," Technical Report July 2021. Cited on page(s): [5]

Bankscope, "Bank Concentration (%)," 2020. World Bank Database, Annual Time Series GFDD.OI.01. Cited on page(s): [1]

Berentsen, Aleksander and Christopher Waller, "Price-Level Targeting and Stabilization Policy," *Journal of Money, Credit and Banking*, 2011, *43* (s2), 559–580. Cited on page(s): [4], [37], [38], [70], [71], [76]

_, Gabriele Camera, and Christopher Waller, "Money, Credit and Banking," *Journal of Economic Theory*, 2007, *135* (1), 171–195. Cited on page(s): [2], [10], [16], [19], [24], [25], [26], [32], [35], [36], [67], [70]

Boel, Paola and Christopher J. Waller, "Stabilization Policy at the Zero Lower Bound," *International Economic Review*, 2019, *60* (4), 1539–1563. Cited on page(s): [4]

Burdett, Kenneth and Kenneth L. Judd, "Equilibrium Price Dispersion," *Econometrica*, 1983, *51* (4), 955–969. Cited on page(s): [3], [11], [17], [19], [20], [21], [49], [55]

Chang, Te-Tsun and Yiting Li, "Banking, Liquidity Effects and Monetary Policy," *Macroeconomic Dynamics*, 2018, *22* (5), 1267–1297. Cited on page(s): [5]

Chiu, Jonathan, Mohammad Davoodalhosseini, Janet Hua Jiang, and Yu Zhu, "Bank Market Power and Central Bank Digital Currency: Theory and Quantitative Assessment," Staff Working Papers 19-20, Bank of Canada May 2019. Cited on page(s): [5]

Choi, Michael and Guillaume Rocheteau, "A Model of Retail Banking and the Deposits Channel of Monetary Policy," Technical Report 2021. Cited on page(s): [4]

Clark, Robert, Jean-François Houde, and Jakub Kastl, "Chapter 15 - The industrial organization of financial markets," in Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, eds., *Handbook of Industrial Organization, Volume 5*, Vol. 5 of *Handbook of Industrial Organization*, Elsevier, 2021, pp. 427–520. Cited on page(s): [5]

Corbae, Dean and Pablo D'Erasmo, "Discussion of "Large Banks, Loan Rate Markup, and Monetary Policy"," *International Journal of Central Banking*, 2015, *June*, 1–14. Cited on page(s): [1]

__ **and** __ , "Capital Requirements in a Quantitative Model of Banking Industry Dynamics," *Unpublished*, 2018. SED2018. Cited on page(s): [1]

__ **and** __ , "Capital Buffers in a Quantitative Model of Banking Industry Dynamics ," *Econometrica*, 2021. Forthcoming. Cited on page(s): [5]

Dong, Mei, Stella Huangfu, Hongfei Sun, and Chenggang Zhou, "A Macroeconomic Theory of Banking Oligopoly," *European Economic Review*, 2021, *138*, 103864. Cited on page(s): [5]

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, "The Deposits Channel of Monetary Policy," *The Quarterly Journal of Economics*, 05 2017, *132* (4), 1819–1876. Cited on page(s): [4], [5]

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, "Over-the-Counter Markets," *Econometrica*, 2005, *73* (6), 1815–1847. Cited on page(s): [11]

Duval, Romain A, Davide Davide Furceri, Raphael Lee, and Marina M. Tavares, "Market Power and Monetary Policy Transmission," Technical Report WP/21/184 July 2021. Cited on page(s): [1]

Executive Order 14036, "Executive Order 14036: Promoting Competition in the American Economy," 2021. Joseph R. Biden, July 9. *Federal Register*, 86 (36987). Cited on page(s): [1]

Gu, Chao, Fabrizio Mattesini, Cyril Monnet, and Randall Wright, "Banking: A New Monetarist Approach," *The Review of Economic Studies*, 01 2013, *80* (2), 636–662. Cited on page(s): [5]

He, Ping, Lixin Huang, and Randall Wright, "Money, Banking, and Monetary Policy," *Journal of Monetary Economics*, 2008, *55* (6), 1013–1024. Cited on page(s): [5]

Head, Allen, Alok Kumar, and Beverly Lapham, "Market Power, Price Adjustment, And Inflation," *International Economic Review*, February 2010, *51* (1), 73–98. Cited on page(s): [37]

__ **and** __ , "A Comment on "Price Dispersion, Inflation, and Welfare"," *International Economic Review*, 2005, *46* (2), 573–578. Cited on page(s): [21]

__ **, Lucy Qian Liu, Guido Menzio, and Randall Wright**, "Sticky prices: A New Monetarist Approach," *Journal of the European Economic Association*, 2012, *10* (5), 939–973. Cited on page(s): [14]

Kashyap, Anil K., Raghuram Rajan, and Jeremy C. Stein, "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-taking," *The Journal of Finance*, 2002, *57* (1), 33–73. Cited on page(s): [5]

Lagos, Ricardo and Randall Wright, "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 2005, *113* (3), 463–484. Cited on page(s): [11], [13]

Lucas, Robert E. and Juan Pablo Nicolini, "On the Stability of Money Demand," *Journal of Monetary Economics*, 2015, *73* (C), 48–65. Cited on page(s): [30], [31]

Martín-Oliver, Alfredo, Vicente Salas-Fumás, and Jesús Saurina, "A Test of the Law of One Price in Retail Banking," *Journal of Money, Credit and Banking*, 2007, *39* (8), 2021–2040. Cited on page(s): [2]

__ **,** __ **, and** __ , "Informational Differentiation, Interest Rate Dispersion and Market Power," *Applied Economics Letters*, 2009, *16* (16), 1645–1649. Cited on page(s): [2]

Productivity Commission, "Competition in the Australian Financial System," Technical Report, Canberra 2018. Report No. 89. Cited on page(s): [1]

Rocheteau, Guillaume and Antonio Rodriguez-Lopez, "Liquidity Provision, Interest Rates, and Unemployment," *Journal of Monetary Economics*, 2014, *65* (C), 80–101. Cited on page(s): [11]

— **and Randall Wright**, "Money in Search Equilibrium, in Competitive Equilibrium, and in Competitive Search Equilibrium," *Econometrica*, 2005, *73* (1), 175–202. Cited on page(s): [3], [18]

Sims, Rod, "Committee Hansard," 2016. Chairman of the Australian Competition and Consumer Commission, 14 October. Cited on page(s): [1]

Wang, Liang, "Endogenous search, price dispersion, and welfare," *Journal of Economic Dynamics and Control*, 2016, *73* (C), 94–117. Cited on page(s): [17]

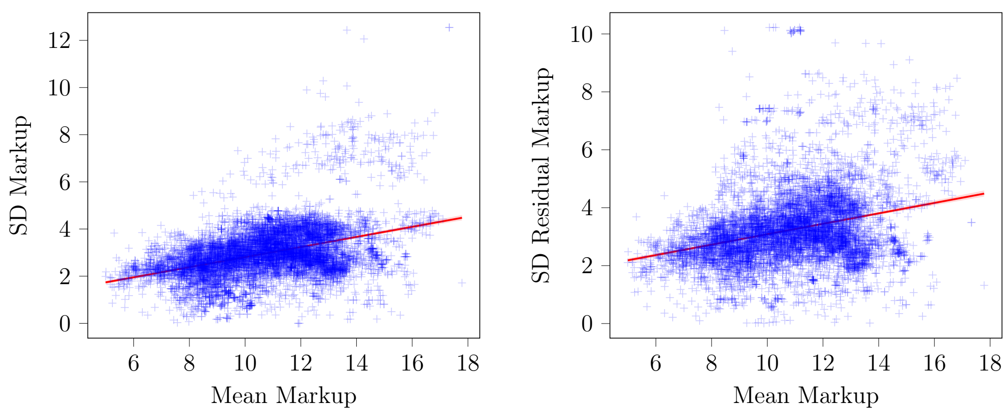
Wilkins, Carolyn, "Why Do Central Banks Care About Market Power?," April 2019. G7 Conference at Banque du France. Cited on page(s): [1]

Online Appendix

A.1 Empirical analysis of markups at the state level

In this section, we calculate the standard deviations and means of loan-rate markups. There are 8,464 usable observations of the variables at the state and month level. This allows us to construct a panel dataset. In Figure A.9, we can see that the markups' standard deviation and average are positively correlated at the state-month level.

Figure A.9: State-month-level relationship between markup dispersion and markup average.
Dispersion measures: SD: *standard deviation* and CV: *coefficient of variation*. Data source: *RateWatch*, "Personal Unsecured Loans (Tier 1)."



Then, we run OLS regressions of markup standard deviation on markup average after controlling for state fixed effects (Z_s) and time fixed effects (Z_t). Specifically, we estimate b_1 in the following sepecification,

$$Dispersion_{s,t} = b_0 + b_1 \overline{Markup}_{s,t} + b_2 Z_s + b_3 Z_t + \epsilon_{s,t} \tag{A.1.1}$$

The index s stands for a particular state and t stands for the month of observation. We cluster standard errors by state and month.

In Table A.1, we report how state-month markup standard deviations are associated with markup average. Column (1) to (3) examine how markup standard deviation is related with markup average using raw markups. Column (4) to (6) examine how markup standard deviation is related with markup average using orthogonalized markups. (See Table A.2 in Section A.2 for the controls used to define the orthogonalized markups.) All columns show a positive and statistically significant relationship between markup standard deviation and markup average. The magnitude of the coefficient is also economically significant. From column (6), the coefficient indicates that a one-percentage-point increase in orthogonalized markup average is associated with a 0.286-percentage-point increase in the standard deviation. This is after controlling for state fixed effects and time fixed effects.

Table A.1: OLS regressions on state markup standard deviation and state markup mean from January 2003 to December 2017.

	Markup dispersion: $Dispersion_{s,t}$					
	Raw markup			Orthogonalized markup		
	(1)	(2)	(3)	(4)	(5)	(6)
	State FE	Time FE	Both FE	State FE	Time FE	Both FE
$\overline{Markup}_{s,t}$	0.179***	0.290***	0.353***	0.220***	0.304***	0.286***
	(0.030)	(0.094)	(0.077)	(0.055)	(0.079)	(0.084)
State fixed effects	X		X	X		X
Time fixed effects		X	X		X	X
N	8237	8237	8237	7463	7463	7463
adj. R^2	0.618	0.178	0.646	0.538	0.203	0.577

Note: Standard errors in parentheses. $^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

A.2 Control variables list

In Table A.2, we detail the controls used in constructing our orthogonalized measures of markups.

Table A.2: Control variables to obtain the orthogonalized markup.

(a) Panel A: County variables

Variable	Data source	Frequency	Details
Real GDP	BEA	Annual	Annual county real GDP
GDP growth	BEA	Annual	Real GDP growth
Establishments	BLS	Annual	Number of establishments within county
Unemployment	BLS	Annual	County unemployment rate
House price	U.S. Census	Annual	Average housing pricing in the county
Median income	U.S. Census	Annual	Median Household Income
Population	U.S. Census	Annual	ln(Total population)
Poverty	U.S. Census	Annual	Proportion of county population under poverty line

(b) Panel B: Local competition

Variable	Data source	Frequency	Details
Within county share	SOD	Annual	Total branch deposits / Total county deposits
County deposit HHI	SOD	Annual	HHI of county's deposit holdings
County branch count	SOD	Annual	Number of branch counts in the county

(c) Panel C: Bank branch network

Variable	Data source	Frequency	Details
Within bank share	SOD	Annual	Total branch deposits / Total bank deposits
Bank deposit HHI	SOD	Annual	HHI of commercial bank's deposit holdings across its branches
Bank branch count	SOD	Annual	Number of branch counts in the commercial bank

(d) Panel D: Commercial bank controls

Variable	Data source	Frequency	Details
Deposit reliance	Call reports	Quarter	Total deposits / Total liabilities
Leverage	Call reports	Quarter	Total equity / Total assets
Credit risk	Call reports	Quarter	(Allowance for Loan and Lease Losses)/(Total Loans)x
Bank size	Call reports	Quarter	ln(Total assets)

A.3 Omitted proofs: Banking with noisy loan search

In this section, we collect the intermediate results and proofs that lead to the characterization of an equilibrium distribution of loan rates in the noisy-search model for loans. Most of the proofs in this section are standard in the Burdett and Judd (1983) model. We revisit them here for completeness.

Remark on notation. In the paper, functions such as Π, Π^m, R, $l^\star$—respectively, *ex-ante* profit, monopoly profit, per-customer profit and optimal loan demand functions—all depend on a vector of individual state m, aggregate state M and policies $\boldsymbol{\tau}$, which we summarize as $(m, \mathbf{s}) = (m, (M, \boldsymbol{\tau}))$. Since the noisy-search banking equilibrium is an intratemporal or static one, in the proofs below, we dispense with explicit dependencies on $(m, \mathbf{s})$ to keep proofs more readable. For example, we will write $l^\star(i)$ in place of the explicit notation $l^\star(i, m, \mathbf{s})$.

Summary of results. We begin by proving for Case 1 in Lemma 1 (since the case where $\alpha_1 \in (0, 1)$ is our main focus of the model). Then we lay out the proof for the remaining cases of a pure monopoly bank in one limit and competitive banks in the other.

The characterization is arrived at in a few intermediate steps. First, in Section A.3.1, we show any bank faced with just one customer ex post will earn strictly positive profit. Second, in Section A.3.2 we show that banks that ex post face more than one customer will also earn strictly positive profit. Third, in Section A.3.3 we show that there is a unique upper bound on loan prices. Fourth, if the upper bound loan rate is the monopoly rate, we show (in Section A.3.4) that this rate is uniquely determined as a function of the state of the economy. There is a natural lower bound on loan rates, which is i_d. These results help establish that the equilibrium support on the distribution of loan rate F is bounded.

In a noisy search equilibrium, the banks will be indifferent between a continuum of pure-strategy price posting outcomes. For example, a bank can choose some lower rate in return for attracting a larger measure of borrowers. Or it can post some higher rate to increase its intensive-margin markup but attract a smaller measure of borrowers. Or it can charge a monopolist price. The intermediate results in Lemmata 11 to 13 (in Section A.3.5 to A.3.6) show that there is a continuum of pure-strategy price posting outcomes that deliver the same maximal monopoly profit. Thus, banks can ex-ante mix over these pure strategies, and in equilibrium, borrowers face a lottery over loan rates, given by a distribution function F. Finally, we can summarize F as an analytical expression in Lemma 1. The proof of this is in Section A.3.7.

A.3.1 Positive monopoly bank profit

Lemma 7. $\Pi^m(i) > 0$ *for* $i > i_d$.

Proof. For any positive markup $i - i_d$,

$$\begin{aligned}\Pi^m(i) &= n\alpha_1 R(i) \\ &= n\alpha_1 l^\star(i)\left[(1+i) - (1+i_d)\right].\end{aligned}$$

Since $l^\star(i) > 0$ and $i - i_d > 0$, then $\Pi^m(i) > 0$. □

A.3.2 All banks earn positive expected profit

Now, we prove that banks will earn strictly positive expected profits:

Lemma 8. $\Pi^\star > 0$.

Proof. Since we are restricting to a class of linear pricing rules, then, for any markup over marginal cost $\mu > 1$, the profit from positing $i = \mu i_d$ is

$$\begin{aligned}\Pi(\mu i_d) &= n\left[\alpha_1 + 2\alpha_2(1 - F(\mu i_d)) + \alpha_2\xi(\mu i_d)\right] R(\mu i_d) \\ &> n\alpha_1 R(\mu i_d) = \Pi^m(\mu i_d) > 0,\end{aligned}$$

where $R(i) = l^\star(m; i, p, \phi, M, \tau_b)\left[(1+i) - (1+i_d)\right]$. The last inequality is from Lemma 7. From the definition of the max operator in (3.24),

$$\begin{aligned}\Pi^\star &= \max_{i \in \text{supp}(F)} \Pi(i) \\ &\geq \Pi(\mu i_d) > \Pi^m(\mu i_d) > 0.\end{aligned}$$

□

A.3.3 Maximal loan pricing

Third, we can also show that:

Lemma 9. *The largest possible price in the support of F is the smaller of the monopoly price and ex-post borrower's maximum willingness to pay:* $\bar{i} := \min\{i^m, \hat{i}\}$.

Although the monopoly rate i^m is the maximal possible price in defining an arbitrary support of F, it may be possible in some equilibrium that this exceeds the maximum willingness to pay by households, $\hat{i}$. We condition on this possibility when characterizing an *equilibrium* support of F later.

Proof. First assume the case that $\hat{i} \geq i^m$. Suppose there is a $\bar{i} \neq i^m$ which is the largest element in $\text{supp}(F)$. Then $\Pi^m(\bar{i}) = n\alpha_1 R(\bar{i})$. Since $F(i^m) \geq 0$ amd $\zeta(i^m) \geq 0$, then

$$\begin{aligned}\Pi(i^m) &= n\left[\alpha_1 + 2\alpha_2(1 - F(i^m)) + \alpha_2\zeta(i^m)\right] R(i^m) \\ &\geq n\alpha_1 R(i^m) = \Pi^m(i^m) \\ &> \Pi^m(\bar{i}).\end{aligned}$$

The last inequality is true by the definition of a monopoly price i^m. Therefore $\Pi(i^m) > \Pi^m(\bar{i})$. The equal profit condition would require that, $\Pi^m\left(\bar{i}\right) = \Pi^\star \geq \Pi^m(i^m)$. Therefore $\bar{i} = i^m$ if $\hat{i} \geq i^m$.

Now assume $\hat{i} < i^m$. In this case, the most that a bank can charge for loans is $\hat{i}$, since at any higher rate, no ex-post buyer will execute his line of credit (i.e., he will not borrow). Thus trivially, $\bar{i} = \hat{i}$ if $\hat{i} < i^m$. □

A.3.4 Unique monopoly loan rate

Fourth, under a mild parametric regularity condition on preferences, we show that there is a unique monopoly loan rate.

Lemma 10. *Assume $\sigma < 1$. For an arbitrarily small constant bounded below by zero, i.e., $\epsilon > 0$, if $\sigma \geq \epsilon/(2+\epsilon)$, then there is a unique monopoly-profit-maximizing price i^m that satisfies the first-order condition*

$$\frac{\partial \Pi^m(i)}{\partial i} = n\alpha_1\left[\frac{\partial l^\star(i)}{\partial i}(1+i) + l^\star(i) - \frac{\partial l^\star(i)}{\partial i}(1+i_d)\right] = 0.$$

Proof. Assume $\hat{i} > i^m$. Using the demand for loans from Equation (3.16) the first-order condition at $i = i^m$ is explicitly

$$\underbrace{-\frac{m+\tau_b M}{p^{\frac{\sigma-1}{\sigma}}\phi^{-\frac{1}{\sigma}}}}_{f(i)} + \underbrace{\frac{1}{\sigma}(1+i)^{-\frac{1}{\sigma}}\left[(\sigma-1) + \frac{1+i_d}{1+i}\right]}_{g(i)} = 0. \tag{A.3.1}$$

Note that given individual state m, aggregate state M, and policy/prices (τ_b, p, ϕ), the term $f(i)$ is a constant with respect to i. Given i_d, the term $g(i)$ has these properties:

1. $g(i)$ is continuous in i;
2. $\lim_{i\searrow 0} g(i) = +\infty$;
3. $\lim_{i\nearrow +\infty} g(i) = 0$, and,

4. the RHS is monotone decreasing, $g'(i) < 0$.

The first three properties are immediate from Equation (A.3.1). Since $\Pi^m(i)$ is twice-continuously differentiable, the last property can be shown by checking for a second-order condition: For a maximum profit at $i = i^m$, we must have $\left.\frac{\partial^2 \Pi^m(i)}{\partial i^2}\right|_{i=i^m} \leq 0$. Observe that the second-derivative function is

$$\frac{\partial^2 \Pi^m(i)}{\partial i^2} = g'(i) = -\underbrace{\frac{1}{\sigma^2}(1+i)^{-\frac{1}{\sigma}-1}}_{>0}\left[(\sigma-1) + \frac{(1+\sigma)(1+i_d)}{(1+i)}\right]. \tag{A.3.2}$$

For (A.3.2) to hold with ≤ 0, Case A ($\sigma < 1$) would require

$$\frac{(1+\sigma)(1+i_d)}{(1+i)} \geq 1 - \sigma$$

for all $i \geq i_d$.

Let $1 + i \equiv (1+\epsilon)(1+i_d)$ since $i^m \geq i > i_d$. The above inequality can be re-written as

$$\frac{1}{1+\epsilon} \geq \frac{1-\sigma}{1+\sigma},$$

which implies

$$1 > \sigma \geq \frac{\epsilon}{2+\epsilon}. \tag{A.3.3}$$

Condition (A.3.3) is a sufficient condition on parameter σ to ensure that a well-defined and unique monopoly profit point exists with monopoly price $i^m \geq \underline{i} > i_d$ if $\frac{\epsilon}{2+\epsilon} \leq \sigma < 1$.

□

A.3.5 Distribution is continuous

In the next two results, we show that the loan pricing distribution is continuous with connected support.

Lemma 11. *F is a continuous distribution function.*

We will prove Lemma 11 in two parts. First, we document a technical observation that the per-customer profit difference is always bounded above:

Lemma 12. *Assume there is an $i' < i$ and an $i'' < i'$, with*

$$\zeta(i) = \lim_{i' \nearrow i} \{F(i) - F(i')\} > 0,$$

and $\zeta(i') = \lim_{i'' \nearrow i'} \{F(i') - F(i'')\} > 0$, *and that* $R(i') > 0$. *The per-customer profit difference is always bounded above:* $\Delta := R(i) - R(i') < \frac{\alpha_2 \zeta(i) R(i)}{\alpha_1 + 2\alpha_2}$.

Proof. The expected profit from posting i is

$$\Pi(i) = n\left[\alpha_1 + 2\alpha_2(1 - F(i)) + \alpha_2 \zeta(i)\right] R(i).$$

The expected profit from posting i' is

$$\Pi(i') = n\left[\alpha_1 + 2\alpha_2(1 - F(i')) + \alpha_2 \zeta(i')\right] R(i').$$

A firm would be indifferent to posting either prices if $\Pi(i) - \Pi(i') = 0$. This implies that

$$\begin{aligned}(\alpha_1 + 2\alpha_2)\left[R(i) - R(i')\right] + \alpha_2 \zeta(i) R(i) - \alpha_2 \zeta(i') R(i') \\ - 2\alpha_2\left[F(i) R(i) - F(i') R(i')\right] = 0.\end{aligned}$$

Rearranging and using the definition of $\zeta(i) = \lim_{i' \nearrow i} \{F(i) - F(i')\} > 0$:

$$\begin{aligned}(\alpha_1 + 2\alpha_2)\left[R(i) - R(i')\right] &= \alpha_2\left[F(i) R(i) - F(i') R(i')\right] - \alpha_2 \zeta(i') R(i') \\ &< \alpha_2\left[F(i) R(i) - F(i') R(i')\right] \\ &\leq \alpha_2 \lim_{i' \nearrow i} \{F(i) - F(i')\} R(i).\end{aligned}$$

The strict inequality is because $R(i') > 0$ and $\zeta(i') > 0$. The subsequent weak inequality comes from the fact that $R(i)$ is continuous, so that we can write

$$\lim_{i' \nearrow i} \{F(i) R(i) - F(i') R(i')\} = \lim_{i' \nearrow i} \{F(i) - F(i')\} R(i).$$

Since $\zeta(i) = \lim_{i' \nearrow i} \{F(i) - F(i')\}$, the last inequality implies that $R(i) - R(i') < \frac{\alpha_2 \zeta(i) R(i)}{\alpha_1 + 2\alpha_2}$. □

The following is the proof of Lemma 11.

Proof. Suppose there is a $i \in \text{supp}(F)$ such that $\zeta(i) > 0$ and

$$\Pi(i) = n\left[\alpha_1 + 2\alpha_2(1 - F(i)) + \alpha_2 \zeta(i)\right] R(i).$$

R is clearly continuous in i. Hence there is a $i' < i$ such that $R(i') > 0$ and from Lemma 12,

$\Delta := R(i) - R(i') < \frac{\alpha_2 \zeta(i) R(i)}{\alpha_1 + 2\alpha_2}$. Then

$$\begin{aligned}\Pi(i') &= n\left[\alpha_1 + 2\alpha_2\left(1 - F(i')\right) + \alpha_2\zeta(i')\right] R(i') \\ &\geq n\left[\alpha_1 + 2\alpha_2\left(1 - F(i)\right) + \alpha_2\zeta(i)\right]\left[R(i) - \Delta\right] \\ &\geq \Pi(i) + n\left\{\alpha_2\zeta(i)\left[R(i) - \Delta\right] - \left(\alpha_1 + 2\alpha_2\right)\Delta\right\}.\end{aligned}$$

The first weak inequality is a consequence of $F(i) - F(i') \geq \zeta(i)$. Since $R(i) > \Delta$ and $\Delta < \frac{\alpha_2 \zeta(i) R(i)}{\alpha_1 + 2\alpha_2}$, then the last line implies $\Pi(i') > \Pi(i)$. This contradicts $i \in \text{supp}(F)$. □

A.3.6 Support of distribution is connected

Lemma 13. *The support of F, $\text{supp}(F)$, is a connected set.*

Proof. Pick two prices i and i' belonging to the set $\text{supp}(F)$, and suppose that $i < i'$ and $F(i) = F(i')$. The expected profit under these two prices are, respectively,

$$\Pi(i) = n\left[\alpha_1 + 2\alpha_2\left(1 - F(i)\right)\right] R(i),$$

and,

$$\Pi(i') = n\left[\alpha_1 + 2\alpha_2\left(1 - F(i')\right)\right] R(i').$$

Since $F(i) = F(i')$, then the first terms in the profit evaluations above are identical:

$$n\left[\alpha_1 + 2\alpha_2\left(1 - F(i)\right)\right] = n\left[\alpha_1 + 2\alpha_2\left(1 - F(i')\right)\right].$$

However, since i and i' belonging to the set $\text{supp}(F)$, then clearly, $i_d < i < i' \leq i^m$. From Lemma 10, we know that $R(i)$ is strictly increasing for all $i \in [i_d, i^m]$, so then, $R(i) < R(i')$. From these two observations, we have $\Pi(i) < \Pi(i')$. This contradicts the condition that if firms are choosing i and i' from $\text{supp}(F)$ then F must be consistent with maximal profit $\Pi(i) = \Pi(i') = \Pi^\star$ (viz. the equal profit condition must hold). □

A.3.7 Proof of Proposition 1

Proof. Consider the case where $\alpha_1 \in (0, 1)$. Since F has no mass points by Lemma 13, and is continuous by Lemma 11, then expected profit from any $i \in \text{supp}(F)$ is a continuous function over $\text{supp}(F)$,

$$\Pi(i) = n\left[\alpha_1 + 2\alpha_2\left(1 - F(i)\right)\right] R(i),$$

where the image $\Pi\left[\text{supp}\left(F\right)\right]$ is also a connected set. From Lemma 9, profit is maximized at $\Pi^m\left(i^m\right) = n\alpha_1 R\left(i^m\right)$. For any $i \in \text{supp}\left(F\right)$, the induced expected profit must also be maximal, i.e.,

$$\Pi\left(i\right) = n\left[\alpha_1 + 2\alpha_2\left(1 - F\left(i\right)\right)\right] R\left(i\right) = n\alpha_1 R\left(i^m\right).$$

Solving for F yields the analytical expression in Equation (3.25).

Proofs for the remaining Case 2 and Case 3 in Lemma 1 follow directly from Lemma 1 and Lemma 2 in Burdett and Judd (1983). The pricing outcomes, $\bar{i}$ and i_d are, respectively, the upper bound (the monopoly price) and the lower bound (Bertrand price) on the support of F. □

A.4 General money demand Euler equation

Evaluating the partial derivation of the value function in Equation (3.17) one period ahead, combining this with First-order Condition (3.7) and the optimal goods demand functions in Equation (3.10) and (3.13), we can derive an Euler functional describing the optimal money demand function. Re-writing this in terms of stationary variables, we have the steady-state Euler equation on z as:

$$\begin{aligned}
\underbrace{\frac{\gamma - \beta}{\beta}}_{\text{MC of extra dollar}} \quad &= \quad \underbrace{\Theta(z, \mathbf{z}) - 1}_{\text{Net MB of extra dollar}} \\
+ \quad & \underbrace{\mathbb{I}_{\{0 \leq \rho < \hat{\rho}\}} \times n\alpha_0 \left[\frac{1}{\rho} \left(\frac{z + \tau_b Z}{\rho} \right)^{-\sigma} - 1 \right]}_{\text{Liquidity cons, no bank contact (self-insure): Net MB of consumption from extra dollar}} \\
& \underbrace{+ n \int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} \mathbb{I}_{\{0 \leq \rho < \tilde{\rho}_i\}} \left[\alpha_1 + 2\alpha_2 \left(1 - F\left(i, z, \mathbf{z}\right)\right) \right] i \mathrm{d}F(i, z, \mathbf{z})}_{\text{Borrow, MB of a dollar's less of borrowing is } i} \\
& \underbrace{+ n \int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} \mathbb{I}_{\{\tilde{\rho}_i \leq \rho < \hat{\rho}\}} \left[\alpha_1 + 2\alpha_2 \left(1 - F\left(i, z, \mathbf{z}\right)\right) \right] \left[\frac{1}{\rho} \left(\frac{z + \tau_b Z}{\rho} \right)^{-\sigma} - 1 \right] \mathrm{d}F(i, z, \mathbf{z})}_{\text{No Borrow, MB of consumption from extra dollar}}
\end{aligned}, \tag{A.4.1}$$

where the net marginal benefit of an extra dollar is decomposable as:

$$\begin{aligned}
\Theta(z, \mathbf{z}) - 1 := \quad & \underbrace{(1 - n)(1 + i_d)}_{\text{Expected gross value of additional deposit}} \\
& + n \left[\alpha_0 + \int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} \mathbb{I}_{\{0 \leq \rho < \tilde{\rho}_i\}} \left[\alpha_1 + 2\alpha_2 \left(1 - F\left(i, z, \mathbf{z}\right)\right) \right] \mathrm{d}F(i, z, \mathbf{z}) \right. \\
& \qquad + \int_{\bar{i}(z,\mathbf{z})}^{i^m(z,\mathbf{z})} \mathbb{I}_{\{\tilde{\rho}_i \leq \rho < \hat{\rho}\}} \left[\alpha_1 + 2\alpha_2 \left(1 - F\left(i, z, \mathbf{z}\right)\right) \right] \mathrm{d}F(i, z, \mathbf{z}) \\
& \qquad \left. + \int_{\bar{i}(z,\mathbf{z})}^{i^m(z,\mathbf{z})} \mathbb{I}_{\{\rho > \hat{\rho}\}} \left[\alpha_1 + 2\alpha_2 \left(1 - F\left(i, z, \mathbf{z}\right)\right) \right] \mathrm{d}F(i, z, \mathbf{z}) \right] - 1,
\end{aligned} \tag{A.4.2}$$

and, the cut-off prices, $\tilde{\rho}_i$ and $\hat{\rho}$, are also functions of $(z, \mathbf{z})$.

A.5 Friedman Rule and the first-best allocation: Proof of Proposition 6

Proof. Suppose that $\gamma = \beta$ but that there is an SME with a non-degenerate distribution of loan interest rates, $F(\cdot, z, \mathbf{z})$.

Since we focus on $\alpha_1 \in (0,1)$, from Lemma 1 (part 1), we know that if there is an SME, then the posted loan-rate distribution $F(\cdot, z, \mathbf{z})$ is non-degenerate and continuous with connected support, $\text{supp}(F(\cdot, z, \mathbf{z})) = [\underline{i}(\cdot, z, \mathbf{z}), \bar{i}(\cdot, z, \mathbf{z})]$.

If there is an SME, then the general Euler condition for money demand (A.4.1) holds. However the marginal cost of holding money—i.e., LHS of (A.4.1)—is zero at the Friedman rule ($\gamma = \beta$). Also, the liquidity premium of carrying more real money balance at the margin into next period is always non-negative—*i.e.*, for any $q > 0$, $u'(q)/c'(q) - 1 \geq 0$. What remains on the RHS of (A.4.1) are all the (net) marginal benefit of borrowing less at the margin when one has additional real balance, *i.e.*, the integral terms. These terms are also non-negative measures. Thus, for an SME to hold, it must be that $F(\cdot, z, \mathbf{z})$ is degenerate on a singleton set.

Since (A.4.1) holds in any SME, then our previous reasoning must further imply that the integral terms reduce to the condition $u'(q^f) = c'(q^f)$. We can compare this with the first best allocation. Given our CRRA preference representation assumption, the first-best allocation solving $u'(q^*) = c'(q^*)$ will yield $q^* = 1$.

Thus if there is an SME at the Friedman rule, then $F(\cdot, z, \mathbf{z})$ must be degenerate. Moreover, at the Friedman rule, the allocation is Pareto efficient: $q^f = q^* = 1$.

□

A.6 Omitted proofs: SME

We provide the intermediate results and proofs for establishing existence and uniqueness of a stationary monetary equilibrium with co-existing money and credit.

The conclusion is arrived at in a few intermediate steps. First, in Section A.6.1 we show that a posted loan-price distribution with lower real money balance first-order stochastic dominance a distribution with higher real money balance, given a monetary policy rule $\gamma > \beta$. Second, in Section A.6.2 we show that the general money demand Euler Equation (A.4.1) can be simplified as Condition (4.5), and the candidate real money balance solution to the money demand Euler equation is bounded. Third, we use results from Section A.6.1 and Section A.6.2 together in section A.6.3 to show there exists a unique real money balance that solves the money demand Euler Equation (4.5). This establishes existence. Finally, we prove for the uniqueness of a SME with co-existing money and credit in Section A.6.4.

A.6.1 First-order stochastic dominance: Proof of Lemma 2.

Proof. The analytical formula for the loan-price distribution $F(i, z, \mathbf{z})$ is characterized in Equation (4.1). Suppose we fix $\bar{i}(z, \mathbf{z}) = \bar{i}(z', \mathbf{z})$, and denote it as $\bar{i}$. In general, the lower and upper support of the distribution F is changing with respect to z and policy γ. By fixing the upper support at both z and z' here, we are checking whether the curve of the cumulative distribution function, $F(\cdot, z, \mathbf{z})$, is lying on top or below for z relative to z'. Next, differentiate $F(i, z, \mathbf{z})$ with respect to z, we

$$\frac{\partial F(i, z, \mathbf{z})}{\partial z} = \underbrace{\frac{\alpha_1}{2\alpha_2}}_{>0} \left[\frac{(\bar{i} - i_d)R(i, z, \mathbf{z}) - (i - i_d)R(\bar{i}, z, \mathbf{z})}{\underbrace{(R(i, z, \mathbf{z}))^2}_{>0}} \right].$$

For $\partial F(i, z, \mathbf{z})/\partial z > 0$ to hold, one needs to show the numerator is positive. Suppose this were not the case. Then we have

$$\begin{aligned}
&(\bar{i} - i_d)R(i, z, \mathbf{z}) - (i - i_d)R(\bar{i}, z, \mathbf{z}) \leq 0 \\
\implies &(\bar{i} - i_d) \underbrace{\left[(1 + i)^{\frac{-1}{\sigma}} - z \right] (i - i_d)}_{=R(i,z,\mathbf{z})} \leq (i - i_d) \underbrace{\left[(1 + \bar{i})^{\frac{-1}{\sigma}} - z \right] (\bar{i} - i_d)}_{=R(\bar{i},z,\mathbf{z})} \\
&\implies \left[(1 + i)^{\frac{-1}{\sigma}} - z \right] \leq \left[(1 + \bar{i})^{\frac{-1}{\sigma}} - z \right]
\end{aligned}$$

The last inequality contradicts the fact that the loan demand curve is downward sloping in i, and $\bar{i}$ is the highest possible loan-price posted by banks (lending agents). Thus, the numerator must be positive and $\partial F(i, z, \mathbf{z})/\partial z > 0$. This shows that a loan-price distribution $F(\cdot, z, \mathbf{z})$ first-order stochastically dominates $F(\cdot, z', \mathbf{z})$, for $z < z'$. □

A.6.2 Money and credit: Proof of Lemma 3

Proof. We want to show equivalence in the three claims in Lemma 3. The proof relies on a CRRA(σ) preference representation and linear cost of producing the DM good $c(q) = q$.

1. We say that the DM relative price ρ is sufficiently low if real money balance z is such that

$$\rho = 1 < \tilde{\rho}_i(z, \mathbf{z}) \equiv (z)^{\frac{\sigma}{\sigma-1}} (1 + i)^{\frac{1}{\sigma-1}}, \qquad 0 < \sigma < 1. \tag{A.6.1}$$

The following is a sufficient requirement: If $z < \left(\frac{1}{1+i}\right)^{\frac{1}{\sigma}}$, then inequality (A.6.1) holds. From Lemma 1, if $\alpha_1 \in (0, 1)$, the distribution $F(\cdot, z, \mathbf{z})$ is non-degenerate and

$\text{supp}(F(\cdot, z, \mathbf{z})) = [\underline{i}(z, \mathbf{z}), \bar{i}(z, \mathbf{z})]$ exists. This implies that for all $i \in \text{supp}(F(\cdot, z, \mathbf{z}))$, the inequality $z < \left(\frac{1}{1+\bar{i}(z,\mathbf{z})}\right)^{\frac{1}{\sigma}}$ is also true. Since SME $z = z^\star$ exists and $z^\star < \left(\frac{1}{1+\bar{i}(z^\star,\mathbf{z})}\right)^{\frac{1}{\sigma}}$, then ρ is sufficiently low and satisfies inequality (A.6.1).

2. From Claim 1 above, the DM relative price ρ satisfies inequality (A.6.1). From (4.14), there is ex-post positive loan demand by the active DM buyers who meet at least one bank. In the opposite direction: If there is ex-post positive loan demand, then condition (A.6.1) must hold, thus implying Claim 1.

3. Combining Claim 2 with agents' first-order condition for optimal money demand, we can reduce their Euler Equation (A.4.1) to Equation (4.5). In reverse, Equation (4.5) implies that there is positive demand for loans and money (Claim 2).

□

A.6.3 Unique real money balance

Lemma 14. *Fix long-run inflation as $\gamma = 1 + \tau > \beta$. Assume $\alpha_0, \alpha_1 \in (0, 1)$. In any SME, there is a unique real money demand, $z^\star \equiv z^\star(\boldsymbol{\tau})$.*

Proof. Consider the case where the long-run inflation target is set away from the Friedman rule, i.e., $\gamma > \beta$. From Lemma 3, the money demand Euler equation is characterized by

$$i_d = \frac{\gamma - \beta}{\beta} = \underbrace{\alpha_0 \left(u'[q_b^0(z^\star, \mathbf{z})] - 1 \right)}_{=:A} + \underbrace{\int_{\underline{i}(z^\star,\mathbf{z})}^{\bar{i}(z^\star,\mathbf{z})} i \mathrm{d}J\left(i, z^\star, \mathbf{z}\right)}_{=:B}, \tag{A.6.2}$$

where

$$\begin{aligned} \mathrm{d}J(i, z^\star, \mathbf{z}) &= \underbrace{\{\alpha_1 + 2\alpha_2(1 - F(i; z^\star))\} f(i, z^\star, \mathbf{z})}_{=:j(i,z^\star,\mathbf{z})} \mathrm{d}i \\ &\equiv \alpha_1 + 2\alpha_2(1 - F(i, z^\star, \mathbf{z}))\mathrm{d}F(i, z^\star, \mathbf{z}). \end{aligned}$$

Recall that $1 \equiv \rho < \tilde{\rho}_i(z^\star, \mathbf{z})$ from Lemma 3, the ex-post DM goods demand function for the event where the active DM buyer failed to meet with a lending bank is given by $q_b^0 = \frac{z}{\rho}$, i.e., she is liquidity constrained with own money balance. Thus, $\partial q_b^0 / \partial z > 0$. Since $u'' < 0$, then $u' \circ q_b^0(z, \mathbf{z})$ is continuous and decreasing in z. Thus, term A is continuous and decreasing in z.

Next, let

$$H(z,\mathbf{z}) := \int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} i \mathrm{d}J(i,z,\mathbf{z}).$$

Applying integration by parts, we obtain

$$H(z,\mathbf{z}) = \bar{i}(z,\mathbf{z}) - \underbrace{\int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} J(i,z,\mathbf{z})\mathrm{d}i}_{\tilde{H}(z)}.$$

Applying Leibniz' Rule to $\tilde{H}(z)$, we have

$$\tilde{H}^{'}(z,\mathbf{z}) = \bar{i}^{'}(z,\mathbf{z}) + \int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} \frac{\partial J(i,z,\mathbf{z})}{\partial z}\mathrm{d}i.$$

Overall, we have

$$H^{'}(z,\mathbf{z}) = \bar{i}^{'}(z,\mathbf{z}) - \tilde{H}^{'}(z,\mathbf{z}) \ = -\int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} \frac{\partial J(i,z,\mathbf{z})}{\partial z}\mathrm{d}i.$$

From Lemma 2, we know that $J(\cdot,z,\mathbf{z})$ first-order stochastically dominates $J(\cdot,z^{'},\mathbf{z})$ for all $z < z^{'}$. Thus, $\partial J(i,z,\mathbf{z})/\partial z > 0$, which implies $H^{'}(z,\mathbf{z}) < 0$. Thus, both terms A and B on the RHS of Equation (A.6.2) are continuous and monotone decreasing in z. Moreover, the LHS of Equation (A.6.2) is constant with respect to z. Therefore, there exists a unique real money demand $z^{\star}(\boldsymbol{\tau})$ that solves the money-demand Euler Equation (A.6.2). Moreover, $z^{\star}(\boldsymbol{\tau})$ is bounded, by Lemma 3. □

A.6.4 SME with money and credit: Proof of Proposition 5

Proof. From Lemmata 2, 3, and 14, we have established existence of solution to both money and credit. In particular, we have shown that there exists a unique money demand $z^{\star} \equiv z^{\star}(\boldsymbol{\tau})$ such that

$$z^{\star} \in \left(0,\ [1+\bar{i}(z^{\star})]^{\frac{-1}{\sigma}}\right),$$

for a given $\gamma > \beta$. This condition ensures that the optimal real money balance $z^{\star}$ is bounded and that the maximal loan interest of the posted loan-price distribution is not too high. Moreover, this guarantees positive loan demand.

To establish a unique SME with both money and credit, what remains is to show that the following equilibrium requirements also hold, when evaluated at $z = z^{\star}$. That is,

1. Deposit interest is feasible (i.e., interest on loans weakly exceeds that on deposits):

$$\begin{aligned}\underbrace{(1-n)\delta^{\star}(z,\mathbf{z})\,i_d}_{:=A} &\equiv (1-n)\,(z+\tau_b Z)\,i_d \\ &\leq \underbrace{n\int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})}\left[\alpha_1+2\alpha_2-2\alpha_2 F(i,z,\mathbf{z})\right] i\xi^{\star}(z,\mathbf{z})\,\mathrm{d}F(i,z,\mathbf{z})}_{:=B},\end{aligned} \tag{A.6.3}$$

   where $z = Z$ at equilibrium, and term A and term B are respectively the interest on deposits and loans.

2. DM (competitive price-taking) goods market clears:

$$\begin{aligned}q_s(z,\mathbf{z}) &= n\alpha_0 q_b^{0,\star}(z,\mathbf{z}) \\ &+ n\left[\int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})}\left[\alpha_1+2\alpha_2-2\alpha_2 F(i,z,\mathbf{z})\right] q_b^{\star}(z;\rho,Z,\gamma)\,\mathrm{d}F(i,z,\mathbf{z})\right].\end{aligned} \tag{A.6.4}$$

3. Both CM goods and labor market clear.

Next, we focus on Condition 1. From the households' problem, we can show that depositors (i.e., *inactive DM buyers*) deposit all of their idle money balance at the perfectly competitive depository institution as long as deposit interest is positive. That is, $\delta(z,\mathbf{z}) > 0$ if and only if $i_d > 0$. The deposit interest rate is pinned down by $i_d = r = \frac{\gamma-\beta}{\beta}$ (where r is given by the Fisher equation under our underlying "small open economy" assumption on depository institutions). Since we focus on monetary policy $\gamma > \beta$, and it follows that the deposit interest is positive. Moreover, n is the probability of agent wanting to consume in the DM (i.e., the measure of *active DM buyers*). Thus, for $n \in [0,1]$, we have

$$0 \leq (1-n)\delta^{\star}(z,\mathbf{z})\,i_d.$$

Now, we have to consider the remaining term on the RHS of the Loans-feasibility Constraint (A.6.3). We have shown there is a positive loan demand evaluated at z, i.e., $\xi^{\star}(z,\mathbf{z}) > 0$, and the loan-price distribution $F(\cdot,z,\mathbf{z})$ is continuous on a connected support, with the lower support being $\underline{i}(z,\mathbf{z}) > i_d$. Thus, for $n \in [0,1]$, we have

$$0 \leq n\int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})}\left[\alpha_1+2\alpha_2-2\alpha_2 F(i,z,\mathbf{z})\right] i\xi^{\star}(z,\mathbf{z})\,\mathrm{d}F(i,z,\mathbf{z}).$$

Since both sides of the loans feasibility constraint are non-negative, then rearranging, we have

$$N(z,\mathbf{z}) := \frac{\delta^{\star}(z,\mathbf{z})\, i_d}{\delta^{\star}(z,\mathbf{z})\, i_d + \int_{\underline{i}(z,\mathbf{z})}^{\bar{i}(z,\mathbf{z})} [\alpha_1 + 2\alpha_2 - 2\alpha_2 F(i,z,\mathbf{z})]\, i\xi^{\star}(z,\mathbf{z})\, \mathrm{d}F(i,z,\mathbf{z})} \leq n.$$

From this, we can deduce the bound on n as follows.

$$0 \leq \lim_{[\underline{i}(z,\mathbf{z}),\bar{i}(z,\mathbf{z})] \nearrow \infty} N(z,\mathbf{z}), \tag{A.6.5}$$

and,

$$\lim_{[\underline{i}(z,\mathbf{z}),\bar{i}(z,\mathbf{z})] \searrow i_d} N(z,\mathbf{z}) \leq 1. \tag{A.6.6}$$

Condition (A.6.5) says if the support of the distribution of posted loan interest rates is sufficiently wide in the sense that

$$\frac{\bar{i}(z,\mathbf{z})}{\underline{i}(z,\mathbf{z})} \to \infty,$$

then the equilibrium restriction on n is *relaxed* as long as it exceeds a small number (closed to zero). Intuitively, if the average (total) revenue from loans is very high, banks do not need to rely on a large share of active DM buyers (i.e., high n) to be able to (weakly) cover the interest paid on deposits.

Likewise, Condition (A.6.6) says if the support of the posted loan interest rate distribution is sufficiently narrow in the sense that

$$\frac{\bar{i}(z,\mathbf{z})}{\underline{i}(z,\mathbf{z})} \to i_d,$$

then the restriction on n is *tight*. That is, the interest from loans is low and n needs to be at least as great as a large number (closed to one) for banks to be able to cover the interest paid on deposits.

In summary, we have derived a sufficient condition that there exists an endogenous lower bound $N(z,\mathbf{z}) \in [0,1]$ such that $n \geq N(z,\mathbf{z})$. This condition requires that the measure of *active DM buyers* not be too small for feasible deposit interest at equilibrium. While these sufficient conditions to ensure positive loan demand and feasible deposit interest above depend on equilibrium objects, they both can be easily verified by numerical calculations.

Now, we turn to the DM goods market clearing requirement in Condition 2. Since the DM firms' optimal production rule is pinned down by a constant marginal cost (due to linear production technology), then the aggregate supply equals to the aggregate demand in the DM goods market.

Finally, we consider Condition 3. In any equilibrium we have constant optimal CM consumption $x^\star$ (due to quasi-linear preference). Given real money balance $z^\star \equiv z^\star(\boldsymbol{\tau})$ and DM allocations $(q_b^{0,\star}(z^\star, \mathbf{z}), q_b^\star(\cdot, z^\star, \mathbf{z}))$, we can verify that the CM goods and labor market also clear. Hence, the details are omitted here. In equilibrium $z = z^\star(\boldsymbol{\tau}) = Z$, so we could further reduce the characterizations above by rewriting $(z, \mathbf{z})$ as just $\mathbf{z}$ in a SME. □

A.7 Omitted proofs - Markup and inflation

Recall that gross inflation is $\gamma = 1 + \tau$. How does average, posted loan-rates markup ($\mu(\gamma)$) change with respect to inflation γ? Also, from a household's perspective, how does *ex-ante* loan-rates markup ($\hat{\mu}(\gamma)$) change with respect to inflation γ? We will show below that successively higher-inflation SME economies having higher average loan rates and higher deposit rates (*i.e.*, banks' common marginal cost of funds). However, in our comparative stationary monetary equilibrium (SME) experiments, higher inflation is associated with successively lower average markups in the banking (loans) sector.

For the result that average loan-rate markup falls with inflation, it must be that average loan rate itself is rising slower than the deposit rate. In this part, we provide a theoretical proof of this result under quite mild regularity conditions. The following proposition says that if the support of an SME loan-rate distribution is not too wide, and, the gap between the lowest posted loan rate and banks' (common) marginal cost of fund is not too large, then one can show that the average markup measure is a decreasing function of long-run inflation.

We will use the notation $f_x(x; y) := \frac{\partial f(x,y)}{\partial x}$ to denote the partial derivative of function $f(x, y)$ with respect to argument x. The results below are with regard to an equilibrium, so we have $z = Z = z^\star(\boldsymbol{\tau})$ and we can also write $\mathbf{z} = (z^\star, \mathbf{z})$.

Proposition 15. *Assume $\gamma = 1 + \tau > \beta$, and $\alpha_1 \in (0, 1)$. Let the average loan-rates markup be*

$$\mu(\gamma) := \frac{\int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} i \, dF(i, \mathbf{z})}{i_d(\gamma)} =: \frac{g(\gamma)}{h(\gamma)},$$

and let the ex-ante *loan-rates markup be*

$$\hat{\mu}(\gamma) := \frac{\int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} i \cdot [\alpha_1 + 2\alpha_2(1 - F(i, \mathbf{z}))] \, dF(i, \mathbf{z})}{i_d(\gamma)} =: \frac{g(\gamma)}{h(\gamma)},$$

where $i_d(\gamma) = \frac{\gamma-\beta}{\beta}$*. If (1):* $\bar{i}(\mathbf{z}) - \underline{i}(\mathbf{z}) < \frac{1}{\beta}$*, and, (2):* $\underline{i}(\mathbf{z}) - i_d(\gamma) < \epsilon(\gamma)$*, where*

$$\epsilon(\gamma) := \sqrt{\frac{1}{\beta}\frac{\alpha_1}{2\alpha_2}\frac{\xi(\bar{i}(\mathbf{z}),\mathbf{z})}{\xi(\underline{i}(\mathbf{z}),\mathbf{z})}\frac{1}{\mu(\gamma)}} > 0,$$

then both the average, posted-loan-rates markup and the ex-ante *loan-rates markup are monotone decreasing in inflation* γ*. Respectively,* $\mu_\gamma(\gamma) < 0$ *and* $\hat{\mu}_\gamma(\gamma) < 0$*.*

We should point out that the sufficient conditions behind Proposition 15 are perhaps not the most general ones, but they suffice practically: For plausible experiments around the empirically calibrated model, the sufficient conditions always hold. For extremely high, hyperinflationary scenarios, these specific sufficient conditions may not hold. Nevertheless, we will see that average loan markup is still decreasing with inflation in our numerical experiments.

Proof. Fix $\gamma > \beta$ (i.e., inflation target away from the Friedman rule) and $\alpha_1 \in (0,1)$ (i.e., agents can meet more than one lending agent). Consider an SME with co-existence of money and bank loans at the given γ. In such an equilibrium, the distribution of loan rates is non-degenerate.

Average, posted loan-rates markup. First, we prove this for $\mu(\gamma)$. At each γ, $g(\gamma) > h(\gamma)$, since average markup is strictly greater than unity $\mu(\gamma) > 1$.

Since the average loan markup function μ is differentiable with respect to γ, then we have

$$\mu_\gamma(\gamma) = \frac{g_\gamma(\gamma)h(\gamma) - g(\gamma)h_\gamma(\gamma)}{[h(\gamma)]^2}. \quad \text{(A.7.1)}$$

To show that average loan-rate markup is decreasing in inflation, $S_\gamma(\gamma) < 0$, it suffices to verify that

$$\frac{g_\gamma(\gamma)}{g(\gamma)} < \frac{h_\gamma(\gamma)}{h(\gamma)}.$$

This requires that the percentage change in average loan rate with respect to inflation is strictly smaller than that of banks' marginal cost of funds.

Using the definition of g and h, we can also rewrite the last inequality as

$$g_\gamma(\gamma) < \frac{1}{\beta}\mu(\gamma). \quad \text{(A.7.2)}$$

Applying integration by parts, we can rewrite average loan-rate $g(\gamma)$ as

$$g(\gamma) = [iF(i,\mathbf{z})]_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} - \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} \frac{\partial i}{\partial i} F(i,\mathbf{z})\,\mathrm{d}i \; = \bar{i}(\mathbf{z}) - \underbrace{\int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} F(i,\mathbf{z})\,\mathrm{d}i.}_{=:\tilde{g}(\gamma)} \tag{A.7.3}$$

Differentiating Expression (A.7.3) with respect to γ yields

$$\begin{aligned} g_\gamma(\gamma) = \bar{i}_\gamma(\gamma) - \tilde{g}_\gamma(\gamma) \; &= \bar{i}_\gamma(\gamma) - \left[\bar{i}_\gamma(\gamma) + \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} F_\gamma(i,\mathbf{z})\,\mathrm{d}i\right] \\ &= -\int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} F_\gamma(i,\mathbf{z})\,\mathrm{d}i, \end{aligned} \tag{A.7.4}$$

where

$$F_\gamma(i,\mathbf{z}) = \frac{\alpha_1}{2\alpha_2}\frac{1}{\beta}\left\{\frac{\xi(\bar{i},\mathbf{z})R(i,\mathbf{z}) - \xi(i;\gamma)R(\bar{i},\mathbf{z})}{[R(i,\mathbf{z})]^2}\right\} \; = \frac{1}{\beta}\frac{\alpha_1}{2\alpha_2}\frac{\xi(\bar{i},\mathbf{z})}{\xi(i,\mathbf{z})}\frac{i-\bar{i}(\mathbf{z})}{[i-i_d(\gamma)]^2} < 0. \tag{A.7.5}$$

The last term $\tilde{g}_\gamma(\gamma)$ in Equation (A.7.4) is obtained by Leibniz' rule:

$$\tilde{g}_\gamma(\gamma) \; = \bar{i}_\gamma(\gamma) + \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} F_\gamma(i,\mathbf{z})\,\mathrm{d}i.$$

Observe that $F_\gamma(\cdot,\mathbf{z})$ is negative valued for all i in the equilibrium support of $F(\cdot,\mathbf{z})$, since $i < \bar{i}$ and since the event that two banks post the same interest rate, $\{i\}$, has zero probability measure in any SME. Thus, from Equations (A.7.4) and (A.7.5), we have that average loan rate is increasing with inflation, or, $g_\gamma(\gamma) > 0$.

Consider Expression (A.7.5). Since loan demand ξ is decreasing in i, $\bar{i}(\mathbf{z}) > \underline{i}(\mathbf{z})$, and, $\underline{i}(z,\mathbf{z}) - i_d(\gamma) < i - i_d(\gamma)$, then the relative demand terms are always bounded in $(0,1)$:

$$0 < \frac{\xi(\bar{i}(\mathbf{z}),\mathbf{z})}{\xi(\underline{i}(\mathbf{z}),\mathbf{z})} < \frac{\xi(\bar{i}(\mathbf{z}),\mathbf{z})}{\xi(i,\mathbf{z})} < 1, \tag{A.7.6}$$

and,

$$0 < \frac{1}{[i-i_d(\gamma)]^2} < \frac{1}{[\underline{i}(z,\mathbf{z}) - i_d(\gamma)]^2} < 1, \tag{A.7.7}$$

for all $i \in \big(\underline{i}(\mathbf{z}), \bar{i}(\mathbf{z})\big)$.

The bounds in Inequalities (A.7.6) and (A.7.7) allow us to look at the extreme case by setting $i = \bar{i}(z,\mathbf{z})$ so that the sufficient bound is independent of the endogenous i. From sufficient condition (1), we can deduce

$$0 < \frac{\bar{i}(z, \mathbf{z}) - i}{\beta} < \frac{\bar{i}(z, \mathbf{z}) - \underline{i}(z, \mathbf{z})}{\beta} < 1. \tag{A.7.8}$$

Using Inequalities (A.7.6), (A.7.7) and (A.7.8), $0 < \alpha_1/2\alpha_2 < 1$, Sufficient Conditions (1) and (2) and Equation (A.7.5), we have an upper bound on how fast the average loan rate varies with inflation:

$$0 < g_\gamma(\gamma) := -\int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} F_\gamma(i, \mathbf{z})\, \mathrm{d}i < \left[\bar{i}(z, \mathbf{z}) - \underline{i}(z, \mathbf{z})\right] \mu(\gamma) < \frac{1}{\beta}\mu(\gamma). \tag{A.7.9}$$

The result above says that the upper bound on $g_\gamma(\gamma)$ is given by the rate of change in the deposit rate with respect to inflation, $1/\beta$, times the average loan rate markup, $\mu(\gamma)$. Therefore, we have that average loan-rate markup decreases with inflation, $\mu_\gamma(\gamma) < 0$.

Note that at any $\gamma > \beta$, the second last term in Condition (A.7.9) gives the area of a rectangle whose height is $\mu(\gamma)$, and width is $\left[\bar{i}(\mathbf{z}) - \underline{i}(\mathbf{z})\right]$. Under sufficient condition (1) and (2), and the fact that $F_\gamma(i, \mathbf{z})$ is monotone decreasing in i, we have that the maximal value of $F_\gamma(i, \mathbf{z})$ is bounded above by $\mu(\gamma)$. Sufficient condition (1) bounds the limits of the integral above by $1/\beta$. Hence the definite integral $g_\gamma(\gamma)$ is bounded: $0 < g_\gamma(\gamma) < \frac{1}{\beta}\mu(\gamma)$. This suffices for the conclusion that average markup is decreasing with inflation, i.e., $\mu_\gamma(\gamma) < 0$ as desired.

***Ex-ante* loan-rates markup.** We now prove the second part. Observe that the only difference between $\mu(\gamma)$ and $\hat{\mu}(\gamma)$ is that in the latter, an additional probability weighting function, $\alpha_1 + 2\alpha_2(1 - F(i, \mathbf{z}))$ appears in the definition of the *ex-ante* or mean transaction rate buyers face. Let this be $\hat{g}(\gamma) := \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} i \cdot [\alpha_1 + 2\alpha_2(1 - F(i, \mathbf{z}))]\, \mathrm{d}F(i, \mathbf{z})$. It is immediate that $0 < \hat{g} \leq g$. Under the same sufficient conditions above, we also have

$$\frac{\hat{g}_\gamma(\gamma)}{\hat{g}(\gamma)} \leq \frac{g_\gamma(\gamma)}{g(\gamma)} < \frac{h_\gamma(\gamma)}{h(\gamma)}.$$

That is, since the integrand in the integral function $\hat{g}$ is dominated by the integrand in g, then $\hat{g}(\gamma)$ can grow no faster than $g(\gamma)$ with respect to inflation γ. Finally, since we concluded that $g(\gamma)$ grows slower than the deposit rate $h(\gamma)$ as γ increases, then so must $\hat{g}(\gamma)$. Thus, $\hat{g}(\gamma)$ is also decreasing with γ under the same sufficient condition.

□

A.8 Equilibrium with banking versus financial autarky

In Section 5 in the paper, we compared welfare gains from banking relative to a no-bank equilibrium. As an alternative exercise, we replace the no-bank equilibrium with an equilibrium where banks exist but agents are restricted in their access to banks. We will consider two flavors of "financial autarky" here.

First, consider a *partial financial autarky* where agents with idle funds cannot deposit. However, lending agents exist and active buyers can still borrow. Here, depository institutions can still source funds from the external market. All else equal, the money demand characterization now becomes

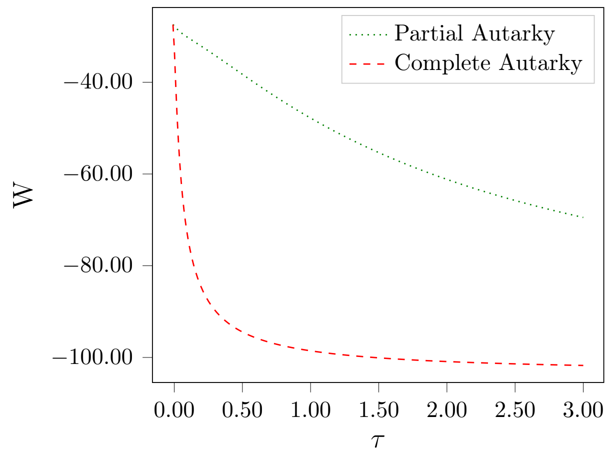
$$\frac{\gamma - \beta}{\beta} = n \Bigg[\alpha_0 \left(u^{'} [q_b^0(\mathbf{z})] - 1 \right) + \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} i \left[\alpha_1 + 2\alpha_2 (1 - F(i, \mathbf{z})) \right] \mathrm{d}F\left(i, \mathbf{z}\right) \Bigg]. \tag{A.8.1}$$

In a similar vein to our discussion in Section 4.2, we now compare the money-demand description in our baseline HKNP economy in Equation (4.5) with that of the partial financial autarky in Equation (A.8.1). The characterization in Equation (A.8.1) is almost identical to our baseline (HKNP) case in Equation (4.5). The only difference is that Equation (A.8.1) is missing the benefit-to-depositors terms $(1-n)i_d$. All else constant, Equation (A.8.1) effectively involves a shift down of the marginal benefit of carrying money in Equation (4.5). That is, HKNP's imperfectly-competitive banks would induce a higher real-money-balance allocation than under this partial financial autarky.

Proposition 16. *Fix inflation away from the Friedman rule, $\gamma > \beta$. Assume a HKNP economy: The probability of contacting two lenders is $\alpha_2 \in (0,1)$ and the measure of active buyer is $n < 1$. Financial intermediation in a HKNP economy improves allocation and welfare relative to partial financial autarky: $z^{p-autarky} < z^{HKNP} < z^{BCW}$ where z^{HKNP} approaches z^{BCW} as the HNKP economy tends to its perfect-competition limit ($\alpha_2 \nearrow 1$).*

Second, consider a *complete financial autarky*: This is the partial autarky case but we further restrict active buyers to be able to meet potential lenders, but they cannot or do not borrow. Hence, complete financial autarky here is to the same degree as the case in Berentsen et al. (2007)).

Figure A.10: Welfare levels—complete versus partial financial autarky.



Money demand in this economy is characterized by

$$\frac{\gamma - \beta}{\beta} = n\Bigg[\alpha_0 \left(u^{'}[q_b^0(\mathbf{z})] - 1\right) + \int_{\underline{i}(\mathbf{z})}^{\bar{i}(\mathbf{z})} \left(u^{'}[q_b^0(\mathbf{z})] - 1\right) [\alpha_1 + 2\alpha_2(1 - F(i, \mathbf{z}))] \, \mathrm{d}F\,(i, \mathbf{z})\Bigg]. \tag{A.8.2}$$

Intuitively, with the further restriction of zero borrowing, we should expect welfare to be worse off. We will show a numerical example here to illustrate this intuition. In Figure A.10, we can see that complete-autarky welfare is dominated by the partial autarky welfare across all long-run inflation settings (except at the Friedman rule).

Let $\hat{z}$ denote the equilibrium money demand in a complete financial autarky. The SME state-policy vector becomes $\hat{\mathbf{z}} = (\hat{z}, \boldsymbol{\tau})$. Welfare in this equilibrium is given by:

$$W^{autarky}(\tau) = \frac{1}{1-\beta}\Bigg[U(x^{\star}) - x^{\star} - c[q_s(\hat{\mathbf{z}})]\Bigg] + \frac{n}{1-\beta}\Bigg[\alpha_0 u[q_b^0(\hat{\mathbf{z}})] + \int_{\underline{i}(\hat{\mathbf{z}})}^{\bar{i}(\hat{\mathbf{z}})} [\alpha_1 + 2\alpha_2 - 2\alpha_2 F(i, \hat{\mathbf{z}})] u[q_b^0(\hat{\mathbf{z}})] \mathrm{d}F(i, \hat{\mathbf{z}})\Bigg]. \tag{A.8.3}$$

Figure 6 shows the welfare gain from banking—relative to complete financial autarky—in both our imperfectly-competitive benchmark (*dotted-green graph*) and the perfectly competitive banking economy (*dash-dotted-red graph*). In either case of perfectly-competitive or our imperfectly-competitive banks, welfare is always higher than that under complete financial autarky.

Also, in both cases, the relationship between trend inflation and welfare gain from banking is non-monotonic: There is little gain from banks in reallocating idle money balances among

Figure A.11: Welfare gain—HNKP or BCW versus financial autarky.

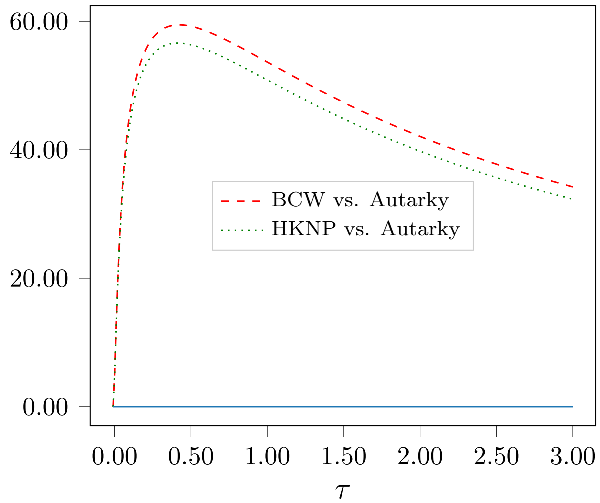
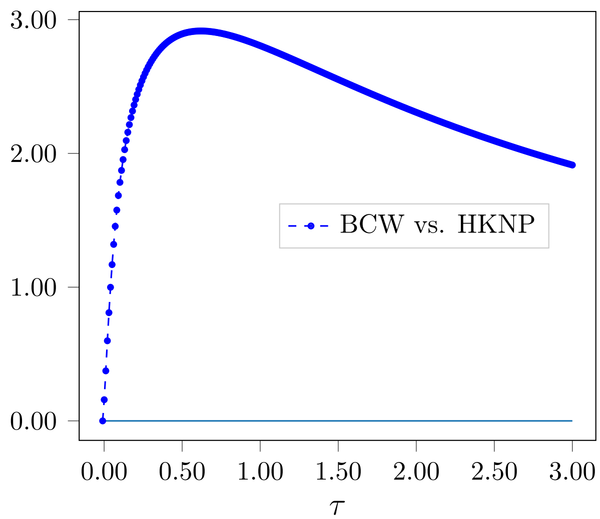


Figure A.12: Welfare difference—BCW versus HNKP.



households near the Friedman rule (or when inflation $\tau \to \infty$) in both cases. This is because the cost of carrying money is small when inflation is low and money has little value in hyperinflationary economies.

We can also directly compare between the BCW, perfect-competition welfare outcomes with those of our model (HKNP). This is shown in Figure A.12.

A.9 Aggregate demand shocks in the baseline model

We provide the details of the stochastic version of the baseline model used in Section 6 of the paper here. In particular, we characterize the SME with aggregate demand shocks and we set up the Ramsey optimal policy problem for aggregate demand stabilization. The optimal policy exercise here is in the same spirit as Berentsen and Waller (2011). In contrast to the perfectly-competitive banking environment of Berentsen et al. (2007) and Berentsen and Waller (2011), our model now has non-trivial consequences for the design of optimal monetary policy in response to aggregate demand shocks.

A.9.1 Shocks

We can consider n and/or ϵ as random variables to capture aggregate demand fluctuation in the DM. The random variable ϵ is interpreted as marginal utility shock of the DM special goods, and the random variable n is interpreted as shock that affects the number of active DM buyers (or we can interpert that as shocks to the probability of *early consumption*). In particular, $n \in [\underline{n}, \overline{n}]$, where $0 < \underline{n} < \overline{n} < 1$, and $\epsilon \in [\underline{\epsilon}, \overline{\epsilon}]$, where $0 < \underline{\epsilon} < \overline{\epsilon} < \infty$. Let $\omega = (n, \epsilon) \in \Omega$ denote the aggregate state (vector).

A.9.2 Monetary policy

The central bank commits to an overall long-run inflation target τ (or equivalently, a price path) and conducts state-contingent liquidity policy. The timing and sequence of actions is as follows. First, monetary injection, τM occurs at the beginning of the period (before shocks are realized). Second, the central bank injects a lump-sum amount of money to the DM agents up to $\tau_1(\omega)M = n\tau_b(\omega) + (1-n)\tau_b(\omega) + \tau_s(\omega)$ after shocks are realized. We assume the central bank can only tax in the CM but not the DM, so only transfers are premitted in the DM, i.e., $\tau_1(\omega) \geq 0$. Third, agents conduct banking arrangement, exchange and production of DM goods take place. Lastly, agents trade goods, work and settle the financial claims (loans and deposits) in the subsequent CM. Here, we have also assumed any state-contingent injection of liquidity received by the DM agents will be undone in CM, i.e., $\tau_2(\omega) = -\tau_1(\omega)$. The state-contingent policy plan here can be thought as a repo agreement made by the central bank: The central bank sells money in DM and promised to buy that back in the CM.

Given the assumption of DM state-contingent policy will be undone in the subsequent CM, the total change to the aggregate money stock is deterministic, which is given by

$$M_{+1} - M = (\gamma - 1)M = \tau M, \tag{A.9.1}$$

where $\gamma = 1 + \tau$ is the gross growth rate of money supply.

A.9.3 Characterization of SME with shocks

The markets structure of the model is the same as in baseline except that ϵ and n are random variables now.[38] We also work with stationary variables and restrict attention to stationary monetary equilibrium (SME). End-of-period real money balances are both time and state invariant

$$\phi M = \phi_{+1} M_{+1} = z, \qquad \text{for all } \omega \in \Omega,$$

and in an SME,

$$\frac{\phi}{\phi_{+1}} = \frac{M_{+1}}{M} = \frac{p_{+1}}{p} = \gamma = 1 + \tau.$$

This means that the central bank engages in price-level targeting by choosing a path for the money stock as in Berentsen and Waller (2011).

Before we go into the summary of a SME in the economy with shocks, we lay out its components.

Ex-post households with at least one lending bank contact. In events with probability measure α_1 and α_2, the buyer's optimal demand for DM consumption and loan is respectively characterized by

$$q_b^{1,\star}(z, \mathbf{z}, \omega) = \begin{cases} \epsilon^{\frac{1}{\sigma}} \left[\rho(1+i)\right]^{-\frac{1}{\sigma}} & \text{if } 0 < \rho \leq \tilde{\rho}_i \text{ and } 0 \leq i \leq \hat{i} \\ \frac{z + \tau_b Z}{\rho} & \text{if } \tilde{\rho}_i < \rho < \hat{\rho} \text{ and } i > \hat{i} \\ \epsilon^{\frac{1}{\sigma}} \rho^{-\frac{1}{\sigma}} & \text{if } \rho \geq \hat{\rho} \text{ and } i > \hat{i} \end{cases}, \tag{A.9.2}$$

and,

$$\xi^{\star}(z, \mathbf{z}, \omega) = \begin{cases} \epsilon^{\frac{1}{\sigma}} \rho^{\frac{\sigma-1}{\sigma}} (1+i)^{-\frac{1}{\sigma}} - (z + \tau_b Z) & \text{if } 0 < \rho \leq \tilde{\rho}_i \text{ and } 0 \leq i \leq \hat{i} \\ 0 & \text{if } \tilde{\rho}_i < \rho < \hat{\rho} \text{ and } i > \hat{i} \\ 0 & \text{if } \rho \geq \hat{\rho} \text{ and } i > \hat{i} \end{cases}, \tag{A.9.3}$$

[38] If we treat ϵ and n as parameter, and set $\epsilon = 1$, then we are back to the deterministic baseline case.

where

$$\hat{\rho} := \hat{\rho}(z, \mathbf{z}, \omega) = \epsilon^{-\left(\frac{1}{\sigma-1}\right)} (z + \tau_b Z)^{\frac{\sigma}{\sigma-1}},$$

$$\tilde{\rho}_i := \hat{\rho}\,(1+i)^{\frac{1}{\sigma-1}},$$

$$\text{and } \hat{i} = \epsilon\,(z + \tau_b Z)^{-\sigma} \rho^{\sigma-1} - 1 > 0.$$

where both ex-post demand functions require to hold for each realization of state $\omega \in \Omega$.

Ex-post households with zero lending bank contact. The buyer's optimal demand for DM consumption (for events with probability measure α_0) is

$$q_b^{0,\star}(z, \mathbf{z}, \omega) = \begin{cases} \frac{z+\tau_b Z}{\rho} & \text{if } \rho \leq \hat{\rho} \\ \epsilon^{\frac{1}{\sigma}} \rho^{-\frac{1}{\sigma}} & \text{if } \rho \geq \hat{\rho} \end{cases}, \tag{A.9.4}$$

where

$$\hat{\rho} := \hat{\rho}(z, \mathbf{z}, \omega) = \epsilon^{-\left(\frac{1}{\sigma-1}\right)} (z + \tau_b Z)^{\frac{\sigma}{\sigma-1}}.$$

which requires to hold for each realization of state $\omega \in \Omega$.

Firms. The firm's optimal production plan satisfies

$$c_q(q_s) = p\phi, \qquad \omega \in \Omega, \tag{A.9.5}$$

where the marginal cost of producing is equal to the real relative price of DM goods for each realization of state ω.

Hypothetical monopolist lending bank. We can derive the closed-form loan-price posting distribution similar to the baseline, except that the distribution is both state and policy dependent now. Given a realization of shock ω, this bank's "monopoly" profit function is

$$\Pi^m(i) = n\alpha_1 R(i).$$

To pin down a monopoly loan price, differentiate the bank's "monopoly" profit function wrt. i, the (stationary variable version) FOC is

$$-\underbrace{z + \tau_b Z}_{f(i)} + \underbrace{\frac{1}{\sigma}\epsilon^{\frac{1}{\sigma}} (1+i)^{-\frac{1}{\sigma}} \left[(\sigma - 1) + \frac{1+i_d}{1+i}\right]}_{g(i)} = 0, \tag{A.9.6}$$

which needs to hold for each realization of state $\omega \in \Omega$.

Observe that in Condition (A.9.6), for a given individual state z, aggregate state Z, trend inflation rate τ, state ω, and $\omega \mapsto \tau_b(\omega)$, $f(i)$ is a constant w.r.t. i, and $g(i)$ is decreasing in i. Thus, as in the earlier, baseline model, there exists a unique monopoly-profit-maximizing price i^m that satisfies the above FOC for each realization of state $\omega \in \Omega$.

Once we pin down this $i^m(\mathbf{z}, \omega)$ in a SME, then we use the equal profit condition combining with the upper support of the distribution $\bar{i}(\omega) := \min\{i^m(\mathbf{z}, \omega), \hat{i}(\mathbf{z}, \omega)\}$ to derive the lower support of the distribution $\underline{i}(\mathbf{z}, \omega)$, which together pin down the closed-form loan-price posting distribution for each realization of state $\omega \in \Omega$.

Real money demand. Similar to the baseline case, we differentiate the DM value function with respect to m, update one period and substitute that into the CM first-order condition. Convert the result using stationary variables and combining that with the *ex-post* optimal goods demand functions in Equations (A.9.2) and (A.9.4) in DM, and we get the Euler equation for real money demand as

$$
\begin{aligned}
\frac{\gamma - \beta}{\beta} &= \theta\left(z, \mathbf{z}, \omega\right) - 1 \\
&+ \int_{\omega \in \Omega} n \mathbb{I}_{\{\rho < \hat{\rho}\}} \alpha_0 \left[\frac{1}{\rho} \epsilon \left(\frac{z + \tau_b(\omega) z}{\rho}\right)^{-\sigma} - 1\right] \psi\left(\omega\right) \mathrm{d}\omega \\
&+ \int_{\omega \in \Omega} n \int_{\underline{i}}^{\bar{i}} \mathbb{I}_{\{\rho < \tilde{\rho}_i\}} \left[\alpha_1 + 2\alpha_2\left(1 - F\left(i, z, \mathbf{z}, \omega\right)\right)\right] i \mathrm{d}F\left(i, z, \mathbf{z}, \omega\right) \psi\left(\omega\right) \mathrm{d}\omega \\
&+ \int_{\omega \in \Omega} n \int_{\underline{i}}^{\bar{i}} \mathbb{I}_{\{\tilde{\rho}_i \leq \rho < \hat{\rho}\}} \left[\alpha_1 + 2\alpha_2\left(1 - F\left(i, z, \mathbf{z}, \omega\right)\right)\right] \\
&\times \left[\frac{1}{\rho} \epsilon \left(\frac{z + \tau_b(\omega) z}{\rho}\right)^{-\sigma} - 1\right] \mathrm{d}F\left(i, z, \mathbf{z}, \omega\right) \psi\left(\omega\right) \mathrm{d}\omega,
\end{aligned}
\tag{A.9.7}
$$

and,

$$
\begin{aligned}
\theta(z,\mathbf{z},\omega)-1 :=& \int_{\omega\in\Omega}(1-n)(1+i_d)\,\psi(\omega)\,\mathrm{d}\omega \\
&+\int_{\omega\in\Omega} n\alpha_0\psi(\omega)\,\mathrm{d}\omega \\
&+\int_{\omega\in\Omega} n\int_{\underline{i}}^{\bar{i}} \mathbb{I}_{\{\rho<\tilde{\rho}_i\}}\left[\alpha_1+2\alpha_2\left(1-F(i,z,\mathbf{z},\omega)\right)\right]\mathrm{d}F(i,z,\mathbf{z},\omega)\,\psi(\omega)\,\mathrm{d}\omega \\
&+\int_{\omega\in\Omega} n\int_{\bar{i}}^{i^{\mathrm{m}}} \mathbb{I}_{\{\tilde{\rho}_i\leq\rho<\hat{\rho}\}}\left[\alpha_1+2\alpha_2\left(1-F(i,z,\mathbf{z},\omega)\right)\right]\mathrm{d}F(i,z,\mathbf{z},\omega)\,\psi(\omega)\,\mathrm{d}\omega \\
&+\int_{\omega\in\Omega} n\int_{\bar{i}}^{i^{\mathrm{m}}} \mathbb{I}_{\{\hat{\rho}\leq\rho\}}\left[\alpha_1+2\alpha_2\left(1-F(i,z,\mathbf{z},\omega)\right)\right]\mathrm{d}F(i,z,\mathbf{z},\omega)\,\psi(\omega)\,\mathrm{d}\omega \\
&-1.
\end{aligned}
$$

Note that the integral limits $(\bar{i}, i^{\mathrm{m}}, \underline{i})$ and cut-off prices $(\tilde{\rho}_i, \hat{\rho})$ are also functions of $(z, \mathbf{z}, \omega)$. The LHS of Condition (A.9.7) captures the marginal cost of accumulating an extra unit of real money balance at the end of each CM, and the RHS captures the expected marginal utility value of that extra unit of money balance (evaluated at the beginning of next DM before shock is realized and before buyer types, matching and trading occurs).

Loan price-posting distribution. We restrict to the case $\alpha_1 \in (0,1)$ for the stochastic version here. The distribution of loan (interest-rate) price posts is given by:

$$
F(i,z,\mathbf{z},\omega) = 1-\frac{\alpha_1}{2\alpha_2}\left[\frac{R\left(\bar{i}(z,\mathbf{z},\omega)\right)}{R\left(i(z,\mathbf{z},\omega)\right)}-1\right], \tag{A.9.8}
$$

and, $\mathrm{supp}\left(F(\cdot,z,\mathbf{z},\omega)\right) = \left[\underline{i}(z,\mathbf{z},\omega), \bar{i}(z,\mathbf{z},\omega)\right]$, and, given $\bar{i}(z,\mathbf{z},\omega) = \min\{i^m(z,\mathbf{z},\omega), \hat{i}(z,\mathbf{z},\omega)\}$, $\underline{i}(z,\mathbf{z},\omega)$ solves:

$$
R\left(\underline{i}(z,\mathbf{z},\omega)\right) = \frac{\alpha_1}{\alpha_1+2\alpha_2}R\left(\bar{i}(z,\mathbf{z},\omega)\right), \tag{A.9.9}
$$

where the (real) bank profit per customer served is

$$
R(i,z,\mathbf{z},\omega) = \left[\epsilon^{\frac{1}{\sigma}}\rho^{\frac{\sigma-1}{\sigma}}(1+i)^{-\frac{1}{\sigma}}-(z+\tau_b Z)\right]\left(i-i^d\right). \tag{A.9.10}
$$

Observe that in Equations (A.9.2) and (A.9.3), all the cutoff functions (in terms of relative price of DM goods or lending interest rate) are all now depend on the optimal policy function, $\omega \mapsto \tau_b(\omega)$ function, and also on the $\omega := (\epsilon, n)$ states of the economy.

Similarly, the support of the posted loan interest rate distribution in Equation (A.9.8) now also depends on a given $\omega \mapsto \tau_b(\omega)$ function, and also on $\omega := (\epsilon, n)$. This can be seen from the optimal monopoly rate that solves the Condition (A.9.6), from households' reservation interest rate $\hat{i}(z, \mathbf{z}, \omega)$, and from the associated lowest possible loan rate of the distribution $\underline{i}(z, \mathbf{z}, \omega)$.

The key difference between ϵ shocks and n shocks is that the former induces one extra moving part—a direct effect of policy outcomes $\tau_b(\omega)$ on the support of $F(\cdot, z, \mathbf{z}, \omega)$. The latter shock implies one less moving part.

Competitive price taking and goods market clearing. DM goods market clears for all ω:

$$
\begin{aligned}
q_s(z, \mathbf{z}, \omega) \equiv c^{'-1}(\rho) &= \int_\Omega n\alpha_0 q_b^{0,\star}(z, \mathbf{z}, \omega)\,\psi(\omega)\mathrm{d}\omega \\
&+ \int_\Omega n \int_{\underline{i}(z,\mathbf{z},\omega)}^{\bar{i}(z,\mathbf{z},\omega)} \left[\alpha_1 + 2\alpha_2\left(1 - F(i, z, \mathbf{z}, \omega)\right)\right] q_b^\star(i, z, \mathbf{z}, \omega)\,\mathrm{d}F(i, z, \mathbf{z}, \omega)\,\psi(\omega)\mathrm{d}\omega.
\end{aligned}
\tag{A.9.11}
$$

We can also verify that the CM labor and goods market clear given the SME solutions $\{z^\star, q_b^{0,\star}, q_b^\star\}$.

Aggregate feasibility of loanable funds in banking market. Interests on total loans weakly exceed that on total deposits

$$
\begin{aligned}
&\int_\Omega n \int_{\underline{i}(z,\mathbf{z},\omega)}^{\bar{i}(z,\mathbf{z},\omega)} \left[\alpha_1 + 2\alpha_2\left(1 - F(i, z, \mathbf{z}, \omega)\right)\right] \xi^\star(i, z, \mathbf{z}, \omega)\, i\mathrm{d}F(i, z, \mathbf{z}, \omega)\,\psi(\omega)\mathrm{d}\omega \\
&\qquad \geq \int_\Omega (1-n)\, i_d \delta^\star(z, \mathbf{z}, \omega)\psi(\omega)\mathrm{d}\omega \qquad \text{(A.9.12)} \\
&\qquad \equiv \int_\Omega (1-n)\, i_d \left(\frac{z + \hat{\tau}_b(\omega)Z}{\rho}\right)\psi(\omega)\mathrm{d}\omega,
\end{aligned}
$$

for each realization of state $\omega \in \Omega$.

We now summarize the description of a SME with aggregate demand shocks below.

Definition 17. Assume $\sigma < 1$. Given money supply growth $\gamma = 1 + \tau$, and redistributive policy plan $\{\boldsymbol{\tau}(\omega)\}_{\omega\in\Omega} := \{\tau_1(\omega), \tau_2(\omega)\}_{\omega\in\Omega}$ a *Stationary Monetary Equilibrium (SME)* is a list of time- and state-invariant CM consumption allocation and residual real money balance outcomes $\{x^\star \equiv 1, z^\star\}$, $z^\star = Z$ so that $\mathbf{z} = (z^\star, \boldsymbol{\tau}(\omega))$ and time-independent allocation functions for DM goods and loans, $\{q^\star(\mathbf{z}^\star, \omega), \xi^\star(\mathbf{z}^\star, \omega)\}$, and distribution, $F(\cdot; \mathbf{z}^\star, \omega)$ such that:

1. household optimization satisfies the money-demand Euler Equation (A.9.7);

2. the distribution of posted loan (interest-rate) price satisfies Equation (A.9.8);

3. DM goods market clearing satisfies Condition (A.9.11);

4. loans feasibility satisfies Requirement (A.9.12); and

5. the government budget constraint holds for each ω, i.e.,

$$\frac{\gamma-\beta}{\beta} = \tau + \tau_1(\omega) + \tau_2(\omega), \quad \tau_1(\omega) = -\tau_2(\omega). \tag{A.9.13}$$

A.9.4 Optimal stabilization policy over SME with shocks

To understand how the stabilization policy in response to demand fluctuation may work, we compare two types of government policy:

1. **Active central bank**. The policymaker commits to an ex-ante, optimal policy plan that maximizes social welfare over a steady-state equilibrium (i.e., a SME). In particular, the active central bank solves

$$\begin{aligned} \max_{\{q_b^0(\cdot,\omega),\ q_b(\cdot,\omega),\tau_b(\omega)\}_{\omega\in\Omega}} \quad & U\left(x\right) - x - c(q_s(\mathbf{z},\omega)) \\ & + \int_{\omega\in\Omega} n\alpha_0 \epsilon u\left[q_b^0\left(\mathbf{z},\omega\right)\right]\psi\left(\omega\right)\mathrm{d}\omega \\ & + \int_{\omega\in\Omega} n \int_{\underline{i}(\mathbf{z},\omega)}^{\bar{i}(\mathbf{z},\omega)} \left[\alpha_1 + 2\alpha_2\left(1 - F\left(i,\mathbf{z},\omega\right)\right)\right] \\ & \times \epsilon u\left[q_b\left(i,\mathbf{z},\omega\right)\right]\mathrm{d}F\left(i,\mathbf{z},\omega\right)\psi\left(\omega\right)\mathrm{d}\omega \end{aligned} \tag{A.9.14}$$

   subject to optimal money demand in Equation (A.9.7) the distribution of loan interest rates in Equation (A.9.8), the DM goods market clearing in Condition (A.9.11), the loans feasibility Condition (A.9.12) and government budget feasibility in Condition (A.9.13), where q_s is given by Equation (A.9.11).

   Note: The policy plan prescribes ω-contingent liquidity injections, i.e., $\tau_1(\omega) = \tau_b(\omega) \geq 0$, and assume $\tau_2(\omega) = -\tau_1(\omega)$ for all $\omega \in \Omega$.

2. **Passive central bank**. In this regime, the policymaker is constrained by $\tau_1(\omega) = \tau_2(\omega) = 0$ for all $\omega \in \Omega$. The outcomes will be very similar to our deterministic, baseline SME.

The objective of the active central bank is similar to Berentsen and Waller (2011). New insights arises from the equilibrium varying dispersion of loan-price markups since $F(i;\omega)$ is now both state and policy dependent. We explain what the new insights are in Section 6 of the paper.