

Essays on Financial Intermediation and Liquidity

Ye Li

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

© 2017

Ye Li

All rights reserved

ABSTRACT

Essays on Financial Intermediation and Liquidity

Ye Li

This dissertation studies the demand and supply of liquidity with a particular focus on the financial intermediation sector. The first essay analyzes the role of financial intermediaries as suppliers of inside money. The demand for money arises from the needs of nonfinancial corporations to buffer liquidity shocks. The dynamic interaction between inside money supply and demand gives rise to a mechanism of financial instability that puts the procyclicality of intermediary leverage at the center. Introducing outside money, in the form of government debt, can be counterproductive, as it may amplify the procyclicality of inside money creation and intermediary leverage, making booms more fragile and crises more stagnant.

The second essay addresses an issue that is left out in the first essay – the interaction between money and credit. It offers a model of macroeconomy where intermediaries are needed for both money and credit creation. Specifically, entrepreneurs hold money to finance new projects, while intermediaries issue money backed by investments in existing projects. The complementarity between money and credit arises from financial frictions and amplifies economic fluctuations.

In the third essay, my coauthors and I model the liquidity demand of banks. To buffer liquidity shocks, banks hold central bank reserves and can borrow reserves from each other. The propagation of liquidity shocks, depend on the topology of interbank credit network, but more importantly, on the type of equilibrium on the network (strategic complementarity vs. substitution). The model is estimated using data on reserves, interbank credit, bank balance sheets, and macroeconomic variables. We propose a method to identify banks that contribute the most to systemic risk, and offer policy guidance by comparing the decentralized outcome with the choice of a benevolent planner.

Content

List of Figures and Tables	iii
Acknowledgments	v
Essay 1: Procyclical Finance: The Money View	1
Introduction	2
Static Model: An Anatomy of Money Shortage	9
Dynamic Model: Procyclical Money Creation	19
Government Debt: Instability, Stagnation, and Welfare	43
Conclusion	52
References	55
Appendix I Proofs	65
Appendix II Calibration and Additional Results	81
Appendix III Preliminary Evidence	83
Essay 2: The Rise of Intangible Capital and Financial Fragility	93
Introduction	94
Model	102
Solution	116
Conclusion	126
References	128
Appendix: Solving the Equilibrium	134
Essay 3: Network Risk and Key Players: A Structural Analysis of Interbank Liquidity	137
Introduction	138
The Network Model	146
Empirical Methodology	159
Description of the Network and Other Data	164

Estimation Results 172
Conclusion 189
References 191
Appendix: Background, Alternative Model, Estimation Method, More Results . . 197

List of Figures and Tables

Essay 1: Procyclical Finance: The Money View

Figures:

Timeline	12
Money Market	27
Intertemporal Feedback and Procyclicalit y	29
Money Market Response to A Positive Shock	31
State Variable Dynamics.	37
Procyclical Leverage.	39
Static and Dynamic Investment Inefficiencies.	42
Outside money: Instability and Stagnation	45
Non-monotonic Effects of Outside Money on Investment and Growth	48
Money Premium Standard Deviation and Government Debt	83
Inside Money Demand and Supply	84
Money Premium and Nonfinancial Firms' Holdings of Money-like Securities	86
Money-like Securities Issued by Financial Intermediaries and Government	87
Intermediary Leverage Cycle and Government Debt	89

Tables:

Calibration	81
Government Debt and Procyclical Leverage	91

Essay 2: The Rise of Intangible Capital and Financial Fragility

Figures:

Money Supply and Demand	98
Equilibrium Capital Price and Allocation	117
Endogenous Risk and Risk Pricing	119
Risk Accumulation and Amplification	121

Inside Money Channel	123
Simulation	124
Financial Instability in an Intangible Economy	125

Essay 3: Network Risk and Key Players: A Structural Analysis of Interbank Liquidity

Figures:

Second Largest Eigenvalue of G_t	166
Average Clustering Coefficient of the Interbank Network	167
Daily Gross Interbank Borrowing in the Interbank Network	168
Weekly Average of Aggregate Liquidity Available at the Beginning of the Day.	169
Intraday Volatility of Aggregate Outflows	170
The Period Before the Norther Rock/Hedge Fund Crisis	176
After the Hedge Fund Crisis but Before QE	179
QE Period.	180
Decomposition of Total Change in the $NIRF_s$ between Periods	186
Spatial Error and Durbin Rolling Estimates of ϕ	187
Velocity of Money in the Payment System	209
Turnover Rate in the Payment System	210
Weekly Average of the Right Kurtosis of Aggregate Payment Times	210
Interest Rates in the Interbank Market	211
Cross-sectional Dispersion of Interbank Rates	211

Tables:

Spatial Error Model Estimation	173
Rank Correlation of Bank Characteristics and $NIRF_i^e$	182
Central Planner vs. Market Equilibria	184
Full Spatial Error Model Estimation	212
Ratio of Network to Idiosyncratic Volatility	213

Acknowledgments

I want to thank my dissertation committee members: Prof. Patrick Bolton (chair), Prof. Olivier M. Darmouni, Prof. Tano Santos (sponsor), Prof. José A. Scheinkman, and Prof. Neng Wang. I am grateful to faculty members, staff, and fellow students at Columbia University.

I want to thank my family, to whom I dedicate this dissertation, for their endless support.

To my beloved family

Procyclical Finance: The Money View*

Ye Li[†]

Abstract

This paper offers a theory of procyclical inside money creation and the resulting instability. Banks hold risky loans and issue safe debt that serves as stores of value and means of payment (“inside money”). Firms hold bank debt to buffer liquidity shocks that cut off financing precisely when resources are needed for growth. Aggregate shocks to loan return trigger boom-bust cycles through banks’ balance-sheet capacity. In booms, abundant supply of inside money helps firms manage liquidity. As firms’ enterprise value rises, they hold even more money in anticipation of growth opportunities. The feedback loop pushes up bank leverage, so downside risk accumulates as a boom prolongs. In crises, the spiral flips. Recovery is sluggish, as banks rebuild their capital slowly under low leverage. Introducing government debt as an alternative money (“outside money”) can be counterproductive. Its competition with inside money amplifies the procyclicality of bank leverage, and slows down the rebuild of bank capital in crises, making booms more fragile and crises more stagnant.

*I am deeply indebted to Patrick Bolton, Tano Santos, and José A. Scheinkman for their invaluable guidance. I also thank Tobias Adrian, Geert Bekaert, Nina Boyarchenko, Markus Brunnermeier, Charles Calomiris, Murillo Campello, Georgy Chabakauri, Guojun Chen, Gilles Chemla, Dong Beom Choi, Nicolas Crouzet, Olivier M. Darmouni, Ian Dew-Becker, Douglas Diamond, John Donaldson, Fernando Duarte, Lars Peter Hansen, Zhiguo He, Harrison Hong, Yunzhi Hu, Gur Huberman, Christian Julliard, Peter Kondor, Arvind Krishnamurthy, Siyuan Liu, Marco Di Maggio, Harry Mamaysky, Konstantin Milbradt, Emi Nakamura, Martin Oehmke, Mitchell Petersen, Tomasz Piskorski, Vincenzo Quadrini, Adriano A. Rampini, Raghuram Rajan, Jon Steinsson, Amir Sufi, Huijun Sun, Suresh Sundaresan, Harald Uhlig, Dimitri Vayanos, S. Viswanathan, Chen Wang, Neng Wang, Kathy Yuan, and Jing Zeng for helpful comments. I am grateful to seminar/conference participants at Becker Friedman Institute (University of Chicago), Chicago Booth, CEPR Credit Cycle Conference, Columbia Finance Seminar, Columbia Economics Colloquium, Federal Reserve Bank of New York, Finance Theory Group, Georgetown McDonough, Imperial College, Johns Hopkins Carey, London School of Economics, Nanyang Technological University, Northwestern Kellogg, NYU Stern PhD Seminar, OSU Fisher, Oxford Financial Intermediation Theory Conference, University of Melbourne, USC Marshall, and Wharton. I acknowledge the generous dissertation support from Macro Financial Modeling Group of Becker Friedman Institute. All errors are mine.

[†]Columbia University. E-mail: yli17@gsb.columbia.edu

1 Introduction

In the years leading up to the Great Recession, the financial sector grew rapidly, setting a favorable liquidity condition that stimulated the real economy. A booming real sector in turn fueled financial intermediaries' expansion and leverage. During the crisis, the spiral flipped. Much progress has been made in recent years to characterize crisis dynamics by incorporating intermediaries in macro models (e.g., He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014)), yet a complete account of procyclical intermediation, and in particular, the run-up to crisis, remains a challenge.

This paper argues that at the heart of this procyclicality is intermediaries' role as money creators. The monetary aspect of financial intermediation is so ubiquitous that we often fail to notice. Intermediary debt (e.g., bank deposits) is a store of value, but more importantly, it supports trade by serving as a means of payment, or "inside money".¹ In an economy where agents' future income is not fully pledgeable, money facilitates spot transactions and resources reallocation. However, the money demand of the real sector feeds leverage to the financial sector, and thus, breeds instability.

I build a continuous-time model of macroeconomy that crystallizes this money view of financial intermediation. A key ingredient is the money demand from firms' liquidity management problem that is similar to Holmström and Tirole (1998). Banks supply money by issuing debt, and thus, build up leverage in the process. The dynamic interaction between money demand and supply generates a rich set of unique predictions, such as the cyclicity of intermediary leverage, money premium dynamics, the accumulation of endogenous risk in booms, stagnant recessions, and real investment inefficiencies driven by cyclical inside money supply.

The idea that financial intermediaries affect the real economy through inside money supply goes back at least to the classic account of the Great Depression by Friedman and Schwartz

¹The term "inside money" is from Gurley and Shaw (1960). From the private sector's perspective, fiat money and government securities are in positive supply ("outside money"), while bank deposits are in zero net supply ("inside money"). Both outside and inside money facilitate transactions. Lagos (2008) briefly reviews the concept.

(1963). One may argue that this money view is less relevant today given the active supply of *outside money* in the form of liquid government securities and central bank liabilities (Woodford (2010)). However, the model shows that the competition between inside and outside money can *destabilize* the intermediary sector by amplifying its leverage cycle, exacerbate the endogenous risk accumulation in booms, and prolong financial crises. These unique results complement the recent literature on outside money supply as a means to financial *stability* (Greenwood, Hanson, and Stein (2015); Krishnamurthy and Vissing-Jørgensen (2015); Woodford (2016)).

The model economy operates in continuous time. It has three types of agents, bankers, entrepreneurs (“firms”), and households who play a limited role. All agents are risk-neutral with the same time discount rate, and consume nonstorable generic goods produced by firms’ capital.

Firms can hold capital and bank deposits, and can borrow from banks and households *as long as* they are not hit by the liquidity shock. Every instant, firms face a constant probability of liquidity shock, and in such an event, their production halts, and their capital can either grow – if further investment is made – or perish, if not. This investment is not pledgeable, so firms obtain goods as inputs through spot transactions with other firms, using deposits as means of payment. Therefore, banks add value because their debt (deposits) facilitates trade and resources reallocation.

Bankers issue short-term risk-free debt (“deposits”), i.e., the inside money, and they extend loans to firms that are backed by designated capital as collateral. Every instant, a stochastic fraction of collateral is destroyed, and thus, the corresponding loans default. There is only one aggregate shock, a Brownian motion, that drives the stochastic destruction of capital. Following bad shocks, more loans default, and bankers lose their equity. Bankers can raise equity by issuing shares to households subject to an issuance cost. This recapitalization friction makes banks effectively risk-averse, and generates persistent effect of shocks on bank equity.

Money creation requires risk-taking, for at the margin, one more dollar of deposits is backed by one more dollar of risky loan. Therefore, inside money supply depends on bank equity. When

undercapitalized banks fail to supply enough money, firms' liquidity management is compromised.

The model has a Markov equilibrium with the ratio of bank equity to firm capital as state variable, which is intuitively the size of money suppliers relative to money demanders. Because banks have leveraged exposure to the capital/collateral destruction shock, the state variable rises following good shocks, and falls following bad ones. It is also bounded by two endogenous boundaries: when the banking sector is huge, the marginal value of equity equals one, so bankers consume and pay out dividends to shareholders (upper boundary); when the banking sector is tiny, bankers raise equity because the marginal value of equity reaches one plus the issuance cost (lower boundary).²

Bank leverage is procyclical.³ Good shocks increase bank equity, but banks issue even more debt to meet firms' procyclical money demand. Following good shocks, fewer loans default than expected. Banks will hoard the windfall instead of paying it out because the issuance cost creates a wedge between the marginal value of equity and one dollar, so the shocks' impact will only dissipate gradually. Expecting banks to be better capitalized and to issue more money going forward, firms expect themselves to carry more money *in the future* that will then finance a faster growth of capital. As a result, capital becomes more valuable, making firms want to hold more money *today* in case the liquidity shock arrives the very next instant. Through the cash-financed growth, firms' money demand exhibits intertemporal complementarity. It rises in the expectation of future money market conditions, feeding leverage to banks. This paper shares with Kiyotaki and Moore (1997) the idea that certain types of intertemporal complementarity amplify economic fluctuations.

Downside risk accumulates through procyclical leverage.⁴ As banks become more levered,

²As in Phelan (2015) and Klimenko, Pfeil, Rochet, and Nicolo (2016), banks issue equity in bad times, and pay out dividends in good times, which is consistent with the evidence in Baron (2014) and Adrian, Boyarchenko, and Shin (2015).

³The leverage is the ratio of book asset to book equity, as will be clearly defined in the model.

⁴Intermediary leverage does not exhibit procyclicality in models that feature a static demand for intermediary debt and focus on the asset-side of intermediary balance sheet to look for channels of shock amplification (e.g. He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014); Phelan (2015); Klimenko, Pfeil, Rochet, and Nicolo

their equity is more sensitive to shocks. And, as the economy approaches bank payout boundary, high leverage only serves to amplify bad shocks, because after good shocks, bank equity cannot rise above the boundary without triggering payout. So the longer a boom lasts, the higher downside risk is. Bad shocks trigger deleveraging, and flip downward the feedback between firms and banks.

Crises are stagnant. As the economy approaches bank issuance boundary, low bank leverage only serves to dampen good shocks, because bank equity never falls below the issuance boundary. Therefore, banks can only rebuild equity after a sequence of sufficiently large good shocks. The calibrated model predicts an eight-year recovery period, during which the economy is stuck with insufficient money creation that compromises firms' liquidity management and investment. In sum, fragile booms and stagnant crises result from a combination of procyclical leverage and the asymmetric impact of shocks near the reflecting boundaries of bank payout and equity issuance.

So far, we have focused on the procyclical quantity of inside money. The model also generates a countercyclical price of money, the *money premium*, a spread between the time discount rate and deposit rate. Being able to borrow at an interest rate lower than the time discount rate, banks earn the money premium. Carrying low-yield deposits, firms pay the money premium, which is a cost of liquidity management, but they optimally do so in anticipation of future transaction needs.

Another important theme of this paper is the financial stability implications of outside money. I model outside money as government debt that offers the same monetary service to firms as bank deposits. Its empirical counterparts include a broad range of liquid government securities, not just central bank liabilities.⁵ To highlight the competition between inside and outside money, I abstract away other fiscal distortions by assuming the debt issuance proceeds are paid to agents

(2016)).

⁵The monetary service of government liabilities is an old theme (e.g., Patinkin (1965); Friedman (1969)). Recent contributions include Bansal and Coleman (1996), Bansal, Coleman, and Lundblad (2011), Krishnamurthy and Vissing-Jørgensen (2012), Greenwood, Hanson, and Stein (2015), Bolton and Huang (2016), and Nagel (2016).

as lump-sum transfer and debt is repaid with lump-sum tax. Outside money decreases the money premium in *every* state of the world, which seems to indicate a more favorable condition for firms and more investments as a result. However, the impact of outside money depends on how banks respond.

By lowering the money premium, outside money increases banks' debt cost, and thereby, decreases their return on equity. This profit crowding-out effect amplifies the bank leverage cycle. In the states where banks' risk-taking capacity is high, banks raise leverage even higher so that over the cycle, profit is still high enough to justify the occasionally incurred cost of equity issuance. The crowding-out effect can also lengthen the crisis. With lower profit, banks become more reluctant to raise equity, which translates into a lower equity issuance boundary. And, with lower return on equity, it takes more time for banks to rebuild equity through retained earnings.

With booms being more fragile and crises more stagnant, the economy spends more time in states where banks are undercapitalized and inside money creation depressed. Unless outside money satiates firms' money demand, the economy still relies on banks as the *marginal* suppliers of money. Therefore, even if outside money increases the total money supply in *every* state of the world, by shifting the probability mass to relatively worse states, it can lead to a lower *average* money supply over the cycles, and thus, hurt resources reallocation and welfare in the long run.

Related literature. Financial intermediaries provide liquidity – the ease of transferring resources over time and between agents. They finance projects (supply credit) and issue securities that facilitate trade (supply money). The cost of liquidity is zero in the frictionless world of Modigliani and Miller (1958), but in reality, we rely on intermediaries to supply money and credit, and due to their limited balance-sheet capacity, they earn a spread. This paper focuses on money supply. It advances an old tradition in macroeconomics by taking a corporate-finance approach, which emphasizes money as a store of value and a means of payment rather than a unit of account (the critical ingredient of models with nominal rigidities, e.g., Christiano, Eichenbaum, and Evans (2005)).

A recent literature has revived the money view of financial intermediation by emphasizing bank liabilities as stores of value and means of payment (Hart and Zingales (2014); Brunnermeier and Sannikov (2016); Donaldson, Piacentino, and Thakor (2016); Piazzesi and Schneider (2017); Quadrini (2017)).⁶ This paper takes this money view of banking to understand the dynamic interaction between firms' liquidity management and banks' choice of leverage, frequency and duration of crises, and how government debt may contribute to financial instability.

Because of the equity issuance cost, the model has a "balance-sheet channel", through which shock impact is persistent (e.g., Bernanke and Gertler (1989)).⁷ Bank net worth is important, which is a feature shared with other models of balance-sheet channel.⁸ My model differs in two aspects. First, asset (capital) price plays a role in shock amplification through the intertemporal complementarity of firms' money holdings, instead of the typical balance-sheet impact on intermediaries (e.g., Brunnermeier and Sannikov (2014)). Second, the demand for intermediary debt is dynamic, contributing to the procyclicality of bank leverage and endogenous risk accumulation. In contrast, many models have a static/passive demand for intermediary debt, so book leverage is countercyclical because equity is more responsive to shocks than assets (e.g., He and Krishnamurthy (2013)).

Models that produce procyclical leverage often focus on the impact of asset price variations on collateral or risk constraints (Brunnermeier and Pedersen (2009); Geanakoplos (2010); Adrian and Boyarchenko (2012); Danielsson, Shin, and Zigrand (2012); Moreira and Savov (2014)). This paper offers a complementary explanation based on corporate money demand. More impor-

⁶There are several branches of literature that provide a microfoundation for bank debt serving as a medium of exchange. Limited commitment (Kiyotaki and Moore (2002)) and imperfect record keeping (Kocherlakota (1998)) limits credit, so trades must engage in *quid pro quo*, involving a transaction medium. Banks overcome such problems and supply money (e.g., Kiyotaki and Moore (2000); Cavalcanti and Wallace (1999)). Ostroy and Starr (1990) and Williamson and Wright (2010) review the monetary theories. Another approach relates asset resalability to information sensitivity. Intermediaries create liquidity by issuing information-insensitive claims that circulate in secondary markets (e.g., Gorton and Pennacchi (1990); DeMarzo and Duffie (1999); Holmström (2012); Dang, Gorton, Holmström, and Odonez (2014)).

⁷The issuance cost limits the sharing of aggregate risk between banks and the rest of the economy (Di Tella (2015)).

⁸Bank equity is common measure of financial slackness. It alleviates agency friction (Holmström and Tirole (1997); Diamond and Rajan (2000)) and facilitates collateralization (Rampini and Viswanathan (2017)). Parlour, Stanton, and Walden (2012) model financial flexibility differently as limited capacity of intermediation expertise.

tantly, it unveils a feedback mechanism between the real and financial sectors that sets the stage for a formal analysis of the financial stability implications of government debt.

A recent literature documents a money premium that lowers the yield on government securities (e.g., Bansal and Coleman (1996); Krishnamurthy and Vissing-Jørgensen (2012); Nagel (2016)). Intermediaries earn the money premium by issuing money-like liabilities, such as asset-backed commercial paper (Sunderam (2015)), deposits (Drechsler, Savov, and Schnabl (2016)), and certificates of deposits (Kacperczyk, Pérignon, and Vuillemeys (2017)). Many have emphasized the risk of excessive leverage (Gorton (2010); Stein (2012)), and pointed out that increasing government debt stabilizes the economy by crowding out intermediary debt (Greenwood, Hanson, and Stein (2015); Krishnamurthy and Vissing-Jørgensen (2015); Woodford (2016)). Advancing this line of research, this paper highlights banks' *dynamic* balance-sheet management under financial constraints, the key ingredient that generates the destabilizing effects of outside money.⁹

As in Woodford (1990b) and Holmström and Tirole (1998), government debt facilitates investment by allowing entrepreneurs to transfer wealth across contingencies. This paper adds to this line of research by emphasizing government as intermediaries' competitor in liquidity supply. By crowding out intermediated liquidity, public liquidity can reduce the overall liquidity and welfare. After the financial crisis, governments increased their indebtedness and central banks expanded balance sheets dramatically in advanced economies, raising concerns such as moral hazard and excessive inflation (Fischer (2009)). This paper highlights a financial instability channel through which an expanding government balance sheet can be counterproductive.¹⁰

A key ingredient of the model is firms' money demand, which is partly motivated by studies on the enormous amount of corporate cash holdings in recent decades (e.g., Bates, Kahle, and Stulz (2009); Eisfeldt and Muir (2016)).¹¹ This paper connects corporate cash holdings to bank

⁹Also highlighting the dilution cost of equity issuance, Bolton and Freixas (2000) analyze the effects of monetary policy on banks' profit and equity capital through changes in the lending spread instead of the money premium.

¹⁰The model does omit other channels through which the government may interact with the banking sector, such as monetary policy (Drechsler, Savov, and Schnabl (2017); Di Tella and Kurlat (2017)).

¹¹Firms' liquidity management problem in the model can be viewed as simplified version of He and Kondor

leverage. In the model, money demand arises from firms' investment needs.¹² R&D is a typical form of investment that heavily relies on internal liquidity, and exhibits strong procyclicality.¹³ A number of studies have shown that the increase of corporate cash holdings in the last few decades is largely driven by the entry of R&D-intensive firms (Begenau and Palazzo (2015); Graham and Leary (2015); Pinkowitz, Stulz, and Williamson (2016)). As the world economy becomes more intangible intensive (Corrado and Hulten (2010)), we would expect the model's mechanism, and more generally, the money view of financial intermediation, to become increasingly relevant.

The remainder is organized as follows. Section 2 lays out key economic forces in a static setting (similar to Holmström and Tirole (1998)). The continuous-time model is in Section 3. Section 4 adds government debt. Section 5 concludes. Appendix I contains proofs and algorithm. Appendix II summarizes calibration and additional results. Appendix III shows preliminary evidence.

2 Static Model: An Anatomy of Money Shortage

This section lays out the key economic forces in a two-period model ($t = 0, 1$). There are goods, capital, and three types of agents, households, bankers, and entrepreneurs (firms). Firms

(2016). By allowing capital to be traded, the model easily aggregates firms' liquidity demand (that is linear in firm wealth, different from the nonlinear liquidity demand in Bolton, Chen, and Wang (2011) who assume capital is not traded).

¹²Eisfeldt (2007) shows that liquidity demand from household consumption smoothing cannot explain the liquidity premium on Treasury bills. Eisfeldt and Rampini (2009) show that the liquidity premium tends to increase when asynchronicity between cash flow and investment opportunities in the corporate sector becomes more severe, which is consistent with the prediction in Holmström and Tirole (2001). Investment need is a key determinant of corporate cash holdings (e.g., Denis and Sibilkov (2010); Duchin (2010)), especially for firms with less collateral (e.g., Almeida and Campello (2007); Li, Whited, and Wu (2016)) and more intensive R&D activities (Falato and Sim (2014)).

¹³The procyclicality of R&D expenditures, as measured by the NSF (U.S. National Science Foundations), has been documented by many studies, including Griliches (1990), Fatas (2000), and Comin and Gertler (2006). Using data from the NSF and Compustat, Barlevy (2007) finds a significant positive correlation between real GDP growth and the growth rate of R&D. Ouyang (2011) documents procyclical R&D at the industry level. Using French firm-level data, Aghion, Askenazy, Berman, Cetto, and Eymard (2012) show that the procyclicality of R&D investment (with respect to sales growth) is found among firms that are financially constrained. Fabrizio and Tsolmon (2014) find that R&D investments are more procyclical in industries with faster obsolescence. The setup of firms' liquidity shock is partly motivated by these findings.

own capital that produces goods at $t = 1$, but liquidity shocks hit some firms at $t = 1$ before production: without further investment, their capital is lost. Because investment inputs have to be purchased in spot transactions, firms carry deposits issued by bankers as means of payment. In turn, bankers back deposits by loans extended to firms. There are two limits on inside money creation: banks' equity and the amount of capital that firms can pledge as collateral for loans. Insufficient supply of inside money compromise firms' liquidity management and investment.

2.1 Setup

Physical structure. All agents consume a non-storable, generic good, and have the same risk-neutral utility with discount rate ρ . At $t = 0$, there are K_0 units of capital endowed to a unit mass of entrepreneurs. One unit of capital produces α units of goods at $t = 1$, and it is only productive in the hands of entrepreneurs. Capital can be traded in a competitive market at $t = 0$, at price q_0^K . Let k_0 denote a representative firm's holdings of capital, so that $K_0 = \int_{s \in [0,1]} k_0(s) ds$. Throughout the paper, I use subscripts for time, and whenever necessary, superscripts for type ("B" for bankers, "H" for households, and "K" for firms who own capital). There is also a unit mass of bankers. Each is endowed with e_0 units of goods, so their aggregate endowment is $E_0 = \int_{s \in [0,1]} e_0(s) ds$. Going forward, the index "s" will be suppressed without loss of clarity. There are a unit mass of households endowed with a large amount of goods per period. Households play a very limited role.

At the beginning of date 1 (i.e., $t = 1$), all firms experience a capital destruction shock, while some also experience a liquidity shock. The economy has one aggregate shock Z_1 , a binary random variable that takes value 1 or -1 with equal probability. After Z_1 is realized, all firms lose a fraction $\pi(Z_1)$ of their capital. This is the capital destruction shock. For simplicity, I assume that $\pi(Z_1) = \delta - \sigma Z_1$ ($\delta - \sigma \geq 0$ and $\delta + \sigma \leq 1$). Later after introducing banks, we will see this aggregate shock makes bank equity essential for liquidity creation. After the capital loss, firms proceed to produce $\alpha [1 - \pi(Z_1)] k_0$ units of goods, if they not are hit by the liquidity

shock.

Independent liquidity shocks hit firms with probability λ , and destroys *all* capital. In the spirit of Holmström and Tirole (1998) and (2001), firms must make further investment; otherwise, they can not produce anything, and thus, exit with zero terminal value. By investing i_1 units of goods per unit of capital, firms can create $F(i_1)k_0$ ($F'(\cdot) > 0$, $F''(\cdot) < 0$) units of new capital that produce goods with productivity α . Homogeneity in k_0 helps reduce the dimension of state variable later in the continuous-time dynamic analysis. I assume that after the investment, firms can revive their old capital, so the post-investment production is $\alpha [1 - \pi(Z_1) + F(i_1)]k_0$. Overall, this liquidity shock is a curse and an opportunity – through investment, firms preserve the existing scale of production and grow. The first-best level of investment, i_{FB} , is defined by:

$$\alpha F'(i_{FB}) = 1, \tag{1}$$

which equates the marginal benefit and the marginal cost. Note that firms making investment at the beginning of $t = 1$ instead of $t = 0$. This backloaded specification gives rise to firms' liquidity holdings later in the presence of financial constraints.

Last but not least, it is assumed that all securities issued by agents in this economy pay out *at the end of date 1*. This timing assumption is particularly relevant for defining what is liquidity from firms' perspective. Assets that firms carry to relax constraints on investment must be *resalable* in exchange for investment inputs *at the beginning of date 1*. Firms are not buy-and-hold investors. It will be shown that this resalability requirement relates the model setup to several strands of literature that study banks as issuers of inside money. Figure 1 shows how events unfold.

Liquidity demand. The model features three frictions, one that gives rise to firms' liquidity demand, and the other two limiting liquidity supply. The first friction is that investment has to be internally financed. In other words, the newly created capital is not pledgeable.¹⁴ As a

¹⁴This can be motivated by a typical moral hazard problem as in Holmström and Tirole (1998).

<i>Date 0</i>	<i>Date 1</i>		
All agents – Raise funds by issuing securities – Consume – Allocate savings	Capital – $\pi(Z_1)$ fraction destroyed	Investing Firms: λ – Acquire goods as investment inputs – Invest Non-investing Firms: $1 - \lambda$ – Produce & sell goods	All agents – Produce & sell goods – Pay out to investors – Consume

Figure 1: Timeline. This figure plots the timeline of static model. Agents set their balance sheets at $t = 0$. At $t = 1$, aggregate shock Z_1 is realized, and $\pi(Z_1)$ fraction of capital destroyed. Then λ fraction of firms are hit by the liquidity shock and need to make investments, while the rest produce. These λ firms produce after investments. At the end of $t = 1$, all agents consume and pay out to liability holders.

result, firms need to carry liquidity (i.e., instruments that transfer wealth from $t = 0$ to $t = 1$). This is a common assumption used to model firms' liquidity demand.¹⁵ To achieve the first-best investment, a firm must have access to liquidity of at least $i_{FB}k_0$ when the λ shock hits.

The objective of this paper is to analyze the *endogenous* supply of liquidity, so I assume that goods cannot be stored (i.e., there is no exogenous storage technology). And, capital cannot transfer wealth to a contingency where itself is destroyed without further spending. So firms must hold financial assets as liquidity buffer. While households receive endowments per period, it is assumed that they cannot sell claims on their future endowments because they can default with impunity; otherwise, there would be no liquidity shortage (as in Holmström and Tirole (1998)). Therefore, the focus is on the asset creation capacity of entrepreneurs themselves and bankers.

Liquidity supply. At date 0, what is a firm's capacity to issue claims that pay out at date 1? I assume firms' *endowed* capital is pledgeable (but not new capital created at $t = 1$, to be consistent with previous assumption). It is collateral that can be seized by investors when default happens.¹⁶ In an equilibrium where firms carry liquidity and invest when hit by the λ shock, a fraction $[1 - \pi(Z_1)]$ of endowed capital is always preserved. Thus, a firm's *expected* pledgeable value at $t = 1$ is $\alpha(1 - \delta)k_0$, and $\alpha(1 - \delta - \sigma)k_0$ when $Z_1 = -1$.

¹⁵See Froot, Scharfstein, and Stein (1993), Holmström and Tirole (1998) and (2001) in the micro literature, and Woodford (1990b), Kiyotaki and Moore (2005), and He and Kondor (2016) in the macro literature among others.

¹⁶This reflects that a firm's mature capital can be relatively easily valued and seized by investors. Allowing capital created at date 1 to serve as collateral complicates the expressions but does not change the main results.

Potentially, firms could hold securities issued by each other as liquidity buffer. If the aggregate pledgeable value always exceeds firms' aggregate liquidity demand, i.e.,

$$\alpha (1 - \delta - \sigma) K_0 \geq i_{FB} K_0, \quad (2)$$

the economy achieves the first-best investment defined in Equation (1).¹⁷ Even better, as long as the liquidity shock is verifiable, firms' liabilities can be pooled into a mutual fund that pays out to investing firms, so given this perfect risk-sharing, the first-best investment is achieved if

$$\alpha (1 - \delta - \sigma) K_0 \geq \lambda i_{FB} K_0, \text{ where } \lambda \in (0, 1). \quad (3)$$

In a very similar setting, Holmström and Tirole (1998) study the question whether entrepreneurs' supply of assets meets their own liquidity demand (i.e., Equation (2) and (3)), and emphasize the severity of liquidity shortage depends on aggregate shock (i.e., σ in my setting).

This paper departs from Holmström and Tirole (1998) by introducing the second friction: firms can hold liquidity only in the form of bank liabilities. Therefore, in the model, banks issue claims to firms that are in turn backed by banks' holdings of firm liabilities.

There are several reasons why firms hold *intermediated liquidity*. Entrepreneurs may simply lack the required expertise of asset management.¹⁸ And, cross holding is regulated in many countries and industries. What motivates this assumption is also the strands of theoretical literature that study banks as inside money creators. Given the timing in Figure 1, entrepreneurs purchase goods as investment inputs by selling their liquidity holdings when the λ shock hits. In other words, firms carry liquidity as a means of payment. Kiyotaki and Moore (2000) and

¹⁷I assume liquidity instruments cannot serve as collateral. Otherwise, the pledgeable value is infinity: firms' issuance of securities enlarge each other's financing capacity, which lead to more securities issued. See a related argument in the Chapter 3 of Holmström and Tirole (2011).

¹⁸While there is a large literature on households' limited participation in financial markets, such as Mankiw and Zeldes (1991), Basak and Cuoco (1998), Vissing-Jørgensen (2002), and He and Krishnamurthy (2013), the portfolio choice of firms is relatively less understood. A recent paper by Duchin, Gilbert, Harford, and Hrdlicka (2017) finds risky securities in some firms' liquidity portfolio, but the dominant component is safe debt issued by financial intermediaries or governments.

(2002) model bankers as agents with superior ability to make multilateral commitment, i.e., to pay *whoever* holds their liabilities, so bank liabilities circulate as means of payment.¹⁹ Taking a step further, money creation may require not only a special set of agents (bankers), but also a particular security design. In Gorton and Pennacchi (1990) and Dang, Gorton, Holmström, and Ordonez (2014), banks create money by issuing information-insensitive claims (safe debt) that suffer less the asymmetric information in secondary markets.

This paper takes the aforementioned literature as a starting point: firms are assumed to hold liquidity in the form of safe debt issued by banks (“deposits”).²⁰ Let m_0 denote a firm’s deposit holdings *per unit of capital*. Investment at $t = 1$ is thus directly tied to deposits carried from $t = 0$:

$$i_1 k_0 \leq m_0 k_0. \quad (4)$$

Equation (4) resembles a money-in-advance constraint (e.g., Svensson (1985); Lucas and Stokey (1987)), except that what firms hold for transaction purposes is not fiat money, but bank debt, or “inside money.” Thus, banks add value to the economy by supplying deposits that can be held by firms to relax this “money-in-advance” constraint on investment. Linking firm cash holdings to bank debt is in line with evidence.²² Pozsar (2011) argue that corporate treasury, as one of the major cash pools, feeds leverage to the financial sector in the run-up to the global financial crisis.

Inside money creation capacity. What determines the safe debt capacity of banks? First, let us shift our focus from the liability side of bank balance sheet to the asset side. To impose more structure on the analysis, I assume that bank loans (assets) take a particular contractual form:

¹⁹In a richer setting with limited commitment and imperfect record keeping (Kocherlakota (1998)), credit is constrained, so trades must engage in *quid pro quo*, involving a transaction medium (Kiyotaki and Wright (1989)). Cavalcanti and Wallace (1999) show that bankers arise as issuers of inside money when their trading history is public knowledge. Ostroy and Starr (1990) and Williamson and Wright (2010) review the literature of monetary theories.

²⁰The concavity of investment technology $F(\cdot)$ also implies that firms prefer safe assets as liquidity buffer.

²¹To be consistent with the continuous-time expressions, deposits’ interest payments are ignored in Equation (4).

²²Based on Financial Accounts of the United States, Figure 11 in Appendix III.1 shows 80% of liquidity holdings of nonfinancial corporate businesses are in financial intermediaries’ debt, with the rest dominated by Treasury securities.

each dollar of loan extended at date 0 is backed by a designated collateral, and is repaid at the end of date 1 with an interest rate R_0 if the collateral is intact. Therefore, a fraction $\pi(Z_1)$ of loans default as the corresponding collateral capital is destroyed. So, bankers' loan portfolio has a return equal to $[1 - \pi(Z_1)](1 + R_0)$. To match the expressions in the continuous-time analysis, I approximate this return with $1 + R_0 - \pi(Z_1)$, ignoring $\pi(Z_1)R_0$, product of the two percentages.

Because of the aggregate shock, banks' safe debt capacity depends on their equity cushion. Let r_0 denote the deposit rate, and x_0 denote the leverage (asset-to-equity ratio). A banker will never default if her net worth is still positive even in a bad state ($Z_1 = -1$), i.e.,

$$\underbrace{x_0 e_0}_{\text{total assets}} \underbrace{[1 + R_0 - \pi_D(-1)]}_{\text{loan return}} \geq \underbrace{(x_0 - 1) e_0}_{\text{total debt}} \underbrace{(1 + r_0)}_{\text{deposit repayment}} .$$

This incentive or solvency constraint can be rewritten as a limit on leverage:

$$x_0 \leq \frac{r_0 + 1}{r_0 + \delta + \sigma - R_0} := \bar{x}_0 \quad (5)$$

Finally, I introduce the third friction – banks' equity issuance cost. At $t = 0$, bankers may raise equity subject to a proportional dilution cost χ . To raise one dollar, a bank needs to give $1 + \chi$ worth of equity to investors.²³ I will consider $\chi < \infty$ in the continuous-time analysis. For now, the static model assumes $\chi = \infty$, i.e., banks do not issue equity. As a result, inside money creation is limited by bankers' wealth: total deposits cannot exceed $(\bar{x}_0 - 1)E_0$. Even if banks are sufficiently capitalized, deposits are ultimately backed by loans, so money creation faces another limit, that is firms' promises to pay cannot exceed $\alpha(1 - \delta)K_0$ in expectation.

Summary. Three key frictions form the three pillars of the model: (1) firms' money demand; (2) bank debt as money; (3) equity constraint on banks. Insufficient inside money creation leads

²³ χ is a reduced form representation of informational frictions in settings such as Myers and Majluf (1984) or Dittmar and Thakor (2007). Moreover, related to the information insensitivity explanation of the monetary services of safe debt, the illiquidity of bank equity may result from banks' intentional choice of balance-sheet opaqueness in order to protect the information insensitivity of deposits as in Dang, Gorton, Holmström, and Ordóñez (2014).

to investment inefficiency by compromising entrepreneurs' liquidity management. There are two limits on liquidity creation: bank equity that buffers the aggregate shock, and the total pledgeable value of firms' capital. The former is the focus of this paper, while the latter is studied by Holmström and Tirole (1998), and echoes the broad literature on asset shortage.²⁴

The next section casts the model into a continuous-time framework, and delivers the main results. Before introducing the dynamic model, I will close this section by showing several features of the static equilibrium that are shared with the continuous-time Markov equilibrium

2.2 Equilibrium

Lemma 1, 2, and 3 below summarize the optimal choices for firms and banks at $t = 0$. We will focus on an equilibrium where firms' liquidity constraint binds (i.e., $i_1 = m_0$). One more unit of deposits can be used to purchase one more unit of goods as inputs to create $F'(m_0)$ more units of capital (with productivity α) when λ shock hits, which has an expected net value of $\lambda[\alpha F'(m_0) - 1]$. This convenience yield makes firms willing to accept a return lower than ρ , the discount rate.²⁵ The spread, $\rho - r_0$, is a "money premium", a cost of carrying money. When $r_0 < \rho$, households do not hold deposits, since they do not face liquidity shocks as firms do.

Lemma 1 (Money Demand) *Firms' equilibrium deposits, m_0 , satisfy the condition*

$$\lambda[\alpha F'(m_0) - 1] = \rho - r_0.$$

Firms also choose the amount they borrow from banks, which is subject to the collateral constraint that the expected repayment cannot exceed the total pledgeable value. Given the ex-

²⁴See Woodford (1990b), Kocherlakota (2009), Kiyotaki and Moore (2005), Caballero and Krishnamurthy (2006), Farhi and Tirole (2012), Giglio and Severo (2012) among others.

²⁵Because bankers' only endowments are goods that cannot be stored, to carry net worth to date 1, bankers must lend some goods to firms in exchange for loans, i.e., the instruments that bankers use to transfer wealth intertemporarily. Since goods cannot be stored, entrepreneurs must consume at $t = 0$ in equilibrium. To make risk-neutral entrepreneurs indifferent between consumption and savings, the price of capital q_0^K adjusts so that acquiring capital delivers an expected return equal to ρ , which is the opportunity cost of holding deposits instead of capital.

pected default probability $\mathbb{E}[\pi(Z_1)] = \delta$, the expected loan repayment is $(1 - \delta)(1 + R_0)$ per dollar of borrowing, approximated by $1 + R_0 - \delta$ (the product of two percentages ignored). When $R_0 - \delta = \rho$, firms are indifferent; when $R_0 - \delta < \rho$, firms borrow to the maximum. The spread, $\rho - (R_0 - \delta)$, is the collateral shadow value ($\kappa_0 \geq 0$), i.e., the Lagrange multiplier of firms' collateral constraint.

Lemma 2 (Credit Demand) *The equilibrium loan rate is given by: $R_0 = \delta + \rho - \kappa_0$.*

Competitive bankers take as given the market loan rate R_0 and deposit rate r_0 . At $t = 0$, a representative banker chooses consumption-to-wealth ratio y_0 (and retained equity $e_0 - y_0 e_0$), and the asset-to-equity ratio x_0 (leverage). Each dollar of retained equity is worth $x_0 [1 + R_0 - \pi(Z_1)] - (x_0 - 1)(1 + r_0)$ at $t = 1$, which is the difference between asset and liability value. Note that $\mathbb{E}[\pi_D(Z_1)] = \delta$, so that the expected return on retained equity is $1 + r_0 + x_0(R_0 - \delta - r_0)$, and the return in a bad state is $1 + r_0 + x_0(R_0 - \delta - \sigma - r_0)$. Let ξ_0 denote the Lagrange multiplier of the solvency constraint, i.e., the shadow value of bank equity at $t = 1$.²⁶ The value function is

$$v(e_0; R_0, r_0) = \max_{y_0 \geq 0, x_0 \geq 0} y_0 e_0 + \frac{(e_0 - y_0 e_0)}{(1 + \rho)} \{1 + r_0 + x_0(R_0 - \delta - r_0) + \xi_0 [1 + r_0 + x_0(R_0 - \delta - \sigma - r_0)]\}.$$

Lemma 3 (Bank Optimization) *The first-order condition (F.O.C.) for bank leverage x_0 is*

$$R_0 - r_0 = \delta + \gamma_0^B \sigma, \tag{6}$$

where $\gamma_0^B = \left(\frac{\xi_0}{1 + \xi_0}\right) = \frac{R_0 - \delta - r_0}{\sigma} \in [0, 1)$ is the banker's "effective risk aversion". The solvency constraint binds if and only if $\gamma_0^B > 0$. Substituting the F.O.C. into the value function, we have

$$v(e_0; q_0^B) = y_0 e_0 + q_0^B (e_0 - y_0 e_0), \text{ where, } q_0^B = \frac{(1 + r_0)(1 + \xi_0)}{(1 + \rho)}. \tag{7}$$

The banker consumes if $q_0^B \leq 1$; if $q_0^B > 1$, $y_0 = 0$ so that the entire endowments are lent out.

²⁶Note that ξ_0 is known at $t = 0$, so its subscript is 0 instead of 1.

The equilibrium credit spread, $R_0 - r_0$, has two components: the expected default probability δ and the risk premium $\gamma_0^B \sigma$. Each dollar lent adds σ units of downside risk at date 1, thus, tightening the capital adequacy constraint. γ_0^B is the price of risk charged by bankers, which is equal to the market Sharpe ratio of risky lending financed by risk-free deposits.

q_0^B is the marginal value of bank equity (Tobin's Q). Retained equity has a compounded payoff of $(1 + r_0)(1 + \xi_0)$ from reducing the external financing (debt) cost and relaxing the solvency constraint, so its present value is $\frac{(1+r_0)(1+\xi_0)}{(1+\rho)}$. When $q_0^B > 1$, bankers lend out all endowments and carry all of their wealth to $t = 1$ in the form of loans.

Substituting the equilibrium loan rate into Equation (6), we can solve for the money premium $\rho - r_0$, as the sum of $\gamma_0^B \sigma$, banks' risk compensation, and κ_0 , the collateral shadow value.

Proposition 1 (Money Premium Decomposition) *The equilibrium money premium is given by*

$$\rho - r_0 = \gamma_0^B \sigma + \kappa_0. \quad (8)$$

Equation (8) decomposes the money premium into an *intermediary wedge*, $\gamma_0^B \sigma$, that measures of the scarcity of bank equity, and a collateral wedge κ_0 . Since the money premium equals the expected value of foregone marginal investment (Lemma 1), Equation (8) offers an anatomy of investment inefficiency. To support the first-best investment, i_{FB} , each firms must carry at least $i_{FB} K_0$ deposits in aggregate, which requires a minimum level of bank equity:

Condition 1 $E_0 \geq E_{FB}$, where $E_{FB} := \frac{i_{FB} K_0}{\bar{x}_{FB} - 1} = \frac{i_{FB}}{1 + \rho - 1} K_0$, and i_{FB} is defined in Equation (1).

\bar{x}_{FB} is solved as follows: under the first-best investment, the money premium is zero, so $\kappa_0 = 0$. Substituting $r_0 = \rho$ and $R_0 = \delta + \rho$ into the solvency constraint yields $\bar{x}_{FB} = \frac{1 + \rho}{\sigma}$. Intuitively, when the size of the aggregate shock is larger (i.e., higher σ), the required bank equity as risk buffer (E_{FB}) is larger, and Condition 1 is more likely to be violated.

First-best deposit creation also requires a minimum stock of collateral to back bank loans. The minimum bank lending that supports the first-best investment is $\bar{x}_{FB} E_{FB}$, so that collateral must be sufficient to cover firms' expected debt repayment: $\alpha(1 - \delta) K_0 \geq \bar{x}_{FB} E_{FB} (1 + \rho)$, or,

$$K_0 \geq K_{FB} := \frac{\bar{x}_{FB} E_{FB} (1 + \rho)}{\alpha (1 - \delta)} = \left(\frac{\frac{1+\rho}{\sigma}}{\frac{1+\rho}{\sigma} - 1} \right) \left(\frac{i_{FB} (1 + \rho)}{\alpha (1 - \delta)} \right) K_0,$$

This condition can be simplified into the following parameter restrictions:

Condition 2 $\frac{\alpha(1-\delta)}{1+\rho} \geq \left(\frac{1}{1-\frac{\sigma}{1+\rho}} \right) i_{FB}$, where i_{FB} is defined in Equation (1).

Condition 2 is more likely to be violated when δ is higher, the expected collateral destruction rate. Thus, δ measures the severity of the collateral shortage problem that is reflected in Condition 2 and studied by Woodford (1990b), Holmström and Tirole (1998), Caballero and Krishnamurthy (2006) and others. As shown in Condition 1, the scarcity of bank equity, which is the focus of this paper, is more severe if σ is larger. Therefore, the two parameters, δ and σ , capture two channels through which inside money creation is constrained. Corollary 1 summarizes the analysis.

Corollary 1 (Sufficient Conditions for a Money Shortage) *The equilibrium money premium is positive, and investment is below the first-best level, if either Condition 1 or 2 is violated.*

3 Dynamic Model: Procyclical Money Creation

The model is recast in continuous time. The analysis focuses on the intermediary wedge, assuming a corresponding version of Condition 2 holds, so the economy has enough collateral to back loans, but banks may not have sufficient equity as risk buffer. While the static model shows key economic forces, the focus of this paper is on the dynamic implications of this setup, such as bank leverage cycle, frequency and duration of crisis, and the financial stability implications of outside money.

3.1 Setup

Continuous-time setup. All agents maximize risk-neutral life-time utility with discount rate ρ . Households consumes the generic goods and can invest in securities issued by firms and banks.²⁷ Firms trade capital at price q_t^K . One unit of capital produces α units of goods per unit of time. They can issue equity to households, promising an expected net return of ρ per unit of time (cost of equity). Given the deposit rate r_t , the deposit carry cost or the money premium, is defined by the spread between ρ and deposit rate r_t as in the static model.²⁸

At idiosyncratic Poisson times (intensity λ), firms are hit by a liquidity shock, and cut off from external financing. A firm either quits or invests. Let k_t denote its capital holdings. Investing $i_t k_t$ units of goods can preserve the existing capital and create $F(i_t)k_t$ units of new capital. Investment is constrained by the firm's deposit holdings, $i_t \leq m_t$, where m_t is the deposits on its balance sheet per unit of capital.²⁹ Holding deposits allows firms' wealth to jump up at these Poisson times through the creation of new capital. I assume the technology $F(\cdot)$ is sufficiently productive, so we focus on an equilibrium where the liquidity constraint always binds.

The aggregate shock Z_t is a standard Brownian motion. Every instant, $\delta dt - \sigma dZ_t$ fraction of capital is destroyed. Firms default on loans backed by the destroyed capital. Let R_t denote the

²⁷Risk-neutral households' required return is fixed at ρ because negative consumption is allowed, which is interpreted as dis-utility from additional labor to produce extra goods as in Brunnermeier and Sannikov (2014). Allowing negative consumption serves the same purpose of assuming large endowments of goods per period in the static model.

²⁸Nagel (2016) emphasizes the variation in illiquid return (i.e. ρ in the model) as a driver of the money premium dynamics in data. This paper provides an alternative model that focuses on the yield on money-like securities, r_t .

²⁹The idiosyncratic nature of liquidity shock and the assumption that firms can excess external funds in normal times imply that firms' money demand does not contain hedging motive that complicates model mechanism. Bolton, Chen, and Wang (2013) model the market timing motive of corporate liquidity holdings in the presence of technological illiquidity and state-dependent external financing costs. He and Kondor (2016) examine how the hedging motive of liquidity holdings amplifies investment cycle through pecuniary externality in the market of productive capital.

loan rate. For one dollar borrowed from banks at t , firms expect to pay back

$$\underbrace{(1 + R_t dt)}_{\text{principal + interest payments}} \underbrace{[1 - (\delta dt - \sigma dZ_t)]}_{\text{default probability}} = 1 + R_t dt - (\delta dt - \sigma dZ_t),$$

where high-order infinitesimal terms are ignored. The default probability is a random variable that loads on dZ_t .³⁰ Both loans and deposits are short-term contracts, initiated at t and settled at $t + dt$.³¹

Let r_t denote the deposit rate, and x_t banks' asset-to-equity ratio. Let c_t^B denote a bank's cumulative dividend. $dc_t^B > 0$ means consumption and paying dividends to outside shareholders (households); $dc_t^B < 0$ means raising equity. As in the static model, we can define $dy_t = \frac{dc_t^B}{e_t}$ as the payout or issuance ratio, which is an impulse variable, so bank equity e_t follows a regulated diffusion process, reflected at payout and issuance (i.e., when $dy_t \neq 0$):

$$de_t = \underbrace{e_t x_t}_{\text{Loan value}} \underbrace{[R_t dt - (\delta dt - \sigma dZ_t)]}_{\text{Loan return}} - \underbrace{e_t (x_t - 1)}_{\text{Debt value}} \underbrace{r_t dt}_{\text{Interest payment}} - \underbrace{e_t dy_t}_{\text{Payout or issue}} - \underbrace{e_t \iota dt}_{\text{Cost of operations}}. \quad (9)$$

Because in equilibrium, banks earn a positive expected return on equity, the operation cost ι is introduced to motivate payout so that banks will not outgrow the economy.³²

Bankers maximize life-time utility, subject to a proportional equity issuance cost:

$$\mathbb{E} \left\{ \int_{t=0}^{\tau} e^{-\rho t} [\mathbb{I}_{\{dy_t \geq 0\}} - (1 + \chi) \mathbb{I}_{\{dy_t < 0\}}] e_t dy_t \right\}.$$

\mathbb{I}_A is the indicator function of event A .³³ The solvency constraint in the static setting boils down to

³⁰Probit transformation can make $\pi(dZ_t) \in (0, 1)$, but complicates expressions. See also Klimenko, Pfeil, Rochet, and Nicolo (2016).

³¹For short-term deposits, in order to highlight the resale of deposits, I assume that banks repay deposits after investment takes place, so that investing firms cannot wait for goods paid by banks until the maturity of deposits, and thus, have to actually pay them out in exchange for goods. Long-term deposits avoid this assumption, but would introduce other sources of instability, such as the Fisherian deflationary spiral in Brunnermeier and Sannikov (2016). Similarly, long-term loan contracts introduce the fire sale mechanism in Brunnermeier and Sannikov (2014).

³²The cost of operations is equivalent to a higher time-discount rate for bankers, common in the literature of heterogeneous-agent models (e.g., Kiyotaki and Moore (1997)). It can also be interpreted as an agency cost.

³³In different settings, Van den Heuvel (2002), Phelan (2015), and Klimenko, Pfeil, Rochet, and Nicolo (2016)

the requirement of non-negative equity. Unlike the static setting, in equilibrium, bankers always preserve a slackness, so $\tau := \inf\{t : e_t \leq 0\} = \infty$. As will be shown later, even in the absence of a binding solvency constraint, bankers are still risk-averse due to the recapitalization friction χ .

State variable. At time t , the economy has K_t units of capital and aggregate bank equity E_t . In principle, a time-homogeneous Markov equilibrium would have both as state variables. Because production has constant return-to-scale and the investment technology is homogeneous of degree one in capital (as in Hayashi (1982)), the Markov equilibrium has only one state variable:

$$\eta_t = \frac{E_t}{K_t}.$$

As the model puts the interaction between money supply and demand at the center, intuitively, η_t measures the size of liquidity suppliers (banks) relative to that of liquidity demanders (firms).

Because there is a unit mass of homogeneous bankers, E_t follows the same dynamics as e_t , so the instantaneous expectation, μ_t^e , and standard deviation, σ_t^e , of $\frac{dE_t}{E_t}$ are $r_t + x_t(R_t - \delta - r_t) - \iota$ and $x_t\sigma$ respectively (from Equation (9)). By Itô's lemma, η_t follows a regulated diffusion process

$$\frac{d\eta_t}{\eta_t} = \mu_t^\eta dt + \sigma_t^\eta dZ_t - dy_t, \tag{10}$$

where μ_t^η is $\mu_t^e - [\lambda F(m_t) - \delta] - \sigma_t^e\sigma + \sigma^2$, with the second term being the expected growth rate of K_t , and σ_t^η is $(x_t - 1)\sigma$, which is positive because banks lever up (i.e. $x_t > 1$). Positive shocks increase η_t , so banks become relatively richer; negative shocks make banks relatively undercapitalized. As η_t evolves over time, the economy repeats the timeline in Figure 1 with date 0 replaced by t and date 1 replaced by $t + dt$. Let intervals $\mathbb{B} = [0, 1]$ and $\mathbb{K} = [0, 1]$ denote the sets of banks and firms respectively. The Markov equilibrium is formally defined as follows.

also introduce issuance frictions in dynamic banking models. Dilution cost is just one form of frictions that give rise to the endogenous variation of intermediaries' risk-taking capacity. He and Krishnamurthy (2012) use to a minimum requirement of insiders' stake that resolves the principal-agent problem between inside and outside equity holders.

Definition 1 (Markov Equilibrium) For any initial endowments of firms' capital $\{k_0(s), s \in \mathbb{K}\}$ and banks' goods (i.e., initial bank equity) $\{e_0(s), s \in \mathbb{B}\}$ such that

$$\int_{s \in \mathbb{K}} k_0(s) ds = K_0, \text{ and } \int_{s \in \mathbb{B}} e_0(s) ds = E_0,$$

a Markov equilibrium is described by the stochastic processes of agents' choices and price variables on the filtered probability space generated by the Brownian motion $\{Z_t, t \geq 0\}$, such that:

- (i) Agents know and take as given the processes of price variables, such as the price of capital, the loan rate, and the deposit rate (i.e., agents are competitive with rational expectation);
- (ii) Households optimally choose consumption and savings that are invested in securities issued by firms and banks;
- (iii) Firms optimally choose capital holdings, deposit holdings, investment, and loans;
- (iv) Bankers optimally choose leverage, and consumption/payout and issuance policies;
- (v) Price variables adjust to clear all markets with goods being the numeraire;
- (vi) All the choice variables and price variables are functions of η_t , so Equation (10) is an autonomous law of motion that maps any path of shocks $\{Z_s, s \leq t\}$ to current state η_t .

3.2 Markov Equilibrium

The risk cost of money creation. In analogy to Proposition 1, I will show a risk cost of money creation that ties inside money supply to bank equity. I start with firms' demand for bank deposits.

Lemma 1' gives firms' optimal deposit demand in analogy to Lemma 1, with one modification that capital is valued at the market price q_t^K instead of the terminal value α in the static setting.³⁴ As will be shown, this difference is critical, as it leads to a unique intertemporal feed-

³⁴To be precise, the liquidity shock hits at $t + dt$, and by then the capital created will be worth $q_{t+dt}^K = q_t^K + dq_t^K$. In equilibrium, q_t^K is a diffusion process with continuous sample paths, so dq_t^K is infinitesimal, and thus, ignored.

back mechanism that amplifies the procyclicality of money creation. As in the static model, households do not hold deposits when $r_t < \rho$, so firms' deposit demand is the aggregate demand for bank debt.

Lemma 1' (Money Demand) *Firms' equilibrium deposits, m_t , satisfy the condition*

$$\lambda [q_t^K F'(m_t) - 1] = \rho - r_t. \quad (11)$$

The static model highlights two limits on inside money creation: the scarcities of firm collateral and bank equity. In the dynamic analysis, I will focus on the latter, and later confirm that in the calibrated equilibrium, firms' external financing (or collateral) constraint never binds. As a result, the expected loan repayment, $R_t - \delta$, is equal to ρ , the discount rate.

Lemma 2' (Credit Demand) *The equilibrium loan rate is given by: $R_t = \delta + \rho$.*

Bankers solve a fully dynamic problem. Following Lemma 3, I conjecture that bankers' value function is linear in equity, $v(e_t; q_t^B) = q_t^B e_t$, where q_t^B summarizes the investment opportunity set. Define ϵ_t^B as the elasticity of q_t^B : $\epsilon_t^B := \frac{dq_t^B/q_t^B}{d\eta_t/\eta_t}$. Intuitively, q_t^B signals the scarcity of bank equity, so I look for an equilibrium in which $\epsilon_t^B \leq 0$. Individual bankers take as given the equilibrium dynamics of q_t^B . Let μ_t^B and σ_t^B denote the instantaneous expectation and standard deviation of $\frac{dq_t^B}{q_t^B}$ respectively. The Hamilton-Jacobi-Bellman (HJB) equation can be written as

$$\begin{aligned} \rho = \max_{dy_t^B \in \mathbb{R}} & \left\{ \frac{(1 - q_t^B)}{q_t^B} \mathbb{I}_{\{dy_t > 0\}} dy_t + \frac{(q_t^B - 1 - \chi)}{q_t^B} \mathbb{I}_{\{dy_t < 0\}} (-dy_t) \right\} \\ & + \mu_t^B + \max_{x_t \geq 0} \{ r_t + x_t (R_t - \delta - r_t) - x_t \gamma_t^B \sigma \} - \iota, \end{aligned} \quad (12)$$

where the effective risk aversion is defined by $\gamma_t^B := -\sigma_t^B$. By Itô's lemma, $\gamma_t^B = -\epsilon_t^B \sigma_t^\eta \geq 0$.

Lemma 3' (Bank Optimization) *The first-order condition for equilibrium leverage x_t is*

$$R_t - \delta - r_t = \gamma_t^B \sigma, \quad (13)$$

The banker pays dividends ($dy_t > 0$) if $q_t^B < 1$, and raises equity ($dy_t < 0$) if $q_t^B > 1 + \chi$.

Bankers' issuance and payout policies imply that η_t is bounded by two reflecting boundaries: the issuance boundary $\underline{\eta}$, given by $q^B(\underline{\eta}) = 1 + \chi$, and the payout boundary $\bar{\eta}$ given by $q^B(\bar{\eta}) = 1$. When η_t falls to $\underline{\eta}$, banks raise equity and η_t never decreases further; When η_t rises to $\bar{\eta}$, banks pay out dividends and η_t never increases further. When $\eta_t \in (\underline{\eta}, \bar{\eta})$, bankers neither issue equity nor pay out dividends, because $q_t^B \in (1, 1 + \chi)$ by the monotonicity of $q^B(\eta_t)$. As η_t rises following good shocks and falls following bad shocks, banks follow a countercyclical equity management strategy, paying out dividends in good times and issuing shares in bad times, which is consistent with the evidence (Baron (2014); Adrian, Boyarchenko, and Shin (2015)).

Lemma 4 (Reflecting Boundaries) *The economy moves within bank issuance boundary $\underline{\eta}$ and payout boundary $\bar{\eta}$. In $[\underline{\eta}, \bar{\eta}]$, the law of motion of state variable η_t is given by Equation (10).*

Bankers are risk-averse because of the recapitalization friction. From an individual banker's perspective, the issuance cost causes her marginal value of equity q_t^B to be negatively correlated with shocks. Following a negative shock, bankers will not raise equity unless q_t^B reaches $1 + \chi$, so the whole industry shrinks (i.e. the *aggregate* bank equity decreases), and intuitively, Tobin's Q, q_t^B , increases. Following a positive shock, bankers will not immediately pay out dividends unless q_t^B drops to 1, so the whole industry expands and q_t^B decreases. Thus, bankers require a risk premium for holding any asset whose return is positively correlated with dZ_t (i.e., negatively correlated with q_t^B). In particular, bankers require a risk compensation for extending loans.

On the left-hand side of Equation (13) is the net interest margin, $R_t - \delta - r_t$, the marginal benefit of issuing deposits backed by risky loans. The right-hand side is the marginal cost, that

is the σ units of risk exposure, priced at γ_t^B per unit.³⁵ We can interpret the equilibrium γ_t^B as the expected profit per unit of risk (i.e. the Sharpe ratio), from creating deposits backed by risky loans:

$$\gamma_t^B = \frac{R_t - \delta - r_t}{\sigma}.$$

Banks face two markets, the loan market and the money market. With the loan rate R_t equal to $\rho + \delta$ (Lemma 2'), there is a one-to-one mapping between the deposit rate r_t and γ_t^B .

Interpreting γ_t^B as the Sharpe ratio or profitability of money creation helps us build an intuitive connection between q_t^B and γ_t^B . As a summary statistic for banks' investment opportunity set, q_t^B reflects the expectation of future profits from money creation (i.e. the future paths of γ_t^B). Intuitively, when the banking sector is relatively large, i.e., η_t is high, banks' profit per unit of risk, γ_t^B , declines. This is a key property of the equilibrium later confirmed by the full solution. Substituting the equilibrium loan rate into Equation (13), we have the dynamic counterpart of Proposition 1.³⁶

Proposition 1' (Money Premium) *The equilibrium money premium is given by*

$$\rho - r_t = \gamma_t^B \sigma. \tag{14}$$

Figure 2 takes a snapshot of the deposit market, given q_t^K . In the Markov equilibrium, these variables vary continuously with η_t . The horizontal axis is m_t , the representative firm's deposits per unit of capital. The vertical axis is the money premium. The investment technology $F(\cdot)$ is concave, so firms' indifference curve from Lemma 1' gives a downward-sloping demand curve. The supply curve is represented by bankers' indifference curve $\rho - r_t = \gamma_t^B \sigma$.

³⁵ $\gamma_t^B \sigma$ opens up a wedge between the credit spread, $R_t - r_t$, and δ the expected default rate. This intermediary premium shares the insight of He and Krishnamurthy (2013), but here, the purpose of intermediation is to create inside money. Bankers need loans to back deposits, and all that households need is to break even as shown in Lemma 2'.

³⁶Recall that for the transparency of the dynamic mechanisms, I assume firms' collateral constraint never binds, so the collateral shadow value disappears. This assumption is confirmed later by the solution of calibrated model.

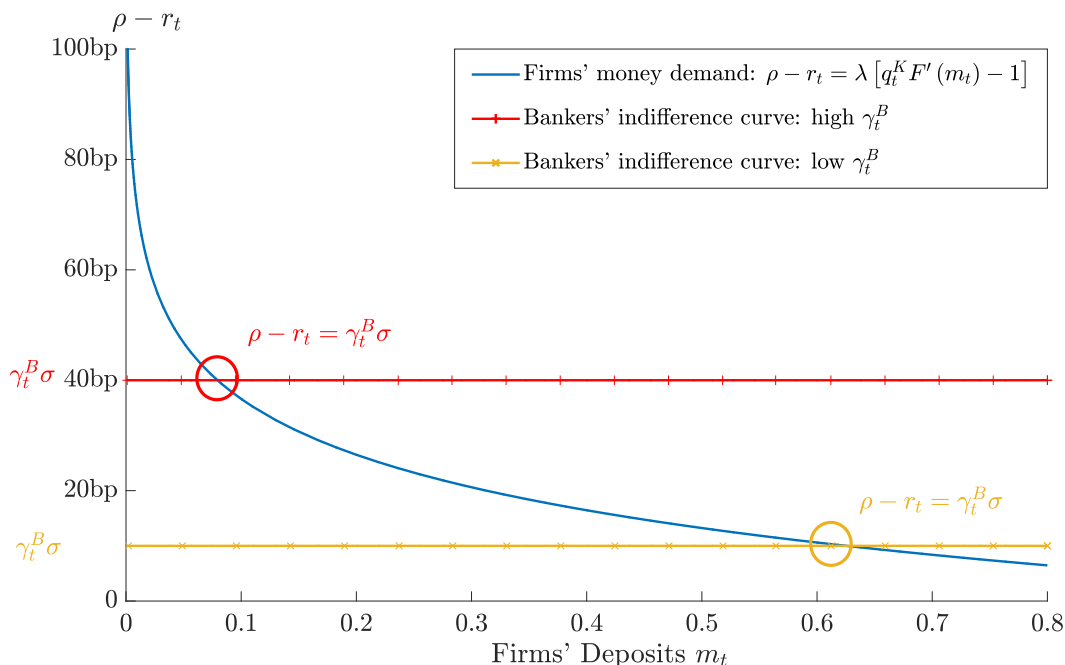


Figure 2: **Money Market.** This figure plots the money demand curve of firms and the indifference curve of banks, taking as given capital price q_t^K and bank risk price γ_t^B . The intersection point is the money market equilibrium $(\rho - r_t, m_t)$, i.e., the money premium (price) and firms' deposit holdings per unit of capital (quantity). Two indifference curves of banks are plotted corresponding to high and low γ_t^B respectively.

Two equilibrium points are circled. When the banking sector is undercapitalized (low η_t), γ_t^B is high and bankers' risk-taking capacity is limited. The equilibrium money premium must compensate bankers' risk exposure. When the banking sector is well capitalized (high η_t), γ_t^B is low, and bankers risk-taking capacity is large. The equilibrium money premium declines.³⁷

The quantity of inside money that bank creates depends on their risk-taking capacity, measured by γ_t^B , so does the price of money, $\rho - r_t$. As repeatedly emphasized throughout this paper, money is created when banks extend loans. Therefore, with money being risk-free and loan being risky, money creation necessarily induces and amplifies the risk mismatch on bank balance sheets, with precisely σ units of risk mismatch per dollar of money created. As γ_t^B varies with η_t , the relative bank equity measure, this *risk cost* of money production links banks' balance-sheet capacity to the real economy through firms' liquidity constraint on investments. By highlighting

³⁷Drechsler, Savov, and Schnabl (2016) provide evidence on banks' unique position in creating money-like securities, such as deposits, but in contrast with this paper, they emphasize banks' market power as a driver of deposit rate instead of banks' balance-sheet capacity under financial frictions.

the simultaneity of money and credit creation, i.e., this risk cost of money creation, the model adds to the literature of balance-sheet channel (Bernanke and Gertler (1989); Kiyotaki and Moore (1997)). And by modeling banks as money creators, this paper offers a bank balance-sheet perspective on literature of liquidity shortage (Woodford (1990b); Holmström and Tirole (1998)).

Corollary 1' (Investment Inefficiency) *From Lemma 1' and Corollary 1', we have*

$$\lambda [q_t^K F'(m_t) - 1] = \gamma_t^B \sigma. \quad (15)$$

The risk compensation charged by bankers is exactly the net present value of the foregone marginal investment. When banks are undercapitalized, and thus, γ_t^B is high, firms tend to hold less liquidity and invest less. This result echoes Corollary 1 of the static model, except that now bank equity evolves endogenously in response to shocks. Follow bad shocks, banks supply less inside money, slowing down the reallocation of resources towards productive agents, i.e., the investing firms under liquidity shocks. Following good shocks, more inside money is created, facilitating reallocation. Eisfeldt and Rampini (2006) document procyclical reallocation among firms. Bachmann and Bayer (2014) find procyclical dispersion of firms' investment rates. In this model, procyclical reallocation is driven by procyclical money creation and transaction volumes.

So far, we have revisited the main results of the static model in a dynamic setting. Next, I will discuss an intertemporal feedback mechanism that amplifies the procyclicality of intermediation.

Intertemporal feedback. The endogenous capital price plays a critical role in generating a feedback mechanism. Proposition 2 shows firms' indifference condition as a capital pricing formula.

Proposition 2 (Capital Valuation) *The equilibrium price of capital satisfies*

$$q_t^K = \frac{\underbrace{\rho}_{\text{Discount rate}} - \left(\underbrace{\alpha}_{\text{Production}} + \underbrace{\lambda [q_t^K F'(m_t) - m_t]}_{\text{Expected net investment gain}} - \underbrace{(\rho - r_t) m_t}_{\text{Deposit carry cost}} + \underbrace{\sigma_t^K \sigma}_{\text{Quadratic covariation}} \right)}{\underbrace{\mu_t^K}_{\text{Expected price appreciation}} - \underbrace{\delta}_{\text{Expected capital destruction}}}, \quad (16)$$

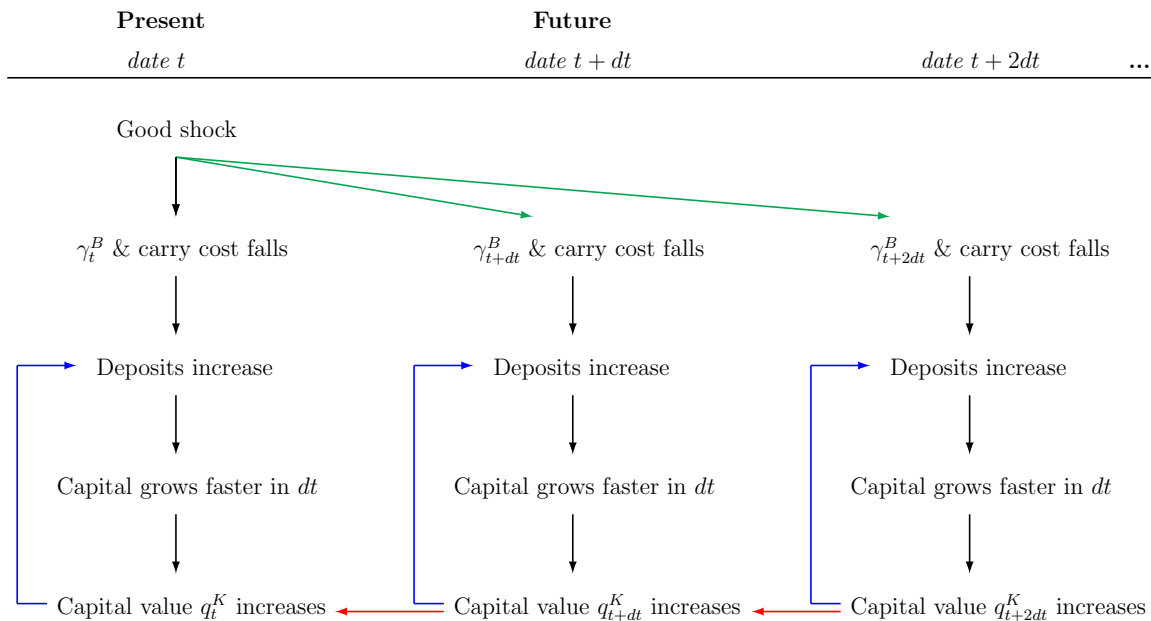


Figure 3: Intertemporal Feedback and Procyclicality. This figure illustrates the mechanism behind the procyclicality of firms' money demand. Following good shocks, bank risk price γ_t^B declines, and due to the persistence of shock impact, the path of γ_t^B in expectation shifts down. Therefore, firms face a lower money premium right now and hold more money, and they expect so in the future, which translates into a higher growth path of capital in expectation (through the cash-financed investments) and higher value of capital.

where μ_t^K and σ_t^K are defined in the equilibrium price dynamics: $dq_t^K = \mu_t^K q_t^K dt + \sigma_t^K q_t^K dZ_t$.

Capital price is procyclical. Consider a positive shock, $dZ_t > 0$, to an interior state, $\eta_t \in (\underline{\eta}, \bar{\eta})$. Since fewer loans default than expected, banks receive a windfall. Given the wedge between q_t^B and 1 that is created by the issuance cost, q_t^B does not immediately jump down to one and trigger payout. So, banks' equity increases, and *in expectation*, the shock's impact on the bank equity will only dissipate gradually into the future. Thus, a positive shock increases current bank equity, and through the persistence of its impact, it lifts up the expectation of future bank equity.

The positive shock increases capital price through two channels. As banks' equity increases, they charge a lower price of risk for deposit creation, so firms pay a lower deposit carry cost, $\rho - r_t$, and hold more deposits from t to $t + dt$. Through more investments financed by these deposits, capital is expected to grow faster in dt , which directly leads to a higher market price of capital. This is the *contemporaneous* channel of procyclicality.

An *intertemporal* channel further increases capital price. Due to the persistent impact of the shock, firms expect the banking sector to be better capitalized for an extended period of time, and thereby, they expect to hold more deposits and capital to grow faster going forward. This lifts up the expectation of future capital prices, which feeds back into an even higher current price through the expected price appreciation μ_t^K . Figure 3 illustrates the two channels of procyclical q_t^K .³⁸

Procyclical bank leverage. Following good shocks, banks' equity increases and they charge a lower price of risk γ_t^B . As illustrated by Figure 4, the money market moves from "1" to "2". The equilibrium quantity of deposits increases, and whether it increases faster or slower than banks' equity determines whether bank leverage is procyclical and countercyclical.

Because capital price is procyclical, firms' money demand will also shift outward, so the equilibrium point moves further from "2" to "3," which increases the equilibrium quantity of deposits even further. This endogenous expansion of firms' money demand allows banks' debt to grow faster than their equity, contributing to the procyclicality of bank leverage.

One aspect of the model that distinguishes itself from other macro-finance models is this endogenous expansion of the demand for intermediaries' debt. A static demand may lead to countercyclical leverage (e.g. He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014)).³⁹

This paper shares with Kiyotaki and Moore (1997) the idea that intertemporal complementarity amplifies fluctuations. Capital becomes more valuable (higher q_t^K) because it grows faster, which is in turn due to more money held by firms in the *future*. Through q_t^K , firms' *current* money demand rises in the expectation of future market conditions. The procyclicality of money demand contributes to the procyclicality of bank leverage and the resulting risk accumulation.

³⁸Note that an additional channel of procyclical q_t^K has been shut down by the assumption that the economy has enough collateral to back loans. Following good shocks, banks expand and become willing to lend more. When firms' borrowing constraint binds, a collateral shadow value (κ in the static model) arises, which increases capital price.

³⁹In these models, intermediaries bet on asset prices and the volatility of asset prices is countercyclical, so a binding value-at-risk constraint on leverage can make leverage procyclical (e.g., Adrian and Boyarchenko (2012); Danielsson, Shin, and Zigrand (2012)). Adrian and Shin (2014) rationalize the risk-based constraint in a principal-agent setup.

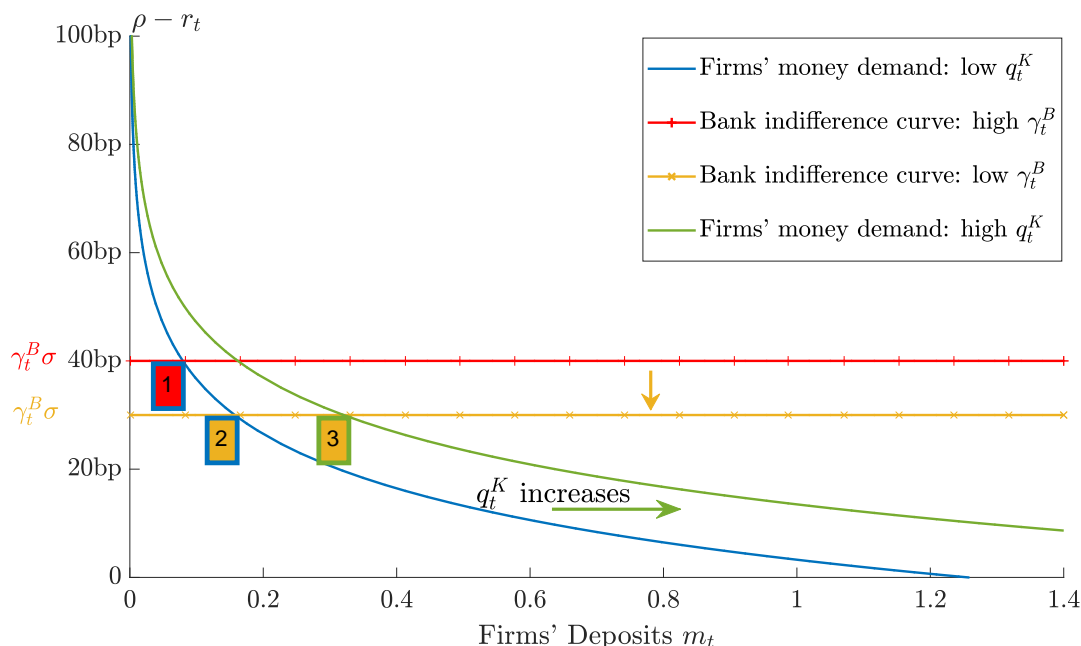


Figure 4: Money Market Response to A Positive Shock. This figure illustrates how the money market responds to a positive aggregate shock. First, the bank indifference curve shifts downward because bank risk price γ_t^B declines (i.e., from (1) to (2)). Second, firms' money demand curve shifts outward because capital price q_t^K rises (i.e., from (2) to (3)), which is in turn due to a higher growth path in expectation as shown in Figure 3.

As leverage rises, bank equity becomes more sensitive to shocks, so does the whole economy through η_t . Asset price here plays a key role in intertemporal feedback, but it differs from a typical balance-sheet effect in Kiyotaki and Moore (1997), and more recently, Brunnermeier and Sannikov (2014).

Even though the model features a specific type of money demand motivated by the corporate cash holdings, the insight that leverage procyclicality results from the procyclicality of money demand is general. We would naturally expect that a booming economy has a stronger transaction demand for money-like securities issued by the financial intermediaries. The firms' money demand in this paper is only one particular characterization of procyclical money demand.

An assumption in the model is that when firms are not experiencing the Poisson liquidity shock, they can immediately respond to changes in the capital price by raising funds from banks or households to build up savings. In reality, firms may not act so swiftly in the constant presence of financing frictions, and as a result, build up savings largely through internal cash flows. To

what extent it weakens the channel of procyclical leverage is an interesting empirical question.⁴⁰

Dynamic investment inefficiency. The endogenous variation in capital price leads to a new form of investment inefficiency that only arises in a dynamic setting. Taking as given q_t^K , Corollary 1' reveals a form of *static inefficiency*, measured by the wedge between firms' deposits m_t and the contemporaneous investment target i_t^* , defined by $q_t^K F'(i_t^*) = 1$. Static inefficiency arises because bankers' *current* capacity to create money is limited. This echoes Corollary 1 of the static model.

In a dynamic setting, the investment target i_t^* also varies with capital price. Due to the necessity and cost of carrying deposits, q_t^K is smaller than q_{FB}^K , the price of capital in an unconstrained economy in which firms finance investment freely:

$$q_{FB}^K = \frac{\alpha + \lambda [q_{FB}^K F(i_{FB}) - i_{FB}]}{\rho + \delta}, \quad (17)$$

where the first-best investment is given by $q_{FB}^K F'(i_{FB}) = 1$. Because $q_t^K < q_{FB}^K$, the investment target i_t^* is below the first-best investment rate i_{FB} . Since q_t^K reflects the *expectation* of future money market conditions, and in particular, future deposit carry cost (i.e., the money premium, $\rho - r_t$, driven by bank equity), the wedge, $i_{FB} - i_t^*$, measures a form of *dynamic inefficiency*.⁴¹

Stagnation and instability. The procyclicality of bank leverage has two critical implications: first, recession lasts a long period of time; second, the risk of recession accumulates in booms.

Recession states are defined as the states where the economy's expected growth rate of K_t , $\lambda F(m_t) - \delta$, is negative. Growth is driven by investment, which is tied to firms' liquidity

⁴⁰Related, since deposits can be regarded as insurance against the Poisson shock and the demand for such insurance is procyclical, this paper shares the insight of Rampini and Viswanathan (2010) that when firms are richer, they hedge more. In Rampini and Viswanathan (2010), hedging competes with investing for limited resources, and thus, richer firms, with more resources at hand, hedge more. In contrast, firms in this paper are always unconstrained outside of the Poisson times, so the procyclical demand for insurance is not driven by more resources available, but rather, by the procyclical benefit of hedging (i.e., more valuable investment), which is in turn due to the procyclicality of q_t^K .

⁴¹Note that this dynamic inefficiency is about the lack of investment or productive reallocation across firms, *not* the dynamic inefficiency in overlapping-generation models that is due to the lack of intergenerational trade.

holdings. Negative shocks lead to recession. As banks' equity is depleted, banks require a higher risk compensation $\gamma_t^B \sigma$, pushing up the money premium. At the same time, the price of capital decreases. As illustrated by Figure 4, we see both an upward shifting of bankers' indifference curve and an inward shifting of firms' money demand. In equilibrium, firms' deposit holdings and investment decrease. A banking crisis affects the real economy through the collapse of inside money creation, which echoes the classic account of the Great Depression by Friedman and Schwartz (1963).

Shocks have *asymmetric* impact near the boundaries. Following bad shocks, firms' money demand contracts, and banks deleverage. As the economy approaches the issuance boundary $\underline{\eta}$, the impact of negative shocks is bounded: bank equity never decreases beyond the issuance boundary. Thus, low leverage offers only a small benefit by making the banks robust to negative shocks, but it limits the impact of positive shocks on the rebuilding of bank equity. Stagnation results from low leverage in bad states. Recovery requires a sequence of *sufficiently large* good shocks.

Consider good shocks. Bank equity increases, and more deposits are issued, facilitating trade and investment. Due to the procyclicality of firms' money demand, banks' leverage grows. As the economy approaches bank payout boundary, the impact of positive shocks is bounded: bank equity never increases beyond the payout boundary. Therefore, high leverage in good states only makes bank equity more sensitive to negative shocks. Therefore, as a boom prolongs, leverage builds up endogenously, so does the downside risk. As a result, the economy becomes increasingly fragile, so even small negative shocks can significantly deplete bank equity and trigger a recession.

Proposition 3 solves the stationary probability density of η_t (i.e., the likelihood of different states) and the expected time to reach $\eta \in [\underline{\eta}, \bar{\eta}]$ from $\underline{\eta}$, the bottom of recession ("recovery time").

Proposition 3 *The stationary probability density of state variable η_t , $p(\eta)$ can be solved by:*

$$\mu^\eta(\eta) p(\eta) - \frac{1}{2} \frac{d}{d\eta} (\sigma^\eta(\eta)^2 p(\eta)) = 0,$$

where $\mu^\eta(\eta)$ and $\sigma^\eta(\eta)$ are defined in Equation (10). The expected time to reach η from $\underline{\eta}$, $g(\eta)$ can be solved by:

$$1 - g'(\eta) \mu^\eta(\eta) - \frac{\sigma^\eta(\eta)^2}{2} g''(\eta) = 0,$$

with the boundary conditions $g(\underline{\eta}) = 0$ and $g'(\underline{\eta}) = 0$.

Solving the equilibrium. The solution of the model is a set of functions defined on $[\underline{\eta}, \bar{\eta}]$. Each function maps the value of state variable η_t to the value of an endogenous variable. These functions are separated into two sets. The first set includes the forward-looking variables $(q^B(\eta_t), q^K(\eta_t))$. The second includes variables, such as banks' leverage x_t , firms' deposits-to-capital ratio m_t , and deposit rate r_t that can be solved directly once we know the first set of functions. And, once these variables are solved as functions of $(q^B(\eta_t), q^K(\eta_t))$ and their derivatives, we can use Itô's lemma to convert Equation (12) and (16) into a system differential equations of $(q^B(\eta_t), q^K(\eta_t))$.

A key step is to solve bank leverage using Equation (15), the intersection of money demand and supply curves. On the left hand side is the marginal benefit of investment, $\lambda [q_t^K F'(m_t) - 1]$. We can use the deposit market clear condition

$$m_t K_t = (x_t - 1) E_t, \text{ i.e., } m_t = (x_t - 1) \eta_t,$$

to substitute out m_t with $(x_t - 1) \eta_t$. On the right hand side is the intermediary wedge, $\gamma_t^B \sigma$. By knowing the function $q^B(\eta_t)$, we know the elasticity ϵ_t^B , so banks' risk aversion, γ_t^B , is directly linked to the equilibrium leverage

$$\gamma_t^B = -\epsilon_t^B \sigma_t^\eta, \text{ where } \sigma_t^\eta = (x_t - 1) \sigma.$$

Therefore, Equation (15) solves bank leverage, x_t , as a function of η_t , q_t^K , and ϵ_t^B :

$$\lambda [q_t^K F'((x_t - 1)\eta_t) - 1] = -\epsilon_t^B (x_t - 1)\sigma^2.$$

With x_t , we can solve m_t from the deposit market clearing condition, and r_t from Equation (11). The details of the algorithm is provided in Appendix I, which also shows the existence and uniqueness of the Markov equilibrium (stated in Proposition 4) in a manner of constructive proof.

Proposition 4 (Markov Equilibrium) *There exists a unique Markov equilibrium with state variable η_t that follows an autonomous law of motion in $[\underline{\eta}, \bar{\eta}]$. Given functions $q^B(\eta_t)$ and $q^K(\eta_t)$, agents' optimality conditions and market clearing conditions solve bank leverage, firms' deposits, and deposit rate as functions of η_t . Substituting these variables into bankers' HJB equation and the capital pricing formula (Equations (12) and (16)), we have a system of two second-order ordinary differential equations that solves $q^B(\eta_t)$ and $q^K(\eta_t)$ under the following boundary conditions:*

$$\begin{aligned} \text{At } \underline{\eta}: \quad (1) \frac{dq^K(\eta_t)}{d\eta_t} = 0; \quad (2) q^B(\bar{\eta}) = 1 + \chi; \quad (3) \frac{d(q_t^B \eta_t)}{d\eta_t} = 0; \\ \text{At } \bar{\eta}: \quad (4) \frac{dq^K(\eta_t)}{d\eta_t} = 0; \quad (5) q^B(\underline{\eta}) = 1; \quad (6) \frac{d(q_t^B \eta_t)}{d\eta_t} = 1. \end{aligned}$$

We need exactly six boundary conditions for two second-order ordinary differential equations and two endogenous boundaries to pin down the solution. (1) and (4) prevent capital price from jumping upon reflection, ruling out arbitrage in the market of capital. (2) and (5) are the value-matching conditions for banks' issuance and payout respectively. (3) and (6) are the smooth-pasting conditions that guarantee that the bank shareholders' value does not jump at the reflecting boundaries. The market value of bank equity is $q_t^B E_t = q_t^B \eta_t K_t$. At $\underline{\eta}$, (3) guarantees the value of existing shares does not jump when banks issue news shares. At $\bar{\eta}$, (6) guarantees the value of bank equity declines by the exact amount of dividends paid out. If either (3) or (6) is violated, taking as given the aggregate issuance and payout, individual banks will have incentive to deviate.

3.3 Solution

Calibration. To numerically solve the differential equations, we need to fix the parameter values. One unit of time is set to one year. δ and σ are set by the mean and standard deviation of loan delinquency rates (source: FRED). The other parameters are chosen to match model moments to data, such as money premium, interest rate, corporate cash holdings, and economic growth. All model moments are based on the stationary distribution. Bank leverage is intentionally left out of the calibration, so that the model's leverage dynamics, and the associated boom-bust cycle, may serve to some extent as an out-of-sample evaluation. Appendix II.1 summarizes the calibration.

Note that as in the static model, firms' external financing capacity cannot exceed their collateral value, i.e., $q_t^K K_t$ in aggregate. The analysis so far has focused on the case that this constraint never binds. This assumption is satisfied by the calibrated solution: the ratio of bank loans to collateral value varies from 4.6% to 83.3% depending on the state of the world (i.e., η_t).⁴²

Bank balance-sheet cycle. The state variable η_t measures the aggregate bank equity (i.e., the size of money suppliers) relative to the size of firms (money demanders). Bank equity affects the real economy through deposit creation. When η_t increases, banks expand balance sheets and issue more deposits, so firms can hold more liquidity for investment, and the economy grows faster. Figure 5 shows the statistical properties of η_t . Panel A plots several sample paths of η_t by simulating the law of motion (Equation (10)).⁴³ The paths are bounded by the issuance and payout boundaries.

Panel B of Figure 5 plots the impulse response function of η_t . It shows that shocks have persistent effects. Because the law of motion of η_t is non-linear and state-dependent, we can-

⁴²The model will have two state variables, η_t and firm equity scaled by capital stock, if in some states of the world, firms' collateral constraint binds. This certainly enriches the model and improves its quantitative performance, but sacrifices the transparency of mechanisms. Rampini and Viswanathan (2017) study the joint dynamics of firm and intermediary equities (see also Elenev, Landvoigt, and Nieuwerburgh (2017)).

⁴³The discretization step is one day so the shock ΔZ_t is drawn from normal distribution $N(0, 1/365)$. Asmussen, Glynn, and Pitman (1995) discuss the discretization error. Simulating regulated diffusion has a weak order of convergence of 1/2, which is slower than the order 1 for simulations of diffusion processes without reflecting boundaries.

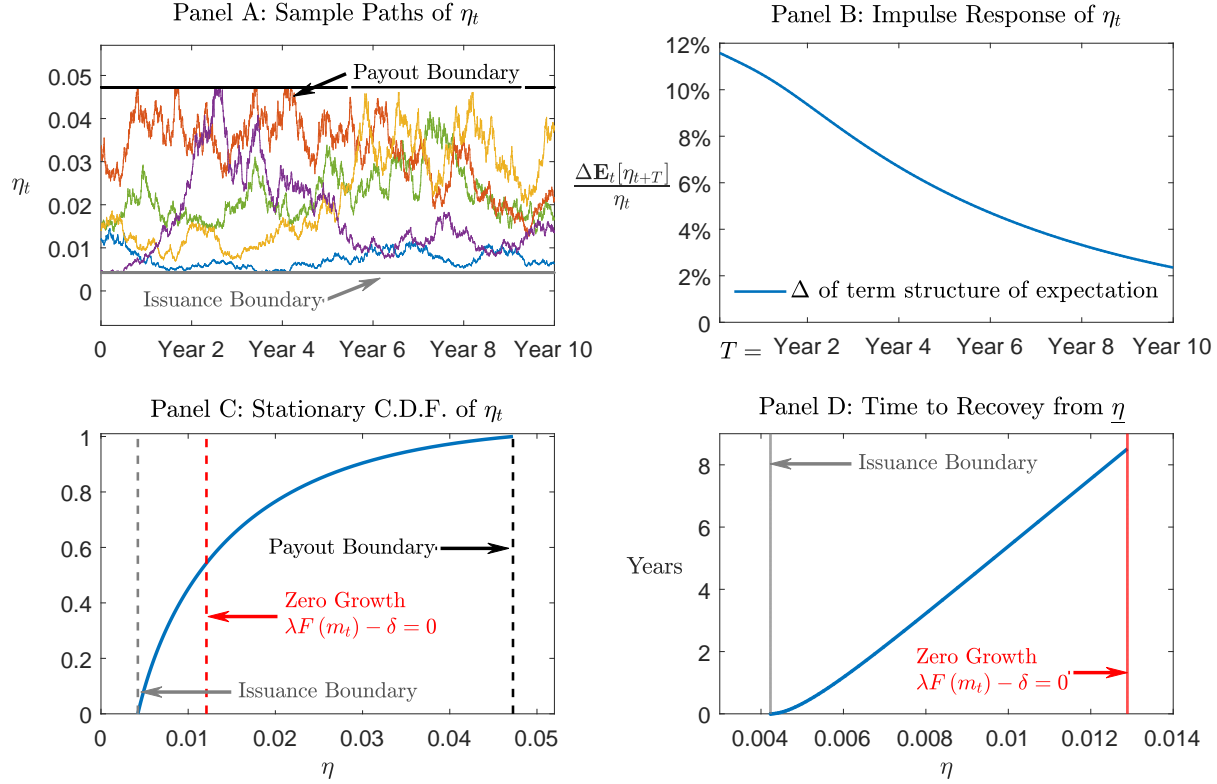


Figure 5: State Variable Dynamics. This figure shows the statistical properties of the state variable η_t , the ratio of banking capital to illiquid, production capital. Panel A plots the simulated paths of η_t . All paths are endogenously bounded by the two (payout and issuance) reflecting boundaries. Panel B plots the percentage change of the expected value of η at different horizons in response to a positive shock. It shows the persistence of shock impact. Panel C plots the stationary cumulative distribution function (C.D.F.) of η_t that starts from the issuance boundary, passes the zero growth point at around 0.5, and ends at the payout boundary. Panel D plots the expected years (vertical axis) to reach a value of η (horizontal axis) when the current state is at the issuance boundary (ending at the zero growth point).

not define impulse response functions as in linear time-series analysis. Thus, to illustrate the persistent impact of shocks, I fix the initial value of η_t to the median value under the stationary distribution, and consider an increase of η_t to the 56th percentile. The figure plots the percentage change of the term structure of expectation, i.e., the expected value of η_{t+T} with T ranging from one month to ten years.⁴⁴ In expectation, the impact of the shock dissipates gradually. The initial increase of 11.6% raises the expected value in ten years by 2.4%.

The persistence is caused by banks' precautionary behavior, which is in turn due to the re-

⁴⁴The expectation is calculated by the Kolmogorov backward equation (i.e. the Feynman–Kac formula), for reflected diffusion processes. The partial differential equations are solved by the Method of Lines (Schiesser and Griffiths (2009)). Borovička, Hansen, and Scheinkman (2014) provide an alternative, systematic framework to define and calculate impulse responses and the term structure of shock elasticity for non-linear diffusion processes.

capitalization cost. If bankers could issue equity freely (i.e. $\chi = 0$), they no longer need to retain equity. Whenever q_t^B is above one, signaling an improvement of the investment opportunity set, bankers raise equity from households; whenever q_t^B is below one, bankers distribute dividends. The dilution cost implies that equity is only raised infrequently when q_t^B reaches $1 + \chi$, signaling severe capital shortage. χ opens up a wedge between q_t^B , the value of one dollar as banks' retained equity, and 1, the value of one dollar paid out as dividends. As a result, banks preserve a financial slackness. Unless the economy hits the payout boundary, banks accumulate equity.

Panel C shows the stationary cumulative probability function (c.d.f.) from Proposition 3. The curve starts from zero at the issuance boundary, and ends at one at the payout boundary. Around 50% of the time, the economy is in a region with negative growth. I calibrate the mean growth rate to a relatively low number, 0.74% per year, which is the growth of output attributed to intangible investment from 1995 to 2007 (Corrado and Hulten (2010)). Intangible investment, such as R&D, relies heavily on internal liquidity, and thus, fits the model specification of investment.⁴⁵

Panel D of Figure 5 plots the expected time to reach different values of η_t from the issuance boundary $\underline{\eta}$ (Proposition 3). The right bound marks the lowest value of η_t that delivers a non-negative growth rate. In expectation, it takes more than eight years to recover from the bottom of a recession. As previously discussed, stagnation results from bank's deleveraging in the bad states and the asymmetric impact of shocks on bank equity near the reflecting boundaries.

Procyclicality. As illustrated by Figure 4, the money market moves with (γ_t^B, q_t^K) , which in turn vary with the state variable, η_t . γ_t^B , the risk price that bankers charge for issuing safe deposits backed by risky loans, drives the money supply, while q_t^K , the capital price, shifts firms' money demand. Panel A of Figure 6 shows γ_t^B as a function of η_t . Because one unit of time is set to one

⁴⁵Since Hall (1992) and Himmelberg and Petersen (1994), it has been well documented that R&D heavily relies on internal financing (see Hall and Lerner (2009) for a survey on innovation financing). A difficulty of external financing is that the knowledge asset created by R&D is intangible, partly embedded in human capital, and often very specialized to the particular firm in which it resides. It is difficult for investors to repossess such intangible assets in case of default.

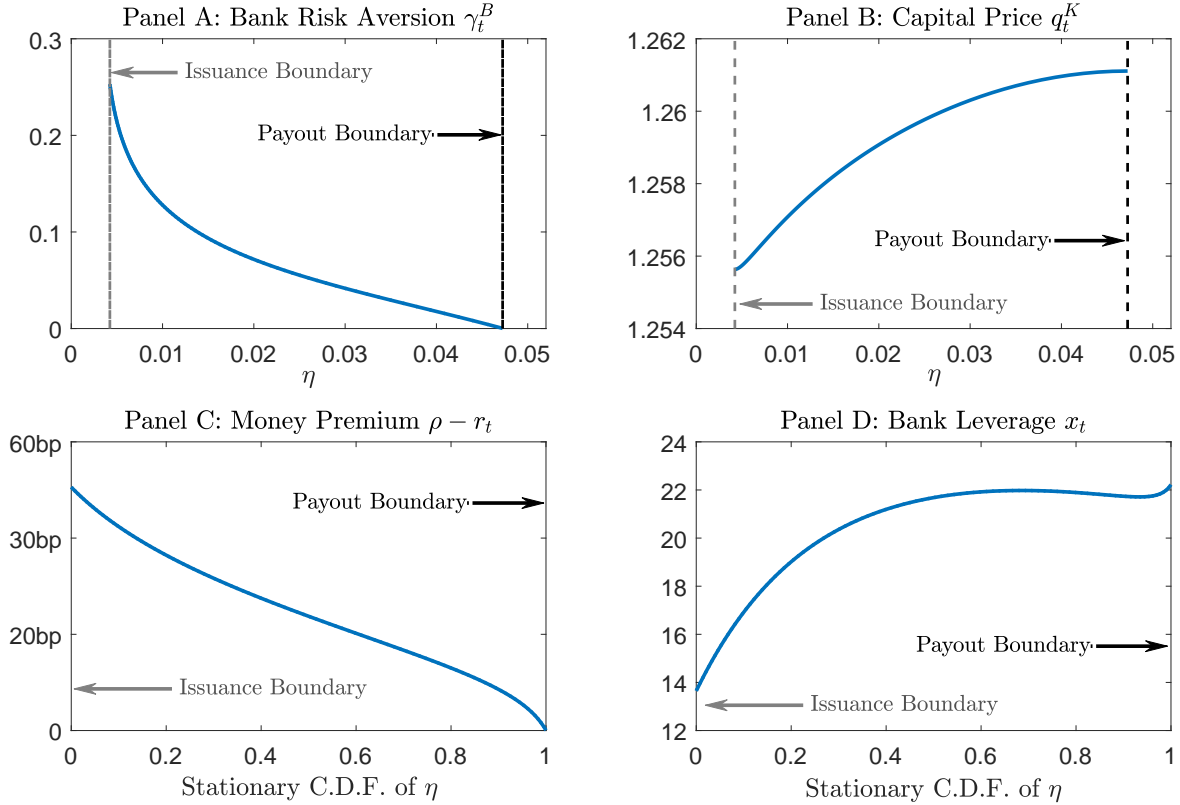


Figure 6: Proccyclical Leverage. This figure plots bank risk price γ_t^B (Panel A) and capital price q_t^K (Panel B) against the state variable η_t . These two variables determine the locale of money market equilibrium (Figure 2). Panel C and D plot money premium and bank book leverage respectively against the stationary cumulative distribution function (C.D.F.). They show how often (horizontal axis) the variable of interest stays in certain regions (vertical axis).

year, γ_t^B is the annual Sharpe ratio of risky lending financed by risk-free deposits. γ_t^B decreases in η_t . When the economy is close to the bank recapitalization boundary, $\underline{\eta}$, banks charge a price of risk close to 0.25; at the payout boundary, γ_t^B is zero. Good shocks increase η_t and decrease γ_t^B , shifting downward the bankers' indifference curve in Figure 4. Panel B shows that q_t^K increases in η_t . Even if the variation of q_t^K is not quantitatively large, it is sufficient to generate strong procyclicality of firms' money demand that leads to procyclical bank leverage.⁴⁶

Panel C plots the equilibrium money premium, $\rho - r_t$, against the stationary cumulative probability of η_t . For instance, 0.2 on the horizontal axis is mapped to a money premium equal

⁴⁶As well documented in the empirical literature, asset price variation is dominated by the variation in discount rate (e.g. Cochrane (2011)). In the model, discount rate is fixed at ρ , so the variation in q_t^K is purely driven by firms' cost of liquidity management (i.e. the money premium), and their choice of liquidity holdings that determines the capital growth. Thus, quantitatively, we would not look for large variation in capital price over the cycle.

to 36 basis points, meaning that 20% of the time, money premium is larger than or equal to 36bp. The width of the curve shows how much time the economy spends in the region. The interval $[0.2, 0.3]$ is mapped to $[32\text{bp}, 36\text{bp}]$, so 10% of the time, the equilibrium money premium is between 32bp and 36bp. Reading the graph from left to right, we follow a path of positive shocks, and see as the banking sector builds up equity, their risk-taking capacity expands, and thus, the equilibrium money premium, i.e., firms' cost of liquidity management, declines.

Panel D plots bank leverage, x_t , against the stationary cumulative probability of η_t . Leverage is procyclical. Reading the graph from left to right, we see that when bank equity increases, banks issue even more deposits, so their leverage increases. The reason is that firms foresee a lower cost of liquidity management going forward, and thus, assign a higher valuation of capital, making the incentive stronger to hoard liquidity in case the investment opportunity arrives the very next instant (3 and 4). Reading the graph from right to left, we see how a crisis unfolds and banks deleverage.

There is a small region near the 80th percentile, where bank leverage rises following bad shocks (i.e., moving the left). In reality, balance-sheet cyclicalities may differ by the types of financial intermediaries. Adrian and Shin (2010) find the book leverage of broker-dealers is procyclical. He, Khang, and Krishnamurthy (2010) show that commercial banks' leverage actually increased in the 2007-09 financial crisis, especially as shadow banks shrank their balance sheets (Krishnamurthy, Nagel, and Orlov (2014)) and assets moved to the commercial banking sector (Acharya, Schnabl, and Suarez (2013)).⁴⁷ Commercial banks' capacity to create deposits depends not only on their financial soundness but also on regulatory constraints. Thus, banks in the model are closer to the financial intermediaries in the shadow banking sector. Many have argued that the demand for money-like securities is one of the major drivers behind the shadow banking development (e.g., Gorton (2010); Gorton and Metrick (2012); Gennaioli, Shleifer, and

⁴⁷He, Kelly, and Manela (2016) discuss several data issues and the choice between book and market leverages that depends on what type of frictions one has in mind. The model leverage, x_t , is the *book* leverage of a generic intermediary, which can be banks or other intermediaries that issue money-like liabilities. Huang (2016) provides a model of banks (under regulation) and shadow banks that captures the movements of assets between these two sectors.

Vishny (2013); Pozsar (2014)). The model does not feature heterogeneous intermediaries, but the non-monotonicity of intermediary leverage with respect to equity reflects some complexity of the leverage cycle.

The procyclicality of bank leverage helps explain the statistical properties of the model in Figure 5. In Panel C, there is a relatively small probability in states where banks' equity is high. In good times, banks' leverage is high, so the economy is very sensitive to shocks. When the economy is close to the payout boundary, the impact of good shocks is limited, because large good shocks trigger dividend distribution, meaning that the banking sector cannot grow beyond $\bar{\eta}$. Thus, high leverage only serves to amplify the impact of negative shocks. As a result, the downside risk accumulates as leverage rises. Because of this fragility, the economy spends less time in boom. Panel D of Figure 5 shows the slow recovery. When the economy is close to the issuance boundary, the impact of negative shocks is bounded. Low bank leverage only serves to reduce the impact of good shocks on bank equity, so banks accumulate equity slowly and the economy tends to get stuck in recessions. Countercyclical leverage would have led to the exactly opposite pattern.⁴⁸

Static and dynamic inefficiencies. In Corollary 1', the money premium is equal to bankers' required risk compensation $\gamma_t^B \sigma$. A higher γ_t^B , and thus, a higher money premium, directly translates into a larger gap between the cash-constrained investment rate and the current target rate i_t^* , defined by $q_t^K F'(i_t^*) = 1$. Panel A of Figure 7 shows the static inefficiency, measured by the percentage deviation of the investment rate, m_t , from the target, i_t^* . As bankers' risk capacity increases, the money premium declines, and firms hold more deposits for investment. Moving from the left to the right, the investment wedge declines from 95% at the depth of recession to 0%.

Panel B of Figure 7 shows the dynamic inefficiency, measured by the percentage deviation

⁴⁸Models with countercyclical leverage generate instability through other mechanisms, such as fire sale in He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014), which need intermediaries to hold long-term assets, and thereby, are exposed to endogenous volatility of asset prices. To highlight the novel link between leverage and money demand, I shut down this channel by restricting banks' investment to short-term loans.

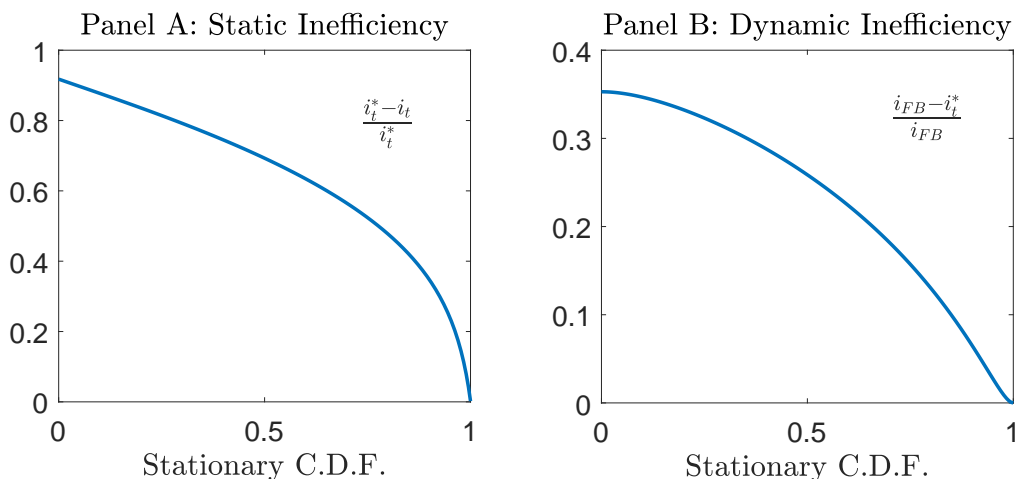


Figure 7: Static and Dynamic Investment Inefficiencies. This figure plots static investment inefficiency (Panel A), measured by the percentage deviation of equilibrium investment rate (i_t) from the target rate implied by the equilibrium capital price (i_t^*), and dynamic investment inefficiency (Panel B), measured by the percentage deviation of target rate (i_t^*) from the first-best investment rate (i_{FB}), against the stationary cumulative distribution function (C.D.F.). The plots show how often (horizontal axis) the variable of interest stays in certain regions (vertical axis).

of the current target from the first-best investment rate defined by Equation (17). The wedge varies with capital price, and declines as η_t increases, because, as shown in Panel B of Figure 6, q_t^K increases in η_t . Around 50% of the time, the current target is 25% or more below the first-best.

Investment inefficiencies decrease welfare. Given the constant productivity α , the aggregate consumption is determined by the capital stock K_t . Under risk-neutral preference, what matters, from a welfare perspective, is the expected growth rate of K_t , i.e., $\lambda F(m_t) - \delta$, the newly created capital net of the expected depreciation of existing capital. Therefore, through firms' liquidity constraint on investment, welfare is tied to money creation. As inside money supply is constrained by banks' balance-sheet capacity, the government has a natural role in supplying *outside money*. But does outside money really benefit investment and growth? How will banks respond? Will outside money make the economy more stable? The next section aims to answer these questions.

4 Government Debt: Instability, Stagnation, and Welfare

It has long been recognized that government debt offers monetary services (Patinkin (1965); Friedman (1969)). The repo market developments since 1980s further enhance the liquidity of Treasury securities (Fleming and Garbade (2003)). In this section, I introduce government debt as an alternative to bank debt as money (“outside money”). While it seems to alleviate the money shortage faced by firms, outside money squeezes banks’ profit through the competition with inside money, and banks respond by adjusting leverage, payout, and equity issuance. As a result, outside money can *destabilize* the economy and exacerbate the investment inefficiencies. These findings complement the literature on government debt as a means to financial *stability* (Greenwood, Hanson, and Stein (2015); Krishnamurthy and Vissing-Jørgensen (2015); Woodford (2016)).

4.1 Setup

Firms can hold both bank debt and government debt to relax the liquidity constraint on investments. Thus, issued at t , government debt pays the same risk-free rate as deposits, r_t , at $t + dt$. To focus on the liquidity provision role of government debt, I abstract away other fiscal distortions: issuance proceeds are distributed as lump-sum payments and debt is repaid with lump-sum tax on households. Moreover, I assume the government faces a debt limit that is proportional to the scale of the economy, i.e., $M^G K_t$, and consider a debt management strategy in line with Friedman’s rule – the government always issues the maximum amount in the presence of money premium.⁴⁹

Firms’ money holdings per unit of capital are now $m_t + M^G$. Substituting it into the optimality condition of deposit holdings in Lemma 1’, we have a new deposit demand curve:

⁴⁹Maximum issuance is in line with Friedman’s rule: individuals’ opportunity cost to hold money should be equal to the social cost of creating money (Friedman (1969); Woodford (1990a)). Firms’ opportunity cost to hold money is $\rho - r_t$, the money premium. The private sector’s marginal cost of money creation is $\gamma_t^B \sigma$, the risk compensation charged by bankers. In contrast, the government’s cost of money creation is zero as I already assume away any fiscal distortions. Therefore, in the current setting, Friedman’s rule suggests that the government should maximize its debt issuance (i.e., outside money supply). However, as will be discussed later, because banks also face financial constraints, Friedman’s rule is not necessarily the *optimal* government debt management strategy.

$$\rho - r_t = \lambda [q_t^K F'(m_t + M^G) - 1].$$

Because $F(\cdot)$ is a concave function, the demand curve is shifted inward. Introducing government debt reduces the marginal benefit of deposit holdings, and the equilibrium money premium. By helping firms manage liquidity, government debt is likely to have a positive effect on investment and growth, which is similar to the investment crowding-in effect in Woodford (1990b) and Holmström and Tirole (1998).⁵⁰ However, the actual effect on investment depends on how banks react.

This setup does not distinguish Treasury securities from central bank liabilities that pay interests, intending to capture outside money supply both in the traditional sense of monetary base expansion and in the broad sense of government issuing liquid securities. Indeed, the setup has an alternative interpretation: government debt is held by a central bank, who, as in reality, issues an equal amount of reserves that pay the same interest rate and are held by banks, and banks in turn issue an equal amount of deposits to firms backed by reserves. On this chain, both the central bank and the banks are just pass-through. Intermediating between risk-free assets and liabilities does not require additional bank equity. The line between short-term Treasury securities and reserves is becoming increasingly blurry in reality.⁵¹ Interest on reserves has been introduced in many countries, such as the U.K. and the U.S. Moreover, the expansion of Federal Reserve reserve repo counterparties allows the public to hold reserves through money market funds.

⁵⁰The traditional crowding-out effect describes how the government debt supply raises interest rate in general, and thereby crowds out private investment through a higher financing cost. My model does not entertain this effect, because by shutting down external financing completely, the model has cash being the only determinant of investment.

⁵¹Accordingly, the liquidity coverage ratio (Basel Committee on Bank Supervision (2013)), which counts government securities as banks' liquidity holdings, is introduced as a modern version of reserves requirement.

4.2 Bank crowding-out effect: Instability and stagnation

Leverage cycle and instability. Figure 8 compares the model’s performances when the government debt-to-output ratio equal to 0%, the benchmark case, and 50%.⁵² Panel A plots the money premium, $\rho - r_t$, against the state variable, η_t . Government debt supply reduces the money premium, which is in line with the evidence (Krishnamurthy and Vissing-Jørgensen (2012); Greenwood and Vayanos (2014); Greenwood, Hanson, and Stein (2015); Sunderam (2015)). However, by raising r_t , outside money increases banks’ debt cost, and thus, reduces their return on equity. In response, banks reset their payout and issuance policies, shifting both boundaries to the left.

This *bank equity crowding-out* effect of government debt is also shown in Panel B, the stationary cumulative distribution of η_t .⁵³ Comparing the solid curve with the dotted curve, we see an increase in government debt shifts probability mass towards the left, i.e., to the states where bank equity is relatively low. Behind this increased fragility is the amplified leverage cycle.

Panel C of Figure 8 shows endogenous risk accumulates faster in booms (i.e., as η_t increases) when government debt-to-output ratio is 50% than when it is 0%. Endogenous risk is measured by σ_t^η , the instantaneous shock elasticity of η_t . σ_t^η is plotted against the stationary cumulative probability of η_t , so we can compare the two cases in the corresponding phases of their cycles. Reading the graph from left to right, we see faster accumulation of endogenous risk when government debt-to-output ratio is 50%. Endogenous risk is directly linked to leverage (Equation (10)),

$$\sigma_t^\eta = (x_t - 1) \sigma,$$

so Panel C also shows that government debt amplifies the bank leverage cycle. Stronger leverage procyclicality makes the economy more sensitive to shocks in good times (high η_t) and less sensitive to shocks in bad times (low η_t). As previously discussed, the impact of shocks is asymmetric

⁵²Comparative statics show the model’s performances in response to *unexpected* and *permanent* changes in the government debt supply. The current practice of U.S. Treasury debt management emphasizes predictability (Garbade (2007)), but since the financial crisis, there has been a considerable amount of uncertainty on fiscal policies, especially on the debt level (e.g. Baker, Bloom, and Davis (2015); Kelly, Pástor, and Veronesi (2016))

⁵³The stationary distribution is calculated using Proposition 4 using the model solutions under different M^G .

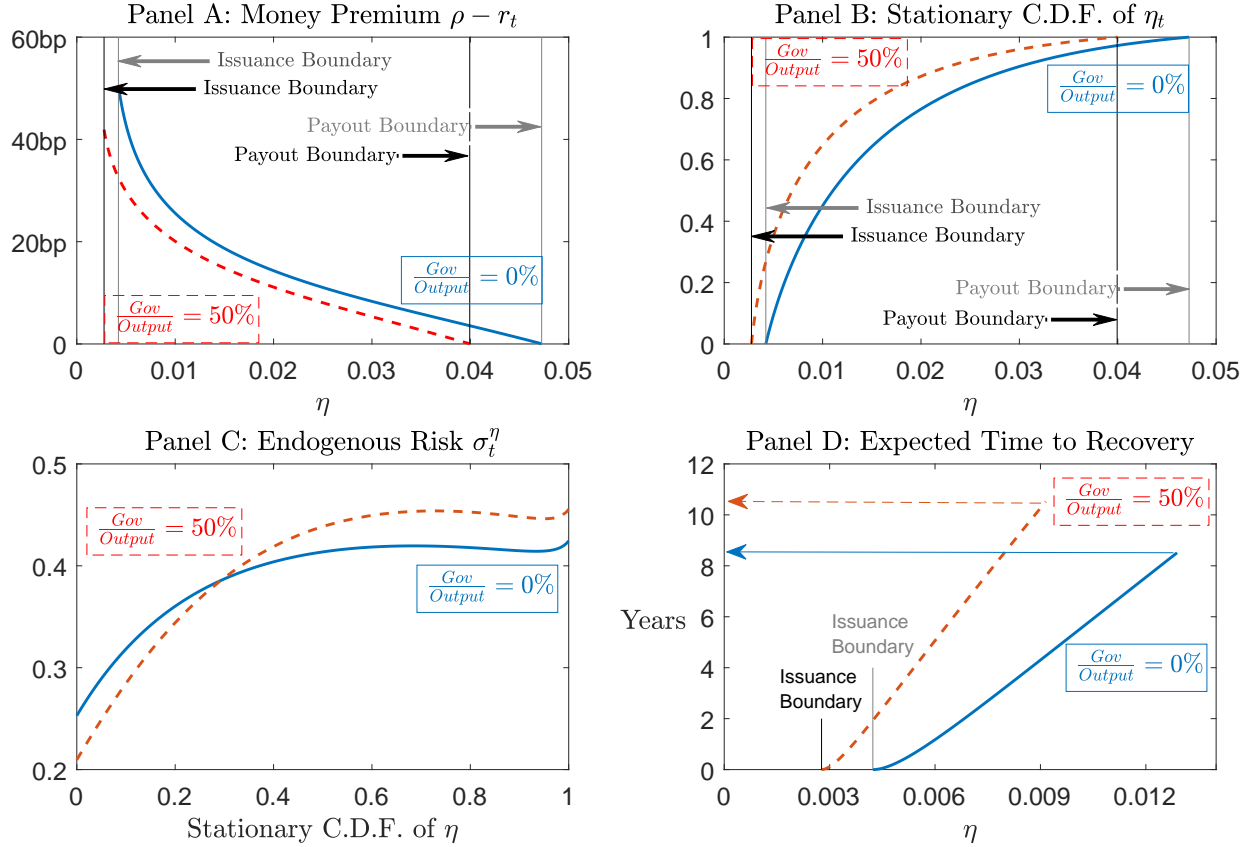


Figure 8: Outside money: Instability and Stagnation. This figure plots variables from the benchmark model (without government debt) and model with 50% government debt-to-output ratio (dotted line). Panel A plots money premium $\rho - r_t$ against state variable η_t . Panel B plots the stationary cumulative distribution function (C.D.F.) of η_t . Panel C plots endogenous risk, σ_t^η , i.e., the instantaneous standard deviation of η_t ($(x_t - 1)\sigma$, where x_t is bank leverage), against the stationary C.D.F. Panel D shows the expected years (vertical axis) to reach a value of η_t (horizontal axis) when the current state is at the issuance boundary (ending at the zero growth point).

near the reflecting boundaries, so the economy is stable near $\underline{\eta}$ and very responsive to negative shocks near $\bar{\eta}$. As a result, probability mass is shifted towards low η_t states, as shown in Panel B.

Outside money amplifies the procyclicality of bank leverage through a simple mechanism: by squeezing banks' profit per dollar of inside money created, i.e., the money premium, government debt forces banks to raise their leverage to sustain a high enough return on equity that justifies the occasionally incurred issuance cost. When the economy hits the bank issuance boundary, banks pay a cost of χ per dollar of shares issued. To compensate for this cost, banks must earn a positive return on equity, by leveraging up and by earning the money premium. The money premium moves in an interval with the minimum always equal to zero at the payout boundary and

the maximum determined by the total money available at the issuance boundary. Outside money reduces the maximum (Panel A), so the *range* of money premium shrinks.⁵⁴ Since return on equity is the product of money premium and leverage, bank leverage must vary in a wider range, given that q_t^B , a forwarding-looking measure of return on equity, has a fixed range of variation, i.e., $[1, 1 + \chi]$.

Greenwood, Hanson, and Stein (2015), Krishnamurthy and Vissing-Jørgensen (2015), and Woodford (2016) also explore the financial stability implications of government debt when it serves as a substitute for bank money. A common prediction is that by squeezing the money premium, government debt crowds out bank debt, and thereby, decreases banks' leverage and stabilizes the economy. However, their models ignore banks' dynamic equity management under equity issuance frictions. My model also features the competition between bank debt and government debt in the money market, but highlights a crowding-out effect on bank equity that works through banks' responses in issuance and payout policies.⁵⁵ By amplifying the bank leverage cycle and expediting the accumulation of downside risk, this bank profit crowding-out effect destabilizes the economy.

Stagnation. Panel D of Figure 8 plots the recovery paths from the bank recapitalization boundary.⁵⁶ The Y-axis shows the expected number of years it takes to travel from $\underline{\eta}$ to the corresponding values of η on the X-axis. For instance, when the government debt-to-output ratio is 0%, it takes a little more than one year to reach $\eta = 0.006$. Both curves end at the lowest value of η that has a non-negative economic growth rate, i.e., $\lambda F(m_t + M^G) - \delta \geq 0$. At the end of the dotted line, the economy has recovered, but it has taken almost eleven years.

Government debt prolongs recessions. Raising government debt-to-output ratio from 0% to

⁵⁴In Appendix II.2, Figure 10 shows that the volatility of money premium decreases in government debt supply.

⁵⁵The bank debt crowding-out effect of government debt supply has been documented by Bansal, Coleman, and Lundblad (2011) for financial commercial papers and banker's acceptance, Greenwood, Hanson, and Stein (2015) also for financial commercial papers, Sunderam (2015) for asset-backed commercial papers, and Krishnamurthy and Vissing-Jørgensen (2015) for the financial sector's debt in general using a long history of data.

⁵⁶The time to recovery is calculated following Proposition 4 using the model solution under different values of M^G .

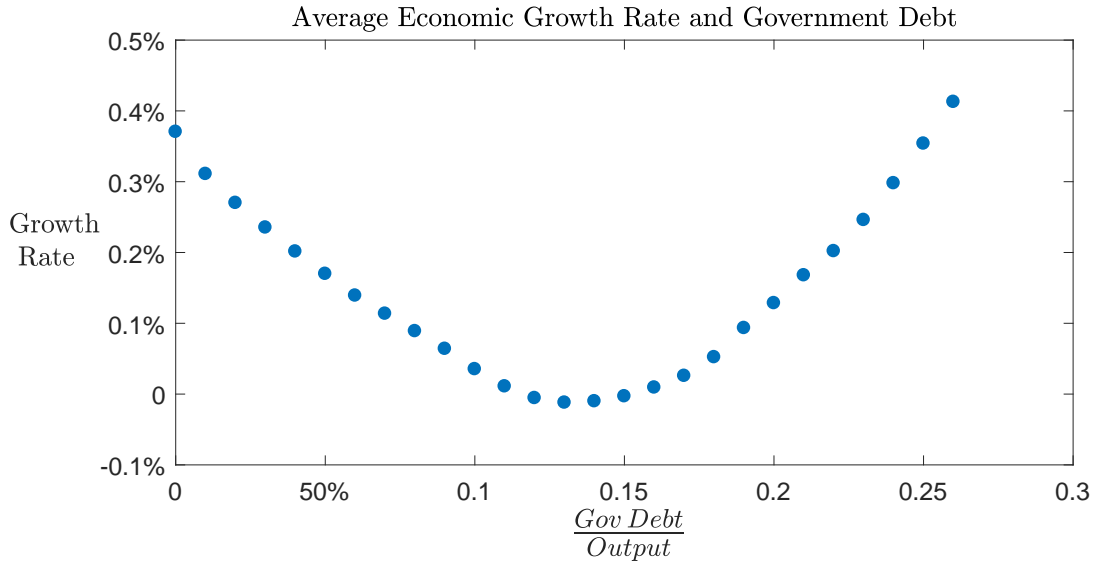


Figure 9: **Non-monotonic Effects of Outside Money on Investment and Growth.** The figure plots the mean growth rate (based on stationary distribution) for models with different levels of government debt-to-output ratio.

50% delays the recovery by two years. The reason is that as the economy still relies on banks as the *marginal* supplier of money, the profit crowding-out effect delays the recovery of the banking sector. Lower return on equity slows down the accumulation of bank equity, and discourages banks from recapitalizing, pushing down the issuance boundary. Admittedly, more government debt also makes bank equity less relevant, because firms already hold enough outside money, and thus, the marginal value of inside money declines. In the extreme case where the government has unlimited debt capacity to satiate firms’ money demand, the economy grows with the first-best investments.

Growth and welfare. Figure 9 shows the mean growth rate under the stationary distribution for different levels of government debt. Since the capital productivity is constant and agents do not care about consumption timing, the mean growth rate of K_t measures welfare. Before the government debt-to-output ratio reaches around 130%, more government debt decreases welfare, because *on average*, one dollar more outside money crowds out more than one dollar of inside money. It might seem difficult to reconcile this with Panel A of Figure 8, which shows more government debt decreases the money premium in *every* state of the world, and thus, must raise firms’

investment rate in *every* state of the world. The key lies in the shift of probability distribution.

Outside money crowds out inside money by amplifying the bank leverage cycle and shifting the probability mass towards states where banks are relatively capitalized and inside money supply is weak (Panel B of 8). Therefore, even if every state of the world has a higher growth rate, the average growth rate can decrease when more probability is assigned to low growth states. This argument no longer holds, when outside money almost satiates firms' money demand, making inside money less relevant. Indeed, once passing the point of 130% government debt-to-output ratio, more government debt leads to more investments, and thus, improves welfare.

The decreasing leg in Figure 9 is particularly relevant for understanding the pre- and post-crisis dynamics in the U.S. From 2001 to 2008, the public debt-to-GDP ratio rose from c.55% to c.70%. This coincided with a period of the strongest procyclicality of leverage in the financial sector. Many argue that the downside risk accumulated (e.g. FSB (2009)). The post-crisis period saw an even more dramatic increase in government debt, partly in response to both fiscal stimulation and the quantitative easing done by the central bank. By the end of 2012, the public debt-to-GDP ratio had reached its current level, around 100%. The recovery from the Great Recession has been slow. Therefore, this paper adds to the contemporaneous debate on the causes of stagnation.⁵⁷

4.3 Discussion

Optimal timing of government debt supply. This paper only considers the simplest case of fixed government debt supply, but the model does reveal an interesting trade-off that calls for an optimal strategy of government debt management. When the government issues more debt, it benefits the firms by reducing the money premium, but at the same time, it hurts the bankers who rely on the money premium as a source of profit. Its decision to increase or decrease its debt should balance the impact on both sectors, and in particular, depend on which sector is more

⁵⁷The existing literature has largely focused on the aggregate demand, demography, or technological progress (e.g., Eggertsson and Mehrotra (2014); Summers (2015)).

constrained.

When η_t is high and banks are already supplying a large amount of money, the marginal benefit of increasing government debt is small. When banks are undercapitalized and not creating enough deposits, raising government debt can significantly alleviate the money shortage. This suggests a countercyclical government debt supply. However, the bank crowding-out effect favors procyclical government debt supply. In good times, banks' leverage rises, making the economy unstable. The government should increase its debt to crowd out bank leverage. And since banks are well capitalized, the government worries less about crowding out banks' profit.⁵⁸ In low η_t states, the equity crowding-out effect becomes a major concern. The government may want to reduce its debt, allowing banks to rebuild equity by earning a high money premium.

Bank equity crowding in. So far, we have only considered the competition between government and banks as money suppliers. Under additional frictions, government debt may be held by banks for their own liquidity needs (Bianchi and Bigio (2014); Drechsler, Savov, and Schnabl (2017)) or as collateral (Saint-Paul (2005); Bolton and Jeanne (2011)). Banks may also hold government securities for regulatory purpose, for example to meet the liquidity coverage ratio (Basel Committee on Bank Supervision (2013)). Therefore, increases in government debt could relax banks' liquidity or regulatory constraints, and thereby, may increase their profit and crowd in equity.

Optimal bailout scale. In bad states, the government may intervene to recapitalize the banking sector, like the Troubled Asset Relief Program ("TARP") in the financial crisis of 2007-09. On the one hand, the banking sector benefits from equity injection. On the other hand, banks' profit from money creation is squeezed by the government debt that finances the bailout program. Balancing the two effects, the government may find an optimal scale of bailout financed by government debt

⁵⁸But by crowding out banks' profit, raising government debt gives banks more incentive to pay out dividends at a lower level of equity. By decreasing the payout boundary, it strengthens the asymmetric impact of shocks, making the impact of negative shocks stronger relative to good shocks. This negative effect needs to be weighed in against the positive effect on stability from the reduction of bank leverage.

in the sense that the expected time to recover is minimized.

Instability from coexistence of regulated and unregulated banks. We can reinterpret M^G as money supplied by a separate banking sector that is fully regulated and backs deposits with 100% reserves in the form of government debt. Regulated banks' balance sheets are just a pass-through. Their money supply grows proportionally with the economy. Such a banking sector was proposed in the Chicago Plan.⁵⁹ The competition between government and banks in supplying money can thus be reinterpreted as the competition between fully regulated banks and unregulated banks, or "shadow banks". Besides many issues surrounding the Chicago Plan, this paper points to a particular concern. If private money cannot be completely forbidden, the leverage cycle in unregulated shadow banking can be amplified by money that is fully backed government securities.

Instability from market liquidity. In reality, firms can and do hold some of other firms' or households' liabilities in their liquidity portfolios, provided that these securities have sufficiently liquid secondary markets. We can define $M^F K_t$ as the maximum amount of liquid securities that firms and households can issue in aggregate, so M^F is almost a measure of financial market development. Since firms and households are willing to issue any securities that promise an expected return less or equal to ρ , firms and households will maximize their issuance in the presence of money premium. To analyze the effects of market liquidity on growth and stability, we can simply follow the analysis of government debt, and all the conclusions carry through. Through the lens of the model, larger and deeper secondary security markets (i.e., higher M^F) can amplify the bank leverage cycle and prolong recessions. From 1975 to 2014, the stock market capitalization in the United States increased from c.40% of GDP to c.140%. The competition between market and intermediated liquidity is not the focus of this paper, but the model does point out a possibility that from a liquidity provision perspective, financial markets can destabilize financial intermediaries.

⁵⁹It is a banking reform initially proposed by Frank Knight and Henry Simons of the University of Chicago and supported by Irving Fisher of Yale University (Phillips (1996)). Benes and Kumhof (2012) revisited the plan in the interest of financial stability.

Relaxing the financial constraint. It has been assumed that when the liquidity shock hits, a firm cannot mobilize any resources other than its money holdings because existing capital has been destroyed and the investment project is not pledgeable. One way to relax the financial constraint is to introduce a liquidation value of the firm, say $M^L k_t$, where M^L is a constant, so creditors can liquidate the firm and seize $M^L k_t$ when the firm defaults. Thereby, when the liquidity shock arrives, the firm has liquidity equal to $(m_t + M^L) k_t$, the sum of money holdings and the pledgeable value of the firm. Now, we have a new money demand curve:

$$\rho - r_t = \lambda [q_t^K F'(m_t + M^L) - 1].$$

The role that M^L plays is the same as M^G (and M^F in the analysis of market liquidity). Therefore, the previous analysis carries through.⁶⁰ Relaxing the financial constraint indeed reduces firms' demand for bank debt, but it may amplify the bank leverage cycle, prolong recessions, and actually make the economy worse. To understand the financial stability implications of liquidity provided by the government (M^G), the secondary market of securities (M^F), and the productive capital (M^L), we must take into account the endogenous response of the banking system. More liquidity is not always beneficial.

5 Conclusion

This paper revisits the money view of financial intermediation in Friedman and Schwartz (1963). The theory has two ingredients: first, a bank balance-sheet channel of inside money creation, and second, procyclical money demand driven by firms' investment needs. The model makes predictions on intermediary leverage cycle, money premium dynamics, investment inefficiencies, frequency and duration of crises, and how government debt may contribute to financial instability.

⁶⁰There is one caveat – more capital holdings (i.e., larger k_t) relaxes the liquidity constraint on future investments by raising the liquidation value, so capital price q_t^K is likely to increase in comparison with the benchmark case. However, this mechanism enhances the procyclicality of q_t^K , and thus, amplifies the procyclicality of bank leverage.

Firms' capital requires complementary money holdings to finance periodic growth opportunities that require transactions with other firms. Money is banks' risk-free debt backed by risky loans. When more loans default following bad shocks, banks become undercapitalized, and the money premium spikes. As a result, firms' liquidity management is compromised, and their investment slumps. Recessions happen when the inside money creation collapses.

The frequency and duration of recession depend on the cyclical nature of bank leverage. Under the recapitalization friction that banks face, the impact of shocks is persistent, and thus, a unique intertemporal complementarity arises in firms' money demand, which contributes to the procyclicality of bank leverage. This money view of procyclicality stands in contrast with the existing literature, and together with banks' endogenous payout and equity issuance decisions, it reveals a new mechanism of downside risk accumulation in booms and slow recovery from crises.

The model also provides a laboratory to explore the financial instability implications of government debt (outside money). By squeezing banks' profit from money creation, government debt crowds out bank equity, amplifies the leverage cycle, and lengthens crises. Banks respond to government debt supply through their dynamic balance-sheet management, which is ignored by the first attempts in the literature. The destabilizing effect has negative welfare consequences, as long as outside money does not satiate the money demand, and thus, the economy still relies on banks as the marginal suppliers of money.

Whether outside and inside money are complements or substitutes is still an open question. This paper focuses on the later, but a richer environment can definitely entertain both possibilities. Banks may hold government debt as a buffer against their own liquidity shocks. Moreover, the proceeds from government debt issuance can be used to recapitalize banks in order to avoid a sudden evaporation of inside money that happens when banks become insolvent.

Another direction of future research is to combine the money view with the credit view (Bernanke (1983)), by allowing firms to partially finance their investment through bank credit. Whether cash and credit are substitutes or complements has critical implications on the cyclicity

of bank leverage and instability. If the investment technology has very strong decreasing returns to scale, cash and credit can be substitutes: when banks have a strong balance sheet and are willing to extend credit, firms demand less bank debt as money. Countercyclical money demand could lead to countercyclical bank leverage, stable booms, and fast recovery from recessions. In contrast, if the investment technology has increasing returns to scale, or the availability of credit depends on firms' internal funds (as a typical moral hazard friction would imply), cash and credit can be complements: access to bank credit amplifies the marginal benefit of money holdings, so in good states when banks are well capitalized and willing to lend, firms' demand for bank debt as money increases, which amplifies the procyclicality of intermediation activities, and leverage in particular.

This paper is partly motivated by two concurring phenomena. In the last few decades, the financial sector has grown exponentially, fueled by short-term debt that is deemed money-like. Meanwhile, nonfinancial corporations in the United States hold an enormous amount of money-like securities, largely driven by the entry of R&D-intensive firms. Behind these two trends is a structural transformation towards a new economy where growth heavily relies on intangible, cash-heavy investment (Corrado and Hulten (2010)). Against this broad background, the money view of financial intermediation may help us understand many other phenomena going forward, far beyond this simple account of procyclical intermediation and financial instability.

References

- ACHARYA, V. V., P. SCHNABL, AND G. SUAREZ (2013): “Securitization without risk transfer,” *Journal of Financial Economics*, 107(3), 515 – 536.
- ADRIAN, T., AND N. BOYARCHENKO (2012): “Intermediary Leverage Cycles and Financial Stability,” Staff Reports 567, Federal Reserve Bank of New York.
- ADRIAN, T., N. BOYARCHENKO, AND H. S. SHIN (2015): “The Cyclicity of Leverage,” Staff Reports 442, Federal Reserve Bank of New York.
- ADRIAN, T., E. ETULA, AND T. MUIR (2014): “Financial Intermediaries and the Cross-Section of Asset Returns,” *The Journal of Finance*, 69(6), 2557–2596.
- ADRIAN, T., AND H. S. SHIN (2010): “Liquidity and leverage,” *Journal of Financial Intermediation*, 19(3), 418 – 437.
- (2014): “Procyclical Leverage and Value-at-Risk,” *Review of Financial Studies*, 27(2), 373–403.
- AGHION, P., P. ASKENAZY, N. BERMAN, G. CETTE, AND L. EYMARD (2012): “Credit Constraints and the Cyclicity of R&D Investment: Evidence from France,” *Journal of the European Economic Association*, 10(5), 1001–1024.
- ALMEIDA, H., AND M. CAMPELLO (2007): “Financial Constraints, Asset Tangibility, and Corporate Investment,” *Review of Financial Studies*, 20(5), 1429–1460.
- ASMUSSEN, S., P. GLYNN, AND J. PITMAN (1995): “Discretization Error in Simulation of One-Dimensional Reflecting Brownian Motion,” *The Annals of Applied Probability*, 5(4), 875–896.
- BACHMANN, R., AND C. BAYER (2014): “Investment Dispersion and the Business Cycle,” *American Economic Review*, 104(4), 1392–1416.
- BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2015): “Measuring Economic Policy Uncertainty,” Working Paper 21633, National Bureau of Economic Research.
- BANSAL, R., AND W. J. COLEMAN (1996): “A Monetary Explanation of the Equity Premium, Term Premium, and Risk-Free Rate Puzzles,” *Journal of Political Economy*, 104(6), 1135–1171.
- BANSAL, R., W. J. COLEMAN, AND C. T. LUNDBLAD (2011): “Endogenous Liquidity Supply,” 2011 Meeting Papers 1200, Society for Economic Dynamics.
- BARLEVY, G. (2007): “On the Cyclicity of Research and Development,” *American Economic Review*, 97(4), 1131–1164.
- BARON, M. (2014): “Countercyclical Bank Equity Issuance,” Working paper, Princeton University.

- BASAK, S., AND D. CUOCO (1998): “An Equilibrium Model with Restricted Stock Market Participation,” *The Review of Financial Studies*, 11(2), 309–341.
- BASEL COMMITTEE ON BANK SUPERVISION (2013): “Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools,” Discussion paper, Bank for International Settlements.
- BATES, T. W., K. M. KAHLE, AND R. M. STULZ (2009): “Why Do U.S. Firms Hold So Much More Cash than They Used To?,” *The Journal of Finance*, 64(5), 1985–2021.
- BEGENAU, J., AND B. PALAZZO (2015): “Firm Selection and Corporate Cash Holdings,” 2015 Meeting Papers 1047, Society for Economic Dynamics.
- BENES, J., AND M. KUMHOF (2012): “The Chicago Plan Revisited,” Working paper WP/12/202, International Monetary Fund.
- BERNANKE, B. S. (1983): “Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression,” *The American Economic Review*, 73(3), 257–276.
- BERNANKE, B. S., AND M. GERTLER (1989): “Agency Costs, Net Worth, and Business Fluctuations,” *The American Economic Review*, 79(1), 14–31.
- BIANCHI, J., AND S. BIGIO (2014): “Banks, Liquidity Management and Monetary Policy,” Working Paper Series 20490, National Bureau of Economic Research.
- BOLTON, P., H. CHEN, AND N. WANG (2011): “A unified theory of Tobin’s q , corporate investment, financing, and risk management,” *The Journal of Finance*, 66(5), 1545–1578.
- (2013): “Market timing, investment, and risk management,” *Journal of Financial Economics*, 109(1), 40 – 62.
- BOLTON, P., AND X. FREIXAS (2000): “Equity, Bonds, and Bank Debt: Capital Structure and Financial Market Equilibrium under Asymmetric Information,” *Journal of Political Economy*, 108(2), 324–351.
- BOLTON, P., AND H. HUANG (2016): “The Capital Structure of Nations,” Working paper, Columbia University.
- BOLTON, P., AND O. JEANNE (2011): “Sovereign Default Risk in Financially Integrated Economies,” *IMF Economic Review*, 59(2), 162–194.
- BOROVÍČKA, J., L. P. HANSEN, AND J. A. SCHEINKMAN (2014): “Shock elasticities and impulse responses,” *Mathematics and Financial Economics*, 8(4), 333–354.
- BRUNNERMEIER, M. K., AND L. H. PEDERSEN (2009): “Market Liquidity and Funding Liquidity,” *Review of Financial Studies*, 22(6), 2201–2238.
- BRUNNERMEIER, M. K., AND Y. SANNIKOV (2014): “A Macroeconomic Model with a Financial Sector,” *American Economic Review*, 104(2), 379–421.

- (2016): “The I Theory of Money,” Working paper, Princeton University.
- CABALLERO, R. J., AND A. KRISHNAMURTHY (2006): “Bubbles and capital flow volatility: Causes and risk management,” *Journal of Monetary Economics*, 53(1), 35 – 53.
- CAVALCANTI, R. D. O., AND N. WALLACE (1999): “A Model of Private Bank-Note Issue,” *Review of Economic Dynamics*, 2(1), 104 – 136.
- CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (2005): “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, 113(1), 1–45.
- COCHRANE, J. H. (2011): “Presidential Address: Discount Rates,” *The Journal of Finance*, 66(4), 1047–1108.
- COCHRANE, J. H., AND M. PIAZZESI (2002): “The Fed and Interest Rates: A High-Frequency Identification,” Working Paper 8839, National Bureau of Economic Research.
- COMIN, D., AND M. GERTLER (2006): “Medium-Term Business Cycles,” *American Economic Review*, 96(3), 523–551.
- COOK, T., AND T. HAHN (1989): “The effect of changes in the federal funds rate target on market interest rates in the 1970s,” *Journal of Monetary Economics*, 24(3), 331 – 351.
- CORRADO, C. A., AND C. R. HULTEN (2010): “How Do You Measure a ”Technological Revolution“?,” *American Economic Review*, 100(2), 99–104.
- DANG, T. V., G. GORTON, B. HOLMSTRÖM, AND G. ORDONEZ (2014): “Banks as Secret Keepers,” Working Paper 20255, National Bureau of Economic Research.
- DANIELSSON, J., H. S. SHIN, AND J.-P. ZIGRAND (2012): “Procyclical Leverage and Endogenous Risk,” Working paper, London School of Economics.
- DEMARZO, P. M., AND D. DUFFIE (1999): “A Liquidity-based Model of Security Design,” *Econometrica*, 67(1), 65–99.
- DENIS, D. J., AND V. SIBILKOV (2010): “Financial Constraints, Investment, and the Value of Cash Holdings,” *Review of Financial Studies*, 23(1), 247–269.
- DI TELLA, S. (2015): “Uncertainty Shocks and Balance Sheet Recessions,” *Journal of Political Economy*, p. forthcoming.
- DI TELLA, S., AND P. KURLAT (2017): “Why are Banks Exposed to Monetary Policy?,” Working paper, Stanford University.
- DIAMOND, D. W., AND R. G. RAJAN (2000): “A Theory of Bank Capital,” *The Journal of Finance*, 55(6), 2431–2465.
- DITTMAR, A., AND A. THAKOR (2007): “Why Do Firms Issue Equity?,” *The Journal of Finance*, 62(1), 1–54.

- DONALDSON, J. R., G. PIACENTINO, AND A. THAKOR (2016): “Warehouse banking,” *Journal of Financial Economics*, p. forthcoming.
- DRECHSLER, I., A. SAVOV, AND P. SCHNABL (2016): “The Deposits Channel of Monetary Policy,” Working Paper Series 22152, National Bureau of Economic Research.
- (2017): “A Model of Monetary Policy and Risk Premia,” *The Journal of Finance*, p. forthcoming.
- DUCHIN, R. (2010): “Cash Holdings and Corporate Diversification,” *The Journal of Finance*, 65(3), 955–992.
- DUCHIN, R., T. GILBERT, J. HARFORD, AND C. M. HRDLICKA (2017): “Precautionary Savings with Risky Assets: When Cash Is Not Cash,” *The Journal of Finance*, 72(2), 793–852.
- EGGERTSSON, G. B., AND N. R. MEHROTRA (2014): “A Model of Secular Stagnation,” Working Paper 20574, National Bureau of Economic Research.
- EISFELDT, A., AND A. RAMPINI (2009): “Financing Shortfalls and the Value of Aggregate Liquidity,” Working paper, Duke University.
- EISFELDT, A. L. (2007): “Smoothing with liquid and illiquid assets,” *Journal of Monetary Economics*, 54(6), 1572 – 1586.
- EISFELDT, A. L., AND T. MUIR (2016): “Aggregate external financing and savings waves,” *Journal of Monetary Economics*, 84(Supplement C), 116 – 133.
- EISFELDT, A. L., AND A. A. RAMPINI (2006): “Capital reallocation and liquidity,” *Journal of Monetary Economics*, 53(3), 369 – 399.
- ELENEV, V., T. LANDVOIGT, AND S. V. NIEUWERBURGH (2017): “A Macroeconomic Model with Financially Constrained Producers and Intermediaries,” Working paper, New York University Stern School of Business.
- FABRIZIO, K. R., AND U. TSOLMON (2014): “An Empirical Examination of the Procyclicality of RD Investment and Innovation,” *The Review of Economics and Statistics*, 96(4), 662 – 675.
- FALATO, A., AND J. W. SIM (2014): “Why Do Innovative Firms Hold So Much Cash? Evidence from Changes in State R&D Tax Credits,” Finance and Economics Discussion Series 2014-72, Board of Governors of the Federal Reserve System (U.S.).
- FARHI, E., AND J. TIROLE (2012): “Bubbly Liquidity,” *Review of Economic Studies*, 79(2), 678–706.
- FATAS, A. (2000): “Do Business Cycles Cast Long Shadows? Short-Run Persistence and Economic Growth,” *Journal of Economic Growth*, 5(2), 147–162.

- FISCHER, S. (2009): “Preparing for future crises,” Speech at the 33rd Annual Symposium of the Federal Reserve Bank of Kansas City, on “Financial Stability and Macroeconomic Stability”, at Jackson Hole, Wyoming, 21 August 2009.
- FLEMING, M. J., AND K. D. GARBADE (2003): “The Repurchase Agreement Refined: GCF Repo,” *Current Issues in Economics and Finance*, 9(6).
- FRIEDMAN, M. (1969): *The Optimum Quantity of Money and Other Essays*. Aldine Aldine Transaction, Chicago, I.L.
- FRIEDMAN, M., AND A. J. SCHWARTZ (1963): *A Monetary History of the United States, 1867-1960*. Princeton University Press, Princeton, N.J.
- FROOT, K. A., D. S. SCHARFSTEIN, AND J. C. STEIN (1993): “Risk Management: Coordinating Corporate Investment and Financing Policies,” *The Journal of Finance*, 48(5), 1629–1658.
- FSB (FINANCIAL STABILITY BOARD) (2009): *Report of the Financial Stability Forum on Addressing Procyclicality in the Financial System*. Basel, Switzerland.
- GARBADE, K. D. (2007): “The Emergence of “Regular and Predictable” as a Treasury Debt Management Strategy,” *Economic Policy Review*, 13(1).
- GEANAKOPOLOS, J. (2010): *The Leverage Cycle* NBER Macroeconomics Annual 2009, Volume 24, pp.1-65. University of Chicago Press.
- GENNAIOLI, N., A. SHLEIFER, AND R. W. VISHNY (2013): “A Model of Shadow Banking,” *The Journal of Finance*, 68(4), 1331–1363.
- GERTLER, M., AND P. KARADI (2015): “Monetary Policy Surprises, Credit Costs, and Economic Activity,” *American Economic Journal: Macroeconomics*, 7(1), 44–76.
- GIGLIO, S., AND T. SEVERO (2012): “Intangible capital, relative asset shortages and bubbles,” *Journal of Monetary Economics*, 59(3), 303 – 317.
- GORTON, G. (2010): *Slapped by the Invisible Hand: The Panic of 2007*. Oxford University Press, New York, N.Y.
- GORTON, G., AND A. METRICK (2012): “Securitized banking and the run on repo,” *Journal of Financial Economics*, 104(3), 425 – 451.
- GORTON, G., AND G. PENNACCHI (1990): “Financial Intermediaries and Liquidity Creation,” *The Journal of Finance*, 45(1), 49–71.
- GRAHAM, J. R., AND M. T. LEARY (2015): “The Evolution of Corporate Cash,” Working paper, Duke University.
- GREENWOOD, R., S. G. HANSON, AND J. C. STEIN (2015): “A Comparative-Advantage Approach to Government Debt Maturity,” *The Journal of Finance*, 70(4), 1683–1722.

- GREENWOOD, R., AND D. VAYANOS (2014): “Bond Supply and Excess Bond Returns,” *Review of Financial Studies*, 27(3), 663–713.
- GRILICHES, Z. (1990): “Patent Statistics as Economic Indicators: A Survey,” Working Paper 3301, National Bureau of Economic Research.
- GURLEY, J. G., AND E. S. SHAW (1960): *Money in a Theory of Finance*. Brookings Institution, Washington, D.C.
- HALL, B. H. (1992): “Investment and Research and Development at the Firm Level: Does the Source of Financing Matter?,” Working Paper 4096, National Bureau of Economic Research.
- HALL, B. H., AND J. LERNER (2009): “The Financing of R&D and Innovation,” Working Paper 15325, National Bureau of Economic Research.
- HANSON, S. G., AND J. C. STEIN (2015): “Monetary policy and long-term real rates,” *Journal of Financial Economics*, 115(3), 429 – 448.
- HART, O., AND L. ZINGALES (2014): “Banks Are Where The Liquidity Is,” Working paper, Harvard University.
- HAYASHI, F. (1982): “Tobin’s Marginal q and Average q: A Neoclassical Interpretation,” *Econometrica*, 50(1), 213–224.
- HE, Z., B. KELLY, AND A. MANELA (2016): “Intermediary Asset Pricing: New Evidence from Many Asset Classes,” Working Paper 21920, National Bureau of Economic Research.
- HE, Z., I. G. KHANG, AND A. KRISHNAMURTHY (2010): “Balance Sheet Adjustments in the 2008 Crisis,” Working Paper 15919, National Bureau of Economic Research.
- HE, Z., AND P. KONDOR (2016): “Inefficient Investment Waves,” *Econometrica*, 84(2), 735–780.
- HE, Z., AND A. KRISHNAMURTHY (2012): “A Model of Capital and Crises,” *The Review of Economic Studies*, 79(2), 735–777.
- HE, Z., AND A. KRISHNAMURTHY (2013): “Intermediary Asset Pricing,” *American Economic Review*, 103(2), 732–70.
- HIMMELBERG, C. P., AND B. C. PETERSEN (1994): “R&D and Internal Finance: A Panel Study of Small Firms in High-Tech Industries,” *The Review of Economics and Statistics*, 76(1), 38–51.
- HOLMSTRÖM, B. (2012): “The nature of liquidity provision: When ignorance is bliss,” in *Presidential Address, Econometric Society, ASSA meetings, Chicago*.
- HOLMSTRÖM, B., AND J. TIROLE (1997): “Financial Intermediation, Loanable Funds, and the Real Sector,” *The Quarterly Journal of Economics*, 112(3), 663–691.

- (1998): “Private and Public Supply of Liquidity,” *Journal of Political Economy*, 106(1), 1–40.
- (2001): “LAPM: A Liquidity-Based Asset Pricing Model,” *The Journal of Finance*, 56(5), 1837–1867.
- (2011): *Inside and Outside Liquidity*. The MIT Press, Cambridge, M.A.
- HUANG, J. (2016): “Banking and Shadow Banking,” Working paper, National University of Singapore.
- KACPERCZYK, M., C. PÉRIGNON, AND G. VUILLEMEY (2017): “The Private Production of Safe Assets,” Working Paper FIN-2017-1212, HEC Paris Research Paper.
- KELLY, B., Ľ. PÁSTOR, AND P. VERONESI (2016): “The Price of Political Uncertainty: Theory and Evidence from the Option Market,” *The Journal of Finance*, p. forthcoming.
- KIYOTAKI, N., AND J. MOORE (1997): “Credit Cycles,” *Journal of Political Economy*, 105(2), 211–248.
- (2000): “Inside money and liquidity,” Working paper, London School of Economics.
- (2002): “Evil Is the Root of All Money,” *American Economic Review*, 92(2), 62–66.
- (2005): “Lawrence R. Klein Lecture: Liquidity and asset prices,” *International Economic Review*, 46(2), 317–349.
- KIYOTAKI, N., AND R. WRIGHT (1989): “On Money as a Medium of Exchange,” *Journal of Political Economy*, 97(4), 927–954.
- KLIMENKO, N., S. PFEIL, J.-C. ROCHET, AND G. D. NICOLO (2016): “Aggregate Bank Capital and Credit Dynamics,” Swiss Finance Institute Research Paper Series 16-42.
- KOCHERLAKOTA, N. R. (1998): “Money Is Memory,” *Journal of Economic Theory*, 81(2), 232 – 251.
- (2009): “Bursting Bubbles: Consequences and Cures,” Unpublished manuscript, Federal Reserve Bank of Minneapolis.
- KRISHNAMURTHY, A., S. NAGEL, AND D. ORLOV (2014): “Sizing Up Repo,” *The Journal of Finance*, 69(6), 2381–2417.
- KRISHNAMURTHY, A., AND A. VISSING-JØRGENSEN (2012): “The Aggregate Demand for Treasury Debt,” *Journal of Political Economy*, 120(2), 233–267.
- KRISHNAMURTHY, A., AND A. VISSING-JØRGENSEN (2015): “The impact of Treasury supply on financial sector lending and stability,” *Journal of Financial Economics*, 118(3), 571 – 600.
- KUTTNER, K. N. (2001): “Monetary policy surprises and interest rates: Evidence from the Fed funds futures market,” *Journal of Monetary Economics*, 47(3), 523 – 544.

- LAGOS, R. (2008): “inside and outside money,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.
- LI, S., T. M. WHITED, AND Y. WU (2016): “Collateral, Taxes, and Leverage,” *Review of Financial Studies*, 29(6), 1453–1500.
- LUCAS, R. E., AND N. L. STOKEY (1987): “Money and Interest in a Cash-in-Advance Economy,” *Econometrica*, 55(3), 491–513.
- MANKIW, N., AND S. P. ZELDES (1991): “The consumption of stockholders and nonstockholders,” *Journal of Financial Economics*, 29(1), 97 – 112.
- MODIGLIANI, F., AND M. H. MILLER (1958): “The Cost of Capital, Corporation Finance and the Theory of Investment,” *The American Economic Review*, 48(3), 261–297.
- MOREIRA, A., AND A. SAVOV (2014): “The Macroeconomics of Shadow Banking,” Working Paper 20335, National Bureau of Economic Research.
- MYERS, S. C., AND N. S. MAJLUF (1984): “Corporate financing and investment decisions when firms have information that investors do not have,” *Journal of Financial Economics*, 13(2), 187 – 221.
- NAGEL, S. (2016): “The Liquidity Premium of Near-Money Assets,” *The Quarterly Journal of Economics*, p. forthcoming.
- NAKAMURA, E., AND J. STEINSSON (2017): “High Frequency Identification of Monetary Non-Neutrality: The Information Effect,” *Quarterly Journal of Economics*, p. forthcoming.
- NEWAY, W. K., AND K. D. WEST (1994): “Automatic Lag Selection in Covariance Matrix Estimation,” *The Review of Economic Studies*, 61(4), 631–653.
- OSTROY, J. M., AND R. M. STARR (1990): “Chapter 1 The transactions role of money,” vol. 1 of *Handbook of Monetary Economics*, pp. 3 – 62. Elsevier.
- OUYANG, M. (2011): “On the Cyclicalities of R&D,” *The Review of Economics and Statistics*, 93(2), 542 – 553.
- PARLOUR, C. A., R. STANTON, AND J. WALDEN (2012): “Financial Flexibility, Bank Capital Flows, and Asset Prices,” *The Journal of Finance*, 67(5), 1685–1722.
- PATINKIN, D. (1965): *Money, Interest, and Prices: An Integration of Monetary and Value Theory*. Harper & Row, New York, N.Y., 2 edn.
- PHELAN, G. (2015): “Financial Intermediation, Leverage, and Macroeconomic Instability,” Working paper, Department of Economics, Williams College.
- PHILLIPS, R. J. (1996): *The ‘Chicago Plan’ and New Deal Banking Reform* Stability in the Financial System, pp. 94–114. Palgrave Macmillan UK.

- PIAZZESI, M., AND M. SCHNEIDER (2017): “Payments, Credit and Asset Prices,” Working paper, Stanford University.
- PINKOWITZ, L., R. M. STULZ, AND R. WILLIAMSON (2016): “Do U.S. Firms Hold More Cash than Foreign Firms Do?,” *Review of Financial Studies*, 29(2), 309–348.
- POZSAR, Z. (2011): “Institutional Cash Pools and the Triffin Dilemma of the U.S. Banking System,” Working paper WP/11/190, International Monetary Fund.
- (2014): “Shadow Banking: The Money View,” Working paper 14-04, Office of Financial Research.
- QUADRINI, V. (2017): “Bank liabilities channel,” *Journal of Monetary Economics*, 89, 25 – 44, Carnegie-Rochester-NYU Conference Series on the Macroeconomics of Liquidity in Capital Markets and the Corporate Sector.
- RAMPINI, A. A., AND S. VISWANATHAN (2010): “Collateral, Risk Management, and the Distribution of Debt Capacity,” *The Journal of Finance*, 65(6), 2293–2322.
- (2017): “Financial Intermediary Capital,” Working Paper 23302, National Bureau of Economic Research.
- SAINT-PAUL, G. (2005): “Fiscal Policy and Economic Growth: the Role of Financial Intermediation,” *Review of International Economics*, 13(3), 612–629.
- SCHIESSER, W. E., AND G. W. GRIFFITHS (2009): *A Compendium of Partial Differential Equation Models*. Cambridge University Press, New York, N.Y.
- STEIN, J. C. (2012): “Monetary Policy as Financial Stability Regulation,” *The Quarterly Journal of Economics*, 127(1), 57–95.
- SUMMERS, L. H. (2015): “Demand Side Secular Stagnation,” *American Economic Review*, 105(5), 60–65.
- SUNDERAM, A. (2015): “Money Creation and the Shadow Banking System,” *Review of Financial Studies*, 28(4), 939–977.
- SVENSSON, L. E. O. (1985): “Money and Asset Prices in a Cash-in-Advance Economy,” *Journal of Political Economy*, 93(5), 919–944.
- VAN DEN HEUVEL, S. J. (2002): “The Bank Capital Channel of Monetary Policy,” Working paper, The Wharton School of the University of Pennsylvania.
- VISSING-JØRGENSEN, A. (2002): “Limited Asset Market Participation and the Elasticity of Intertemporal Substitution,” *Journal of Political Economy*, 110(4), 825–853.
- WILLIAMSON, S., AND R. WRIGHT (2010): “Chapter 2 - New Monetarist Economics: Models,” vol. 3 of *Handbook of Monetary Economics*, pp. 25 – 96. Elsevier.

- WOODFORD, M. (1990a): “Chapter 20 The optimum quantity of money,” vol. 2 of *Handbook of Monetary Economics*, pp. 1067 – 1152. Elsevier.
- (1990b): “Public Debt as Private Liquidity,” *The American Economic Review*, 80(2), 382–388.
- (2010): “Financial Intermediation and Macroeconomic Analysis,” *Journal of Economic Perspectives*, 24(4), 21–44.
- (2016): “Quantitative Easing and Financial Stability,” Working Paper 22285, National Bureau of Economic Research.

Appendix I Proofs

I.1 Static Model

Firms' problem. Let $\underline{k}_0(s)$ denote the capital endowments of the representative firm s at $t = 0$, and $k_0(s)$ denote the representative firm's capital *demand*. The aggregate capital stock is $K_0 = \int_{s \in [0,1]} \underline{k}_0(s) ds$, and in equilibrium, the capital market clears, so we have $K_0 = \int_{s \in [0,1]} k_0(s) ds$. To save notations, I will suppress the firm index s going forward. Given q_0^K , the market price of capital, the representative firm's wealth is $w_0 = q_0^K \underline{k}_0$. At $t = 0$, the firm chooses capital holdings k_0 , deposits held per unit of capital m_0 , consumption c_0 , the total value of bank loan l_0 , and the funds raised from issuing securities to households h_0 . At $t = 1$, firm chooses investment rate i_1 .

Before laying out the firm's problem, we can simplify the analysis by noticing that raising funds at $t = 0$ by issuing securities to households is equivalent to negative consumption at $t = 0$. Let v_1^H denote the firm's *expected* payments to households at $t = 1$. By delivering an expected value of v_1^H at $t = 1$, the firm raises h_0 from households at $t = 0$. Given households' discount rate ρ (i.e., their required rate of return), with a competitive market for firms' securities, we have

$$v_1^H = h_0(1 + \rho).$$

Let v_0 denote the firm's value function. The firm's choices of consumption and financing from households can be solved by the following program:

$$v_0 = \max_{c_0 \geq 0, h_0 \geq 0} c_0 + \frac{1}{1 + \rho} (v_1^* - v_1^H),$$

where v_1^* is the *maximized expected* value of the firm before repaying households, which is a function of firms' other choices, such as capital holdings, deposit holdings, bank loan, and in-

vestment. Substituting $v_1^H = h_0(1 + \rho)$ into the value function, we have

$$v_0 = \max_{c_0 \geq 0, h_0 \geq 0} c_0 - h_0 + \frac{1}{1 + \rho} v_1^*.$$

Therefore, what matters is the net consumption ($c_0 - h_0$) or net financing ($h_0 - c_0$). The firm's problem has this property because entrepreneurs and households have the same discount rate ρ . Going forward, we allow c_0 to take positive or negative values, so when $c_0 > 0$, the entrepreneur consumes, and when $c_0 < 0$, the entrepreneur raises funds from households.⁶¹ We can write the firms' value function as

$$v_0 = \max_{c_0^B \in \mathbb{R}} c_0 + \frac{1}{1 + \rho} v_1^*.$$

The firm can also issue securities to banks. Let v_1^B denote the firm's *expected* repayment to banks at $t = 1$. Note that in the analysis of the firm's problem, we do not need to specify the contractual form of securities issued to banks. All that matters is the expected repayment. The same holds true for securities issued to households. Let v_1^{**} denote the *maximized expected* value of the firm before repaying both households and banks, so $v_1^{**} = v_1^* + v_1^B$ by definition. We can rewrite the value function as

$$v_0 = \max_{c_0^B \in \mathbb{R}} c_0 + \frac{1}{1 + \rho} (v_1^{**} - v_1^B).$$

Because bank debt earns a money premium, the required rate of return of banks can be different from that of households. Let ρ_0^B denote banks' required expected return in equilibrium. For the firm to raise l_0 from banks at $t = 0$, it must deliver an expected repayment to competitive banks equal to $v_1^B = l_0(1 + \rho_0^B)$. Thus, we can rewrite the value function as

$$v_0 = \max_{c_0^B \in \mathbb{R}} c_0 - \frac{1 + \rho_0^B}{1 + \rho} l_0 + \frac{1}{1 + \rho} v_1^{**}. \quad (18)$$

⁶¹The same logic applies to the dynamic analysis in continuous time.

Apparently, if $\rho_0^B < \rho$ (i.e., the cost of bank financing is lower than the cost of household financing), the firm only issue securities to banks; likewise, if $\rho < \rho_0^B$, the firm only issue securities to households if at all. Therefore, the firm's external financing cost is $\min\{\rho, \rho_0^B\}$. Moreover, since we study an equilibrium where firms do borrow from banks, it must be true that

$$\rho_0^B \leq \rho.$$

To determine the firm's financing capacity, we can calculate the firm's total pledgeable value at $t = 1$. Newly created capital is not pledgeable, and a random fraction $\delta - \sigma Z_1$ of existing capital will be gone by $t = 1$ (with $\mathbb{E}_0 [Z_1] = 0$), so the expected pledgeable value is

$$\underbrace{\alpha}_{\text{productivity}} \underbrace{k_0 (1 - \delta)}_{\text{expected surviving capital}}.$$

Therefore, the firm faces the following external financing constraint at $t = 0$:

$$l_0 (1 + \rho_0^B) + \mathbb{I}_{\{c_0 < 0\}} (-c_0) (1 + \rho) \leq \alpha k_0 (1 - \delta), \quad (19)$$

where $\mathbb{I}_{\{\cdot\}}$ is an indicator function. Note that when $-c_0 < 0$, the firm raises $|c_0|$ from households.

The firm also faces the budget constraint

$$c_0 + q_0^K k_0 + m_0 k_0 \leq w_0 + l_0, \text{ where } c_0 \in \mathbb{R}, \quad (20)$$

and the liquidity constraint

$$i_1 \leq m_0. \quad (21)$$

The firm maximizes its objective function in Equation (18) subject to constraints in Equation

(19), (20), and (21), with the expected total firm value at $t = 1$ equal to

$$v^{**} = \underbrace{\alpha k_0 (1 - \delta)}_{\text{expected surviving capital value}} + \underbrace{(1 + r_0) m_0 k_0}_{\text{deposits}} + \underbrace{\lambda (\alpha F(i_1) - i_1) k_0}_{\text{expected new capital net value}}, \quad (22)$$

and taking as given the market price of capital q_0^K , the interest rate on deposits r_0 , and the required expected return of banks ρ_0^B .

Let κ_0 denote the Lagrange multiplier of the financing constraint (Equation (19)), ψ_0 the Lagrange multiplier of the budget constraint, and θ_0 the Lagrange multiplier of the liquidity constraint. We can write down the Lagrange (omitting the non-negativity constraints):

$$\begin{aligned} v_0 = \max_{c_0^I \in \mathbb{R}, l_0 \geq 0, k_0 \geq 0, m_0 \geq 0, i_1 \geq 0} & c_0 - \frac{1 + \rho_0^B}{1 + \rho} l_0 + \frac{1}{1 + \rho} [\alpha k_0 (1 - \delta) + (1 + r_0) m_0 k_0 \\ & + \lambda (\alpha F(i_1) - i_1) k_0] \\ & + \kappa_0 [\alpha k_0 (1 - \delta) - l_0 (1 + \rho_0^B) - \mathbb{I}_{\{c_0 < 0\}} (-c_0) (1 + \rho)] \\ & + \psi_0 (w_0 + l_0 - c_0 - q_0^K k_0 - m_0 k_0) \\ & + \theta_0 (m_0 - i_1). \end{aligned}$$

Proof of Lemma 2. We can use the firm's optimization and other equilibrium conditions to tightly characterize the equilibrium. First, we solve ψ_0 . Notice that because c_0 can take either positive or negative values, we must have the coefficient of c_0 equal to zero

$$1 + \kappa_0 \mathbb{I}_{c_0 < 0} (1 + \rho) - \psi_0 = 0. \quad (23)$$

Because we are characterizing an equilibrium where banks lend out at least some of their goods endowments to firms, and because goods cannot be stored, in aggregate, entrepreneurs must consume, so $c_0 > 0$, and from Equation (23), $\psi_0 = 1$.

Next, we solve κ_0 , the shadow value of external financing. In equilibrium $l_0 > 0$ (i.e., not a corner solution), so locally the entrepreneur must be indifferent. Thus, the coefficient of l_0 is

equal to zero

$$-\frac{1 + \rho_0^B}{1 + \rho} - \kappa_0 (1 + \rho_0^B) + \psi_0 = 0. \quad (24)$$

Substituting $\psi_0 = 1$ into Equation (24), we have

$$\kappa_0 = \frac{1}{1 + \rho_0^B} - \frac{1}{1 + \rho}. \quad (25)$$

If the firm promises one unit of goods in expectation at $t = 1$, it can obtain $\frac{1}{1 + \rho_0^B}$ bank financing and $\frac{1}{1 + \rho}$ household financing at $t = 0$. Intuitively, κ_0 is exactly the difference between the price of securities issued to banks and the price of securities issued to households. And apparently, $\kappa_0 > 0$ if and only if $\rho_0^B < \rho$.

Multiplying both sides of Equation (25) by $(1 + \rho_0^B)(1 + \rho)$, we have

$$\kappa_0 (1 + \rho_0^B)(1 + \rho) = \rho - \rho_0^B$$

Expanding the left hand side we have,

$$\kappa_0 + \kappa_0 \rho_0^B + \kappa_0 \rho + \kappa_0 \rho_0^B \rho = \rho - \rho_0^B$$

The current time interval is 1, so if we let Δ denote an arbitrary length of time between date 0 and date 1, the product terms on the left-hand side are of the order of Δ^2 or higher. As Δ shrinks to zero, these terms approach to zero at a faster pace than Δ . Thus, to approximate the continuous-time expression, we ignore those product terms, and arrive at the expression in Lemma 2:

$$\kappa_0 = \rho - \rho_0^B = \rho - (R_0 - \delta). \quad (26)$$

Proof of Lemma 1. The firm's first order condition with respect to m_0 is the following

$$\frac{1}{1+\rho} (1+r_0) k_0 - \psi_0 k_0 + \theta_0 = 0, \quad (27)$$

and the first order condition with respect to i_1 is

$$\frac{1}{1+\rho} \lambda (\alpha F' (i_1) - 1) k_0 - \theta_0 = 0. \quad (28)$$

Summing up Equation (27) and (28), we have

$$\frac{1}{1+\rho} (1+r_0) k_0 - \psi_0 k_0 + \frac{1}{1+\rho} \lambda (\alpha F' (i_1) - 1) k_0 = 0 \quad (29)$$

We focus on the situation where the investment technology $F(\cdot)$ is so productive that the liquidity constraint always binds, so $i_1 = m_0$. Substituting $\psi_0 = 1$ and rearranging the equation, we have

$$r_0 - \rho + \lambda (\alpha F' (m_0) - 1) = 0. \quad (30)$$

Proof of Lemma 3. The expected return on bank equity is:

$$x_0 (1 + R_0 - E[\pi(Z_1)]) - (x_0 - 1)(1 + r_0) = 1 + r_0 + x_0 (R_0 - E[\pi(Z_1)] - r_0).$$

Note that $E[\pi_D(Z_1)] = \delta$. When $Z_1 = -1$, the realized return on bank equity is:

$$1 + r_0 + x_0 (R_0 - \delta - \sigma - r_0).$$

The representative bank starts with equity e_0 and takes the market loan rate and deposit rate as

given, so the value function can be written as:

$$v(e_0; R_0, r_0) = \max_{y_0 \geq 0, x_0 \geq 0} y_0 e_0 + \frac{e_0(1-y_0)}{(1+\rho)} \{1 + r_0 + x_0(R_0 - \delta - r_0) + \xi_0 [1 + r_0 + x_0(R_0 - \delta - \sigma - r_0)]\},$$

where y_0 is the banker's consumption-to-wealth ratio.

The first order condition for x_0 is:

$$R_0 - r_0 = \delta + \gamma_0^B \sigma, \tag{31}$$

where $\gamma_0^B = \frac{\xi_0}{1+\xi_0} \in [0, 1)$ because $\xi_0 \geq 0$. Rearranging the equation, we have γ_0^B equal to the Sharpe ratio of loans:

$$\gamma_0^B = \frac{R_0 - \delta - r_0}{\sigma}.$$

When $\gamma_0^B > 0$, the capital adequacy constraint binds. Substituting the F.O.C. for x_0 into the value function, we have:

$$v(e_0; R_0, r_0) = y_0 e_0 + q_0^B (e_0 - y_0 e_0),$$

where $q_0^B = \frac{(1+r_0)(1+\xi_0)}{(1+\rho)}$. The bank chooses $y_0 > 0$ only if $q_0^B \leq 1$.

Proof of Proposition 1. Proof is provided in the main text.

Proof of Corollary 1. Proof is provided in the main text.

I.2 Continuous-time Model

Firms' problem: proof of Lemma 1', Lemma 2', and Proposition 2. First, I need to prove firms' marginal value of equity is equal to one. Conjecture that a scalar diffusion process ζ_t

summarizes firms' equilibrium investment opportunity set.

$$d\zeta_t = \zeta_t \mu_t^\zeta dt + \zeta_t \sigma_t^\zeta dZ_t,$$

where μ_t^ζ and σ_t^ζ are scalar diffusion processes of the instantaneous expectation and standard deviation of $\frac{d\zeta_t}{\zeta_t}$ respectively. Let $V(w_t; \zeta_t)$ denote the value function. Following previous notations, w_t is the representative firm's wealth. Note that by the same logic in the analysis of firms' problem in the static model, firms' negative consumption is equivalent to raising funds from households.

The Hamilton-Jacobi-Bellman (HJB) equation is

$$\begin{aligned} \rho V(w_t; \zeta_t) = & \max_{dc_t \in \mathbb{R}, k_t \geq 0, m_t \geq 0, l_t \geq 0} \frac{dc_t}{dt} - \frac{\partial V}{\partial w_t} \frac{dc_t}{dt} + \frac{\partial V}{\partial \zeta_t} \zeta_t \mu_t^\zeta + \frac{\partial V}{\partial w_t} \mu_t^w w_t \\ & + \frac{1}{2} \frac{\partial^2 V}{\partial \zeta_t^2} (\sigma_t^\zeta \zeta_t)^2 + \frac{1}{2} \frac{\partial^2 V}{\partial w_t^2} (\sigma_t^w w_t)^2 + \frac{\partial^2 V}{\partial w_t \partial \zeta_t} (\sigma_t^\zeta \zeta_t) (\sigma_t^w w_t) \\ & + \lambda [V(\hat{w}_t; \zeta_t) - V(w_t; \zeta_t)], \end{aligned}$$

where μ_t^w , σ_t^w , and \hat{w}_t are defined by the following dynamics of firm equity

$$dw_t = -dc_t + \mu_t^w w_t dt + \sigma_t^w w_t dZ_t + (\hat{w}_t - w_t) dN_t,$$

where dN_t is the increment of the idiosyncratic counting (Poisson) process ($dN_t = 1$ when the investment opportunity arrives), and after the Poisson shock, firm equity jumps to

$$\hat{w}_t = w_t + q_t^K F(m_t) k_t - m_t k_t.$$

Note that firms' marginal value of wealth, ζ_t , is a summary statistic of firms' investment opportunity set, which depends on the overall industry dynamics, so it does not jump when an individual firm is hit by investment opportunities.

Conjecture the value function is linear in equity, $V(w_t; \zeta_t) = \zeta_t w_t$. The HJB equation can be simplified as:

$$\rho V(w_t; \zeta_t) = \max_{dc_t \in \mathbb{R}, k_t \geq 0, m_t \geq 0, l_t \geq 0} \frac{dc_t}{dt} - \zeta_t \frac{dc_t}{dt} + w_t \zeta_t \mu_t^\zeta + w_t \zeta_t \mu_t^w + \sigma_t^\zeta \zeta_t \sigma_t^w w_t + \lambda \zeta_t [\widehat{w}_t - w_t].$$

Firms can choose any $dc_t \in \mathbb{R}$, so ζ_t must be equal to one, and thus, I have also confirmed the value function conjecture.

Next, I solve the optimality condition for deposits holdings in Lemma 1', and the capital pricing formula in Proposition 2. Since ζ_t is a constant equal to one, μ_t^ζ and σ_t^ζ are both zero. Therefore, the HJB equation can be further simplified:

$$\rho V(w_t; \zeta_t) = \max_{k_t \geq 0, m_t \geq 0, l_t \geq 0} \mu_t^w w_t + \lambda [q_t^K F(m_t) - m_t] k_t. \quad (32)$$

Because $V(w_t; \zeta_t) = w_t$, the maximized expected growth rate of wealth is equal to ρw_t . The expected changed of wealth after consumption contains the production flow from capital, the change of capital value (from both price change and potential destruction), the return on bank deposits, the expected loan repayment, and the net gain from potential investment opportunities:

$$\underbrace{\alpha k_t dt + \mathbb{E}_t (q_{t+dt}^K k_{t+dt} - q_t^K k_t)}_{\mu_t^w w_t} + r_t m_t k_t dt - l_t (R_t - \delta) dt + \lambda dt [q_t^K F(m_t) - m_t] k_t.$$

The firm sets its balance sheet to maximize the expected growth of wealth under the budget constraint:

$$\begin{aligned} \max_{k_t \geq 0, m_t \geq 0, l_t \geq 0} & \alpha k_t dt + \mathbb{E}_t (q_{t+dt}^K k_{t+dt} - q_t^K k_t) + r_t m_t k_t dt - l_t (R_t - \delta) dt \\ & + \lambda dt [q_t^K F(m_t) - m_t] k_t + (w_t + l_t - q_t^K k_t - m_t k_t) d\psi_t, \end{aligned}$$

where $d\psi_t$ is the Lagrange multiplier of the budget constraint, $q_t^K k_t + m_t k_t \leq w_t + l_t$.

The optimal deposit holdings per unit of capital are given by the first-order condition (F.O.C.):

$$m_t \{r_t dt + \lambda dt [q_t^K F'(m_t) - 1] - d\psi_t\} = 0, \text{ and } m_t \geq 0.$$

A fraction $(\delta dt - \sigma dZ_t)$ of capital is to be destroyed, so the capital evolves as

$$k_{t+dt} = k_t - (\delta dt - \sigma dZ_t) k_t.$$

Combining with the equilibrium price dynamics, we have, under Itô calculus,

$$q_{t+dt}^K k_{t+dt} - q_t^K k_t = q_t^K k_t [-(\delta dt - \sigma dZ_t) + \mu_t^K dt + \sigma_t^K dZ_t + \sigma \sigma_t^K dt].$$

The optimal capital holdings is given by the first-order condition (F.O.C.): $k_t \geq 0$, and

$$k_t \{ \alpha dt + q_t^K (-\delta + \mu_t^K + \sigma \sigma_t^K) dt + r_t m_t dt + \lambda dt [q_t^K F'(m_t) - m_t] - (q_t^K + m_t) d\psi_t \} = 0.$$

The optimal borrowing from banks is given by the following F.O.C.:

$$-(R_t - \delta) dt + d\psi_t = 0.$$

Finally, we have the complementary slackness condition: $d\psi_t \geq 0$, and

$$(w_t + l_t - q_t^K k_t - m_t k_t) d\psi_t = 0.$$

Substituting these optimality conditions into the objective function, we have

$$\max_{k_t \geq 0, m_t \geq 0, l_t \geq 0} \mu_t^w w_t dt + \lambda dt [q_t^K F'(m_t) - m_t] k_t = w_t d\psi_t.$$

From Equation (32), the left hand side is equal to $\rho w_t dt$, so the Lagrange multiplier $d\psi_t = \rho dt$.

Substituting $d\psi_t = \rho dt$ into the F.O.C. for m_t , we have

$$r_t + \lambda [q_t^K F'(m_t) - 1] = \rho.$$

Substituting $d\psi_t = \rho dt$ into the F.O.C. for k_t and rearranging the equation, we have

$$q_t^K = \frac{\alpha - (\rho - r_t) m_t + \lambda [q_t^K F'(m_t) - m_t]}{\rho - (\mu_t^K - \delta + \sigma \sigma_t^K)}.$$

Substituting $d\psi_t = \rho dt$ into the F.O.C. for l_t , we have

$$R_t = \rho + \delta.$$

Banks' problem: proof of Lemma 3'. Conjecture that a scalar diffusion process q_t^B summarizes banks' equilibrium investment opportunity set:

$$dq_t^B = q_t^B \mu_t^B dt + q_t^B \sigma_t^B dZ_t.$$

Following the notations in the main text, the representative bank's value function is $V(e_t; q_t^B)$, and the HJB equation is

$$\begin{aligned} \rho v(e_t; q_t^B) = & \max_{dy_t \in \mathbb{R}, x_t \geq 0} \left(1 - \frac{\partial v}{\partial e_t} \right) \mathbb{I}_{\{dy_t > 0\}} e_t dy_t + \left(\frac{\partial v}{\partial e_t} - 1 - \chi \right) \mathbb{I}_{\{dy_t < 0\}} e_t (-dy_t) + \\ & \frac{\partial v}{\partial q_t^B} q_t^B \mu_t^B + \frac{1}{2} \frac{\partial^2 v}{\partial (q_t^B)^2} (q_t^B \sigma_t^B)^2 + \frac{\partial V}{\partial e_t} \mu_t^e e_t + \frac{1}{2} \frac{\partial^2 V}{\partial (e_t)^2} (\sigma_t^e e_t)^2 + \frac{\partial^2 V}{\partial e_t \partial q_t^B} (q_t^B \sigma_t^B) (\sigma_t^e e_t). \end{aligned}$$

From the bank's budget constraint, Equation (9), we have:

$$\mu_t^e = r_t + x_t (R_t - \delta - r_t), \quad \text{and} \quad \sigma_t^e = x_t \sigma.$$

Conjecture that the bank's value function takes the linear form: $v(e_t; q_t^B) = q_t^B e_t$. Substituting

this conjecture into the bank's HJB equation and dividing both sides by $q_t^B e_t$, we have

$$\rho = \max_{dy_t \in \mathbb{R}} \left\{ \frac{(1 - q_t^B)}{q_t^B} \mathbb{I}_{\{dy_t > 0\}} dy_t + \frac{(q_t^B - 1 - \chi)}{q_t^B} \mathbb{I}_{\{dy_t < 0\}} (-dy_t) \right\} + \mu_t^B + \max_{x_t \geq 0} \{ r_t + x_t (R_t - \delta - r_t) - x_t \gamma_t^B \sigma \} - \iota,$$

where $\gamma_t^B = -\sigma_t^B$.

q_t^B is the marginal value of equity. Paying out one dollar of dividend, the bank's shareholders receive 1, but lose q_t^B . Only when $q_t^B \leq 1$, $dy_t > 0$. When the bank issues equity, it incurs a dilution cost. From the existing shareholders' perspective, one dollar equity is sold to outside investors at price $\frac{q_t^B}{1+\chi}$. To raise $(-dy_t) e_t$ that is worth $q_t^B (-dy_t) e_t$, the bank must issue $\frac{(1+\chi)(-dy_t)e_t}{q_t^B}$ shares, and thus, the existing shareholders give up total value of $q_t^B \frac{(1+\chi)(-dy_t)e_t}{q_t^B} = (1 + \chi) (-dy_t) e_t$. Therefore, the bank raises equity only if $q_t^B \geq 1 + \chi$.

Finally, the indifference condition for x_t is $R_t - \delta - r_t = \gamma_t \sigma$. If $R_t - \delta - r_t < \gamma_t \sigma$, the bank sets $x_t = 0$. If $R_t - \delta - r_t > \gamma_t \sigma$, the bank sets x_t to infinity.

Proof of Proposition 1'. Proof is provided in the main text.

Proof of Corollary 1'. Proof is provided in the main text.

Proof of Lemma 4. First, I derive equation (10). Because individual banks share the same μ_t^e , σ_t^e , and payout/issuance rate dy_t , aggregating over banks, the law of motion of E_t is

$$dE_t = \mu_t^e E_t dt + \sigma_t^e E_t dZ_t - dy_t E_t.$$

Given the expected growth rate, $\lambda F(m_t) - \delta$, which is the investment net of expected deprecia-

tion, the aggregate capital stock, K_t , evolves as

$$dK_t = [\lambda F(m_t) - \delta] K_t dt + \sigma K_t dZ_t,$$

where the diffusion term comes from the stochastic fraction of capital destroyed.

By Itô's lemma, the ratio, $\eta_t = \frac{E_t}{K_t}$, has the following law of motion:

$$d\eta_t = \frac{1}{K_t} dE_t - \frac{E_t}{K_t^2} dK_t + \frac{1}{K_t^3} \langle dK_t, dK_t \rangle - \frac{1}{K_t^2} \langle dE_t, dK_t \rangle,$$

where $\langle dX_t, dY_t \rangle$ denotes the quadratic covariation of diffusion processes X_t and Y_t , so we have $\langle dK_t, dK_t \rangle = \sigma^2 K_t^2 dt$, and $\langle dE_t, dK_t \rangle = \sigma_t^e \sigma E_t K_t dt$. Dividing both sides by η_t , we have

$$\frac{d\eta_t}{\eta_t} = \frac{dE_t}{E_t} - \frac{dK_t}{K_t} + \sigma^2 dt - \sigma_t^e \sigma dt.$$

Substituting the law of motions of E_t and K_t , we have Equation (10). The boundaries are given by banks' optimal payout and issuance policies in Proposition 3'.

Proof of Proposition 3. The proof follows Brunnermeier and Sannikov (2014). First, I derive the stationary probability density function. The Kolmogorov forward equation that characterizes $p(\eta, t)$, the probability density of η_t at time t , is:

$$\frac{\partial}{\partial t} p(\eta, t) = -\frac{\partial}{\partial \eta} (\eta \mu^\eta(\eta) p(\eta, t)) + \frac{1}{2} \frac{\partial^2}{\partial \eta^2} (\eta^2 \sigma^\eta(\eta)^2 p(\eta, t)).$$

Note that in a Markov equilibrium, μ_t^η and σ_t^η are functions of η_t .

A stationary density is a function that solves the forward equation and does not change with time (i.e. $\frac{\partial}{\partial t} p(\eta, t) = 0$). So I suppress the time variable, and denote stationary density as $p(\eta)$. Integrating the forward equation over η , $p(\eta)$ solves the following first-order ordinary

differential equation within the two reflecting boundaries:

$$0 = C - \eta \mu^\eta (\eta) p (\eta) + \frac{1}{2} \frac{d}{d\eta} (\eta^2 \sigma^\eta (\eta)^2 p (\eta)), \quad \eta \in [\underline{\eta}, \bar{\eta}].$$

Note that the integration constant C is zero because of the reflecting boundaries. The boundary condition is the requirement that the probability density is integrated to one (i.e. $\int_{\underline{\eta}}^{\bar{\eta}} p (\eta) d\eta = 1$).

Next, I solve the expected time to reach from $\underline{\eta}$. Define $f_{\eta_0} (\eta)$ the expected amount of time it takes to reach a point η_0 starting from $\eta \leq \eta_0$. Define $g (\eta_0) = f_{\eta_0} (\underline{\eta})$ the expected time to reach η_0 from $\underline{\eta}$. One has to reach $\eta \in (\underline{\eta}, \eta_0)$ first and then reach η_0 from η . Therefore, $g (\eta) + f_{\eta_0} (\eta) = g (\eta_0)$. Since $g (\eta_0)$ is a constant, we can differentiate both sides to have $g' (\eta) = -f'_{\eta_0} (\eta)$ and $g'' (\eta) = -f''_{\eta_0} (\eta)$.

For any η_t , $f_{\eta_0} (\eta_t)$, the expected time to reach η_0 can be decomposed into $s - t$, and $E_t [f_{\eta_0} (\eta_s)]$, the expected time to reach η_0 from η_s ($s \geq t$) after $s - t$ amount of time has passed. We get:

$$f_{\eta_0} (\eta_t) = E_t [f_{\eta_0} (\eta_s)] + s - t.$$

Therefore, $t + f_{\eta_0} (\eta_t)$ is a martingale, so f_{η_0} satisfies the ordinary differential equation (i.e. zero drift)

$$1 + f'_{\eta_0} (\eta) \mu^\eta (\eta) + \frac{\sigma^\eta (\eta)^2}{2} f''_{\eta_0} (\eta) = 0.$$

Therefore, $g (\eta)$ must satisfy

$$1 - g' (\eta) \mu^\eta (\eta) - \frac{\sigma^\eta (\eta)^2}{2} g'' (\eta) = 0.$$

It takes no time to reach $\underline{\eta}$, so $g (\underline{\eta}) = 0$. Moreover, since $\underline{\eta}$ is a reflecting boundary, $g' (\underline{\eta}) = 0$.

Proof of Proposition 4. Because the Markov equilibrium is time-homogeneous, I suppress the time subscripts. In the main text, I have shown how to solve x , m , and r as functions of

$(q^K(\eta), q^B(\eta))$ and their derivatives. Once we know these variables, we solve the dynamics of E_t , and in particular,

$$\mu^e = r + x^B (R - \delta - r), \text{ and } \sigma^e = x^B \sigma,$$

and the economic growth rate, $\lambda F'(m_t) - \delta$. So, we have the drift and diffusion of η_t :

$$\mu^\eta = \mu^e - [\lambda F'(m_t) - \delta] - \sigma^e \sigma + \sigma^2, \quad \text{and} \quad \sigma^\eta = (x - 1) \sigma.$$

Next, we can use bankers' HJB equation and the capital pricing formula (i.e. Equations (12) and (16)) to form a system of differential equations for $(q^K(\eta), q^B(\eta))$, i.e., a mapping from $(\eta, q^B, q^K, \frac{dq^B}{d\eta}, \frac{dq^K}{d\eta})$ to $(\frac{d^2 q^B}{d\eta^2}, \frac{d^2 q^K}{d\eta^2})$. In stead of the first derivatives, we can work with elasticities of (q^B, q^K) , $\epsilon^X = \frac{dq^X/q^X}{d\eta/\eta}$, $X \in \{B, K\}$ to simplify the expressions. Using Itô's lemma, we know

$$\mu^X = \epsilon^X \mu^\eta + \frac{1}{2q^X} (\sigma^\eta \eta)^2 \frac{d^2 q^X}{d\eta^2}, \text{ i.e., } \frac{d^2 q^X}{d\eta^2} = 2q^X \frac{(\mu^X - \epsilon^X \mu^\eta)}{(\sigma^\eta \eta)^2}, X \in \{B, K\}.$$

To calculate μ^K and μ^B , we use banks' HJB equation (Equation (12)):

$$\mu^B = \rho + \iota - r,$$

and the capital pricing formula (Equation (16)),

$$\mu^K = \rho + \delta - \sigma^K \sigma - \frac{\alpha}{q^K} + (\rho - r) m - \lambda \left[F(m) - \frac{m}{q^K} \right], \text{ where } \sigma^K = \epsilon^K \sigma^\eta.$$

As long as the polynomial equation given by the deposit market clearing condition (i.e. Equation (15)) has a unique solution of x in $[0, \infty)$, the mapping from $(\eta, q^B, q^K, \frac{dq^B}{d\eta}, \frac{dq^K}{d\eta})$ to $(\frac{d^2 q^B}{d\eta^2}, \frac{d^2 q^K}{d\eta^2})$ is unique. Given the boundary conditions explained in the main text, the system of differential equations uniquely pins down a solution $(q^B(\eta), q^K(\eta))$ under proper parameter

range, so we have a unique Markov equilibrium with state variable η_t .

To solve the problem with government debt, we note that the only change is firms' demand for bank debt. The whole procedure carries through.

Appendix II Calibration and Additional Results

II.1 Calibration

Table A.1: Calibration.

This table summarizes the parameter values of the solution and corresponding model and data moments (including the sources and sample size) used in calibration. Model moments are calculated using the stationary distribution of the model solution.

	Parameters		Model Moments	Data		Source
(1)	ρ	4.00%	Interest rate $\mathbb{E}[r_t]$	3.78%	3.77%	Average of MZM and T-bill rate FRED (1974Q1-2015Q4)
(2)	Inv. tech.	$F(i) = \omega_0 i^{\omega_1}$				
	ω_0	0.801	Expected capital growth	0.74%	0.74%	Corrado and Hulten (2010)
	ω_1	0.99	Cash to net assets $\mathbb{E}[m_t]$	29.3%	29.2%	Compustat (1971-2015)
(3)	λ	1/5	Expected liquidity premium $\mathbb{E}[\rho - r_t]$	24.59bp	23.65bp	GC Repo/T-bill spread Nagel (2016)
(5)	χ	1	Liquidity premium s.d. $\text{std}[\rho - r_t]$	12.35bp	18.19bp	GC Repo/T-bill spread Nagel (2016)
(6)	α	0.1	Firms' equity P/E ratio	25.6	24.9	S&P 500 (Jan1991-Dec2015)
(7)	ι	2.55%	Operation cost / bank total income $\mathbb{E}\left[\frac{\iota e_t}{(R_t - \delta - r_t)x_t e_t}\right]$	90.3%	91.4%	Opex / (Opex + Net Income) Call Reports (1976Q1-2015Q4)
(8)	δ	4.00%			3.67%	Average Loan delinquency rate FRED (1985Q1-2015Q4)
(9)	σ	2.00%			1.62%	Loan delinquency rate s.d. FRED (1985Q1-2015Q4)

Table A.1 summarizes the calibration. All model moments are calculated using the stationary distribution. ρ is set to 4%, so the mean of r_t matches the average yield of MZM (money

of zero maturity) and three-month Treasury bills. The investment technology is $F(i) = \omega_0 i^{\omega_1}$. Since in reality, the economic growth is driven by both cash-intensive investments and investments that can largely rely on external financing, such as those in construction and manufacturing industries, I set ω_0 to 0.801, so the mean growth rate matches the U.S. economic growth from *intangible investment* (Corrado and Hulten (2010)). $\omega_1 = 0.99$, so the mean of firms' money holdings matches the mean of Compustat firms.⁶² I use the mean and standard deviation of money premium to calibrate λ , the arrival rate of liquidity shocks, and χ , the dilution cost that governs the tail behavior of the economy. The money premium data is the GC repo/T-bill spread in Nagel (2016). Acknowledging that in reality, the money premium varies due to forces beyond the model mechanism, I set χ to a conservative value that generates a standard deviation of money premium that is two thirds of the data standard deviation. α is set to 0.1, so the mean of price-to-earnings ratio of firms' equity matches that of S&P 500 (1990-2015). ι is set to 2.55%, because the empirical counterpart of operation cost is the total operation expenses in Call reports. δ and σ are calibrated using the time series mean and standard deviation of loan delinquency rate.

II.2 Money Premium Volatility and Government Debt

Figure 10 helps explain why government debt amplifies the leverage cycle. It shows the standard deviation of money premium declines when government debt rises. Given that q_t^B , a forward-looking measure of return on equity, always moves in a fixed interval $[1, 1 + \chi]$, and given that return on equity is a product of the money premium and leverage, bank leverage must vary in a wider range to sustain the variation of q_t^B in response to increases in government debt supply.

⁶²I choose Compustat data from 1971 to 2015, because before 1971, money market funds hadn't developed, and thus, under Regulation Q, firms' holdings of money or money-like securities (e.g., deposits) are not interest-paying, which is different from the model setup.

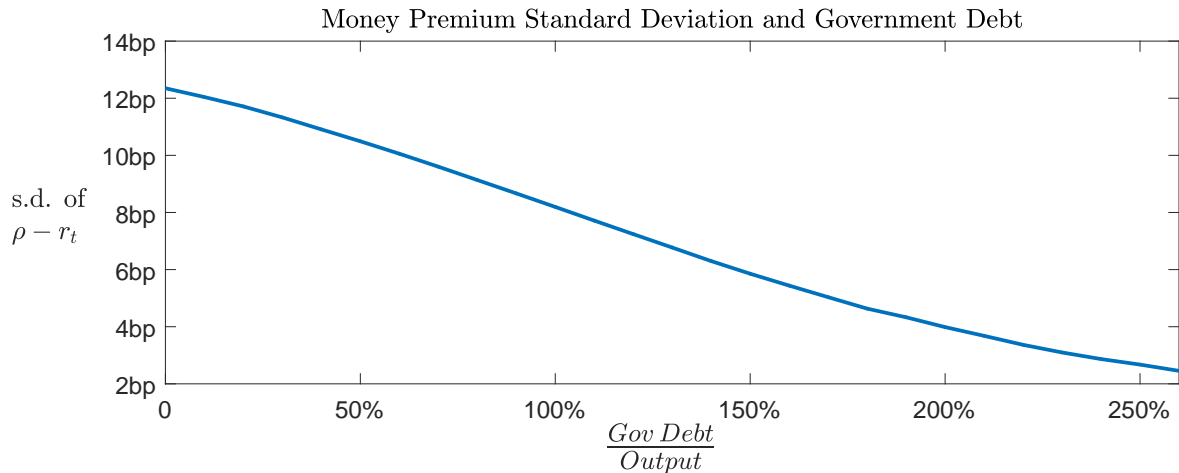


Figure 10: **Money Premium Standard Deviation and Government Debt.** This figure plots the standard deviation (of the stationary distribution) of money premium $\rho - r_t$ against different levels of government debt-to-output ratio.

Appendix III Preliminary Evidence

III.1 The Structure of Inside Money

Figure 11 shows that the model setup and its mechanism are quantitatively important. The left panel shows that to nonfinancial firms, financial intermediaries are the most important suppliers of money-like securities. The right panel shows that nonfinancial firms are among the most important buyers of intermediaries' money-like liabilities. Money-like assets include various financial instruments. Accordingly, “deposits” in the model should be interpreted broadly, including short-term debt issued by financial intermediaries that either serves as a means of payment, such as deposits at commercial banks, or close substitutes (usually held through money market funds by corporate treasuries), such as repurchase agreements and high-quality asset-backed commercial papers.

The left panel of Figure 11 decomposes the money-like assets of U.S. non-financial corporations by the types of securities, and for each security, by the types of its issuers. I use the data in December 2015 from the March 10, 2016 release of the Financial Accounts of the United States (previously known as the “Flow of Funds”). From this graph, we can understand who are supplying money to the U.S. nonfinancial corporations and in what forms. Foreign deposits and

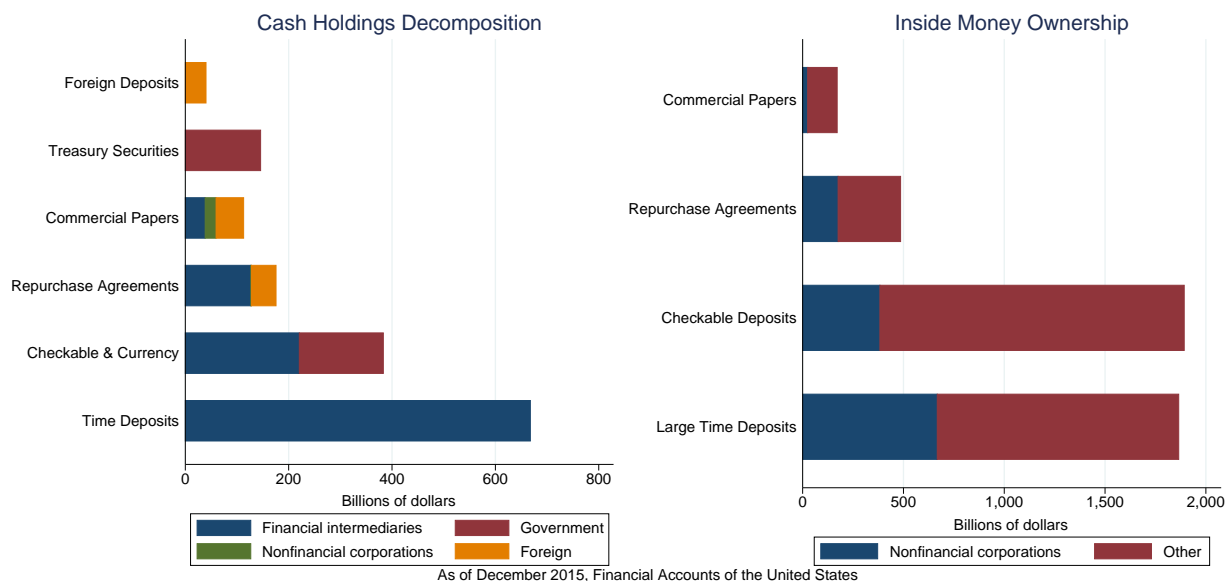


Figure 11: Inside Money Demand and Supply. This figure plots the issuers (by value) of different components of liquidity holdings of nonfinancial businesses (Panel A) and the holders (by value) of each type of money-like liabilities issued by the financial intermediation sector as of December 2015 (source: the Financial Accounts of the United States). The financial intermediation sector includes U.S.-chartered depository institutions, foreign banking offices in U.S., banks in U.S. affiliated areas, credit unions, issuers of asset-backed securities, finance companies, mortgage real estate investment trusts, security brokers and dealers, holding companies, and funding corporations.

time deposits are issued by depository institutions, and Treasury securities by the government. 19% of commercial papers are issued by the nonfinancial corporations themselves, 34% by domestic financial intermediaries, and 47% by foreign entities, of which 90% are issued by “foreign financial firms” (defined in the Financial Accounts). 72% of repurchase agreements are issued by the financial sector, and 27% by the foreign entities. Checkable deposits and currency are reported together in the Financial Accounts, of which 42% issued by the government are “currency outside banks”, and the remaining 58% are the liabilities of depository institutions. Given that firms usually do not hold currency directly, 58% underestimates the contribution of financial intermediaries.

The right panel of Figure 11 decomposes the outstanding money-like securities issued by financial intermediaries by the types of owners. Mainly through money market funds, nonfinancial corporations hold a little less than a third of outstanding repurchase agreements and some commercial papers that account for a smaller fraction of the outstanding amount. They hold a

significant share of checkable deposits and large time deposits.

Here are some details on how to construct this graph. Table L.103 records nonfinancial corporations' assets and liabilities. I break down the holdings of money market fund shares and mutual fund shares into financial instruments using Table L.121 and L.122 respectively, under the assumption that funds held by nonfinancial firms invest in the same portfolio as the aggregate sector does.⁶³ Corporate and foreign bonds, loans, and miscellaneous assets, all held indirectly through money market mutual funds and mutual funds, are excluded. Agency- and GSE-backed securities are excluded because of the potential spikes of repo haircuts and the secondary market illiquidity during crisis times. Municipal securities are excluded because of secondary market illiquidity.

Next, for each financial instrument, I calculate the net supply by each type of issuers using the instrument-level tables. Financial intermediaries are defined as in Krishnamurthy and Vissing-Jørgensen (2015), including the following financial institutions: U.S.-chartered depository institutions, foreign banking offices in U.S., banks in U.S. affiliated areas, credit unions, issuers of asset-backed securities, finance companies, mortgage real estate investment trusts, security brokers and dealers, holding companies, and funding corporations. Insurance companies are not included as financial intermediaries because their liabilities are long-term and usually held until maturity by insurance policy holders instead of resold in secondary market. Their liabilities are not money-like.

III.2 Cyclical money creation

The model predicts procyclical quantity of money and countercyclical price (i.e., the money premium). The banking sector expands its supply of inside money in goods times, allowing firms to better hedge against their liquidity shocks. In the model, all quantity variables are proportional to K_t , the capital stock, and thus, the output αK_t , so I scale all quantities variables in data by

⁶³It is implicitly assumed that funds are just pass-through and do not provide any liquidity services beyond the securities they hold. This ignores the potential sharing of idiosyncratic liquidity shocks through funds.

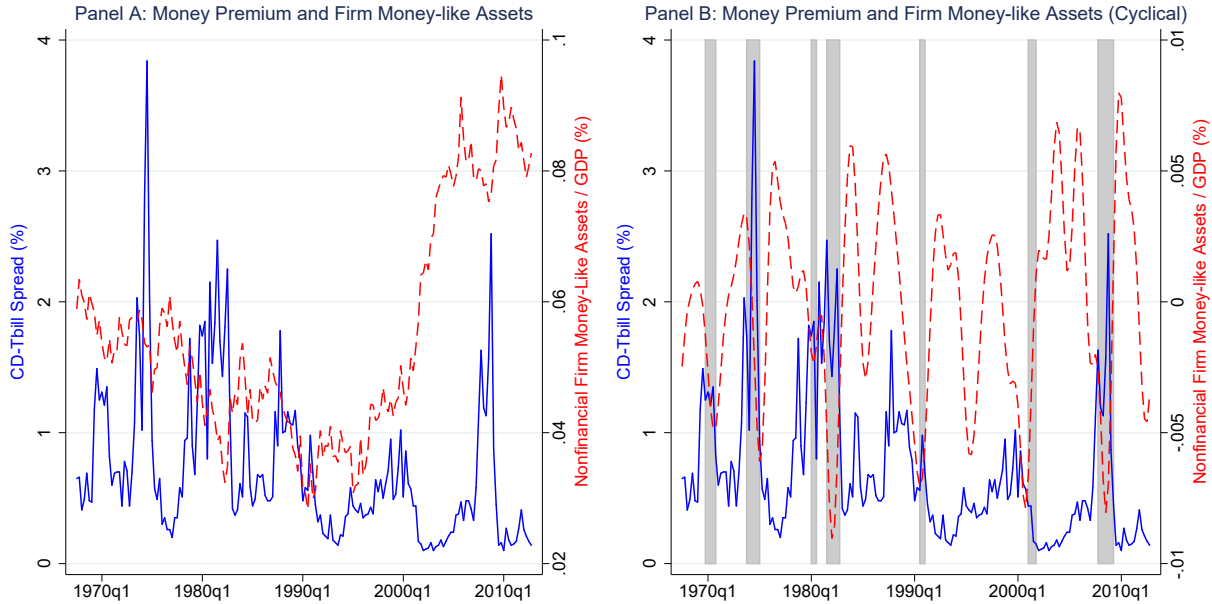


Figure 12: **Money Premium and Nonfinancial Firms' Holdings of Money-like Securities.** Panel A plots money-like assets held by U.S. nonfinancial business (defined in Appendix III.1) and the money premium (solid line), measured by the spread between the three-month Certificate of Deposit rate and three-month Treasury Bill rate. Panel B plots the cyclical components of these two variables. I use the Baxter-King filter that removes frequencies longer than 69 months, the average length of US business cycle after World War II. Plots start from 1967Q3 and ends in 2012Q4.

the nominal GDP. Figure 12 plots the quarterly data of money premium (the solid line) and the U.S. nonfinancial firms' money-like assets (defined in Appendix III.1). Following Nagel (2016), I define money premium as the spread between the three-month Certificate of Deposit rate and three-month Treasury Bill rate.⁶⁴ Panel B plots the cyclical component of firms' money holdings applying the Baxter-King filter.⁶⁵ NBER recession periods are marked by gray shades.

Figure 12 shows that firms' money holdings tend to be procyclical, and the money premium tends to be countercyclical, consistent with the equilibrium behavior of the model. The correlation between the cyclical component of firms' money holdings and the recession dummy is -47.3% , and the correlation between the money premium and the recession dummy is 53.4% .

⁶⁴In Nagel (2016), the preferred measure of money premium is the spread between general collateral repo rate (GC) and Treasury bill rate. However, repo rates are available only since 1991. Nagel (2016) argues that outside the crisis periods (savings and loans crisis of 1980s; the financial crisis of 2007-09), the credit risk of CD component is small, and shows that when both CD rates and GC repo rates are available, there is only a small difference between the two.

⁶⁵The filter remove frequencies longer than 69 months, the average length of NBER business cycle in 1945–2009.

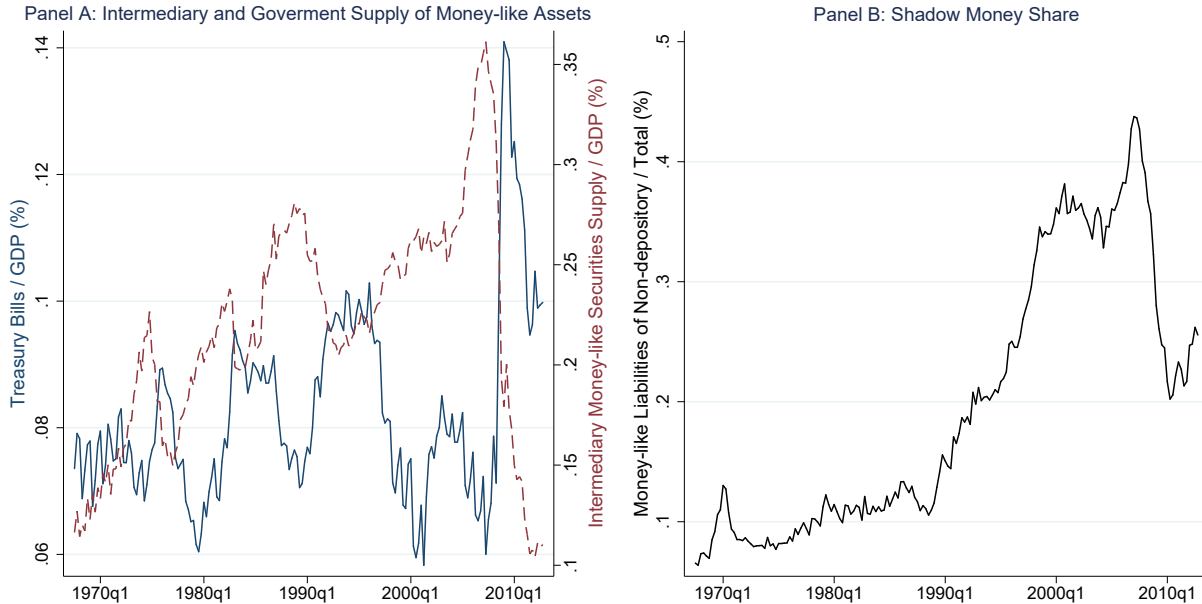


Figure 13: **Money-like Securities Issued by Financial Intermediaries and Government.** Panel A plots the outstanding Treasury bills (solid line) scaled by nominal GDP and the money-like liabilities of financial intermediaries scaled by GDP (defined in Figure 11 in Appendix III.1). Panel B plots the share of inside money, i.e., total money-like liabilities of financial intermediaries, that is from the non-depository intermediaries (“shadow banks”). The data source is the Financial Accounts of the United States from 1967Q3 to 2012Q4.

The correlation between the cyclical component of firms’ money holdings and the money premium is -35.0% .

There is a large literature on the secular trends in corporate liquidity holdings (e.g., Bates, Kahle, and Stulz (2009)), but relatively few on its cycle. Eisfeldt and Rampini (2009) document a positive correlation between the money premium and the asynchronicity between the sources and uses of funds in the productive sector. The model by He and Kondor (2016) also predicts a negative correlation between the money premium and firms’ cash holdings, but they focus on the pecuniary externality in the market of productive capital and the resulting inefficient investment waves.

Using data from the Financial Accounts of the United States, Panel A of Figure 13 plots the outstanding Treasury bills (the solid line) and the *net* supply of money-like securities by intermediaries, i.e., the sum of intermediaries’ net liabilities in each instrument listed in Figure

11 minus their holdings of government securities.⁶⁶ A key message is the collapse of inside money supply in the Great Recession, a more than 50% decline from the second quarter of 2007 (the peak) to end of 2009 and a continuing contraction afterwards to a level lower than 1960s. This secular depression of inside money creation is consistent with the model's prediction on stagnant crises.

Treasury Bill supply increased by more than 100% during the Great Recession. The model predicts that such an increase prolongs the crisis by crowding out intermediaries' profit, and thus, delaying the repairment of intermediaries' balance sheet. To some extent, Treasury Bill supply mitigates the money shortage that the productive sector faces, so we do not see a slump in nonfinancial firms' money holdings (Figure 12) that is as large as the slump of inside money supply. However, the yield on money-like securities has been extremely low in the post-crisis period, which signals a persistent scarcity very likely due to the depressed inside money creation.

Panel B of Figure 13 plots the share of inside money supplied by non-depository financial intermediaries, i.e. the "shadow banks", such as broker-dealers, finance companies, and issuers of asset-backed securities. Many have argued that other sectors' demand for money-like securities is a major driver of the growth of shadow banking (Gorton (2010), Stein (2012), and Pozsar (2011) and (2014)). In comparison with traditional banks (depository institutions), shadow banks seemed to be more responsive to rise of money demand in the productive sector in the last two decades, and until the global financial crisis, took an increasingly large share of inside money supply.

III.3 Government debt and intermediary leverage cycle

This section provides evidence on the impact of outside money on the intermediary leverage cycle. I focus on one type of intermediaries, broker-dealers (or investment banks). Adrian and Shin (2010) document the procyclicality of broker-dealer leverage: when their assets grow, their

⁶⁶Government securities include Treasury securities and vault cash held by depository institutions, and Treasury securities indirectly held through money market funds by other financial intermediaries.

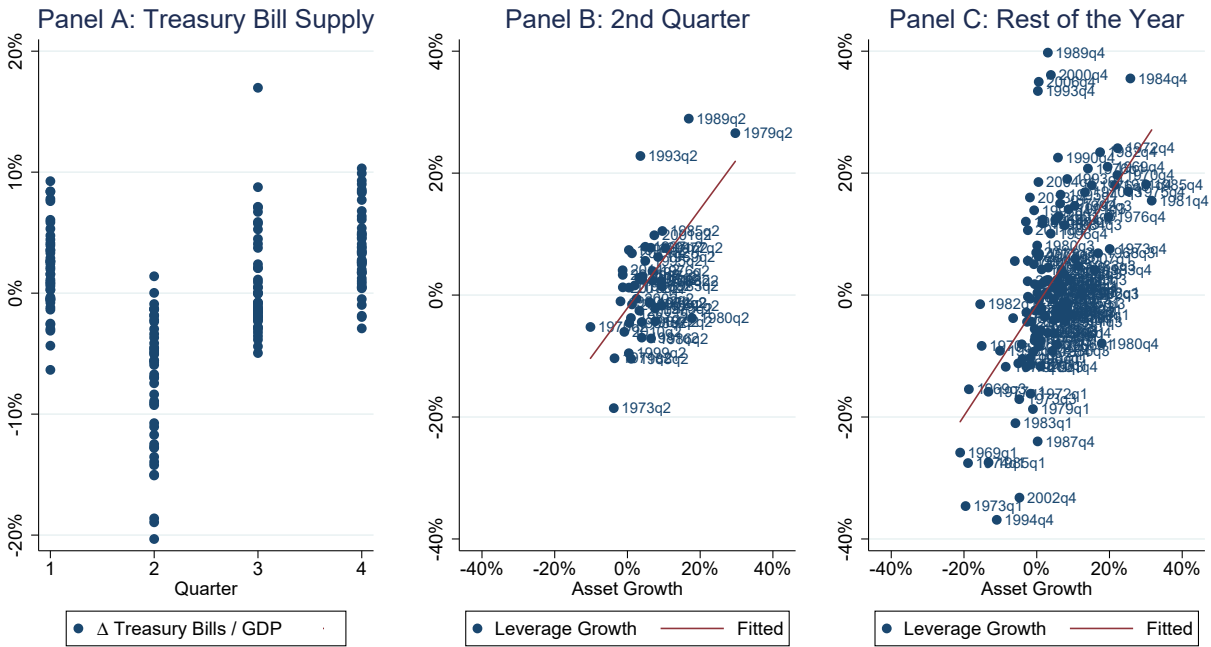


Figure 14: **Intermediary Leverage Cycle and Government Debt.** Panel A plots the change of Treasury bills outstanding scaled by nominal GDP by quarter. Panel B plots the quarterly growth rate (log difference) of broker-dealer book leverage (Adrian and Shin (2010)) against the growth rate of broker-dealer book assets in the second quarterly, and Panel C plots the quarterly growth rate of leverage against the growth rate of assets in the first, third, and fourth quarters. The data source is the Financial Accounts of the United States from 1968Q3 to 2015Q3.

debt grows faster, leading to higher leverage. Broker-dealers issue money-like securities (mainly repurchase agreements) that are held by firms and other entities through money market mutual funds. I focus on investment banks because relative to commercial banks, their choice of leverage is subject to less regulatory constraints, and thereby, they are closer to the laissez-faire banks in the model, and as shown in Figure 13, shadow banks play a more important role than traditional banks in the pre-crisis boom of inside money creation. Leverage is defined as the ratio of total book assets to book equity, using the data from the Financial Accounts, following Adrian and Shin (2010). The sample is 1968Q3–2015Q3.⁶⁷ He, Kelly, and Manela (2016) discuss issues related to internal capital markets that may induce measurement errors in leverage. For government debt, I use the ratio of Treasury Bills to nominal GDP. Short-term debt is more money-like in the sense that their value is more stable, their secondary market more liquid, and repo haircuts smaller.

⁶⁷ Adrian, Etula, and Muir (2014) discuss the data quality concerns in the pre-1968 sample.

To identify the impact of government debt on intermediary leverage, one of the challenges is the endogeneity of government debt supply. Greenwood, Hanson, and Stein (2015) argue that Treasury Bill supply has seasonality. It expands ahead of statutory tax deadlines (e.g., April 15th) to meet its ongoing needs, and contracts following these deadlines. Greenwood, Hanson, and Stein (2015) and Nagel (2016) use week and month dummies respectively to instrument Treasury Bill supply. Panel A of Figure 14 shows that the strongest seasonality is in the second quarter of the year. Because my data is quarterly, I use the Q2 dummy as an instrument for T-bill supply.

Panel B and C of Figure 14 plot the log difference of book leverage (i.e., quarterly growth) against the book asset growth, extending the main result of Adrian and Shin (2010) to this longer sample. Panel B separates out the observations from the second quarters, and Panel C shows the rest of the year. The fitted line is steeper in Panel C, which seems to be consistent with the model's prediction – increases in government debt amplify the procyclicality of intermediary leverage.

Another identification challenge is the endogeneity of asset growth. In the model, the cyclicity of leverage is defined with respect to *exogenous* shocks that increase or decrease banks' asset value through their impact of collateral quality. Such structural shocks are rarely observed in reality. However, for the purpose of empirical analysis, any exogenous shocks that affect banks' asset value can be used as instruments. I use monetary policy shocks as such instruments, and consider two measures: the unanticipated changes in the Fed Funds Rate around the FOMC (Federal Open Market Committee) announcements ("MP-FFR"), and a composite measure of unanticipated changes in several interest rates around the FOMC announcements proposed by Nakamura and Steinsson (2017) ("MP-Comp").⁶⁸ Shocks are aggregated to quarterly level. Calculated from high-frequency data, these policy shocks arguably reflect the changes in interest rates only from the unexpected content of FOMC announcements, and thus, tend to be orthogo-

⁶⁸A monetary policy shock is calculated using a 30-minute window from 10 minutes before the FOMC announcement to 20 minutes after it. The calculation follows Nakamura and Steinsson (2017). For earlier contributions, please refer to Cook and Hahn (1989), Kuttner (2001), and Cochrane and Piazzesi (2002) among others.

Table A.2: Government Debt and Procyclical Leverage.

This table reports the evidence on the impact of government debt on the relation between broker-dealer asset growth and leverage growth. Column (1) reports the results of regressing leverage growth (quarterly log difference) on asset growth (as in Adrian and Shin (2010)). The sample is from 1968Q3 to 2015Q3. Column (2) and (3) repeat the regression in Column (1) with two types of monetary policy shocks, MP-FFR and MP-Comp, as instrument variables (IVs) for broker-dealer asset growth. MP-FFR and MP-Comp are the quarterly sums of unexpected Fed Funds rate change and composite rate changes respectively around FOMC announcements (available from Nakamura and Steinsson (2017) from 1995 to 2014). Column (4) reports the results of regressing broker-dealer leverage growth on asset growth and the interaction (product) between asset growth and the growth rate of Treasury bills scaled by nominal GDP. Column (5) and (6) repeat the regression in Column (4) with the second quarter dummy as IV for Treasury bill growth, and with MP-FFR (Column 5) or MP-Comp (Column 6) as IV for broker-dealer asset growth. All coefficients are estimated using GMM with Newey-West HAC standard errors reported in parentheses (optimal number of lags chosen following Newey and West (1994)). *, **, *** represent $p < 0.10$, $p < 0.05$, $p < 0.01$ respectively.

$\Delta \ln (Leverage)$	(1)	(2)	(3)	(4)	(5)	(6)
IV for $\Delta \ln (Assets)$		MP-FFR	MP-Comp		MP-FFR	MP-Comp
IV for $\Delta \ln \left(\frac{T-Bill}{GDP} \right)$					Q2	Q2
$\Delta \ln (Assets)$	0.910*** (0.113)	3.873*** (1.451)	2.527*** (0.855)	0.203 (1.294)	0.991** (0.449)	0.867** (0.348)
$\Delta \ln \left(\frac{T-Bill}{GDP} \right) \cdot \Delta \ln (Assets)$				24.40 (38.57)	10.28*** (2.174)	8.999*** (2.220)
Observations	189	77	77	188	77	77

nal to contemporaneous variations of other economic variables. Through the impact on interest rates and various spreads, monetary policy shocks affect the value of banks' assets.⁶⁹

Table A.2 reports the results. The first column replicates the main regression in Adrian and Shin (2010). The left-hand side is leverage growth and the right-hand side is asset growth. A positive coefficient shows the procyclicality of leverage. The fourth column of Table A.2 shows the coefficients from regressing leverage growth on both asset growth and the interaction between asset and Treasury bill growth. A positive coefficient on the interaction term suggests when government debt increases, leverage becomes more procyclical, consistent with the model's prediction.

Column 2 and 3 show leverage procyclicality using different measures of monetary policy

⁶⁹Gertler and Karadi (2015) show that these high-frequency monetary policy shocks affect term premia and credit spreads, and Hanson and Stein (2015) show the strong effect on forward real rates even in the distant future.

shocks as instruments for asset growth. Both estimates are positive and significant. Column 5 and 6 use the Q2 dummy as an instrument for Treasury Bill supply, and use two measures of monetary policy shocks respectively to instrument asset growth. The coefficient on the interaction term is positive and significant. These confirm the findings in baseline specifications. Table A.2 provides supporting evidence, but it is far from conclusive. Future research based on longer samples, international data, and alternative identification strategies, shall provide a better evaluation of the model's predictions on intermediary leverage cycle, the impact of government debt on it, and other results such as how banks' payout and issuance policies respond to government debt supply.

The Rise of Intangible Capital and Financial Fragility*

Ye Li[†]

Abstract

This paper studies financial instability in an economy where growth is driven by intangible investment. Firms' intangible investment (R&D) creates new productive capital. Once created, capital can be sold to financial intermediaries. Since intangible investment is not pledgeable, firms carry cash, which is inside money issued by intermediaries (short-term safe debt). In good times, well capitalized intermediaries push up the price of capital. This motivates firms to create more capital, but to do so, they must build up cash holdings. As firms' money demand expands, the yield on inside money (i.e. intermediaries' debt cost) declines, so intermediaries increase leverage and push up capital price even further. Thus, the model generates booms that share several features with the U.S. experience before the Great Recession: the rise of corporate cash holdings, the expansion of the financial sector through leverage, increases in asset price, and the decline of interest rate. The model generates endogenous risk accumulation: a longer period of boom and expansion of the financial sector predict a more severe crisis. In crises, the spiral flips, leading to sudden deleveraging of intermediaries and depressed intangible investment.

*I would like to thank Tobias Adrian, Patrick Bolton, Guojun Chen, Itay Goldstein, Gur Huberman, Christian Julliard, Siyuan Liu, Adriano Rampini, Tano Santos, José A. Scheinkman, Suresh Sundaresan, and Neng Wang for helpful discussions and comments. I am also grateful to seminar participants at Columbia Financial Economics Colloquium, Columbia Finance PhD Seminar, Finance Theory Group Workshop, and LBS Trans-Atlantic Doctoral Conference. This paper was previously circulated under the title "Fragile New Economy".

[†]Columbia University. E-mail: yli17@gsb.columbia.edu

1 Introduction

The U.S. economy exhibited five trends in the two decades leading up to the Great Recession:

Fact 1: The U.S. economy was transforming from a manufacturing-based economy to a more intangible-intensive economy. The production of goods and services increasingly relied on intangible capital, such as knowledge capital, organizational capital, and brand names. According to Corrado and Hulten (2010), intangible investment overtook physical investment as the largest source of economic growth in the United States in the period of 1995-2007.

Fact 2: An increasing share of non-financial corporations' assets were cash holdings (Bates, Kahle, and Stulz (2009)). "Cash" includes bank deposits and shares of money market mutual funds that are in turn portfolios of financial intermediaries' liabilities, such as repurchase agreements and asset-backed commercial papers. The rise of corporate cash holdings came from the growing R&D-intensive sectors (Falato and Sim (2014); Begenau and Palazzo (2015); Pinkowitz, Stulz, and Williamson (2015); Graham and Leary (2015)).

Fact 3: In the United States, the assets held by the financial sector increased dramatically (Gorton, Lewellen, and Metrick (2012)). More generally in advanced economies, Schularick and Taylor (2012) find that the bank loan-to-GDP ratio doubled in the last two decades. The growth of the financial sector was fueled by high leverage, especially through the issuance of money-like securities (or short-term safe debt) (Adrian and Shin (2010); Gorton (2010); Gorton and Metrick (2012); Pozsar (2014)).

Fact 4: Interest rate declined steadily.

Fact 5: The price of risky assets increased across asset classes.

Despite extensive debates on the forces behind these trends and their macroeconomic implications, there are relatively few theories that analyze these phenomena jointly. Motivated by Fact 1, this paper builds a model that reproduces Fact 2, 3, 4, and 5. Beyond explaining these facts, the model reveals a new mechanism of endogenous risk accumulation that helps under a surging empirical literature: a longer period of bank expansion precedes a sharper decline of bank

equity (Baron and Xiong (2016)) and a more severe economic recession (Jordà, Schularick, and Taylor (2013)). At the center of the mechanism is financial intermediaries' role as inside money creators.

Before diving into the theoretical setup and mechanism, let us connect these five facts from the perspective of financial friction. The rising share of productive capital that is intangible implies a shrinkage of pledgeable assets (i.e., physical capital, such as properties, plants, and equipments). As a result, the corporate sector hoards more cash in anticipation of liquidity needs, for instance, investments in intangibles (e.g., R&D). Cash takes the form of the financial sector's short-term safe debt, such as deposits, which are directly used as means of payment, and money-like securities that are close substitutes to money.

Therefore, a rising corporate money demand feeds leverage to the financial sector, and as a result, financial intermediaries are able to acquire more assets. Moreover, because the financial sector tends to be better suited to manage risky assets than other sectors, and thus, assign a higher asset value, the expansion of their purchasing power pushes up asset prices. The rising corporate money demand also pushes down the interest rate. In fact, because of the introduction of money market mutual funds and the repeal of Regulation Q, money is interest-paying (Lucas and Nicolini (2015)). When the demand rises, the yield on money (i.e., the interest rate) declines.

I build a continuous-time model of dynamic economy that crystallizes this narrative. At the center are two key ingredients. First, the cash that firms desire is short-term safe debt issued by bankers (i.e., inside money).¹ Second, the risky asset price depends on intermediaries balance-sheet capacity, because intermediaries assign a higher value to risky assets than the rest of the economy, and thus, are the natural buyers in the sense of Shleifer and Vishny (1992). These two ingredients lead to two reinforcing mechanisms of instability, the inside money channel and the balance-sheet channel.

The economy has two sectors, bankers and entrepreneurs. Both types consume generic

¹The term, inside money, is borrowed from Gurley and Shaw (1960). From the private sector's perspective, gold, fiat money, or government securities are in positive supply ("outside money"), while as bank liabilities, deposits are in zero net supply ("inside money"). See Lagos (2008) for a brief review of the related literature.

goods that are produced by capital. Capital is traded in a competitive market. The aggregate shock (“macro shock”) arrives at Poisson times with a constant intensity. When the shock hits, a fraction of capital is destroyed. However, less capital is destroyed when held by bankers. This expertise to rescue value from failed assets echoes Bolton and Freixas (2000). Due to this expertise, bankers assign a higher value to capital than entrepreneurs. In equilibrium, bankers borrow from entrepreneurs, holding a levered position in capital.

Following a macro shock, asset price declines, because bankers lose wealth (equity) and purchasing power. I assume that bankers cannot recapitalize by issuing equity to entrepreneurs. This friction gives rise to the following balance-sheet channel of shock amplification. As capital price declines, bankers’ equity is further eroded due to the downward reevaluation of their assets, and a lower level of bankers’ equity leads to a further decrease in capital price. This mechanism has been well explored in the macro finance literature (e.g., Brunnermeier and Sannikov (2014)).

The theoretical contribution of this paper is the inside money channel, which is built upon the balance-sheet channel and amplifies it. Entrepreneurs have opportunities to create new capital from goods. These opportunities arrive at idiosyncratic Poisson times with constant intensity. I assume the creation of capital is immediate. Investment can be interpreted as an R&D or prototype stage that happens instantaneously. Once it is done, new capital is created (or matures), and can be readily sold in the market. Entrepreneurs can only issue up to θ fraction of the shares of investment project ($\theta < 1$), and thus, they need to invest out of their own pocket. The requirement of insiders’ stake can be motivated by the fact that entrepreneurs’ inalienable human capital is needed for capital creation (Hart and Moore (1994)).

Moreover, I assume that only $\phi (< 1)$ fraction of entrepreneurs’ holdings of existing capital contributes to internal liquidity (i.e., can be readily pledged or sold), which intends to capture the limited pledgeability of intangible capital. To justify this assumption, consider a technology of creating worthless counterfeits out of thin air that comes along with the investment technology to entrepreneurs. Outsiders can only examine up to ϕ fraction of inventory. This information

friction captures the difficulty to measure intangible capital, such as organizational capital, as its effectiveness tends to be better appraised by insiders. Thus, ϕ is called “tangibility”. In contrast, money (i.e. bank debt) is perfectly liquid: every dollar of it can be used to purchase investment inputs. The superior liquidity of bank debt (or inside money) is a key ingredient. Because of this, entrepreneurs assign a liquidity premium to bank debt, which lowers bankers’ borrowing cost. Because liquidity is used for investment, this liquidity premium depends on the endogenous capital price, leading to the following feedback mechanism.

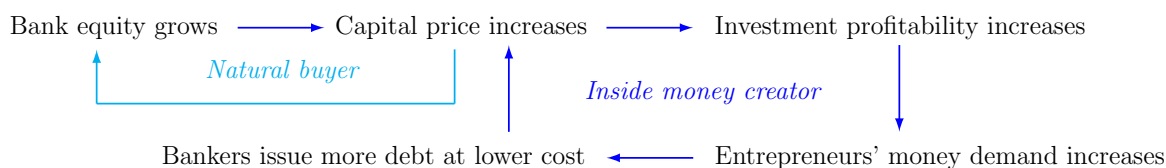
Consider a calm period without any macro shock (“boom”). Bankers accumulate wealth through their levered position on capital. As capital price increases, the balance-sheet channel implies a feedback loop: an upward reevaluation of capital increases bankers’ wealth even further, which in turn leads to an even higher capital price.

However, the story does not end here. Because capital is worth more, entrepreneurs want to create more capital, but to do so, they must build up savings, and ideally in the most liquid form (i.e., inside money). A higher marginal value of money holdings from investment translate into a higher liquidity premium assigned to bank debt, which lowers bankers borrowing cost and encourages them to expand balance sheets by taking on more debt. As the banking sector grows via both equity accumulation and debt issuance, asset price increases even further. This feedback loop is the inside money channel.

In a boom, the mutual reinforcement of the balance-sheet channel and the inside money channel leads to higher capital price, lower interest rate (bankers’ debt cost), the growth of the banking sector, and the growth of entrepreneurs’ money holdings. Along the process, endogenous risk accumulates. In particular, endogenous risk is measured by the difference between the current capital price and the post-shock capital price (i.e., the new equilibrium price after the macro shock hits). This wedge widens as the boom prolongs.

Endogenous risk accumulates in a boom precisely because of the increasing capital price. What differentiates bankers and entrepreneurs as capital holders are two-fold: first, the macro

Good times: no macro shocks



Bad times: triggered by macro shocks

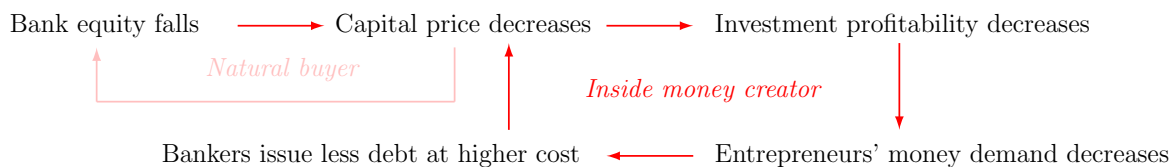


Figure 1: Money Supply and Demand.

shock destroys a smaller fraction of bankers' holdings; second, bankers do not face the capital illiquidity problem that entrepreneurs face at the times of investment ($\phi < 1$). When capital price is higher and investment more profitable, capital illiquidity is more costly for entrepreneurs. Therefore, the second difference between bankers and entrepreneurs as capital holders widens as capital price increases. When the macro shock hits, destroying bankers' wealth and triggering capital reallocation towards entrepreneurs (the second-best capital holders), the downward reevaluation of capital is more severe if the economy has not experienced any shocks for a long period of time and capital price is at a very high level.

The positive feedback loops turn into vicious cycles when the macro shock hits. The initial destruction of capital reduces bankers' wealth, and triggered the balance-sheet channel that amplifies the negative effect on capital price. The decline of capital price discourages entrepreneurs' investment, which in turn leads to a lower money demand and a small liquidity premium assigned to bank debt. Therefore, bankers have to borrow at a relatively higher cost, so their balance sheet shrinks via both equity destruction and deleveraging. As natural buyers retreat from the market, capital price decreases even further. The downward spiral of capital price leads to lower investments by entrepreneurs. Figure 1 illustrates the mechanisms.

Literature. The structural change towards a new economy that is intangible-intensive has attracted a lot of attention (Corrado and Hulten (2010)). Several studies explore the implications of this structural change in different areas from both theoretical and empirical perspectives, such as Atkeson and Kehoe (2005) (productivity accounting), McGrattan and Prescott (2010) (current account dynamics), Eisfeldt and Papanikolaou (2014) (asset pricing), and Peters and Taylor (2016) (corporate investment). This paper looks into the financial stability issue of an intangible economy. While there is a large literature on how financial development affects industrial structure (e.g., Levine (1997); Rajan and Zingales (1998)), the reverse question, how industrial structure affects the financial system, has largely been unexplored.

This paper contributes to the literature of heterogeneous-agent models of macroeconomy. A key friction in this literature is the limited risk-sharing between sectors, which in this paper, corresponds to the assumption that bankers cannot recapitalize by issuing equity to entrepreneurs. Noted by Di Tella (2014), perfect risk-sharing shuts down the balance-sheet channel (e.g., Kiyotaki and Moore (1997); Bernanke and Gertler (1989)). A large literature explores the macroeconomic and asset pricing implications of dynamic wealth distribution among heterogeneous agents (e.g., Basak and Cuoco (1998); Krusell and Smith (1998); Longstaff and Wang (2012); He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014); Moll (2014)). This paper contributes to this literature by highlighting a particular dimension of heterogeneity (i.e., bankers as money suppliers and entrepreneurs as money demanders) that leads to a new feedback loop that interacts with the balance-sheet channel.

A recent literature revives the money view of financial intermediation that emphasizes the liabilities of financial intermediaries as inside money that lubricates the economy by facilitating trades (Kiyotaki and Moore (2000); Hart and Zingales (2014); Quadrini (2014); Piazzesi and Schneider (2016)). This paper takes a step further by embedding this money view in an entrepreneurial economy that suffers from capital and investment illiquidity.² The $\theta - \phi$ framework

²Safe debt serves as money, which echoes the literature that links the information insensitivity of assets' payout to assets' monetary services (Gorton and Pennacchi (1990); Holmström (2012); Dang, Gorton, Holmström, and Ordóñez (2014)).

that leads to entrepreneurs' money demand is inspired by Kiyotaki and Moore (2012) who study the outside money (or fiat money) whose supply is intervened by the government instead of the inside money issued within the private sector by bankers.

In this paper, bankers add value to the economy by providing liquidity – securities that can be hoarded by financially constrained entrepreneurs to meet liquidity needs. Woodford (1990), Holmström and Tirole (1998), and Farhi and Tirole (2011) study the private sector's capacity to create liquidity and potential government intervention in liquidity supply. This paper highlights a particular sector, the banking sector, as liquidity providers, and thereby, links banks' balance-sheet capacity to the liquidity available for entrepreneurs to hoard for capital creation, and also links entrepreneurs' liquidity demand to the leverage of banks who are the natural buyers of risky capital created by entrepreneurs.

From the empirical side, this paper is inspired by the literature of corporate cash holdings (e.g., Opler, Pinkowitz, Stulz, and Williamson (1999); Bates, Kahle, and Stulz (2009)). Many have found a secular increase of nonfinancial firms' cash holdings in the last few decades that was driven by cash held in the R&D-intensive sectors (e.g., Falato and Sim (2014); Begenau and Palazzo (2015); Pinkowitz, Stulz, and Williamson (2015); Graham and Leary (2015)). Physical capital, such as properties, plants, and equipments, support firms' borrowing by serving as collateral (Almeida and Campello (2007)). The structural transformation towards an intangible economy implies a shrinkage of pledgeable assets in the productive sector, and thus, a stronger liquidity hoarding incentives of firms. As pointed out by Pozsar (2011), corporate treasuries have become a prominent component of “institutional cash pools” that feed leverage to the financial sector, particularly through the money markets.³

The theoretical literature on corporate cash holdings commonly focuses only on firms' decision to save (e.g., Bolton, Chen, and Wang (2011); Froot, Scharfstein, and Stein (1993)). This partial equilibrium approach allows modeling a rich environment of corporate decision making,

³Several papers have made the point that the growth of the shadow banking sector is fueled by the demand of money-like securities (e.g., Gorton (2010); Gorton and Metrick (2012); Stein (2012); Pozsar (2014).

but by assuming a perfectly elastic supply of storage, the models sever the link between firms' money demand dynamics and the equilibrium yield on money (i.e., the interest rate paid by money-like securities). By endogenizing money supply, my model predicts that the growth of corporate money demand leads to low interest rate and the growth of the financial sector who supplies money, and thereby, linking these three phenomena in a coherent framework that puts inside money at the center.

Finally, the model produces the endogenous risk accumulation in a boom, which has been documented by a surging literature. Schularick and Taylor (2012) find the expansion of bank asset (loans) precedes financial crises in advanced economies.⁴ In the model, a longer period of bank expansion predicts more severe crises. Jordà, Schularick, and Taylor (2013) document a close relationship between the build-up of bank credit and the severity of the subsequent recessions. Moreover, when the economy is hit by the macro shock, the decrease of bank equity is larger when the preceding boom is longer, which is consistent with the findings by Baron and Xiong (2016) that bank credit expansion predicts bank equity crashes. The key to risk accumulation is the combination of the balance-sheet channel and the inside money channel: due to the investment-driven liquidity demand of entrepreneurs, the difference between the first-best and second-best buyers of capital widens in booms, and thus, triggered by the macro shock, the reallocation from the first-best to the second-best buyers has a stronger impact on capital price.

The rest of the paper is organized as follows. Section 2 introduces the model and discusses the main mechanisms. In section 3, the performances of the fully solved model demonstrate the mechanisms. Section 4 concludes.

⁴In line with Bordo, Eichengreen, Klingebiel, Martinez-Peria, and Rose (2001) and Reinhart and Rogoff (2009), Schularick and Taylor (2012) define financial crises as events during which a country's banking sector experiences bank runs, sharp increases in default rates accompanied by large losses of capital that result in public intervention, bankruptcy, or forced merger of financial institutions. Using the information from credit spreads, Krishnamurthy and Muir (2016) are able to achieve a sharper differentiation between financial and non-financial crises, and thus, provide a granular identification of financial crises. They also find that bank credit expansion precedes financial crises. Using both the information from credit spreads and the composition change of corporate bond issuance, López-Salido, Stein, and Zakrajšek (2015) find that when the issuance of high-yield ("junk") bond outpaces the total bond issuance, and when corporate bond credit spreads are narrow relative to their historical norms, the subsequent real GDP growth, investment, and employment tend to slow down.

2 Model

2.1 Setup

Consider a continuous-time, infinite-horizon economy with two types of agents, bankers and entrepreneurs. Each type has a continuum of representative agents with measure equal to one.

Preferences. Both types of agents maximize expected logarithm utility with discount rate ρ :

$$E \left[\int_{t=0}^{\infty} e^{-\rho t} \ln (c_t^i) dt \right], i \in \{f, nf\}. \quad (1)$$

Throughout this paper, subscripts denote time, and superscripts denote types or sectors with “ f ” being the financial sector (bankers) and “ nf ” being the non-financial sector (entrepreneurs). For example, c_t^{nf} is the representative entrepreneur’s consumption at time t .

Capital and aggregate shock. At time t , the economy has K_t units of capital that produces non-durable generic goods for consumption and investment. One unit of capital produces a units of goods per unit of time (a is a positive constant). Both entrepreneurs and bankers can own capital, and capital is traded in a competitive market at price q_t (denominated in goods).

At Poisson times with intensity equal to λ , a certain fraction of capital is destroyed. The random arrival of capital destruction is the only source of aggregate uncertainty in this economy (i.e., the “macro shock”). Bankers and entrepreneurs differ in their ability to deal with the capital destruction shocks. When capital is owned by bankers, a fraction $\underline{\delta}$ of capital is destroyed when the macro shock hits. When capital is owned by entrepreneurs, a fraction $\bar{\delta}$ is destroyed. I assume $\underline{\delta} < \bar{\delta}$ so that bankers are better at dealing with capital destruction.

We can interpret capital as efficiency units or production projects, so the assumption $\underline{\delta} < \bar{\delta}$ captures bankers’ expertise in restructuring distressed projects (Bolton and Freixas (2000)). Since bankers assign a higher value to capital, when they are undercapitalized, and thus, less willing to hold risky capital, the equilibrium capital price declines. In other words, the market fails

to reflect the full value of capital when bankers' risk-taking capacity is limited. In this paper, the term "market illiquidity" captures the constrained capacity of natural buyers in the sense of Shleifer and Vishny (1992) and Brunnermeier and Sannikov (2014) instead of transaction costs or other trading frictions (reviewed by Vayanos and Wang (2013)).⁵

Investment and inside money. Entrepreneurs have opportunities to create new capital. Investment opportunities arrive at idiosyncratic Poisson times with constant intensity λ_I , where the subscript "I" stands for both idiosyncratic and investment.⁶ Thus, every instant, a constant fraction $\lambda_I dt$ of entrepreneurs make investments. They can convert one unit of goods into one unit of capital instantaneously. This investment is scalable, so when $q_t > 1$, entrepreneurs want to invest as much as possible. The scale of investment is limited by financial frictions.

To create one unit of capital that is worth q_t , entrepreneurs can issue up to θ fraction of equity to competitive investors including non-investing entrepreneurs and bankers ("outside equity"). They must invest $(1 - \theta q_t)$ out of their own pocket ("inside equity"). Once the capital is created, entrepreneurs deliver θ units of capital to outside investors and retain $(1 - \theta)$ units that are worth $(1 - \theta) q_t$. Therefore, the investment return on internal equity ("ROI") is

$$ROI_t = \frac{(1 - \theta) q_t}{1 - \theta q_t}. \quad (2)$$

It increases in the capital price q_t through the unlevered return $(1 - \theta) q_t$ and the leverage on internal equity $\frac{1}{1 - \theta q_t}$. The limited pledgeability of investment can arise because entrepreneurs' inalienable human capital is required in the R&D or prototype stage, and thus, entrepreneurs must hold a stake to credibly commit to work (Hart and Moore (1994)). Once they input their effort and the prototype becomes mature or commercialized, productive capital is created. Capital creation

⁵Other specifications of financial intermediaries' "specialness" may serve the same purpose of making bankers the natural buyer of capital, such as intermediaries' lower risk aversion (Longstaff and Wang (2012)), their collateralization expertise (Rampini and Viswanathan (2015)), their unique ability to hold risky assets (He and Krishnamurthy (2013)), their unique ability to monitor projects (Diamond (1984) and Holmström and Tirole (1997)), and their advantage in diversification (Brunnermeier and Sannikov (2016)).

⁶These idiosyncratic investment shocks are not insurable; otherwise, entrepreneurs will not have liquidity needs, which is a critical ingredient of this model.

take place sequentially within an instant.

Internal funds come from two sources: entrepreneurs' holdings of existing capital and their holdings of money. Only a fraction ϕ of their existing capital can be readily sold in exchange for goods as investment inputs ($\phi < 1$). The illiquidity arises from an information friction. I assume that in addition to the capital creation technology, investing entrepreneurs can also create counterfeits that are indistinguishable from productive capital. Investors can verify the quality of capital up to ϕ fraction of entrepreneurs' inventory.⁷

In contrast, entrepreneurs can exchange all of their money holdings for goods. Money is short-term safe debt issued by bankers. At time t , bankers' issues debt that matures at $t + dt$ and pays a risk-free market interest rate r_t . Entrepreneurs carry bank debt ("inside money") as a perfect means of payment in anticipation of investment needs.⁸ Money lubricates the economy by facilitating goods reallocation towards investing entrepreneurs. The moneyness of safe debt echoes the theories that links assets' information insensitivity to their monetary services (e.g., Gorton and Pennacchi (1990); Holmström (2012); Dang, Gorton, Holmström, and Ordóñez (2014)).

Discussion: intangible economy. The $\theta - \phi$ framework is motivated by advanced economies' increasing reliance on intangibles in the production, such as design, brand names, firms' proprietary technology, and entrepreneurs' human capital.⁹ The financial stability implications of this structural change has largely remained unexploited. The assumption $\theta < 1$ captures the fact that firms' intangible investment is largely financed by internal liquidity.¹⁰ The assumption $\phi < 1$ captures the fact that it is difficult to measure intangible capital.¹¹ Capital represents production

⁷Note that non-investing entrepreneurs do not have access to the counterfeit technology, so they do not face the illiquidity of capital. Whether an entrepreneur is investing or not is public information.

⁸The term is borrowed from Gurley and Shaw (1960). From the private sector's perspective, fiat money and government securities are in positive supply ("outside money"), while deposits, as bank liabilities, are in zero net supply ("inside money"). See Lagos (2008) for a brief review of the related literature.

⁹According to Corrado and Hulten (2010), intangible investment has overtaken physical investment as the largest source of economic growth in the United States in the period of 1995-2007.

¹⁰In particular, R&D heavily relies on internal financing, which has been well documented since Hall (1992) and Himmelberg and Petersen (1994). Hall and Lerner (2009) review the literature on innovation financing.

¹¹Corrado, Hulten, and Sichel (2005) and McGrattan and Prescott (2010) discuss the measurement of technology

(revenue-generating) units, which could simply be a particular design of common products. It is often difficult to measure the quality of such intangible capital.

From a theoretical perspective, this paper adopts the $\theta - \phi$ framework in Kiyotaki and Moore (2012) to analyze the dynamics and stability implications of inside money creation.¹² Kiyotaki and Moore (2012) focuses on outside money (fiat money) that is supplied by government instead of private entities (banks) that engage in dynamic balance-sheet management.

2.2 Markov equilibrium

Wealth evolution. Let n_t^i denote the net worth (wealth) of a representative agent in sector i ($i \in \{f, nf\}$). I characterize a Markov equilibrium where η_t , the fraction of aggregate wealth in the banking sector is the only aggregate state variable:

$$\eta_t = \frac{n_t^f}{n_t^f + n_t^{nf}}, \quad (3)$$

which has a deterministic growth rate μ_t^η ($:= \frac{d\eta_t/dt}{\eta_t}$) before the macro shock hits. Note that because both production and investment have constant return to scale, the economy is scale-free, so the aggregate capital stock K_t is not a state variable.¹³

In a Markov equilibrium, all endogenous variables are functions of η_t . Their dynamics are linked to the dynamics of the state variable η_t , for example, before the macro shock hits,

$$\mu_t^q := \frac{dq_t/dt}{q_t} = \frac{q'(\eta_t)}{q(\eta_t)} \mu_t^\eta \eta_t,$$

As will be explained in detail later, in equilibrium, capital price increases in η_t (i.e., $q'(\eta_t) > 0$).

capital. Atkeson and Kehoe (2005) and Eisfeldt and Papanikolaou (2014) measure organization capital.

¹²The money or liquidity demand is investment-driven, in line with Holmström and Tirole (1998, 2001) and Eisfeldt and Rampini (2009). Eisfeldt (2007) show that the liquidity premium of securities such Treasury bills cannot be explained by a liquidity demand driven by consumption smoothing under standard preferences.

¹³Wealth or consumption share is a common state variable in the literature of heterogeneous-agent models, such Basak and Cuoco (1998), Longstaff and Wang (2012), and Brunnermeier and Sannikov (2014).

Moreover, $q_t > 1$, because investing entrepreneurs cannot invest unlimited resources to close the wedge between q_t , the value of capital, and 1, the cost of creating capital.

Bankers and entrepreneurs allocate wealth between safe debt and risky capital. A short position in safe debt means borrowing.¹⁴ Let x_t^i denote the fraction of wealth invested in capital ($i \in \{f, nf\}$). Before the macro shock hits, the flow of funds constraint (i.e. the evolution of wealth) is represented by a deterministic growth rate of wealth

$$\mu_t^{n^i} := \frac{dn_t^i/dt}{n_t^i} = (1 - x_t^i) r_t + x_t^i \left(\frac{a}{q_t} + \mu_t^q \right) - \frac{c_t^i}{n_t^i}, i \in \{f, nf\}. \quad (4)$$

The pre-shock return of capital holdings includes a dividend gain $\frac{a}{q_t}$ and a capital gain μ_t^q . In equilibrium, bankers borrows from entrepreneurs, so we have $x_t^f > 1$ and $x_t^{nf} < 1$.

A constant fraction $\lambda_I dt$ of entrepreneurs can make investments and experience a jump in their wealth. Because when $q_t > 1$, each dollar of inside equity will be worth $ROI_t (> 1)$ after investment, investing entrepreneurs want to use all of these resources to maximize the scale of investment. Given their internal funds $\left[\phi x_t^{nf} + (1 - x_t^{nf}) \right] n_t^{nf}$, which is the sum of entrepreneurs' holdings of bank debt and their resalable or pledgeable capital holdings, their investment is given by the following Proposition.

Proposition 1 *At time t , if investment opportunities arrive, entrepreneurs will convert*

$$i_t = \left(\frac{1}{1 - \theta q_t} \right) \left[1 - (1 - \phi) x_t^{nf} \right] n_t^{nf}, \quad (5)$$

units of goods into new capital. The investment depends on internal liquidity, and through the leverage on internal equity $\frac{1}{1 - \theta q_t}$, depends on capital price (Tobin's q). After investment, the

¹⁴Under log-utility, optimal consumption is proportional to wealth, and the Inada condition implies infinite marginal utility when wealth equals to zero. As a result, the agents always try to avoid bankruptcy and keep a positive net worth, and agents never default on debt.

investing entrepreneur's wealth jumps up to

$$\widehat{n}_t^{nf} = \left\{ ROI_t \left[1 - (1 - \phi) x_t^{nf} \right] + (1 - \phi) x_t^{nf} \right\} n_t^{nf}. \quad (6)$$

The liquid part of wealth (fraction $\left[1 - (1 - \phi) x_t^{nf} \right]$) is multiplied by the investment return on internal equity (i.e., ROI_t), and the value of the illiquid part of wealth (fraction $(1 - \phi) x_t^{nf}$) stays the same. By holding more money instead of capital (i.e., decreasing x_t^{nf}), entrepreneurs can increase the positive jump in wealth when investment opportunities arrive. This is the benefit of holding inside money.

From bankers' and entrepreneurs' pre-shock wealth evolution, we can derive the dynamics of η_t . Before the macro shock hits, η_t follows a deterministic path with the following growth rate:

$$\begin{aligned} \mu_t^\eta &= \frac{d\eta_t/dt}{\eta_t} = \frac{n_t^{nf}}{(n_t^f + n_t^{nf})} \left(\frac{dn_t^f/dt}{n_t^f} - \frac{dn_t^{nf}/dt}{n_t^{nf}} \right) \\ &= (1 - \eta_t) \left\{ \left(\frac{a}{q_t} + \mu_t^q - r_t \right) (x_t^f - x_t^{nf}) - \lambda_I \left\{ ROI_t \left[1 - (1 - \phi) x_t^{nf} \right] + (1 - \phi) x_t^{nf} - 1 \right\} \right\}, \end{aligned} \quad (7)$$

where the last line represents the wealth jump of the $\lambda_I dt$ measure of investing entrepreneurs. As will be confirmed by the model solution, the Markov equilibrium has a steady state η^* at which $\mu_t^\eta = 0$. In equilibrium, bankers' wealth has a higher loading on the pre-shock excess return of capital ($x_t^f > x_t^{nf}$), but entrepreneurs' wealth grows via the creation of new capital. When the two forces balance each other ($\mu_t^\eta = 0$), the economy reaches its steady state, and stays there until the macro shock hits.

When the shock hits, both sectors lose wealth. Denote the post-shock net worth as \widetilde{n}_t^i ($i \in \{f, nf\}$). Let “ \sim ” denote the post-shock value. Bankers' wealth changes from n_t^f to \widetilde{n}_t^f :

$$\widetilde{n}_t^f = \left[\frac{\widetilde{q}_t}{q_t} (1 - \underline{\delta}) x_t^f + (1 - x_t^f) \right] n_t^f.$$

A fraction $1 - \underline{\delta}$ of risky capital holdings remain and get re-evaluated at the post-shock price \tilde{q}_t . Entrepreneurs' wealth changes from n_t^{nf} to \tilde{n}_t^{nf} :

$$\tilde{n}_t^{nf} = \left[\frac{\tilde{q}_t}{q_t} (1 - \underline{\delta}) x_{t,k}^{nf} + (1 - x_t^{nf}) \right] n_t^{nf}.$$

Therefore, when the macro shock hits, η_t jumps to $\tilde{\eta}_t$:

$$\tilde{\eta}_t = \frac{\tilde{n}_t^f}{\tilde{n}_t^f + \tilde{n}_t^{nf}} = \frac{\left[\frac{\tilde{q}_t}{q_t} (1 - \underline{\delta}) x_t^f + (1 - x_t^f) \right] \eta_t}{\left[\frac{\tilde{q}_t}{q_t} (1 - \underline{\delta}) x_t^f + (1 - x_t^f) \right] \eta_t + \left[\frac{\tilde{q}_t}{q_t} (1 - \underline{\delta}) x_t^{nf} + (1 - x_t^{nf}) \right] (1 - \eta_t)}. \quad (8)$$

Since all endogenous variables are functions of η_t , when η_t jumps to the new value $\tilde{\eta}_t$, all the other variables also jump. For example, the capital price jumps from q_t to \tilde{q}_t .

Proposition 2 *The Markov equilibrium has a single state variable η_t , whose dynamics are summarized in Equations (7) and (8).*

Optimization. Next, I will solve agents' optimal decisions. Let $V^f(n_t^f, \eta_t)$ denote a banker's value function, which depends on the individual's wealth and the aggregate state. The Hamilton–Jacobi–Bellman equation (HJB) is:

$$\rho V^f(n_t^f, \eta_t) = \max_{c_t^f \geq 0, x_t^f \geq 0} \ln(c_t^f) + \frac{\partial V^f}{\partial n_t^f} \mu_t^{nf} n_t^f + \frac{\partial V^f}{\partial \eta_t} \mu_t^\eta \eta_t + \lambda \left[V^f(\tilde{n}_t^f, \tilde{\eta}_t) - V^f(n_t^f, \eta_t) \right].$$

The last term reflects the jump in value function when the net worth of jumps from n_t^f to \tilde{n}_t^f after the macro shock. The log utility implies the functional form of $V^f(n_t^f, \eta_t)$ is

$$V^f(n_t^f, \eta_t) = \frac{1}{\rho} \ln(n_t^f) + U^f(\eta_t).$$

After substituting this function into the HJB equation, we have the following proposition.

Proposition 3 *Bankers' optimal consumption (c_t^f) and leverage (i.e., capital-to-wealth ratio x_t^f)*

satisfy the following first-order conditions (F.O.C.) respectively:

$$c_t^f = \rho n_t^f,$$

and

$$\frac{a}{q_t} + \mu_t^q - r_t \leq \lambda \left(\frac{1 - \frac{\tilde{q}_t}{q_t} (1 - \underline{\delta})}{1 - \left[1 - \frac{\tilde{q}_t}{q_t} (1 - \underline{\delta})\right] x_t^f} \right), \quad (9)$$

which hold in equality if $x_t^f > 0$.

In equilibrium, the F.O.C. condition for x_t^f holds in equality, because bankers always hold some risky capital (i.e., $x_t^f > 0$). Risk is measured by $1 - \frac{\tilde{q}_t}{q_t} (1 - \underline{\delta})$, the total loss per dollar of risky capital holdings (i.e., the different before the pre-shock value, 1, and the post-shock value, $\frac{\tilde{q}_t}{q_t} (1 - \underline{\delta})$). Risk has two components, the exogenous risk $\underline{\delta}$ and the endogenous risk $\frac{\tilde{q}_t}{q_t}$ from capital reevaluation. Increasing one dollar investment in capital brings a potential loss of wealth equal to $1 - \frac{\tilde{q}_t}{q_t} (1 - \underline{\delta})$. Due to log utility, the marginal value of wealth is the reciprocal of wealth level, so the right-hand side of Equation (9) is the expected marginal loss of value by increasing leverage x_t^f . The left-hand side is the expected marginal gain, the excess return (i.e. the difference between the return on capital and the debt cost r_t). Holding constant the pre-shock excess return (i.e., the left-hand side of Equation (9)), the fraction of bankers' wealth invested in capital (or leverage), x_t^f , decreases in risk.

Let $V^{nf}(n_t^{nf}, \eta_t)$ denote an entrepreneur's value function. The HJB equation is:

$$\begin{aligned} \rho V^{nf}(n_t^{nf}, \eta_t) &= \max_{c_t^{nf} \geq 0, x_t^{nf} \geq 0} \ln(c_t^{nf}) + \frac{\partial V^{nf}}{\partial n_t^{nf}} \mu_t^{nf} n_t^{nf} + \frac{\partial V^{nf}}{\partial \eta_t} \mu_t^\eta \eta_t \\ &\quad + \lambda \left[V^{nf}(\tilde{n}_t^{nf}, \tilde{\eta}_t) - V^{nf}(n_t^{nf}, \eta_t) \right] \\ &\quad + \lambda_I \left[V^{nf}(\hat{n}_t^{nf}, \eta_t) - V^{nf}(n_t^{nf}, \eta_t) \right]. \end{aligned}$$

The last line reflects the jump in value function when the net worth jumps from n_t^{nf} to \hat{n}_t^{nf} . The

log utility implies the functional form of $V^{nf} (n_t^{nf}, \eta_t)$ is

$$V^{nf} (n_t^{nf}, \eta_t) = \frac{1}{\rho} \ln (n_t^{nf}) + U^{nf} (\eta_t).$$

Substituting this conjecture into the HJB equation, we have the following proposition.

Proposition 4 *Entrepreneurs' optimal consumption (c_t^{nf}) and capital-to-wealth ratio (x_t^{nf}) satisfy the following first-order conditions (F.O.C.) respectively:*

$$c_t^{nf} = \rho n_t^{nf},$$

and

$$\begin{aligned} \frac{a}{q_t} + \mu_t^q - r_t \leq & \lambda \left(\frac{1 - \frac{\tilde{q}_t}{q_t} (1 - \bar{\delta})}{1 - \left[1 - \frac{\tilde{q}_t}{q_t} (1 - \bar{\delta}) \right] x_t^{nf}} \right) \\ & + \lambda_I \left\{ \frac{(ROI_t - 1) (1 - \phi)}{ROI_t \left[1 - (1 - \phi) x_t^{nf} \right] + (1 - \phi) x_t^{nf}} \right\}, \end{aligned} \quad (10)$$

which holds in equality if $x_t^{nf} > 0$.

In comparison with bankers' F.O.C. condition (Equation (9)), the right-hand side of Equation (10) has a similar risk compensation term and a new illiquidity compensation term. The last line of Equation (10) reflects the illiquidity of capital. Switching one dollar from money to capital, entrepreneurs lose $(1 - \phi)$ dollars of internal liquidity for investments, which in turn reduces the positive jump in wealth by $(ROI_t - 1) (1 - \phi)$ as shown in Equation (6). Due to log utility, the marginal value of wealth is the reciprocal of wealth level, so the last line of Equation (10) is the expected marginal loss of value by switching one dollar of wealth from money to capital. For entrepreneurs to hold capital (i.e. $x_t^{nf} > 0$ and the F.O.C. condition holds in equality), the excess return of risky capital holdings must be sufficient to compensate entrepreneurs for both the risk exposure and the illiquidity.

Aggregation. In aggregate, investing entrepreneurs converts $i_t \lambda_I dt$ units of goods into new capital, where individual entrepreneur's investment, i_t , is given by Equation (5). Thus, the aggregate amount of goods invested is

$$\left(\frac{1}{1 - \theta q_t} \right) \left[1 - (1 - \phi) x_t^{nf} \right] N_t^{nf} \lambda_I dt,$$

where N_t^{nf} is the aggregate wealth of entrepreneurs. Similarly, let N_t^f denote the aggregate wealth of bankers. The aggregate consumption is given by $\rho N_t^f dt + \rho N_t^{nf} dt$. Note that because debt is in zero net supply, the aggregate wealth of the economy is $q_t K_t$ (i.e., the total value of productive capital). Therefore, the aggregate consumption is $\rho q_t K_t dt$. In equilibrium, the aggregate production, $a K_t dt$, is equal to the sum of consumption and investment:

$$a K_t dt = \rho N_t^f dt + \rho N_t^{nf} dt + \left[\frac{1 - (1 - \phi) x_t^{nf}}{1 - \theta q_t} \right] N_t^{nf} \lambda_I dt.$$

Dividing both sides by $K_t dt$, we have the goods market clearing condition:

$$a = \rho q_t + \left[\frac{1 - (1 - \phi) x_t^{nf}}{1 - \theta q_t} \right] (1 - \eta_t) q_t \lambda_I. \quad (11)$$

The capital market clearing conditions is

$$x_t^f N_t^f + x_t^{nf} N_t^{nf} = q_t K_t,$$

which, after dividing both sides by the aggregate wealth (i.e., $q_t K_t$), can be simplified to

$$x_t^f \eta_t + x_t^{nf} (1 - \eta_t) = 1. \quad (12)$$

The debt market clears automatically by Walras' law. Given any initial capital stock K_0 , the Markov equilibrium is formally defined as follows.

Proposition 5 *For any initial capital endowments of bankers and entrepreneurs, there is a Markov equilibrium that is described by the stochastic processes of bankers' consumption (c_t^f) and portfolio choice (x_t^f), entrepreneurs' consumption (c_t^{nf}), portfolio choice (x_t^{nf}), and investment (i_t), and the price variables (i.e., capital price q_t and interest rate r_t) on the filtered probability space generated by the Poisson process of macro shock, such that*

(1) *agents know and take as given the processes of price variables;*

(2) *agents make optimal choices as stated in Proposition 1, 3, and 4;*

(3) *price variables adjust to clear the markets;*

(4) *all the choice and price variables are functions of η_t , so Proposition 2 gives an autonomous law of motion that maps any path of macro shocks to the current state η_t .*

2.3 Mechanism and empirical implications

The model has two mechanisms that amplify each other and lead to financial instability. Bankers add value to the economy in two ways. First, the macro shock destroys a smaller fraction of capital, when capital is held by bankers instead of entrepreneurs ($\underline{\delta} < \bar{\delta}$). This is the specialness of the asset side of bankers' balance-sheet.¹⁵ Second, bankers are money creators. By issuing short-term safe debt that serves as inside money, they provide the most liquid securities that entrepreneurs can hold against liquidity shocks. Inside money allows investing entrepreneurs to purchase goods that are used to create new capital, and thus, directly affects economic growth. To what extent bankers can provide these two services depends on their wealth (i.e., equity).

Consider a calm period without any macro shocks. As shown in the wealth evolution equation (Equation (4)), bankers accumulate equity by earning the spread between return on capital and debt cost. As bankers get richer, they hold more capital, and as capital gets reallocated

¹⁵If we interpret capital as efficiency units or productive projects, bankers' superiority in maintaining existing capital is related to the credit view on banking that emphasizes bankers' expertise in extending credit and collect repayments.

towards its natural buyers, the market price of capital increases. An upward reevaluation of capital increases the value of assets on bankers' balance-sheet, and through leverage, it increases bankers' equity even further. This positive feedback loop, or a balance-sheet channel, echoes the one considered by Brunnermeier and Sannikov (2014). In good times, the increases in asset price and in natural buyers' risk-taking capacity reinforce each other. However, the story does not end here.

When asset price increases, return on internal liquidity increases ($dROI_t/dq_t > 0$), so entrepreneurs require a higher compensation for capital illiquidity (equivalently, assign a larger liquidity premium on money) as shown by Proposition 4. As a result, the entrepreneurs' portfolio is more biased towards bank debt. Intuitively, when capital becomes more valuable, entrepreneurs want to create more capital, but to do so, they need to build up savings, ideally in the most liquid form (i.e., money). The increasing money demand pushes down the market interest rate r_t , which is banks' debt cost, and pushes up the equilibrium quantity of bank debt, which allows bankers to expand their balance-sheet even faster than what is induced by the increases in bank equity. As a result, bankers have more ammunition to push up the asset price even further, which in turn feeds into the aforementioned balance-sheet channel.

The mutually reinforcing mechanisms are illustrated by Figure 1 in Introduction. During a calm period without macro shocks, the economy exhibits the following features:

- 1 *Asset (capital) price increases;*
- 2 *Capital is reallocated towards the financial sector (banks);*
- 3 *The financial sector grows, fueled by increasing amount of debt;*
- 4 *Firms (entrepreneurs) hold more cash (inside money issued by banks);*
- 5 *Interest rate declines.*

The model's calm period reproduces several key features of the two decades before the 2007-09 financial crisis in the United States. During that time, asset prices rose across different

asset classes. The financial sector grows dramatically (Schularick and Taylor (2012); Greenwood and Scharfstein (2013)). The growth is funded by the issuance of short-term safe assets (Adrian and Shin (2010); Gorton and Metrick (2012); Pozsar (2014); Carlson, Duygan-Bump, Natalucci, Nelson, Ochoa, Stein, and Van den Heuvel (2016)).

There is a large empirical literature that documents the increases of cash holdings in the non-financial corporate sector (e.g., Opler, Pinkowitz, Stulz, and Williamson (1999); Bates, Kahle, and Stulz (2009)), and many have attributed the increase to the firms in the new-economy industries that are R&D intensive (Falato and Sim (2014); Begenau and Palazzo (2015); Pinkowitz, Stulz, and Williamson (2015); Graham and Leary (2015)).

Another salient feature during this period is the secular decline of interest rate. A prominent explanation is the global imbalance (Caballero, Farhi, and Gourinchas (2008)). This paper turns the attention to domestic money demand, and in particular, the corporate liquidity hoarding for investment needs.¹⁶ The repeal of Regulation Q essentially introduces interest-paying money (Lucas and Nicolini (2015)). Therefore, when the money demand is strong, the yield on money (i.e., the interest rate) tends to decline. The low interest rate allows banks to borrow at a low cost, and thus, to take on more risks.

Many have argued that the development of shadow banking sector is driven by the strong demand for money-like securities.¹⁷ The financial intermediaries' privilege to create money is a double-edged sword: on the one hand, inside money facilitates transaction and resources reallocation; on the other hand, it fuels the financial sector's risk-taking. When the macro shock hits, the feedback loops in Figure 1 turn into vicious cycles.

Consider the economy hit by the macro shock. As a fraction of capital holdings are destroyed, bankers lose their equity. As the natural buyers retreat from the market, capital price declines, which erodes bankers' equity even further. Moreover, as capital price decreases, en-

¹⁶Investment need is a key determinant of the cross-sectional variation in corporate cash holdings (e.g., Denis and Sibilkov (2010); Duchin (2010)), especially for firms with less collateral (e.g., Almeida and Campello (2007); Li, Whited, and Wu (2016)) and more intensive R&D activities (Falato and Sim (2014)).

¹⁷As pointed out by Gorton (2010), Gorton and Metrick (2012), and Stein (2012), circumventing regulations on leverage to capture the money premium is one of driving forces behind the growth of shadow banking.

trepreneurs find investments less profitable (i.e., the investment return on internal liquidity, ROI_t , is lower), which decreases entrepreneurs' incentive to hoard money. Their money demand contracts, leading to higher debt cost for the bankers and a smaller equilibrium quantity of bank debt, which precipitate the balance-sheet contraction in the banking sector and result in even lower capital price.

A key assumption is that bankers cannot raise equity from entrepreneurs. Otherwise, the effect of macro shock will be dissipated across the whole economy instead of concentrated on the bankers. In fact, if bankers can raise equity from entrepreneurs, they will hold all the capital, so capital price will be a constant at the first-best level and the amplification mechanisms are muted.¹⁸ The model imposes the restriction that the only financial contract between bankers and entrepreneurs is risk-free debt. Allowing banks to issue equity can improve the model's quantitative performance and may introduce new mechanisms like He and Krishnamurthy (2013), but as long as there are some frictions on bankers' equity issuance, the model's qualitative implications carry through.

When the macro shock hits, the state variable η_t will jump downward to $\tilde{\eta}_t$, because in equilibrium, bankers have a larger exposure to macro shocks (i.e., $x_t^f > x_t^{nf}$). As a result, the capital price decreases, jumping from q_t to \tilde{q}_t . In contrast to the exogenous capital destruction, capital reevaluation is an endogenous risk. The strength of the model's shock amplification mechanism is measured by the following ratio,

$$\text{Endogenous amplification strength: } \frac{1 - \frac{\tilde{q}_t}{q_t} (1 - \delta)}{\underline{\delta}}. \quad (13)$$

From bankers' perspective, it is the ratio of total loss to exogenous loss per dollar of risky capital holdings. A large value indicates the model's shock amplification mechanisms are strong. Next, I will calibrate the parameters and fully solve the model. The amplified impact can be three times

¹⁸Equity issuance is mathematically equivalent to perfect risk-sharing (i.e., bankers and entrepreneurs can freely bet on the macro shock). Di Tella (2014) show that perfect risk-sharing shuts down the balance-sheet channel of asset price variation.

larger.

3 Solution

Based on the numeric solution (details in the Appendix), this section shows the model's properties graphically, and in particular, how the model generates the stylized facts in booms and how the feedback mechanisms lead to endogenous risk that materializes in crises.

3.1 Calibration

One unit of time corresponds to a month. I set λ_I equal to 0.0125, which implies 5% arrival rate at quarterly frequency in line with the calibration of Kiyotaki and Moore (2012). λ is $\frac{1}{60}$, so the macro shock arrives every five years, which is roughly consistent with the NBER business cycle frequency in the post-war period. I set $\underline{\delta} = 0.025$ and $\bar{\delta} = 0.05$, which is broadly consistent with the estimates of default loss in Andrade and Kaplan (1998) and Chen (2010). ρ is set to 0.0025 (annualized to 3%). a set to 0.0167, which implies a capital price-to-production ratio around 8 over the cycle. θ , the external financing capacity of investment projects, is set to 10%.

ϕ is set to 0.9 in the baseline calibration. It is the key parameter in this model, as it measures the fraction of existing capital that can be pledged or readily sold when an investment opportunity arrives. Lower value of ϕ corresponds higher illiquidity of capital, which I interpret as a higher level of intangibility. Fixing other parameters, we can think of the economy as indexed by ϕ . Decreasing ϕ corresponds to an unexpected and permanent increase in the intangibility of productive capital in this economy.

If $\phi = 1$, the inside money channel is shut down. An increase in the capital price and the profitability of investment does not increase entrepreneurs' preference for bank debt, because the holdings of bank debt and existing capital are equally liquid when investment opportunities arrive. The inside money channel is shut down. As will be shown at the end of this section, a

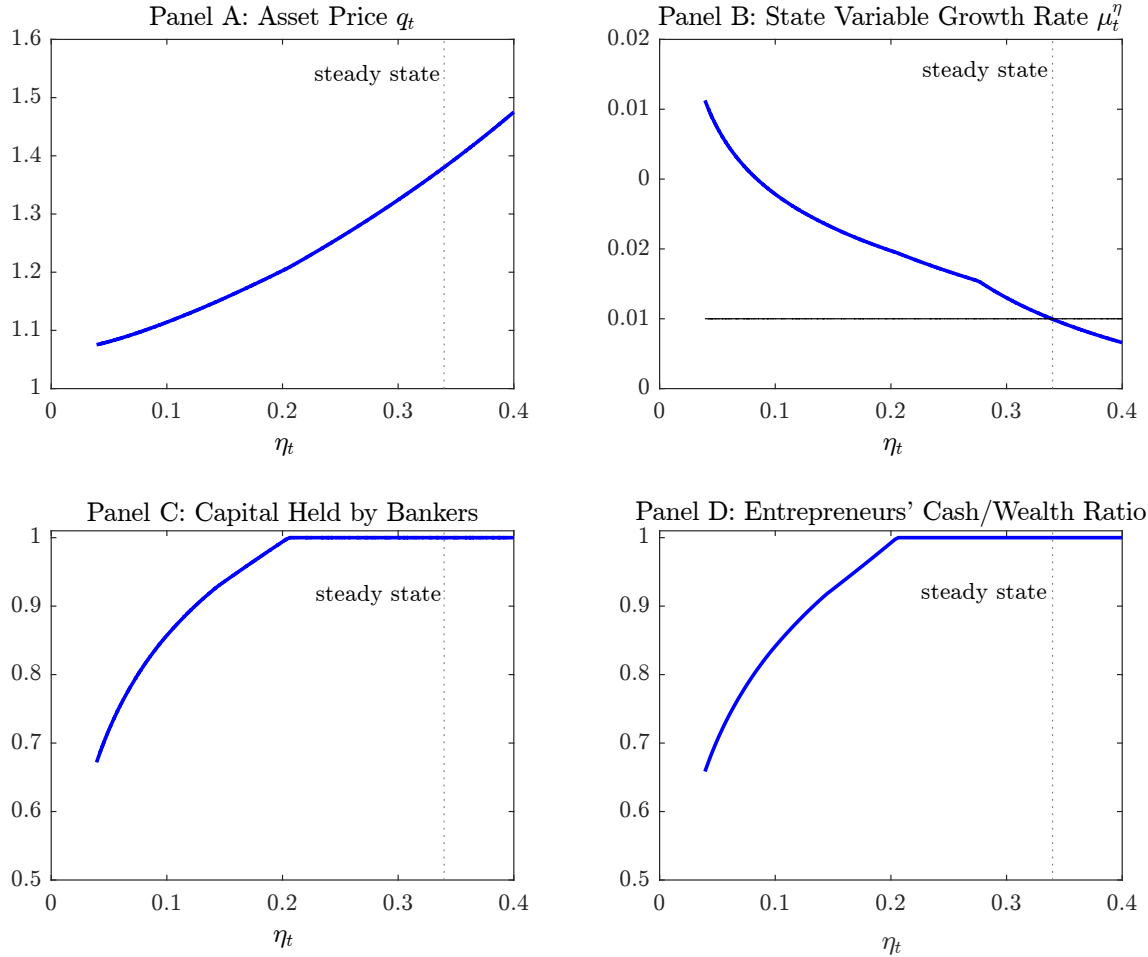


Figure 2: Equilibrium Capital Price and Allocation.

lower value of ϕ strengthens the model's shock amplification mechanism, making the economy more unstable.

3.2 Equilibrium properties

Asset price. Panel A of Figure 2 shows the equilibrium asset price as a function of bankers' wealth share η_t . Capital price increases in bankers' wealth and risk-taking capacity for two reasons. First, when hit by the macro shock, it suffers less loss than entrepreneurs ($\underline{\delta} < \bar{\delta}$). Second, bankers' marginal cost of financing (i.e. their debt cost) is lower than entrepreneurs', because of the liquidity premium assigned to bank debt.

State variable dynamics. The drift of state variable η_t is shown in Panel B of Figure 2. A steady state exists around $\eta_t = 0.34$, which is marked by the dotted vertical lines in the four panels of Figure 2 and the upcoming figures. When the financial sector's wealth share reaches 34%, the drift of state variable is equal to zero, so the economy will stay there until hit by a macro shock. This steady state is stable, because to the left, the growth rate of η_t is positive (i.e., $\mu_t^\eta > 0$), and to the right, $\mu_t^\eta < 0$. The growth rate is large where η_t is small. This is because the bankers are earning a higher spread between the return on capital and their debt cost, and the return of capital is large because capital price is low when η_t is low.

According to simulation results, if the economy starts at $\eta_t = 5\%$, it takes an average of 125 years to reach the steady state $\eta = 34\%$. During the process, we observe the expansion of the financial sector (i.e., increases in η_t), increasing asset price (q_t), and an increasing share of capital being held by the financial sector ($x_k^f \eta_t$). Therefore, the calibrated model speaks to the relatively long-run dynamics of the economy, for instance, the last two decades of the U.S. economy, instead of business-cycle variations.

Capital allocation. Panel C of Figure 2 shows the fraction of capital held by bankers. When the banking sector is still small, entrepreneurs have to hold capital, which happens when $\eta_t < 0.2$. Entrepreneurs' first-order condition for $x_t^{n,f}$ holds in equality (Equation (10)). The expected return on capital must compensate for both the exposure to macro shock and the illiquidity. Because of $\underline{\delta} < \bar{\delta}$, it is inefficient for the entrepreneurs to hold existing capital, but it has to when the natural buyer's risk-taking capacity is limited. Close to the steady state, all the existing capital is held by bankers, so entrepreneurs focus on creating new capital and invests all of its wealth in inside money issued by bankers. In other words, when η_t is high, the economy achieves an efficient specialization: bankers hold capital, and entrepreneurs hold money and create capital (Panel D of Figure 2). The subsequent figures mark the inefficiency point, below which entrepreneurs hold capital.

Endogenous risk and risk pricing. Panel A of Figure 3 shows the pre-shock excess return of

capital

$$\frac{a}{q_t} + \mu_t^q - r_t.$$

Panel B shows bankers' total loss faced per dollar of risky capital holdings when the macro shock hits

$$1 - \frac{\tilde{q}}{q} (1 - \underline{\delta}).$$

I use bankers' total loss as a measure of risk, because bankers are always the marginal investor, while entrepreneurs exit the market when η_t is large enough. As the banking sector grows and the economy moves towards the steady state, endogenous risk accumulates. The bankers' total loss per dollar increases from around 2.5% (i.e. without any endogenous risk) to around 9%. As a consequence, bankers require a higher pre-shock excess return as compensation.

When entrepreneurs hold risky asset, Equation (10) holds in equality and it decomposes the pre-shock excess return into two components: the macro risk compensation

$$\lambda \left(\frac{1 - \frac{\tilde{q}_t}{q_t} (1 - \bar{\delta})}{1 - \left[1 - \frac{\tilde{q}_t}{q_t} (1 - \bar{\delta}) \right] x_t^{nf}} \right),$$

and the illiquidity compensation

$$\lambda_I \left(\frac{ROI_t (1 - \phi) - (1 - \phi)}{ROI_t (1 - (1 - \phi) x_t^{nf}) + (1 - \phi) x_t^{nf}} \right). \quad (14)$$

Because entrepreneurs can only invest in two assets, the risky and illiquid capital and the safe debt, the illiquidity compensation also reflects the liquidity premium that assigned to the safe debt issued by bankers. The red dotted line in Panel A of Figure 3 shows the macro risk compensation required by entrepreneurs. The gap between this line and the pre-shock excess return is the illiquidity compensation. It grows as η_t increases and capital price increases, because capital asset price leads to a higher return on internal liquidity.

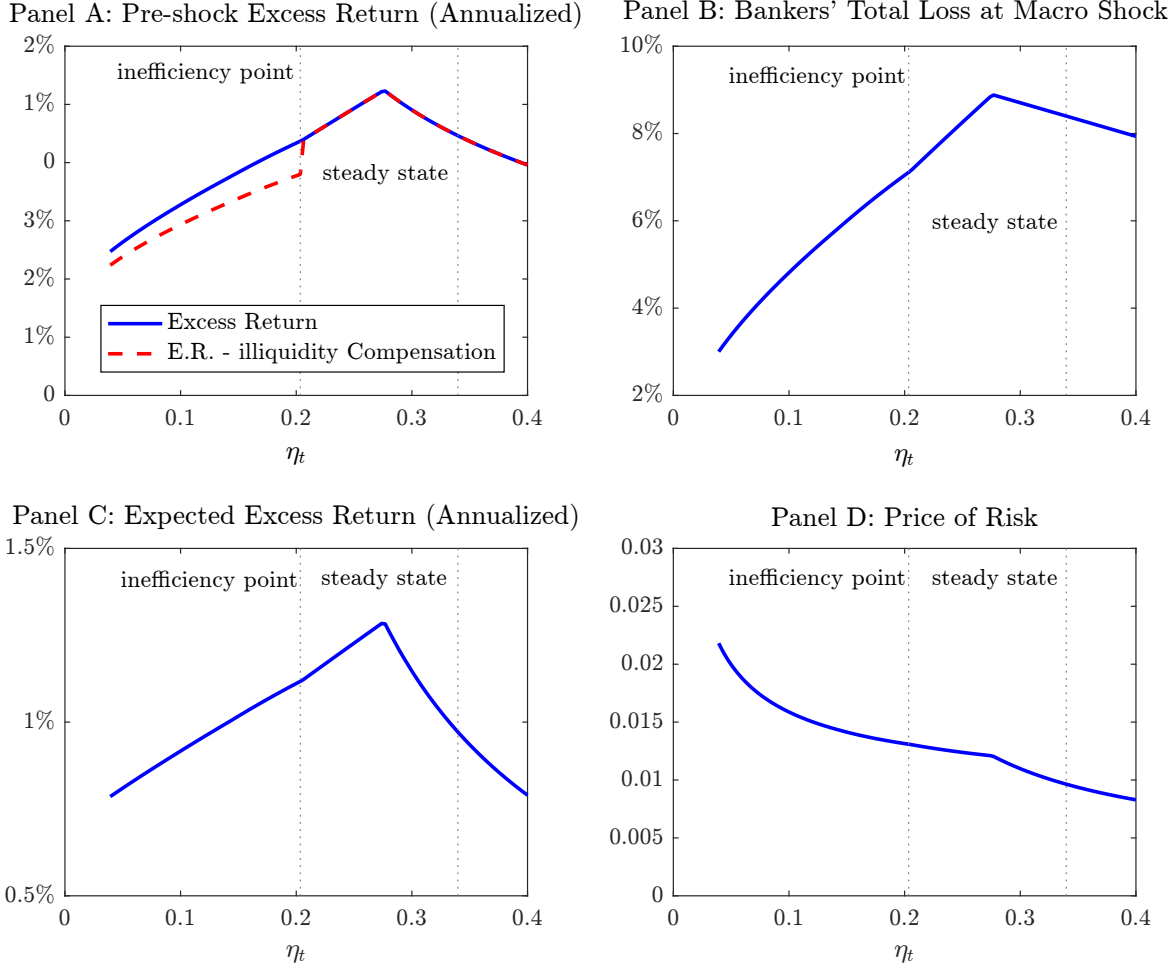


Figure 3: Endogenous Risk and Risk Pricing.

The expected excess return is the pre-shock excess return adjusted by the expected loss

$$\frac{a}{q_t} + \mu_t^q - r_t + \lambda(1 - \underline{\delta}) \left(\frac{\tilde{q}_t}{q_t} - 1 \right) + \lambda \underline{\delta} (-1).$$

It follows the same pattern as the pre-shock excess return (Panel C of Figure 3).

Panel D shows the “price of risk”, which is the ratio of expected excess return divided by total loss per dollar of investment. It decreases as the bankers’ wealth share increases. Thus, the model economy generates counter-cyclical price of risk, which is well documented by the empirical asset pricing literature (e.g., Lettau and Ludvigson (2010)). The dynamics of risk price is dominated by the dynamics of quantity of risk (Panel B). As a result, the dynamics of expected

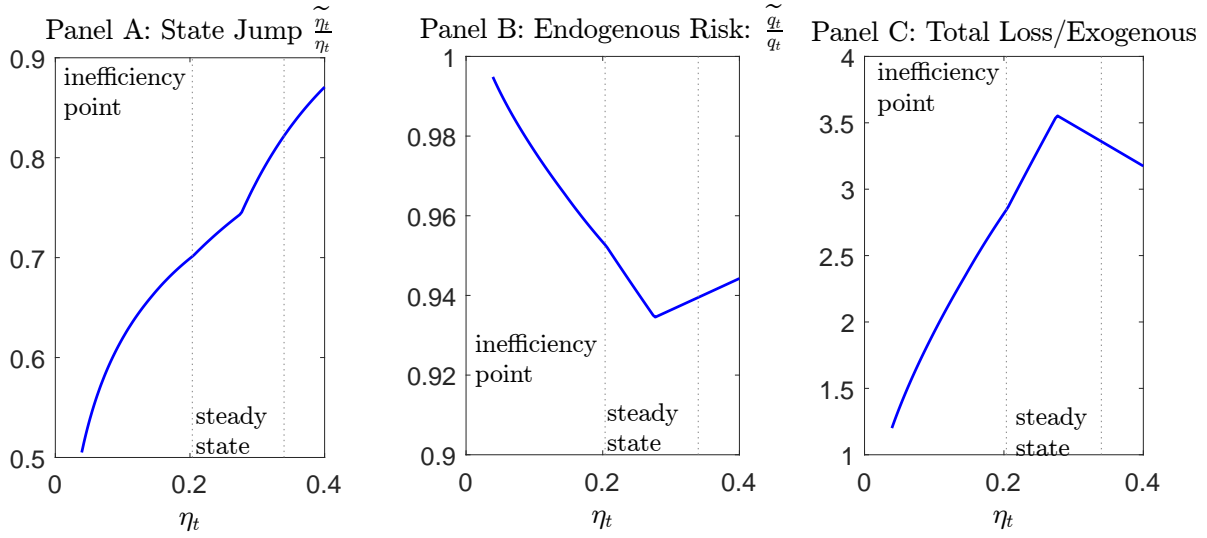


Figure 4: Risk Accumulation and Amplification.

excess return (Panel C) follows the quantity of risk instead of risk price.

Financial instability. Panel A of Figure 4 shows the ratio of post-shock state to the pre-shock state (i.e., $\frac{\tilde{\eta}_t}{\eta_t}$). It shows the impact of macro shock on the state variable. This ratio increases as η_t increases. Therefore, as the banking sector grows, it becomes more stable (or less sensitive to shocks). However, this does not necessarily lead to smaller endogenous risk.

Panel B of Figure 4 shows the ratio of post-shock capital price to pre-shock asset price (i.e., $\frac{\tilde{q}_t}{q_t}$). When η_t is relatively small, this ratio decreases as η_t increases, which means endogenous risk accumulates as the banking sector grows. Going through a long calm period, bankers accumulate equity, which has two effects: first, it makes the banking sector more robust to shocks (direct effect); second, it increases asset price (equilibrium effect).

A higher asset price increases endogenous risk through the inside money channel. When capital price q_t increases, entrepreneurs' return on internal liquidity ROI_t increases, which makes them want to bias their wealth portfolio more towards the inside money (i.e. bank debt). This reduces bankers' debt cost, giving bankers a larger advantage over entrepreneurs in holding capital. In other words, through entrepreneurs' investment-driven money demand, an increase in

asset price amplifies the difference between the first-best buyers of capital (i.e., bankers) and the second-best buyers of capital (i.e., entrepreneurs). As a result, when the macro shock hits and the reallocation towards second-best buyers takes place, capital price decreases more. When the banking sector is sufficiently well-capitalized, this mechanism (i.e., the equilibrium effect) is overwhelmed by the direct effect.

Panel C of Figure 3 shows the strength of shock amplification in the model economy. The ratio of total loss (i.e., $1 - \frac{\tilde{q}_t}{q_t} (1 - \delta)$) to exogenous loss (i.e., δ) increases to above 3.5, and then decreases as the direct effect of bank equity accumulation dominates the equilibrium effect that amplifies the shock through the inside money channel. This amplification effect is quantitatively large in comparison with the existing macro finance literature (e.g., Brunnermeier and Sannikov (2014)).

Inside money channel. Panel A of Figure 5 shows an entrepreneur's return on internal liquidity. As η_t grows, capital price grows, which leads to higher ROI_t , and thus, a higher illiquidity compensation required by entrepreneurs to hold capital (Panel B). Note that after the inefficiency point, entrepreneurs no longer hold capital, so the plot stops. This confirms the previous explanation of endogenous risk accumulation through the widening difference between the first-best (bankers) and second-best holders (entrepreneurs) of capital that is due to entrepreneurs' aversion to capital illiquidity at investment times.

As shown in Panel C of Figure 5, a stronger money demand of entrepreneurs pushes down the interest rate (i.e. the risk-free rate), which reduces the leverage cost for bankers. Panel D plots $(1 - \eta_t) (1 - x_t^{nf})$ (i.e., the inside money to total wealth ratio). $(1 - \eta_t)$ share of wealth held by entrepreneurs, of which $(1 - x_t^{nf})$ fraction is invested in risk-free debt issued by bankers. When the banking sector is small, its money creation capacity is restricted, because it does not have enough net worth to buffer the macro shock. The liquidity supply increases as the banking sector grows and its leverage cost decreases due to a rising liquidity premium assigned to bank debt. After the inefficiency point, bankers already hold all the capital, so asset side of their

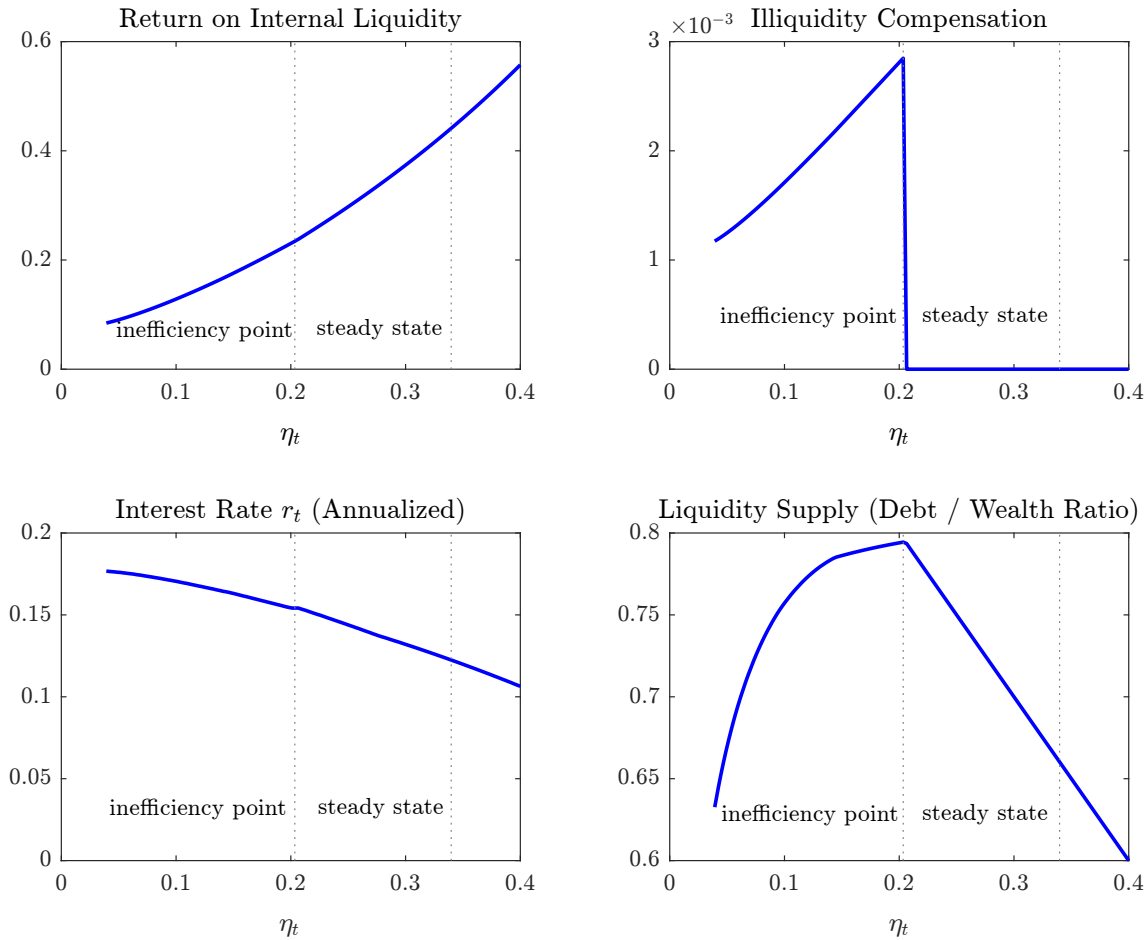


Figure 5: Inside Money Channel.

balance-sheet can only grow through the upward reevaluation of capital instead of reevaluation and reallocation from entrepreneurs to bankers. As a result, a further increase of bank equity crowds out bank debt.

Panel C and D of Figure 5 reproduce the experience of the United States and other advanced economies in the last three decades. Through the inside money channel, the interest rate declines, and at the same time, the indebtedness of the financial sector increases (Schularick and Taylor (2012)). As shown in Figure 4, endogenous risk accumulates along with the increasing capital price, setting the stage of severe crises when the macro shock hits.

Simulation. Figure 6 shows a simulated path of the economy for 30 years (360 months). The

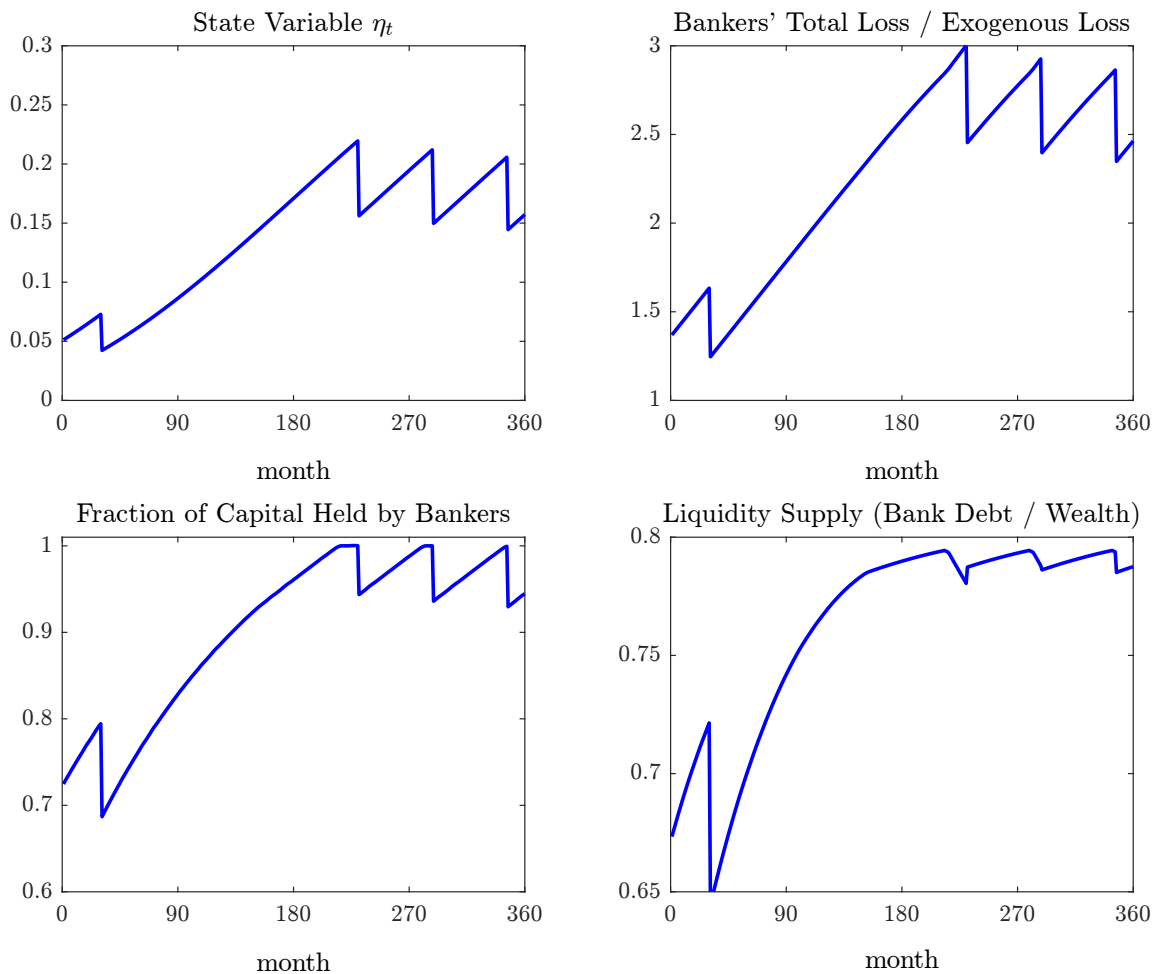


Figure 6: Simulation.

economy starts from $\eta_t = 5\%$. After being hit by the macro shock at the 40th month, the economy experiences a prolonged period of boom. Panel A shows the trajectory of the state variable η_t . The banking sector's wealth share grows to more than 20%. Bankers increase risk-taking, holding a larger share of capital (Panel C). By issuing more inside money, bankers provide liquidity to entrepreneurs (Panel D). Along the process, the economy accumulates endogenous risk (Panel B). The boom is followed by three recessions. Each recession lasts for three to four years, during which the banking sector cannot hold all the capital, so entrepreneurs have to hold capital and the economy deviates from the efficient specialization.

Intangibility and instability. When $\phi = 1$, the inside money channel is shut down, because

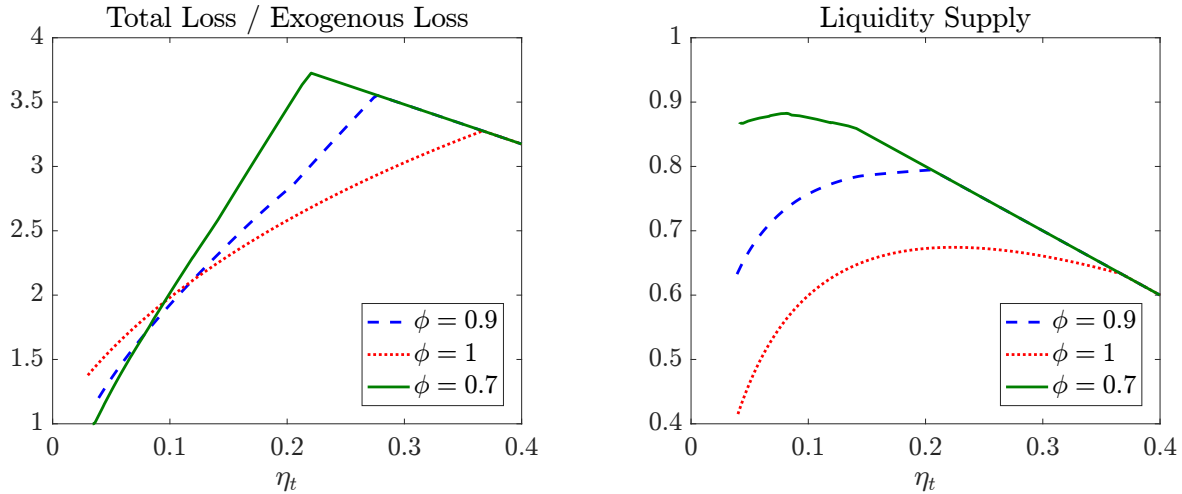


Figure 7: Financial Instability in an Intangible Economy.

investment opportunities arrive, bank debt and capital are equally liquid and both can be freely exchanged for investment inputs. The investing entrepreneurs can invest all of their wealth in creating new capital.¹⁹ When $\phi < 1$, bank debt is more liquid than capital, and thus, entrepreneurs bias their wealth allocation towards bank debt, giving rise to the inside money channel. A lower value of ϕ means capital is more illiquid at the investment times. Note that investing entrepreneurs can either sell existing capital for investment inputs, or equivalently, pledging existing capital to borrow investment inputs. The latter form of capital liquidity emphasizes collateralizability, which is closely related to the capital tangibility. Thus, I interpret a decrease in ϕ as a structural transformation towards a more intangible economy.

Figure 7 compares the model's performances, and in particular, the strength of the shock amplification mechanism across different values of ϕ ($= 0.7, 0.9$, and 1). Moving from $\phi = 1$ to $\phi = 0.9$ and 0.7 , the shock amplification mechanism is strengthened (Panel A). When capital becomes more intangible, entrepreneurs assign a higher liquidity premium to bank debt, leading to a stronger inside money channel of financial instability. In the case of $\phi = 1$, the only shock amplification mechanism is the standard balance-sheet channel (Brunnermeier and

¹⁹They are still financially constrained, in the sense that the creation of new capital cannot fully rely on external funding (i.e., $\theta < 1$).

Sannikov (2014)). Therefore, by comparing it with the cases of $\phi < 1$, we see how much the inside money channel contributes to the endogenous risk.

The economy's indebtedness decreases in ϕ . When the economy becomes more intangible, bankers cater to entrepreneurs' money demand by issuing more safe debt. Panel B of Figure 7 shows the bank debt to total wealth ratio. As the economy becomes more intangible-intensive, the capital illiquidity problem becomes more severe. This drives up the production sector's demand for inside money in anticipation of investment needs, and thereby, pushes down the leverage cost for banks and leads to a more indebted banking sector.

4 Conclusion

This paper aims to provide a unified and coherent explanation of several trends in the U.S. economy in the two decades leading up to the Great Recession, such as the rising corporate money demand, the expansion of the financial sector, the declining interest rate, and the increases in risky asset prices. At the center of my model is the endogenous dynamics of money supply (bankers create inside money) and money demand (entrepreneurs hoard liquidity for investments). Inspired by the U.S. experience of structural transformation towards an intangible economy, this paper proposes an inside money channel of financial instability that reinforces the standard balance-sheet channel and leads to the endogenous risk accumulation in booms. This new channel helps explain the recent empirical findings that a long period of bank expansion precedes severe crises.

The model leaves out the active provision of outside money for future research. By expanding money supply in booms, the government can crowd out bank leverage, and thereby, weaken the upward spiral in asset price that leads to endogenous risk accumulation. However, since entrepreneurs' incentive to invest is tied to asset prices, the government faces the trade-off between growth and stability. In contrast to the current practice of monetary expansion in recessions, an optimal strategy of outside money supply tends to be procyclical in this setting.

The model also leaves out banks' default. The empirical literature on financial crises com-

monly use banks' default or a high possibility of default as a crisis indicator. A theoretical model of crisis should ideally accommodate default, and by doing so, it opens up the question of optimal government intervention, for instance, through equity injection into the banking sector, in order to prevent a sudden evaporation of inside money. To finance the intervention in bad times, the government may increase the issuance of outside money (e.g., short-term, money-like government debt), which suggests countercyclical outside money supply.

References

- ADRIAN, T., AND H. S. SHIN (2010): “Liquidity and leverage,” *Journal of Financial Intermediation*, 19(3), 418 – 437, Risk Transfer Mechanisms and Financial Stability.
- ALMEIDA, H., AND M. CAMPELLO (2007): “Financial Constraints, Asset Tangibility, and Corporate Investment,” *Review of Financial Studies*, 20(5), 1429–1460.
- ANDRADE, G., AND S. N. KAPLAN (1998): “How Costly is Financial (Not Economic) Distress? Evidence from Highly Leveraged Transactions that Became Distressed,” *The Journal of Finance*, 53(5), 1443–1493.
- ATKESON, A., AND P. J. KEHOE (2005): “Modeling and Measuring Organization Capital,” *Journal of Political Economy*, 113(5), 1026–1053.
- AZAR, J., J.-F. KAGY, AND M. C. SCHMALZ (2016): “Can Changes in the Cost of Carry Explain the Dynamics of Corporate Cash Holdings?,” *Review of Financial Studies*, p. forthcoming.
- BARON, M., AND W. XIONG (2016): “Credit Expansion and Neglected Crash Risk,” *Quarterly Journal of Economics*, p. forthcoming.
- BASAK, S., AND D. CUOCO (1998): “An Equilibrium Model with Restricted Stock Market Participation,” *Review of Financial Studies*, 11(2), 309–341.
- BATES, T. W., K. M. KAHLE, AND R. M. STULZ (2009): “Why Do U.S. Firms Hold So Much More Cash than They Used To?,” *The Journal of Finance*, 64(5), 1985–2021.
- BEGENAU, J., AND B. PALAZZO (2015): “Firm Selection and Corporate Cash Holdings,” 2015 Meeting Papers 1047, Society for Economic Dynamics.
- BERNANKE, B. S., AND M. GERTLER (1989): “Agency Costs, Net Worth, and Business Fluctuations,” *The American Economic Review*, 79(1), 14–31.
- BOLTON, P., H. CHEN, AND N. WANG (2011): “A unified theory of Tobin’s q , corporate investment, financing, and risk management,” *The Journal of Finance*, 66(5), 1545–1578.
- BOLTON, P., AND X. FREIXAS (2000): “Equity, Bonds, and Bank Debt: Capital Structure and Financial Market Equilibrium under Asymmetric Information,” *Journal of Political Economy*, 108(2), 324–351.
- BORDO, M., B. EICHENGREEN, D. KLINGEBIEL, M. S. MARTINEZ-PERIA, AND A. K. ROSE (2001): “Is the Crisis Problem Growing More Severe?,” *Economic Policy: A European Forum*, 16(32), 53–82.
- BRUNNERMEIER, M. K., AND Y. SANNIKOV (2014): “A Macroeconomic Model with a Financial Sector,” *American Economic Review*, 104(2), 379–421.
- (2016): “The I Theory of Money,” Working paper, Princeton University.

- CABALLERO, R. J., E. FARHI, AND P.-O. GOURINCHAS (2008): “An Equilibrium Model of “Global Imbalances” and Low Interest Rates,” *American Economic Review*, 98(1), 358–93.
- CARLSON, M., B. DUYGAN-BUMP, F. NATALUCCI, B. NELSON, M. OCHOA, J. STEIN, AND S. VAN DEN HEUVEL (2016): “The Demand for Short-Term, Safe Assets and Financial Stability: Some Evidence and Implications for Central Bank Policies,” *International Journal of Central Banking*, p. forthcoming.
- CHEN, H. (2010): “Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure,” *The Journal of Finance*, 65(6), 2171–2212.
- CORRADO, C., C. HULTEN, AND D. SICHEL (2005): “Measuring Capital and Technology: An Expanded Framework,” in *Measuring Capital in the New Economy*, ed. by C. Corrado, J. Haltiwanger, and D. Sichel, pp. 11–46. University of Chicago Press, San Diego.
- CORRADO, C. A., AND C. R. HULTEN (2010): “How Do You Measure a “Technological Revolution”?” *American Economic Review*, 100(2), 99–104.
- DANG, T. V., G. GORTON, B. HOLMSTRÖM, AND G. ORDONEZ (2014): “Banks as Secret Keepers,” Working Paper 20255, National Bureau of Economic Research.
- DEMARZO, P. M., AND D. DUFFIE (1999): “A Liquidity-based Model of Security Design,” *Econometrica*, 67(1), 65–99.
- DENIS, D. J., AND V. SIBILKOV (2010): “Financial Constraints, Investment, and the Value of Cash Holdings,” *Review of Financial Studies*, 23(1), 247–269.
- DI TELLA, S. (2014): “Uncertainty Shocks and Balance Sheet Recessions,” Working paper, Stanford University Graduate School of Business.
- DIAMOND, D. W. (1984): “Financial Intermediation and Delegated Monitoring,” *The Review of Economic Studies*, 51(3), 393–414.
- DUCHIN, R. (2010): “Cash Holdings and Corporate Diversification,” *The Journal of Finance*, 65(3), 955–992.
- EISFELDT, A., AND A. RAMPINI (2009): “Financing Shortfalls and the Value of Aggregate Liquidity,” Working paper, Duke University.
- EISFELDT, A. L. (2007): “Smoothing with liquid and illiquid assets,” *Journal of Monetary Economics*, 54(6), 1572 – 1586.
- EISFELDT, A. L., AND D. PAPANIKOLAOU (2014): “The Value and Ownership of Intangible Capital,” *American Economic Review*, 104(5), 189–94.
- FALATO, A., AND J. W. SIM (2014): “Why Do Innovative Firms Hold So Much Cash? Evidence from Changes in State R&D Tax Credits,” Finance and Economics Discussion Series 2014-72, Board of Governors of the Federal Reserve System (U.S.).

- FARHI, E., AND J. TIROLE (2011): “Bubbly Liquidity,” *The Review of Economic Studies*.
- FROOT, K. A., D. S. SCHARFSTEIN, AND J. C. STEIN (1993): “Risk Management: Coordinating Corporate Investment and Financing Policies,” *The Journal of Finance*, 48(5), 1629–1658.
- GORTON, G. (2010): *Slapped by the Invisible Hand: The Panic of 2007*. Oxford University Press, New York, N.Y.
- GORTON, G., S. LEWELLEN, AND A. METRICK (2012): “The Safe-Asset Share,” *American Economic Review*, 102(3), 101–06.
- GORTON, G., AND A. METRICK (2012): “Securitized banking and the run on repo,” *Journal of Financial Economics*, 104(3), 425 – 451, Market Institutions, Financial Market Risks and Financial Crisis.
- GORTON, G., AND G. PENNACCHI (1990): “Financial Intermediaries and Liquidity Creation,” *The Journal of Finance*, 45(1), 49–71.
- GRAHAM, J. R., AND M. T. LEARY (2015): “The Evolution of Corporate Cash,” Working paper, Duke University.
- GREENWOOD, R., AND D. SCHARFSTEIN (2013): “The Growth of Finance,” *Journal of Economic Perspectives*, 27(2), 3–28.
- GURLEY, J. G., AND E. S. SHAW (1960): *Money in a Theory of Finance*. Brookings Institution, Washington, D.C.
- HALL, B. H. (1992): “Investment and Research and Development at the Firm Level: Does the Source of Financing Matter?,” Working Paper 4096, National Bureau of Economic Research.
- HALL, B. H., AND J. LERNER (2009): “The Financing of R&D and Innovation,” Working Paper 15325, National Bureau of Economic Research.
- HART, O., AND J. MOORE (1994): “A Theory of Debt Based on the Inalienability of Human Capital,” *The Quarterly Journal of Economics*, 109(4), 841–879.
- HART, O., AND L. ZINGALES (2014): “Banks Are Where The Liquidity Is,” Working paper, Harvard University.
- HE, Z., AND A. KRISHNAMURTHY (2013): “Intermediary Asset Pricing,” *American Economic Review*, 103(2), 732–70.
- HIMMELBERG, C. P., AND B. C. PETERSEN (1994): “R&D and Internal Finance: A Panel Study of Small Firms in High-Tech Industries,” *The Review of Economics and Statistics*, 76(1), 38–51.
- HOLMSTRÖM, B. (2012): “The nature of liquidity provision: When ignorance is bliss,” in *Presidential Address, Econometric Society, ASSA meetings, Chicago*.

- HOLMSTRÖM, B., AND J. TIROLE (1997): “Financial Intermediation, Loanable Funds, and the Real Sector,” *The Quarterly Journal of Economics*, 112(3), 663–691.
- (1998): “Private and Public Supply of Liquidity,” *Journal of Political Economy*, 106(1), 1–40.
- (2001): “LAPM: A Liquidity-Based Asset Pricing Model,” *The Journal of Finance*, 56(5), 1837–1867.
- JORDÀ, O., M. SCHULARICK, AND A. M. TAYLOR (2013): “When Credit Bites Back,” *Journal of Money, Credit and Banking*, 45(s2), 3–28.
- KIYOTAKI, N., AND J. MOORE (1997): “Credit Cycles,” *Journal of Political Economy*, 105(2), 211–248.
- (2000): “Inside money and liquidity,” Working paper, London School of Economics.
- (2005): “Lawrence R. Klein Lecture: Liquidity and asset prices,” *International Economic Review*, 46(2), 317–349.
- (2012): “Liquidity, Business Cycles, and Monetary Policy,” Working paper, Princeton University.
- KRISHNAMURTHY, A., AND T. MUIR (2016): “How Credit Cycles across a Financial Crisis,” Working paper, Stanford University Graduate School of Business.
- KRISHNAMURTHY, A., AND A. VISSING-JORGENSEN (2015): “The impact of Treasury supply on financial sector lending and stability,” *Journal of Financial Economics*, 118(3), 571 – 600.
- KRUSELL, P., AND A. A. SMITH (1998): “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 106(5), 867–896.
- LAGOS, R. (2008): “inside and outside money,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.
- LETTAU, M., AND S. C. LUDVIGSON (2010): “Chapter 11 - Measuring and Modeling Variation in the Risk-Return Trade-off,” in *Handbook of Financial Econometrics: Tools and Techniques*, ed. by Y. Aït-Sahalia, and L. P. Hansen, vol. 1 of *Handbooks in Finance*, pp. 617 – 690. North-Holland, San Diego.
- LEVINE, R. (1997): “Financial Development and Economic Growth: Views and Agenda,” *Journal of Economic Literature*, 35(2), 688–726.
- LI, S., T. M. WHITED, AND Y. WU (2016): “Collateral, Taxes, and Leverage,” *Review of Financial Studies*, 29(6), 1453–1500.
- LONGSTAFF, F. A., AND J. WANG (2012): “Asset Pricing and the Credit Market,” *Review of Financial Studies*, 25(11), 3169–3215.

- LÓPEZ-SALIDO, D., J. C. STEIN, AND E. ZAKRAJŠEK (2015): “Credit-Market Sentiment and the Business Cycle,” Working paper, Harvard University.
- LUCAS, R. E. J., AND J. P. NICOLINI (2015): “On the stability of money demand,” *Journal of Monetary Economics*, 73, 48 – 65.
- MCGRATTAN, E. R. (2016): “Intangible Capital and Measured Productivity,” Working paper, University of Minnesota.
- MCGRATTAN, E. R., AND E. C. PRESCOTT (2010): “Technology Capital and the US Current Account,” *American Economic Review*, 100(4), 1493–1522.
- MOLL, B. (2014): “Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?,” *American Economic Review*, 104(10), 3186–3221.
- OPLER, T., L. PINKOWITZ, R. STULZ, AND R. WILLIAMSON (1999): “The determinants and implications of corporate cash holdings,” *Journal of Financial Economics*, 52(1), 3 – 46.
- PETERS, R. H., AND L. A. TAYLOR (2016): “Intangible Capital and the Investment-q Relation,” *Journal of Financial Economics*, p. forthcoming.
- PIAZZESI, M., AND M. SCHNEIDER (2016): “Payments, Credit and Asset Prices,” Working paper, Stanford University.
- PINKOWITZ, L., R. M. STULZ, AND R. WILLIAMSON (2015): “Do U.S. Firms Hold More Cash than Foreign Firms Do?,” *Review of Financial Studies*.
- POZSAR, Z. (2011): “Institutional Cash Pools and the Triffin Dilemma of the U.S. Banking System,” Working paper WP/11/190, International Monetary Fund.
- (2014): “Shadow Banking: The Money View,” Working paper 14-04, Office of Financial Research.
- QUADRINI, V. (2014): “Bank Liabilities Channel,” CEPR Discussion Papers 10265.
- RAJAN, R. G., AND L. ZINGALES (1998): “Financial Dependence and Growth,” *The American Economic Review*, 88(3), 559–586.
- RAMPINI, A. A., AND S. VISWANATHAN (2015): “Financial Intermediary Capital,” Working paper, Duke University.
- REINHART, C. M., AND K. S. ROGOFF (2009): *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton, NJ.
- SCHULARICK, M., AND A. M. TAYLOR (2012): “Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008,” *American Economic Review*, 102(2), 1029–61.

- SHLEIFER, A., AND R. W. VISHNY (1992): “Liquidation Values and Debt Capacity: A Market Equilibrium Approach,” *The Journal of Finance*, 47(4), 1343–1366.
- STEIN, J. C. (2012): “Monetary Policy as Financial Stability Regulation,” *The Quarterly Journal of Economics*, 127(1), 57–95.
- VAYANOS, D., AND J. WANG (2013): “Chapter 19 - Market Liquidity – Theory and Empirical Evidence,” vol. 2, Part B of *Handbook of the Economics of Finance*, pp. 1289 – 1361. Elsevier.
- WOODFORD, M. (1990): “Public Debt as Private Liquidity,” *The American Economic Review*, 80(2), 382–388.

Appendix: Solving the Equilibrium

The fully solved Markov equilibrium is a set of functions that map η_t to the values of endogenous variables, such as capital price, interest rate, and bank leverage. Time subscripts are suppressed to save notations. First, we need to solve the capital price $q(\eta)$ as the fixed point of a contraction mapping in the space of \mathbb{C}^1 . Once we know $q(\eta)$, the other endogenous variables can be solved easily.

We pick any starting candidate capital price function $q^1(\eta) \in \mathbb{C}^1$, and give it an index equal to “1” (in the superscript). Constructed from individual optimality and market clearing conditions, the following contraction functional maps $q^1(\eta)$ to $q^2(\eta)$ in \mathbb{C}^1 . Then, we apply the functional to $q^2(\eta)$ to get $q^3(\eta)$, and repeat until convergence. The fixed point $q^\infty(\eta)$ is the equilibrium capital price function $q(\eta)$ that satisfies all the equilibrium conditions.

The contraction mapping can be constructed in five steps. First, from the goods market clearing condition,

$$a - \rho q^1 = \lambda_I \left(\frac{1 - (1 - \phi) x^{nf,1}}{1 - \theta q^1} \right) (1 - \eta) q^1,$$

we can solve the implied $x^{nf,1} = x^{nf,1}(\eta)$, i.e. the entrepreneurs’ capital-to-wealth ratio as a function of η .²⁰ Second, using the capital market clearing condition,

$$x^{f,1}\eta + x^{nf,1}(1 - \eta) = 1,$$

we can solve the implied $x^{f,1} = x^{f,1}(\eta)$, i.e. the bankers’ capital-to-wealth ratio as a function of η . Third, we can solve $\tilde{\eta}^1 = \tilde{\eta}^1(\eta)$, which is the state to which the economy jumps from η when the macro shock hits, using the equation

$$\tilde{\eta}^1 = \frac{\tilde{n}^f}{\tilde{n}^f + \tilde{n}^{nf}} = \frac{\left[\frac{\tilde{q}^1}{q^1} (1 - \underline{\delta}) x^{f,1} + (1 - x^{f,1}) \right] \eta}{\left[\frac{\tilde{q}^1}{q^1} (1 - \underline{\delta}) x^{f,1} + (1 - x^{f,1}) \right] \eta + \left[\frac{\tilde{q}^1}{q^1} (1 - \bar{\delta}) x^{nf,1} + (1 - x^{nf,1}) \right] (1 - \eta)},$$

²⁰The starting candidate $q^1(\eta)$ should be chosen such that the implied $x^{nf,1}(\eta) \geq 0$.

where $\tilde{q}^1 = q^1(\tilde{\eta}^1)$. Since the function $q^1(\eta)$ is known, the equation above solves $\tilde{\eta}^1$ as a function of η . Fourth, we can solve the *pre-shock excess return* of capital using bankers' F.O.C. condition for x^f (which holds in equality),

$$\lambda \left(\frac{1 - \frac{\tilde{q}^1}{q^1} (1 - \underline{\delta})}{1 - \left[1 - \frac{\tilde{q}^1}{q^1} (1 - \underline{\delta})\right] x^{f,1}} \right).$$

The final step is to check entrepreneurs' F.O.C. condition for x^{nf} at every value of η (i.e. the inequality (10)). If it is violated, we update q^1 to q^2 ; if it is not violated, the value of q^2 at η is set to equal q^1 . We have already solved the *pre-shock excess return* of capital, which is the left-hand side of the inequality (10). The right-hand side can be solved by substituting in q^1 , \tilde{q}^1 and $x^{nf,1}$,

$$\lambda \left(\frac{1 - \frac{\tilde{q}^1}{q^1} (1 - \bar{\delta})}{1 - \left[1 - \frac{\tilde{q}^1}{q^1} (1 - \bar{\delta})\right] x^{nf,1}} \right) + \lambda_I \left(\frac{(ROI^1 - 1) (1 - \phi)}{ROI^1 [1 - (1 - \phi) x^{nf,1}] + (1 - \phi) x^{nf,1}} \right),$$

where $ROI^1 = \frac{(1-\theta)q^1}{1-\theta q^1}$, the return on internal liquidity given q^1 .

To construct $q^2(\eta)$, consider the following four scenarios for every value of η . First, the left-hand side of the inequality (10) is larger than the right-hand side. In this case, entrepreneurs prefer to hold more capital at the current price $q^1(\eta)$. We solve a new (and higher) $x^{nf,2}$ by setting entrepreneurs' F.O.C. to equality, and then update the value of q^1 at η to $q^2(\eta)$ that is solved by substituting $x^{nf,2}$ into the goods market clearing condition.²¹ Second, the left-hand side of the inequality (10) is smaller than the right-hand side. If $x^{nf,1} = 0$, we do not need to update q^1 , so $q^2(\eta) = q^1(\eta)$. If $x^{nf,1} > 0$, we solve a new $x^{nf,2}$ by setting entrepreneurs' F.O.C. to equality, and then update the value of q^1 at η to $q^2(\eta)$ that is solved by substituting $x^{nf,2}$ into the goods market clearing condition. Finally, if the left-hand side of the inequality (10) is equal

²¹This means to adjust q^1 upward. In particular, the root in $(1, \frac{1}{\theta})$ is selected. When $q < 1$, aggregate investment is zero. Goods market clearing implies $q = \frac{\alpha}{\rho}$, which is larger than one under the calibrated parameter values, a contradiction. Therefore, $q \geq 1$. q has to be small than $\frac{1}{\theta}$. Otherwise, the investment project becomes self-financing, in which case the scale of investment is positive infinite.

to the right-hand side, we do not need to update q^1 , so $q^2(\eta) = q^1(\eta)$.

Therefore, we have constructed a functional that maps $q^1(\eta)$ to $q^2(\eta)$. The proof of this mapping being a contraction mapping in \mathbb{C}^1 is beyond the scope of the paper, but under the calibrated parameter values, the recursive algorithm converges, starting from $q^1(\eta)$ that is given by the goods market clearing condition with $x^{nf} = 0$ for all η in $(0, 1)$.

After $q(\eta)$ is solved, we can solve x^{nf} , x^f , $\tilde{\eta}$, \tilde{q} , and the pre-shock excess return of capital as previously discussed. Since $q(\eta)$ is known, we can solve $\mu^q(\eta)$ using Itô's lemma as follows,

$$\mu^q(\eta) = \frac{dq(\eta)/q(\eta)}{d\eta/\eta} \mu^n(\eta),$$

where $\mu^n(\eta)$ is given by Equation (7). The interest rate $r(\eta)$ is solved as the difference between the pre-shock return on capital (i.e. $\frac{a}{q(\eta)} + \mu^q(\eta)$) and pre-shock excess return.

Network Risk and Key Players: A Structural Analysis of Interbank Liquidity*

Edward Denbee

Christian Julliard

Ye Li

Kathy Yuan

Abstract

We estimate the liquidity multiplier in an interbank market and study systemic risk using a structural network model. In the model, banks hold liquidity to buffer shocks. They borrow liquidity from neighbours and update their valuation based on neighbours actions. When the former (latter) motive dominates, the equilibrium exhibits strategic substitution (complementarity) of holdings, and a reduced (increased) liquidity multiplier dampening (amplifying) shocks. Empirically, we find substantial and procyclical network-generated risks driven mostly by changes of equilibrium type rather than network topology. We identify the banks generating most aggregate risk and solve the planner's problem, providing guidance to macro-prudential policies.

Keywords: financial networks; liquidity; interbank market; key players; systemic risk.

*We thank the late Suddipto Bhattacharya, Yan Bodnya, Douglas Gale, Michael Grady, Charles Khan, Seymon Malamud, Farzad Saidi, Mark Watson, David Webb, Anne Wetherilt, Kamil Yilmaz, Peter Zimmerman, and seminar participants at the Bank of England, Cass Business School, Duisenberg School of Finance, Koc University, the LSE, the Stockholm School of Economics, and the WFA for helpful comments and discussions. Denbee (edward.denbee@bankofengland.co.uk) is from the Bank of England; Julliard (c.julliard@lse.ac.uk) and Yuan (K.Yuan@lse.ac.uk) are from the London School of Economics, SRC, and CEPR; Li (yli17@gsb.columbia.edu) is from Columbia University. This research was started when Yuan was a senior Houblon-Norman Fellow at the Bank of England. The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England. The support of the Fondation Banque de France, of ERC, and of the Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre is gratefully acknowledged [grant number ES/K002309/1].

I Introduction

The 2007-09 financial crisis has stimulated strong interest in understanding financial intermediation and its role in creating liquidity.¹ As emphasized by Bianchi and Bigio (2014) and Piazzesi and Schneider (2017) among others, intermediaries face their own liquidity management problem, and in particular, banks hold central bank reverses to buffer liquidity shocks. Their choices are crucial for liquidity production, payment activities, and asset prices in the macroeconomy. Another area that has drawn increasing attention is financial networks. The interbank network, where banks borrow and lend reverses, has been at the heart of studies of systemic risk.² These two themes merge in our paper. We study both theoretically and empirically how the interbank network affects banks' liquidity holding decisions, and its implications for systemic risk.

This paper structurally estimates a liquidity-holding game where banks obtain credit from an interbank network and their liquidity management objective incorporates different economic sources of network externality. Applying our framework to U.K. banks, we find that the dominant type of network externality varies over the business cycle. In the boom before 2008, banks' liquidity holdings exhibit strategic complementarity, and thereby, the network amplifies liquidity shocks. As the financial crisis unfolds, the degree of strategic complementarity declines, and after the introduction of Quantitative Easing (QE) in the U.K., banks' liquidity holdings exhibit strategic substitution, and the interbank network turns from a shock amplifier to a shock buffer. To the best of our knowledge, we provide the first evidence of a procyclical interbank network externality.

Our framework also offers several novel metrics to guide the monitoring of banks and the design of policy intervention during a crisis. Specifically, we identify the banks that contribute the most to the volatility of aggregate liquidity holdings. Banks' contribution to

¹In the recent macro-finance literature, intermediaries play the role of marginal investor in asset markets (Brunnermeier and Sannikov (2014); He and Krishnamurthy (2013)), credit supplier (Gertler and Kiyotaki (2010); Klimenko, Pfeil, Rochet, and Nicolo (2016)), and money issuer (Brunnermeier and Sannikov (2016); Hart and Zingales (2014); Li (2017); Quadrini (2017)).

²Recent theories on interbank network and systemic risk include Freixas, Parigi, and Rochet (2000), Allen, Carletti, and Gale (2008), and Freixas, Martin, and Skeie (2011).

systemic risk varies significantly over time. We show that such variation is mostly driven by changes of the type of equilibrium on the network (i.e., strategic complementarity or substitution) rather than changes of the network topology. Furthermore, we compare the decentralized equilibrium with the planner’s solution that achieves constrained efficiency.

We model banks’ decision of holding reserves to manage their exposure to liquidity shocks in a linear-quadratic framework (à la Ballester, Calvo-Armengol, and Zenou (2006)), assuming a predetermined interbank network where banks borrow and lend reserves. Bank characteristics and the macroeconomic environment affect banks’ decision, but reserve holdings also depend on the network topology and the nature of network externality. The interbank network generates two counteracting effects for the banks’ liquidity management problem. First, since banks can borrow reserves from their neighbours, the marginal benefit of holding reserves on their own decreases when neighbours hold more reserves. This free-riding incentive gives rise to strategic substitution (Bhattacharya and Gale (1987)). Consequently, the network acts as a risk buffer for liquidity shocks since neighbouring banks’ liquidity holdings are negatively correlated. Second, when banks see neighbouring banks holding more reserves, they positively update their belief on the value of liquidity (e.g., DeMarzo, Vayanos, and Zwiebel (2003)). Due to such informational spillover, the marginal benefit of holding liquidity increases in neighbours’ liquidity, which leads to strategic complementarity. In this case, the network amplifies the liquidity shocks originating from individual banks due to the positive correlation among neighbouring banks’ liquidity holdings. Another channel through which banks’ liquidity holdings may increase in their neighbours’ is the leverage stack mechanism in Moore (2012).³ At the (unique interior) Nash equilibrium of our model, the overall impact of network on banks’ liquidity holdings is summarized by a parameter ϕ , the network attenuation factor. If strategic substitution dominates, ϕ is negative. If strategic complementarity dominates, ϕ is positive.

In equilibrium, the individual banks’ reserve holding decision is affected by the magnitude of the liquidity shocks of all banks in the network. However, not all shocks are equally

³In Moore (2012), by signalling its credit worthiness through a higher liquidity buffer, banks can borrow more from other banks, and thereby, finance more positive NPV projects.

important. In particular, network features such as the dominant type of network externality, ϕ , and bank i 's *indegree* Katz-Bonacich centrality (network topology) determine how a bank weights the importance of all liquidity shocks in the network to optimize its own liquidity management problem. The network indegree centrality gives a direct metric of such a weight since it counts the direct and indirect links from other banks towards bank i , weighting connections by ϕ^k , where k is the number of steps needed to reach bank i . In other words, the liquidity holding decision of a bank is related to its own shocks, the shocks of its neighbours, of the neighbours of its neighbours, etc., with distant shocks becoming increasingly less important as they are weighted by ϕ^k .⁴ It is important to emphasize that the network topology (e.g. Katz-Bonacich centrality measures) is *not* the only determinant of banks' liquidity holding decision. The magnitude of idiosyncratic shocks of all banks in the network and the network attenuation factor ϕ are also crucial in the equilibrium liquidity holding decisions of the banks. For example, banks that receive extremely large liquidity shocks, regardless of their network location, may have a large impact on all banks' liquidity holding decisions in equilibrium. Similarly, whether the network effect is dominated by substitution or complementarity determines the nature of shock transmission in the network.

Based on the network equilibrium results, we conduct further welfare analysis for potential policy interventions to remedy the negative impact of network externalities. We characterise the volatility of the aggregate liquidity and identify key banks that contributes the most to the systemic risk – i.e. the *risk key players*. We find that the contribution by each bank to the network risk is related to 1) the network attenuation factor ϕ , 2) the bank specific liquidity risks (captured by bank-specific standard deviations), and 3) its *outdegree* Katz-Bonacich centrality measure. The outdegree centrality is similarly defined as the indegree centrality but the connections are outbound from bank i to measure the impact of bank i ' on its neighbours, neighbours of its neighbours, etc. Moreover, we introduce

⁴This centrality measure takes into account the number of both immediate and distant connections in a network. For more on the Bonacich centrality measure, see Bonacich (1987) and Jackson (2003). For other economic applications, see Ballester, Calvo-Armengol, and Zenou (2006) and Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012). For an excellent review of the literature, see Jackson and Zenou (2012).

the concept of network impulse response function (NIRF), and show that the contribution of each bank to aggregate risk is measured by the NIRF to that bank’s individual shock. The NIRFs are true impulse response function in that the total volatility of aggregate liquidity can be decomposed into NIRFs to bank specific shocks, and the risk key player is precisely the bank with the largest NIRF. Furthermore, we solve for the planner’s solution and contrast it with the decentralised equilibrium level of systemic liquidity level and risk. This analysis allows us to characterise the sources of the externalities and pinpoint possible interventions to achieve the social optimum.

Using daily data from the Bank of England, we apply our model to study the reserve holding decisions of the member banks of the sterling large payment system, CHAPS, in the period of January 2006 to September 2010. Member banks conduct transaction for their own purpose and on behalf of their clients and hundreds of other non-member banks. The stability of this system is crucial for supporting the real economic activities. On average in 2009, £272 billions of transactions in the U.K. were settled every day in CHAPS (U.K. nominal GDP every 5.5 days). CHAPS banks regularly have intraday liquidity exposures in excess of £1 billion to individual counterparties, and they hold reserves to buffer payment imbalances.⁵ Variation in payment imbalances is as close as we can get to a pure liquidity shocks, because CHAPS transactions are settled in real time and on *gross* terms (“RTGS”) to eliminate counterparty credit risks.⁶ In addition to banks’ own reserve holdings, banks can borrow reserves from each other on an unsecured basis in the overnight market. These interbank connections form a network – a link between two banks is quantified by the fraction of borrowing by one bank from another in the recent past, so the network is directional and its adjacency matrix is weighted (i.e., right stochastic). Note that the links between two banks can be interpreted as the (frequentist) probabilities of their borrowing-

⁵The U.K. monetary framework leaves reserves management largely at individual banks’ discretion (even during the QE period). In Appendix A.1, we provide background information on the monetary framework (i.e. reserve regimes) including QE, the payment system, and the overnight interbank markets.

⁶CHAPS uses RTGS instead of DNS (deferred net settlement). The DNS model is more liquidity efficient but creates credit risk exposure for recipient banks until the end of a clearing cycle. Such risks do not exist under RTGS since all payments are settled individually and on a gross basis. A detailed description of the RTGS in the U.K. is provided by Dent and Dison (2012) of the Bank of England. In this report, the BoE maintains that RTGS improves financial stability by minimizing credit exposures between banks.

lending relationships per unit of currency. We study the impact of this interbank network on banks' reserve holdings.⁷

For our empirical analysis we exploit the fact that the equilibrium of our model maps exactly into the spatial error framework, which naturally separates the hypothetical liquidity holdings of a standalone bank and the network induced component. We allow the non-network component to load on bank characteristics and macro variables, and confine the network component of liquidity holdings only in the spatial error term. This conservative approach leaves a minimal amount of variation in liquidity holdings to be driven by the interbank network.⁸ Yet, we are able to uncover rich, pro-cyclical, dynamics for the network externality: the network amplifies shocks in good times, and dampens them during the crisis and its aftermath. To show the importance of modelling the role of the network to analyze aggregate risks, we compute the ratio of the volatility of the aggregate liquidity implied by our estimate of network multiplier to the counterfactual volatility where the network and its externalities play no role (i.e. the network attenuation factor is zero). We find that during the boom, this ratio reaches 559%, during the crisis, it is 125% and during the QE, it is reduced to 89%.

Our finding of time-varying network externality sheds light on the relative importance of different economic forces over the business cycle. Because it is costly to hold liquidity at the expense of forgoing other investments, banks free ride their peers by borrowing reserves in interbank network in response to liquidity shocks (Bhattacharya and Gale (1987)). This common wisdom is only part of the story, because strategic complementarity arises from informational spillovers and/or a “leverage stack” type mechanism. Our framework accommodates all these facets of network externality, and allows us to identify from the data the dominant forces at different points in time.

⁷In addition to central bank reserves, member banks may pledge government bonds as collateral to borrow reserves from the Bank of England. Therefore, in our definition of reserve holdings, we include both actual reserves and the holdings of collateral eligible for repo with the Bank of England. Our results are robust if we use actual reserves instead.

⁸We also estimated a spatial Durbin model, in which the network not only propagates shocks in the error terms but also from bank characteristics. This more general model also serves as a specification test of our benchmark framework.

In addition we empirically characterize the shock propagation mechanism, quantify the individual banks' contribution to aggregate liquidity risk, and identify the risk key players. We find that most of the volatility of aggregate liquidity in the banking system is driven by a small group of banks, and that each bank's contribution varies substantially over time. Moreover, we find that the risk key player is typically *not* the largest net borrower – even net lenders can generate substantial risk in the system. These finding is particularly relevant for monitoring and regulating the banking system, and policy interventions during crisis.

Since in our sample the network topology changes over time, we decompose the time variation of banks' risk contributions into two components: the changes ascribable to time variation in ϕ and in the network topology. We find that the former is clearly the main driver. This suggests that endogenous network formation plays a limited role in the variation of network effect in our context. It is the type of equilibrium on the network (i.e., strategic complementarity or substitution) that matters most, not the network itself.⁹

Finally, we compute the planner's solution based on our estimates of network multiplier, the sizes of bank-level shocks and other parameters. We find that during the boom period, the amount of aggregate liquidity held by banks is not too far from the planner's equilibrium but the network generates too much systemic liquidity risk through shock amplification. During the crisis period, the decentralized equilibrium generates smaller aggregate liquidity than the planner's solution, and the systemic liquidity risk is still too high. After the introduction of QE, banks hold too much liquidity and the volatility of aggregate liquidity is below the level obtained in planner's solution. These findings may guide policy makers in monitoring banks, designing crisis interventions, and assessing the impact of QE.

This paper is related to the literature on bank liquidity management and monetary policy, and more broadly, the interaction between financial intermediation and the macroe-

⁹Our finding is related to the empirical literature that critically examines the systemic consequence of network linkages. While there is an extensive literature on network contagion, simulation studies based on reasonably realistic network shows small impact of contagion on systemic bank failure (summarized in Upper (2011)). Related, using a unique dataset of all Austrian banks, Elsinger, Lehar, and Summer (2006) find that contagion happens rarely and that the necessary funds to prevent contagion are surprisingly small. By applying an Eisenberg and Noe (2001) style model to German banks, Chen, Wang, and Yao (2016) find that the lack of bank capital is the key contributor to bank failure rather than the network contagion.

economy. Bernanke and Blinder (1988) embed bank reserve management in an IS-LM model. Kashyap and Stein (2000) find that the impact of monetary policy on bank lending is stronger for banks with less liquid assets. Bianchi and Bigio (2014) study monetary policy transmission in a dynamic model of banks' liquidity management. In a model where risk tolerant bankers are the marginal investor in asset markets, Drechsler, Savov, and Schnabl (2014) link risk premia to monetary policy by highlighting bankers' need to hold reserves as a buffer against liquidity shocks. Piazzesi and Schneider (2017) provide a macroeconomic model of the interaction between interbank transactions and payment activities of the rest of the economy. The linkages between payment system, reserves, and interbank credit are crucial elements in liquidity production that affects asset prices, money and credit supply, and gives roles to monetary policy intervention. Our paper complements this line of inquiry by providing evidence on these linkages, especially the state-dependent impact of interbank credit network on banks' liquidity management.

We contribute to the literature on bank liquidity regulation by providing an empirical framework to attribute systemic risk to individual banks, and by characterizing the wedge between decentralised outcome and the planner's solution. Liquidity regulation has attracted a lot of attention after the financial crisis. Stein (2012) argues that reserves requirement may serve as a tool for financial stability regulation. Diamond and Kashyap (2016) study bank liquidity regulation in the setting of Diamond and Dybvig (1983). Allen and Gale (2017) review earlier theories that may provide foundations (i.e., sources of market failures) for bank liquidity regulations, such as liquidity coverage ratio and net stable funding ratio in Basel III. Our findings of pro-cyclical network externality and banks' time-varying contribution to systemic risk lend support to a macro-prudential perspective on liquidity regulation.

Our work also advances the literature on interbank market dynamics and banks' liquidity demand. Fecht, Nyborg, and Rocholl (2010) find that the prices of liquidity depend on counterparties' liquidity levels. Acharya and Merrouche (2010) document evidence of precautionary liquidity demands of U.K. banks during the subprime crisis. Ashcraft, McAndrews, and Skeie (2010) find that in response to heightened payment uncertainty, banks

hold excess reserves in the Fed fund market. There is also a related theoretical literature pioneered by Bhattacharya and Gale (1987). Recent theoretical works in this area highlight the externalities in interbank markets and the associated inefficiencies (e.g. Freixas, Parigi, and Rochet (2000); Allen, Carletti, and Gale (2008); Freixas, Martin, and Skeie (2011); Moore (2012); Castiglionesi, Feriozzi, and Lorenzoni (2017) among others). Our paper differs by modeling banks' liquidity holdings as outcome of a network game, and estimate the time-varying network externality.¹⁰

Our structural estimation contributes to the broad literature of network and systemic risks. Networks have proved to be a useful analytical tool for studying financial contagion and systemic risk from both theoretical and empirical perspectives. Starting from Allen and Gale (2000), recent theories feature increasingly sophisticated networks and shock transmission mechanisms.¹¹ Recent empirical works also cover a wide range of economic networks.¹² We differ from these papers by building upon the linear-quadratic approach of Ballester, Calvo-Armengol, and Zenou (2006) to analyze the impact of economic agents' optimal liquidity holding decision on a network game for systemic liquidity risk and estimate its time-varying properties.

The remainder of the paper is organised as follows. In Section II we present and

¹⁰There is a another line of research that focuses on the topology and formation of linkages. Afonso and Lagos Afonso and Lagos (2015) use a search theoretical framework to study the interbank market and banks' trading behavior. The empirical literature on the topology of interbank networks starts with Furfine (2000, 2003). Other earlier empirical studies of the interbank network topology include Upper and Worms (2004); Boss, Elsinger, Summer, and Thurner (2004); Soramaki, Bech, Arnold, Glass, and Beyeler (2007); Becher, Millard, and Soramaki (2008); and Bech and Atalay (2008). Recent works study the impact of the crisis on the structure of these networks, which include (but are not limited to): Gai and Kapadia (2010); Wetherilt, Zimmerman, and Soramaki (2010); Benos, Garratt, and Zimmerman (2010); Ball, Denbee, Manning, and Wetherilt (2011); and Afonso, Kovner, and Schoar (2011). We differ from this literature by studying, rather than network formation, the types of equilibria on a predetermined network. We empirically show that the variation of network externality is driven by the type of equilibrium on network instead of the changes in network topology.

¹¹This line research includes but is not limited to Leitner (2005); Babus (2009); Babus and Allen (2009) (for a review); Afonso and Shin (2011); Zawadowski (2012); Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012); Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015); Elliott, Golub, and Jackson (2015); Atkeson, Eisdeldt, and Weill (2015); Ozdagli and Weber (2015); Glasserman and Young (2015); Cabrales, Gale, and Gottardi (2015); Cabrales, Gottardi, and Vega-Redondo (2016).

¹²The recent empirical network literature include but is not limited to Diebold and Yilmaz (2009, 2014); Billioa, Getmanskyb, Lo, and Pelizzona (2012); Hautsch, Schaumburg, and Schienle (2012); naki Aldasoro and Angeloni (2013); Kelly, Lustig, and Nieuwerburgh (2013b); Duarte and Eisenbach (2013); Greenwood, Landier, and Thesmar (2015); and Gofman (2017).

solve a liquidity holding decision game in a network, and define key players in terms of level and risk. Section III casts the equilibrium of the liquidity network game in the spatial econometric framework, and outlines the estimation methodology. In Section IV, we describe the data, the construction of the network, and the basic network characteristics throughout the sample period. In Section V, we present and discuss the estimation results. Section VI concludes.

II The Network Model

In this section, in order to study how aggregate liquidity risk is generated within the interbank system, we present a network model of interbank liquidity holding decisions, where the network reflects bilateral borrowing and lending relationships.

The network: there is a finite set of n banks. The time t network, denoted by g_t , is endowed with an n -square adjacency matrix \mathbf{G}_t where $g_{ii,t} = 0$ and $g_{ij \neq i,t}$ is the fraction of borrowing by bank i from bank j . The network g_t is therefore weighted and directed.¹³ Banks i and j are directly connected (in other words, they have a direct lending or borrowing relationship) if $g_{ij,t}$ or $g_{ji,t} \neq 0$. The coefficient $g_{ij,t}$ can be interpreted as the frequentist estimate of the probability of bank i 's receiving one pound from bank j via direct borrowing.

The matrix \mathbf{G}_t is a (right) stochastic (hollow) matrix by construction, is not symmetric, and keeps track of all direct connections – links of order one – between network players. That is, it summarises all the paths of length one between any pair of banks in the network. Similarly, the matrix \mathbf{G}_t^k , for any positive integer k , encodes all links of order k between banks, that is, the paths of length k between any pair of banks in the network. For example, the coefficient in the (i, j) th cell of \mathbf{G}_t^k – i.e. $\{\mathbf{G}_t^k\}_{ij}$ – gives the amount of exposure of bank i to bank j in k steps. Since, in our baseline construction, \mathbf{G}_t is a right stochastic matrix,

¹³We also explore other definitions of the adjacency matrix, where $g_{ij,t}$ is either the sterling amount of borrowing by bank i from bank j , or 1 (0) if there is (no) borrowing or lending between Bank i and j . Note that, in this latter case, the adjacency matrix is unweighted and undirected. In the theoretical part of the paper, we provide results and intuitions for the case when \mathbf{G}_t is a right stochastic matrix. However, the results are easily extendable to other forms of adjacency matrices with some restrictions on the parameter values which we will highlight when needed.

\mathbf{G}_t can also be interpreted as a Markov chain transition kernel, implying that \mathbf{G}_t^k can be thought of as the k -step transition probability matrix, i.e. the matrix with elements given by the probabilities of reaching bank j from bank i in k steps. Simply put, the matrix \mathbf{G}_t measures how liquidity travels in the interbank network.

Banks and their liquidity preferences in a network: we study the amount of liquidity buffer stocks (reserves) banks choose to hold at the beginning of day t when they have access to the interbank borrowing and lending network g_t . We define the total liquidity holding by bank i , denoted by $l_{i,t}$. In the main text, we model $l_{i,t}$ as the sum of two components: bank i 's liquidity holdings absent of any bilateral effects (i.e. the level of liquidity that a bank would hold if it were not part of a network), and bank i 's level of liquidity holdings made available to the network, which depends (potentially) on its neighbouring banks' liquidity contributions to the network. We use $q_{i,t}$ and $z_{i,t}$ to denote these two components respectively, and $l_{i,t} = q_{i,t} + z_{i,t}$. Before modelling the network effect on banks' liquidity choices, we assume $q_{i,t}$ load on all bank-specific and macro variables as follows:

$$q_{i,t} = \alpha_i + \sum_{m=1}^M \beta_m x_{i,t}^m + \sum_{p=1}^P \beta_p x_t^p, \quad (1)$$

where α_i is a bank fixed-effect, $x_{i,t}^m$ belongs to a set of M variables accounting for observable bank characteristics, and x_t^p belongs to a set of P variables controlling for time-series variation in systematic factors. Thus, $q_{i,t}$ captures the liquidity need specific to an individual bank from its balance sheet and fundamental conditions (e.g., leverage ratio, lending and borrowing rate), and its exposure to macro shocks (e.g., aggregate activities, monetary policy, etc.).

The network component $z_{i,t}$ is thus modelled as a residual term. This specification can be interpreted as a conservative approach to model the network effects since, empirically, we allow the residual variation of $l_{i,t}$ to be driven by the network component. As a robustness check, we also model and estimate the network effect when $l_{i,t}$ is not decomposed into a standalone and a bilateral component.¹⁴

¹⁴In the Appendix A.2, we report the results of this alternative formulation where the network effect on

To study a bank's choice of $z_{i,t}$, that is, its liquidity holdings in response to network access, we need to model the various sources of bilateral effects. We assume that banks are situated in different locations of the interbank relationship network g_t . The network allows banks to borrow and lend reserves, and may also transmit information relevant for liquidity management (more on this later). Each bank decides simultaneously how much liquid stock $z_{i,t}$ to hold given a predetermined g_t .¹⁵ The vector \mathbf{z}_t records all banks' choices.

We assume that banks derive utility from having an accessible buffer stock of liquidity, but at the same time they dislike the variability of this quantity. The accessible network liquidity for bank i has two components: direct holdings, $z_{i,t}$, and what can be borrowed from other banks connected through the network. This second component is proportional to the neighbouring banks direct holdings, $z_{j,t}$, weighted by the network linkage, $g_{ij,t}$, and a technological parameter ψ , that is, $\psi \sum_j g_{ij,t} z_{j,t}$. This component can be thought as unsecured borrowing through the interbank network. The marginal benefit of accessible liquidity for bank i is $\tilde{\mu}_{i,t}$ per unit. In sum, banks' liquidity management objective is represented by the following quadratic utility function:

$$u_i(\mathbf{z}_t|g_t) = \underbrace{\tilde{\mu}_{i,t}}_{\text{Per Unit Value}} \underbrace{\left(z_{i,t} + \psi \sum_{j \neq i} g_{ij,t} z_{j,t} \right)}_{\text{Accessible Liquidity}} - \frac{1}{2} \underbrace{\gamma}_{\text{Risk Aversion}} \left(z_{i,t} + \psi \sum_{j \neq i} g_{ij,t} z_{j,t} \right)^2, \quad (2)$$

where γ is the banks' risk aversion parameter. By establishing bilateral relationships in the banking network g_t , a bank also exposes itself to the shocks from its neighbouring banks. We assume that banks dislike the volatility of their own liquidity and of the liquidity they can access given their links, hence the last term in the objective function (2) arises.¹⁶

We further decompose $\tilde{\mu}_{i,t}$ into a bank-specific stochastic component $\hat{\mu}_{i,t}$, and a network

banks' liquidity holding decisions is not modelled a residual. The empirical counterpart of this model is a spatial Durbin model, which serves as a specification test for our baseline model.

¹⁵Since we model and estimate banks' liquidity holdings at the daily level, it is intuitive to take g_t as predetermined as the beginning of the day, as it is unlikely that the interbank relationships change significantly within a day.

¹⁶This objective function has the same spirit as a mean-variance utility in portfolio theory.

component, so

$$\tilde{\mu}_{i,t} := \hat{\mu}_{i,t} + \delta \sum_{j \neq i} g_{ij,t} z_{j,t}, \quad (3)$$

where

$$\hat{\mu}_{i,t}/\gamma =: \bar{\mu}_i + \nu_{i,t},$$

the parameter $\bar{\mu}_i$ denotes the bank specific average valuation of liquidity (absent any valuation spillovers and scaled by γ), and $\nu_{i,t}$ denotes the bank-specific shock to the valuation. The randomness of $\hat{\mu}_{i,t}$ is the ultimate source of uncertainty in this system. As it is part of banks' valuation of liquidity, we interpret this randomness as capturing banks' revision of belief on the forthcoming intraday payment imbalance (i.e., the liquidity shock). The network component of $\tilde{\mu}_{i,t}$ is motivated by potential informational spillover. Even though banks may value liquidity differently (due to *private* value), neighbours' liquidity holdings can be informative about the *common* value of reserves. We assume bank i follows a simple updating rule that adds $\delta \sum_j g_{ij,t} z_{j,t}$ to the standalone valuation $\hat{\mu}_{i,t}$. This updating rule is in the spirit of the boundedly-rational model of opinion formation in DeMarzo, Vayanos and Zwiebel (2003) (see also DeGroot (1974)).¹⁷ Therefore, a smaller coefficient δ reflects a larger informational discount on neighbouring banks' holdings, and the network linkages direct information flows via the interbank network.

The bilateral network influences are captured by the following cross derivatives for $i \neq j$:

$$\frac{\partial^2 u_i(\mathbf{z}_t | g_t)}{\partial z_{i,t} \partial z_{j,t}} = (\delta - \gamma\psi) g_{ij,t}.$$

When the cross derivative is negative, i.e. when $\delta < \gamma\psi$, banks' liquidity holdings exhibit strategic substitution. That is, an individual bank sets aside a smaller amount of liquid assets when its neighbouring banks hold more liquidity, which it can draw upon. This

¹⁷Note that this updating rule is not Bayesian. We choose this updating rule for expositional clarity in capturing two opposing network bilateral effects, as shown later. There is a separate but growing literature that studies the role of information aggregation in network settings (DeMarzo, Vayanos, and Zwiebel (2003); Babus and Kondor (2013)).

reflects the typical free-riding incentive as in Bhattacharya and Gale (1987).¹⁸ In our model, strategic substitutability arises from the fact that banks dislike volatility in their accessible liquidity, and therefore prefer to hold buffer stocks of liquidity that are less correlated with the ones of the neighbouring banks. Since the degree of accessibility of neighbours' liquidity increase in ψ , and the dislike of uncertainty is captured by γ , the degree of strategic substitutability is increasing in these two parameters.

Strategic complementarity arises when $\delta > \gamma\psi$. Through our interbank network not only flows liquidity (via borrowing and lending) but also the information on the common value of reserves. Strategic complementarity arises precisely from the informational spillover. We would expect a higher δ , and stronger strategic complementarity, when the common value of reserves is more prominent than the private value among banks.

Even if we restrict the interbank network to be only relevant for fund flows rather than information flow, the strategic complementarity may still arise as a result of leverage stack as in Moore (2012). Moore (2012) models a chain of borrowing/lending relationships that starts from the bank who borrows from households and ends at the bank with investment project. Interbank loans can be pledged to upstream lenders as collateral, so δ is lower if the collateral haircut is higher. Under this alternative formulation, we may posit bank i 's objective function as follows:

$$u_i(\mathbf{z}_t|g_t) = \underbrace{\hat{\mu}_{i,t} \left(z_{i,t} + \psi \sum_{j \neq i} g_{ij,t} z_{j,t} \right)}_{\text{Accessible Liquidity}} - \frac{1}{2} \gamma \underbrace{\left(z_{i,t} + \psi \sum_{j \neq i} g_{ij,t} z_{j,t} \right)^2}_{\text{Volatility of Accessible Liquidity}} + \underbrace{z_{i,t} \delta \sum_{j \neq i} g_{ij,t} z_{j,t}}_{\text{“Collateralised” Liquidity}} \quad , \quad (4)$$

The “collateralised” liquidity term, $z_{i,t} \delta \sum_j g_{ij,t} z_{j,t}$ has two parts: the available reserves that could be borrowed from neighbours, $\sum_j g_{ij,t} z_{j,t}$, and the multiplication factor $z_{i,t} \delta$, which can be thought of as collateral posted by bank i with the parameter δ reflecting a haircut. Since in the empirical context banks borrow and lend reserves on an unsecured

¹⁸Bhattacharya and Gale (1987) show that banks' liquidity holdings are strategic substitutes, because liquidity holdings come at a cost of forgoing higher interest revenue from long-term investments. Banks would like to free-ride their neighbouring banks for liquidity rather than conducting precautionary liquidity savings themselves.

basis, we may also interpret the multiplication factor as “information collateral,” i.e., by holding liquidity $z_{i,t}$, bank i signals its creditworthiness to neighbouring banks in the inter-bank network. Note that whether banks’ liquidity management objective is from Equation (2) or Equation (4) does not change the equilibrium outcome (i.e., their first-order conditions, best response functions, and equilibrium, stay the same).¹⁹

Equilibrium behaviour: now, we solve banks’ optimal reserve holdings in the Nash equilibrium. Banks choose their liquidity level $z_{i,t}$ simultaneously. A representative bank i maximises (2), and we obtain the following best response function for each bank:

$$z_{i,t}^* = \frac{\hat{\mu}_{i,t}}{\gamma} + \left(\frac{\delta}{\gamma} - \psi \right) \sum_{j \neq i} g_{ij,t} z_{j,t} = \mu_{i,t} + \phi \sum_j g_{ij,t} z_{j \neq i,t} \quad (5)$$

where $\phi := \delta/\gamma - \psi$, $\mu_{i,t}$ is defined earlier in equation (3) and $\nu_{i,t}$ denotes the bank-specific shock of valuation with variance denoted by σ_i^2 .

The “network attenuation factor” ϕ is the key parameter that determines the type of equilibrium on network: i.e., strategic substitution if $\phi < 0$ or complementarity if $\phi > 0$. In our paper, we are agnostic about the the sign of ϕ and we instead estimate it empirically.

Proposition 1 *Suppose that $|\phi| < 1$. Then, there is a unique interior solution for the individual equilibrium outcome given by*

$$z_{i,t}^*(\phi, g) = \{\mathbf{M}(\phi, \mathbf{G}_t)\}_i \mu_t, \quad (6)$$

where $\{\}_i$ is the operator that returns the i -th row of its argument, $\mu_t := [\mu_{1,t}, \dots, \mu_{n,t}]'$, $z_{i,t}$ denotes the bilateral liquidity holding by bank i , and

$$\mathbf{M}(\phi, \mathbf{G}_t) := \mathbf{I} + \phi \mathbf{G}_t + \phi^2 \mathbf{G}_t^2 + \phi^3 \mathbf{G}_t^3 + \dots \equiv \sum_{k=0}^{\infty} \phi^k \mathbf{G}_t^k = (\mathbf{I} - \phi \mathbf{G}_t)^{-1}. \quad (7)$$

¹⁹The only difference between these two objective functions is that Equation (2) has an additional second-order term $\left(\delta \sum_{j \neq i} g_{ij,t} z_{j,t} \right) \left(\psi \sum_{j \neq i} g_{ij,t} z_{j,t} \right)$, but it only contains *other* banks’ choice of liquidity holdings, not bank i ’s, so this additional term does not affect bank i ’s first-order condition. In the planner’s problem that we discuss later, since the planner internalizes any spillover effect, the solutions differ slightly depending on whether we take Equation (2) or Equation (4) as banks’ objective function.

where \mathbf{I} is the $n \times n$ identity matrix.

Proof. Since $\gamma > 0$, the first order condition identifies the individual optimal response. Applying Theorem 1, part b, in Calvo-Armengol, Patacchini, and Zenou (2009) to our problem, the necessary equilibrium condition becomes $|\phi \lambda^{max}(\mathbf{G}_t)| < 1$ where the function $\lambda^{max}(\cdot)$ returns the largest eigenvalue. Since \mathbf{G}_t is a stochastic matrix, its largest eigenvalue is 1. Hence, the equilibrium condition requires $|\phi| < 1$, and in this case the infinite sum in equation (7) is finite and equal to the stated result (Debreu and Herstein (1953)). ■

To gain intuition about the above result, note that a Nash equilibrium in pure strategies $\mathbf{z}_t^* \in \mathbb{R}^n$, where $\mathbf{z}_t := [z_{1,t}, \dots, z_{n,t}]'$, is such that equation (5) holds for all $i = 1, 2, \dots, n$. Hence, if such an equilibrium exists, it solves $(\mathbf{I} - \phi \mathbf{G}_t) \mathbf{z}_t = \mu_t$. If the matrix is invertible, we obtain $\mathbf{z}_t^* = (\mathbf{I} - \phi \mathbf{G}_t)^{-1} \mu \equiv \mathbf{M}(\phi, \mathbf{G}_t) \mu_t$. The rest follows by simple algebra. The condition $|\phi| < 1$ states that network externalities must be small enough in order to prevent the feedback triggered by such externalities to escalate without bounds.

The matrix $\mathbf{M}(\phi, \mathbf{G}_t)$ has an important economic interpretation: it aggregates all direct and indirect links among banks using an attenuation factor, ϕ , that penalises (as in Katz (1953)) the contribution of links between distant nodes at the rate ϕ^k , where k is the length of the path between nodes. In the infinite sum in equation (7), the identity matrix captures the (implicit) link of each bank with itself, the second term in the sum captures all the direct links between banks, the third term in the sum captures all the indirect links corresponding to paths of length two, and so on. The elements of the matrix $\mathbf{M}(\phi, \mathbf{G}_t)$, given by $m_{ij}(\phi, \mathbf{G}_t) := \sum_{k=0}^{+\infty} \phi^k \{\mathbf{G}_t^k\}_{ij}$, aggregates all the exposures in the network of i to j , where the contribution of the k th step is weighted by ϕ^k .

In equilibrium, the matrix $\mathbf{M}(\phi, \mathbf{G}_t)$ contains information about the *centrality* of network players. Multiplying the rows (columns) of $\mathbf{M}(\phi, \mathbf{G}_t)$ by a unit vector of conformable dimensions, we recover the indegree (outdegree) Katz–Bonacich centrality measure.²⁰ The indegree centrality measure provides the weighted count of the number of ties directed to

²⁰Newman (2004) shows that weighted networks can in many cases be analysed using a simple mapping from a weighted network to an unweighted multigraph. Therefore, the centrality measures developed for unweighted networks apply also to the weighted cases.

each node, while the outdegree centrality measure provides the weighted count of ties that each node directs to the other nodes. That is, the i -th row of $\mathbf{M}(\phi, \mathbf{G}_t)$ captures how bank i loads on the network as whole, while the i -th column of $\mathbf{M}(\phi, \mathbf{G}_t)$ captures how the network as a whole loads on bank i .

However, as equation (6) shows, the matrix $\mathbf{M}(\phi, \mathbf{G}_t)$ (which includes the network topology and the network attenuation factor ϕ) is *not* enough to determine the importance of a bank from a systemic risk perspective. Banks' equilibrium reserve holdings depend on \mathbf{M} and μ , suggesting that bank-specific shocks are equally important. Intuitively, when deciding its optimal liquidity holding level, bank i , weights its own shock, neighbouring and more centrally located banks' shocks relatively more heavily. Banks that receive large liquidity shocks, regardless of their network location, may have a larger influence on the other banks' liquidity holding in the network.

Equilibrium properties: we can decompose the network contribution to the total bilateral liquidity into a level effect and a risk effect. To see this, note that the total network generated liquidity, $Z_t := \sum_i z_{i,t}$, can be written at equilibrium as

$$Z_t^* = \underbrace{\mathbf{1}'\mathbf{M}(\phi, \mathbf{G}_t)}_{\text{level effect}} \bar{\mu} + \underbrace{\mathbf{1}'\mathbf{M}(\phi, \mathbf{G}_t)}_{\text{risk effect}} \nu_t \quad (8)$$

where $\bar{\mu} := [\bar{\mu}_1, \dots, \bar{\mu}_n]'$, $\nu_t := [\nu_{1,t}, \dots, \nu_{n,t}]'$. The first component captures the network level effect, and the second component (that aggregates bank-specific shocks) captures the network risk effect. It is clear that if $\bar{\mu}$ has only positive entries, both the network liquidity level and liquidity risk will be increasing in ϕ . That is, a higher network multiplier leads the interbank network to produce more liquidity and also generate more risk.

The equilibrium solution in equation (8) implies that bank i 's marginal contribution to the volatility of aggregate liquidity can be summarised as

$$\frac{\partial Z_t^*}{\partial \nu_{i,t}} \sigma_i = \mathbf{1}' \{\mathbf{M}(\phi, \mathbf{G}_t)\}_{.i} \sigma_i =: b_i^{out}(\phi, \mathbf{G}_t). \quad (9)$$

The above expression is the *outdegree* centrality for bank i weighted by the standard devi-

ation of its own shocks. Moreover, the conditional volatility of the aggregate liquidity level in our model is

$$Var_t(Z_t^*(\phi, \mathbf{G}_t)) = vec(\{b_i^{out}(\phi, \mathbf{G}_t)\}_{i=1}^n) vec(\{b_i^{out}(\phi, \mathbf{G}_t)\}_{i=1}^n)' \quad (10)$$

$$= \mathbf{1}' \mathbf{M}(\phi, \mathbf{G}_t) diag(\{\sigma_i^2\}_{i=1}^n) \mathbf{M}(\phi, \mathbf{G}_t)' \mathbf{1}. \quad (11)$$

Therefore, equation (9) provides a clear ranking of the riskiness of each bank from a systemic perspective. This allows us to define the concept of “systemic risk key player”.

Definition 1 (Risk key player) *The risk key player i_t^* , given by the solution of*

$$i_t^* = \arg \max_{i=1, \dots, n} b_i^{out}(\phi, \mathbf{G}_t), \quad (12)$$

is the one that contributes the most to the volatility of the overall network liquidity.

Similarly, we can identify the bank that can cause the maximum expected level of reduction in the network liquidity when removed from the system.²¹

Definition 2 (Level key player) *The level key player is the player that, when removed, causes the maximum expected reduction in the overall level of bilateral liquidity. We use $\mathbf{G}_{\setminus \tau, t}$ to denote the new adjacency matrix obtained by setting to zero all of \mathbf{G}_t 's τ -th row and column coefficients. The resulting network is $g_{\setminus \tau, t}$. The level key player τ_t^* is found by solving*

$$\tau_t^* = \arg \max_{\tau=1, \dots, n} \mathbb{E} \left[\sum_i z_i^*(\phi, g_t) - \sum_{i \neq \tau} z_i^*(\phi, g_{\setminus \tau, t}) \middle| g_t, \tau \right] \quad (13)$$

where \mathbb{E} is the rational expectation operator.

In this definition, the level key player is the one with the largest impact on the total expected bilateral liquidity, under the assumption that when the player τ is removed, the

²¹This definition is in the same spirit as the concept of the key player in the crime network literature as defined in Ballester, Calvo-Armengol, and Zenou (2006). There, it is important to target the key player for maximum crime reduction. Here, it is useful to consider the ripple effect on the network liquidity when a bank fails and exits from the system. Bailouts for key level players might be necessary to avoid major disruptions to whole interbank network.

remaining other banks do not form new links – i.e., we consider the short-run effect of removing a player from the network.

Using Proposition 1, we have the following corollary.

Corollary 1 *A player τ_t^* is the level key player that solves (13) if and only if*

$$\tau_t^* = \arg \max_{\tau=1,\dots,n} \{ \mathbf{M}(\phi, \mathbf{G}_t) \}_\tau \bar{\mu} + \sum_{i \neq \tau} m_{i\tau}(\phi, \mathbf{G}_t) \bar{\mu}_\tau \quad (14)$$

This follows from the fact that when bank τ is removed, the expected reduction in the total bilateral liquidity can be written as

$$\begin{aligned} \mathbb{E} \left[\sum_i z_i^*(\phi, g_t) - \sum_{i \neq \tau} z_i^*(\phi, g_{\setminus \tau, t}) \middle| g_t, \tau \right] &= \underbrace{\{ \mathbf{M}(\phi, \mathbf{G}_t) \}_\tau \bar{\mu}}_{\text{Indegree effect}} + \underbrace{\mathbf{1}' \{ \mathbf{M}(\phi, \mathbf{G}_t) \}_\tau \bar{\mu}_\tau}_{\text{Outdegree effect}} \\ &\quad - \underbrace{m_{\tau\tau}(\phi, \mathbf{G}_t) \bar{\mu}_\tau}_{\text{Double count correction}} \end{aligned} \quad (15)$$

That is, the removal of the level key player results in a direct (indegree) effect on its own liquidity generation and an indirect (outdegree) bilateral effect on other banks' liquidity generation. Instead of being the bank with the largest amount of liquidity buffer stock (captured by the first term on the right-hand side of equation (15)), the level key bank is the one with the largest expected contribution to its own and as well as its neighbouring banks' liquidity. This discrepancy exists because, in the decentralised equilibrium, no bank internalises the effect of its own liquidity holding level on the utilities of the other banks in the network. That is, no bank internalises the spillover of its choice of liquidity on other banks' liquidity valuation. Therefore, a relevant metric for a planner to use when deciding whether to bail out a failing bank should not be merely based on the size of the bank's own liquidity, but should also include its indirect network impact on other banks' liquidity.

In summary, the two measures (defined in equations (12) and (14)) help to identify the key players in the determination of aggregate liquidity levels and systemic liquidity risk in the network. However, the network topology alone is not enough. Both network multiplier as well banks' idiosyncratic shocks and their variabilities are important inputs

in computing the key players.

The Planner's Solution: this discussion leads us to analyse formally a planner's problem in this interconnected system. A planner that equally weights the utility of each bank (in equation (2)) chooses the network liquidity holdings by solving the following problem:

$$\max_{\{z_{i,t}\}_{i=1}^n} \sum_{i=1}^n u_i(\mathbf{z}_t|g_t) \quad (16)$$

where $u_i(\mathbf{z}_t|g_t)$ is bank's i utility from holding liquidity in the network defined in equation (2). The first order condition for the liquidity holding of the i -th bank ($z_{i,t}$) yields

$$\begin{aligned} z_{i,t} = & \underbrace{\mu_{i,t} + \phi \sum_{j \neq i} g_{ij,t} z_{j,t}}_{\text{decentralised f.o.c.}} + \underbrace{\psi \sum_{j \neq i} g_{ji,t} \mu_{j,t}}_{\substack{\text{the value of liquidity} \\ \text{available to neighbours}}} + \underbrace{\phi \sum_{j \neq i} g_{ji,t} z_{j,t}}_{\substack{\text{impact on neighbors'} \\ \text{valuation of liquidity}}} \\ & - \underbrace{\psi \left(\psi - \frac{2\delta}{\gamma} \right) \sum_{j \neq i} \sum_{m \neq j} g_{ji,t} g_{jm,t} z_{m,t}}_{\substack{\text{volatility of neighbors'} \\ \text{accessible network liquidity}}} \end{aligned} \quad (17)$$

In the above equation, the first two (indegree) terms are exactly the same as in the decentralised case, while the last three (outdegree) terms reflect the fact that the planner internalises a bank's contribution to its neighbouring banks' utilities. The third term captures the neighbours' idiosyncratic valuation of the liquidity provided by agent i . The fourth term reflects bank i 's impact on bank j 's network-dependent valuation of liquidity, so the outbound linkage $g_{ji,t}$ is weighted by bank j 's liquidity holdings, $z_{j,t}$, to arrive at the overall impact on bank j 's utility level.²² The fifth term measures bank i 's contribution to the volatility of network liquidity accessible by neighbouring banks.

²²The total impact on bank j 's valuation of liquidity also depends on ϕ , which is equal to $\frac{\delta}{\gamma} - \psi -$ the positive impact on valuation through informational spillover is captured by δ , and the availability of liquidity through interbank borrowing decreases the benefit of holding liquidity on its own (captured by ψ).

Rewriting equation (17) in matrix form, we obtain $\mathbf{z}_t = (\mathbf{I} + \psi \mathbf{G}'_t) \mu_t + \mathbf{P}(\phi, \psi, \delta, \mathbf{G}_t) \mathbf{z}_t$ where $\mathbf{P}(\phi, \psi, \delta, \mathbf{G}_t) := \phi (\mathbf{G}_t + \mathbf{G}'_t) - \psi (\psi - 2\delta/\gamma) \mathbf{G}'_t \mathbf{G}_t$. This allows us to formally state the planner's solution.

Proposition 2 *Suppose $|\lambda^{\max}(\mathbf{P}(\phi, \psi, \delta, \mathbf{G}_t))| < 1$. Then, the planner's optimal solution is uniquely defined and given by*

$$z_i^p(\phi, \psi, \delta, g_t) = \{\mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t)\}_i \mu_t, \quad (18)$$

where $\mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t) := [\mathbf{I} - \mathbf{P}(\phi, \psi, \delta, \mathbf{G}_t)]^{-1} (\mathbf{I} + \psi \mathbf{G}'_t)$.

Proof. The proof follows the same argument as in the proof of Proposition 1. ■

As in the decentralised solution, one can solve for the aggregate network liquidity level and risk in the planner's problem. We can write the level and volatility as follows.

$$Z_t^p = \mathbf{1}' \mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t) \bar{\mu} + \mathbf{1}' \mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t) \nu_t \quad (19)$$

$$\text{Var}_t(Z^p(\phi, \psi, \delta, \mathbf{G}_t)) = \mathbf{1}' \mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t) \text{diag}(\{\sigma_i^2\}_{i=1}^n) \mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t)' \mathbf{1}. \quad (20)$$

To see what drives the difference between the network liquidity in the decentralised equilibrium (z^*) and in the planner's solution (z^p), one can rewrite the planner's first order condition at the equilibrium as:

$$\mathbf{z}_t^p = \mathbf{z}_t^* + \mathbf{M}(\phi, \mathbf{G}_t) \left[\psi \mathbf{G}'_t \mu_t + \left(\phi \mathbf{G}'_t - \psi \left(\psi - \frac{2\delta}{\gamma} \right) \mathbf{G}'_t \mathbf{G}_t \right) \mathbf{z}_t^p \right]. \quad (21)$$

The extra terms (in the square brackets) of the planner's solution arise from the banks' failure, in the decentralized equilibrium, to internalise the network externalities they generate. Intuitively, among these terms: the first one reflects the contribution to the neighbours' valuations of liquidity holdings; the second one measures the contribution to the neighbouring nodes' indegree centrality and hence their network liquidity production level; and the last one is the contribution to their neighbouring nodes' volatility.²³ Therefore, the

²³Note that the term $\phi \mathbf{G}'_t - \psi \left(\psi - \frac{2\delta}{\gamma} \right) \mathbf{G}'_t \mathbf{G}_t$ vanishes only in the unlikely case of $\frac{\phi}{\psi \left(\psi - \frac{2\delta}{\gamma} \right)}$ being an

discrepancy between the planner's optimum and the decentralised equilibrium rests on the planner's tradeoff between the liquidity level and the liquidity risk in the network. When the planner cares more about the level of liquidity production than the liquidity risk in the network, the first two terms are more pronounced relative to the last term. In this case, banks that have higher outdegree centralities tend to hold less than the socially optimal amount of liquidity. The planner might subsidise or inject liquidity to these banks to increase the liquidity generated by the network. Conversely, when the planner cares more about the liquidity risk in the network (which happens when $\psi \gg 2\delta/\gamma$, e.g. very large ψ or γ and small δ), banks that have higher second-degree centralities tend to hold more than the socially optimal amount of liquidity. The planner might impose a tax on these banks to reduce the risk in the banking network.

The following corollary offers a closed-form characterisation of the wedge between the planner's solution and the decentralised outcome.

Corollary 2 *Let $\mathbf{H}_t := \phi \mathbf{G}'_t - \psi (\psi - 2\delta/\gamma) \mathbf{G}'_t \mathbf{G}_t$. Then, the aggregate network liquidity in the planner's solution can be expressed as*

$$Z_t^p = Z_t^* + \mathbf{1}' \left[\psi \mathbf{M}_t \mathbf{G}'_t + \mathbf{M}_t \mathbf{H}_t \mathbf{M}_t (\mathbf{I} - \mathbf{H}_t \mathbf{M}_t)^{-1} (\mathbf{I} + \psi \mathbf{G}'_t) \right] \mu_t \quad (22)$$

where Z_t^* denotes the aggregate bilateral liquidity in the decentralised equilibrium in equation (8) and $\mathbf{M}_t := \mathbf{M}(\phi, \mathbf{G}_t)$. Moreover, if \mathbf{H}_t is invertible, we have

$$Z_t^p = Z_t^* + \mathbf{1}' \left[\psi \mathbf{M}_t \mathbf{G}'_t + \mathbf{M}_t (\mathbf{H}_t^{-1} - \mathbf{M}_t)^{-1} \mathbf{M}_t (\mathbf{I} + \psi \mathbf{G}'_t) \right] \mu_t. \quad (23)$$

Proof. If \mathbf{H}_t is invertible, observing that

$$\mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t) \equiv \left[\mathbf{M}(\phi, \mathbf{G}_t)^{-1} - \phi \mathbf{G}'_t + \psi (\psi - 2\delta/\gamma) \mathbf{G}'_t \mathbf{G}_t \right]^{-1} (\mathbf{I} + \psi \mathbf{G}'_t)$$

eigenvalue of \mathbf{G}_t .

and using the Woodbury matrix identity (see, e.g. Henderson and Searle (1981)) gives

$$\mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t) = \mathbf{M}_t + \mathbf{M}_t(\mathbf{H}_t^{-1} - \mathbf{M}_t)^{-1}\mathbf{M}_t,$$

hence the result is immediate. If \mathbf{H}_t is not invertible, using equation (26) in Henderson and Searle (1981), we obtain

$$\mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}_t) = \mathbf{M}_t + \mathbf{M}_t\mathbf{H}_t\mathbf{M}_t(\mathbf{I} - \mathbf{H}_t\mathbf{M}_t)^{-1}$$

and the result follows. ■

The above implies that the discrepancy in the planner and market solutions for both expected total liquidity in the system ($\mathbb{E}[Z_t^p - Z_t^*|g_t]$), as well as for the individual expected liquidity holdings ($\{\mathbb{E}[\mathbf{z}_t^p - \mathbf{z}_t^*|g_t]\}_i$), might be positive or negative depending on the parameters and the topology of the network. In particular, one can show that the sign of the discrepancy between the solution of the planner and the decentralised solution depends on the parameters and the eigenvalues of the canonical operator of \mathbf{G}_t (see, e.g. Gorodentsev (1994) for a definition of the canonical operator).²⁴

III Empirical Methodology

In order to estimate the network model presented in Section II, we need to map the observed total liquidity holding of a bank at time t , $l_{i,t}$, into its two components: the liquidity holding absent of any bilateral effects (defined in equation (1)) and the network-dependent component (defined in equation (6)). This can be done by reformulating the theoretical model in the fashion of a spatial error model (SEM). That is, we decompose the total bank liquidity holdings into a function of the observables and a latent term that captures the

²⁴The proof of this result is very involved, hence we present it in an appendix available upon request.

spatial dependence generated by the network:

$$l_{i,t} = \alpha_t^{time} + \alpha_i^{bank} + \sum_{m=1}^M \beta_m^{bank} x_{i,t}^m + \sum_{p=1}^P \beta_p^{time} x_t^p + z_{i,t} \quad (24)$$

$$z_{i,t} = \bar{\mu}_i + \phi \sum_{j=1}^n g_{ij,t} z_{j,t} + \nu_{i,t} \sim iid(0, \sigma_i^2), \quad i = 1, \dots, n, \quad t = 1, \dots, T. \quad (25)$$

The only differences between the theoretical model and the econometric reformulation above are that: *i*) we have made explicit that one of the aggregate factors is a set of common time dummies, α_t^{time} , meant to capture potential trends in the size of the overall interbank market; *ii*) we allow the network links, g_{ij} , to potentially vary over time (but we construct them, as explained in the data description section below, in a fashion that makes them pre-determined with respect to the information set for time t).²⁵ The coefficients β_m^{bank} capture the effect of observable bank characteristics while the coefficients β_p^{time} capture the effects of systematic risk factors on the choice of liquidity.

Equation (25) describes the process of z_i , which is the residual of the individual bank i 's level of liquidity in the network that is not due to bank specific characteristics or systematic factors. Moreover, defining ϵ_i as the demeaned version of z_i , we have that $\sum_{j=1}^n g_{ij,t} \epsilon_{j,t}$ is a standard spatial lag term and ϕ is the canonical spatial autoregressive parameter. That is, the model in equations (24)–(25) is a variation of the Anselin (1988) spatial error model (see also Elhorst (2010a, 2010b)). This specification makes clear the nature of the network as a shock propagation mechanism: the shock to the liquidity of any bank, $\epsilon_{i,t}$, is a function of all the shocks to the other banks' liquidity; the intensity of the shock spillover is a function of the intensity of the network links between banks captured by the network weights g_{ij} ; and whether the network amplifies or damps the effect of the individual liquidity shocks on aggregate liquidity depends, respectively, on whether the banks in the network act as strategic complements ($\phi > 0$) or strategic substitutes ($\phi < 0$). To illustrate this point,

²⁵To allow for potential time variation in ϕ instead we also perform estimations in subsamples and over a rolling window.

note that the vector of shocks to all banks at time t can be rewritten as

$$\epsilon_t = (\mathbf{I} - \phi \mathbf{G}_t)^{-1} \nu_t \equiv \mathbf{M}(\phi, \mathbf{G}_t) \nu_t \quad (26)$$

where $\epsilon_t = [\epsilon_{1,t}, \dots, \epsilon_{n,t}]'$ and $\nu_t = [\nu_{1,t}, \dots, \nu_{n,t}]$. This implies that if \mathbf{G}_t is a right stochastic matrix²⁶ (and this is the case when we model the network weights $g_{ij,t}$ as the fraction of borrowing by bank i from bank j), then a unit shock to the system equally spread across banks (i.e. $\nu_t = (1/n) \mathbf{1}$) would imply a total change in aggregate liquidity equal to $(1 - \phi)^{-1}$ – that is, ϕ captures the ‘average’ network multiplier effect of liquidity shocks.

Moreover, equation (26) implies that any time variation in the network structure, \mathbf{G}_t , or in the network multiplier, $1/(1 - \phi)$, would result in a time variation in the volatility of total liquidity since the variance of the shocks to the total network liquidity ($\mathbf{1}'\epsilon_t$) is

$$\text{Var}_t(\mathbf{1}'\epsilon_t) = \mathbf{1}'\mathbf{M}(\phi, \mathbf{G}_t)\Sigma_\nu\mathbf{M}(\phi, \mathbf{G}_t)'\mathbf{1}.$$

Here we have used the fact that \mathbf{G}_t is pre-determined with respect to the time t information, $\Sigma_\nu := \mathbb{E}[\nu_t\nu_t']$ is a diagonal matrix with the variances of the idiosyncratic shocks $\{\sigma_i^2\}_{i=1}^n$ on the main diagonal.

As outlined in Section A.3.1 of the Appendix, we can estimate the parameters of the spatial error model jointly using a quasi-maximum likelihood approach. In order to elicit the time variation in the network coefficient ϕ , we perform subsample and rolling window estimates. The estimation frequency is daily, with a \mathbf{G}_t network matrix based on a rolling monthly average lagged by one day.

An estimation issue for network models is the well-known reflection problem (Manski (1993)): the neighbouring banks’ decisions about their liquidity holdings affect each other, so that we cannot distinguish between whether a given bank’s action is the cause or the

²⁶If \mathbf{G}_t is a right stochastic matrix, then $\mathbf{G}_t\mathbf{1} = \mathbf{1}$, and therefore

$$\mathbf{1} = (\mathbf{I} - \phi \mathbf{G}_t)^{-1} (\mathbf{I} - \phi \mathbf{G}_t) \mathbf{1} = (\mathbf{I} - \phi \mathbf{G}_t)^{-1} \mathbf{1} (1 - \phi) \implies \mathbf{M}(\phi, \mathbf{G}_t) \mathbf{1} = (1 - \phi)^{-1} \mathbf{1}.$$

effect of its neighbouring banks' actions. To address this problem, Bramoullé, Djebbari and Fortin (2009) have shown that the network effect ϕ can be identified if there are two nodes in the network with different average connectivities of their direct connected nodes. This condition is satisfied in our data.²⁷

As a test of the model specification of our theory-driven formulation, we also consider a more general specification that allows for a richer set of network interactions. That is, we model liquidity holding as a spatial Durbin model (SDM – see, e.g. LeSage and Pace (2009)) where bank specific liquidity is allowed to depend directly on other banks' liquidity and characteristics, and pairwise control variables

$$l_{i,t} = \alpha_t^{time} + \alpha_i^{bank} + \sum_{m=1}^M \beta_m^{bank} x_{i,t}^m + \sum_{p=1}^P \gamma_p^{time} x_t^p + \rho \sum_{j=1}^n g_{i,j,t} l_{j,t} + \sum_{j=1}^n g_{i,j,t} x_{i,j,t} \theta + \nu_{i,t} \sim iid(0, \sigma_i^2), \quad (27)$$

where $x_{i,j,t}$ denotes match specific control variables and the characteristics of other banks. The above formulation²⁸ allows a specification test of our structural model since, setting $x_{i,j,t} := vec(x_{j \neq i,t}^m)'$, and restricting $\theta = -\phi vec(\beta_m^{bank})$, $\gamma_p^{time} = (1 - \phi) \beta_p^{time} \forall p$, and most importantly $\rho = \phi$, we are back to the SEM specification implied by our structural model. These restrictions are tested formally in Section V. Such test is not only for validation of the empirical specification: given the close relation between our empirical and theoretical models, we effectively test our theoretical framework.

With the SEM estimated parameter at hand, we can also identify the risk key players of the interbank liquidity market. To do so, we define the network impulse response function as follows.

²⁷The separate identification of the fixed effects $\bar{\mu}_i$ and α_i^{bank} is more complex, and is discussed in detail in Appendix A.3.1. In particular, when \mathbf{G}_t is a right stochastic matrix, the identification of $\bar{\mu}_i$ and α_i^{bank} requires at least one bank to not borrow from any other bank at some point in the sample (in our data, this happens 13.5% of the time spread over all subsamples and rolling windows we consider). Alternatively, one can normalise one of the μ_i to zero and identify the remaining ones in deviation from it. But note that the separate identification of the fixed effects does not affect the identification of ϕ .

²⁸In Appendix A.2 we show that the above formulation can be obtained as the equilibrium of a network game in which bank specific liquidity is allowed to depend directly (rather indirectly via $z_{i,t}$) on other banks liquidity and characteristics, and we don't impose the decomposition $l_{i,t} = q_{i,t} + z_{i,t}$.

Definition 3 (Network Impulse-Response Functions) Let $L_t \equiv \mathbf{1}'l_t = [l_{1t}, \dots, l_{Nt}]$ denote the total liquidity in the interbank network. The network impulse response function of total liquidity, L_t , to a one standard deviation shock to a given bank i , is given by

$$NIRF_i(\phi, \sigma_i, \mathbf{G}_t) \equiv \frac{\partial L_t}{\partial v_{i,t}} \sigma_i = \mathbf{1}' \{\mathbf{M}(\phi, \mathbf{G}_t)\}_{.i} \sigma_i \quad (28)$$

where the operator $\{\}_{.i}$ returns the i -th column of its argument.

The network impulse response is identical to the shock size weighted *outdegree* centrality of bank i defined in equation (9). Note that $NIRF_i(\phi, 1, \mathbf{G}_t)$ is less than or greater than 1 depending on whether ϕ is positive or negative – that is, if $\phi > 1$ (< 1) individual bank shocks are amplified (reduced) through the system.

The network impulse response provides a metric to identify which bank's shocks have the largest impact on the overall liquidity. Moreover, it does so taking into account both the size of the bank (via σ_i), the network multiplier, ϕ , and all the direct and indirect links between banks, since $\mathbf{1}' \{\mathbf{M}(\phi, \mathbf{G}_t)\}_{.i}$ is the solution, for $|\phi| < 1$, of

$$\mathbf{1}' \{\mathbf{M}(\phi, \mathbf{G}_t)\}_{.i} = \mathbf{1}' \{\mathbf{I} + \phi \mathbf{G}_t + \phi^2 \mathbf{G}_t^2 + \dots\}_{.i} = \mathbf{1}' \left\{ \sum_{k=0}^{\infty} \phi^k \mathbf{G}_t^k \right\}_{.i}$$

where the first element in the series captures the direct effect of a unit idiosyncratic shock to bank i , the second element captures the effects through the first order network links, the third element captures the effect through the second order links, and so on. This also implies that $\{\mathbf{M}(\phi, \mathbf{G}_t)\}_{ji}$ measures the total (direct and indirect) effect of a shock to bank i on the liquidity of bank j .

Furthermore, the network impulse response functions provide a natural decomposition of the variance of the total liquidity in the network system, since

$$Var_t(\mathbf{1}'\epsilon_t) \equiv vec(\{NIRF_i(\phi, \sigma_i, \mathbf{G}_t)\}_{i=1}^n)' vec(\{NIRF_i(\phi, \sigma_i, \mathbf{G}_t)\}_{i=1}^n).$$

where Var_t denotes the time t variance conditional on time $t - 1$ information.²⁹

We can also isolate the purely network part of the impulse response function, that is, the liquidity effect in excess of the direct effect of a shock to a bank (which we call “excess NIRF”):

$$NIRF_i^e(\phi, \sigma_i, \mathbf{G}_t) \equiv NIRF_i(\phi, \sigma_i, \mathbf{G}_t) - \sigma_i = \mathbf{1}' \{(\mathbf{I} - \phi \mathbf{G}_t)^{-1} \phi \mathbf{G}_t\}_{.i} \sigma_i, \quad (29)$$

and the above, setting $\sigma_i = 1$, i.e. considering a unit shock, is exactly the centrality measure of Katz (1953). Note that $NIRF_i^e(\phi, \sigma_i, \mathbf{G}_t)$ has by construction the same sign as ϕ .

Note also that it is straightforward to compute confidence bands for the estimated network impulse response functions (using the delta method), since they are simply a function of the distribution of $\hat{\phi}$, and $\hat{\phi} - \phi_0$ has the canonical Quasi-MLE asymptotic Gaussian distribution (see Section A.3.2 in the Appendix).

IV Description of the Network and Other Data

We study the sterling interbank network over the sample period January 2006 to September 2010. The estimation frequency is daily, but we also use higher frequency data to construct several of the control variables defined below. The network we consider comprises of all banks in the CHAPS system during the sample – a set of 11 banks. These banks play a key role in the sterling large value payment system since they make payments both on their own behalf and on behalf of banks that are not direct members of CHAPS. The banks in the network are: Halifax Bank of Scotland (owned by Lloyds Banking Group); Barclays; Citibank (the consumer banking arm of Citigroup); Clydesdale (owned by National Australia Bank); Co-operative Bank (owned by The Co-operative Group); Deutsche Bank; HSBC (that incorporated Midland Bank in 1999 – one of the historical “big four” sterling clearing banks³⁰); Lloyds TSB; Royal Bank of Scotland (including Natwest); Santander

²⁹Note that, by construction, \mathbf{G}_t is in the time $t - 1$ information set.

³⁰For most of the 20th Century, the phrase “the Big Four” referred to the four largest sterling banks, which acted as clearing houses for bankers’ cheques. These were: Barclays Bank; Midland Bank (now part of HSBC); Lloyds Bank (now Lloyds TSB Bank and part of Lloyds Banking Group); and National

(formerly Abbey, Alliance & Leicester and Bradford & Bingley, owned by Banco Santander of Spain); and Standard Chartered.

We split our sample into three sub-samples of similar length: the Pre-crisis period (1 January 2006 to 9 August 2007); the Post Northern Rock/ Hedge Fund Crisis period (10 August 2007 to 19 September 2009); the Post Asset Purchase Programme period (20 September 2009 to 30 September 2010). This is explained in more detail below.

Our proxies for the intensity of network links are the interbank overnight borrowing relations. This data is identified from overnight payment data between banks by applying an algorithm developed by Furfine (2000). This is an approach which is common to most papers on the interbank money market. The algorithm identifies pairs of payments between two payment system counterparties where the outgoing payment (the loan) is a multiple of 100,000 and the incoming payment (the repayment) happens the following day and is equivalent to the outgoing payment plus a plausible interest rate. This algorithm has been tested thoroughly, and tracks accurately the LIBOR rate on the whole.³¹ Furfine (2000) showed that the algorithm accurately identifies the Fed Funds rate when applied to Fedwire data.³²

The loan data are compiled to form an interbank lending and borrowing network. In particular, the elements $g_{ij,t}$ of the adjacency matrix \mathbf{G}_t are given by the fraction of bank's

Westminster Bank ("NatWest", now part of The Royal Bank of Scotland Group). Currently, the largest four U.K. banks are Barclays, HSBC, Lloyds Banking Group and The Royal Bank of Scotland Group (with a combined market capitalization of more than £254bn) closely followed by Standard Chartered (with a market cap of over £37bn) – and all of these banks are part of our network.

³¹The data are only for banks which are participants in the payment systems. This creates two problems. First, some loans may be attributed to the settlement bank involved when in fact the payments are made on behalf of one of their customers. Second, where a loan is made between one customer of a settlement bank and another, this transaction will not be settled through the payment system but rather across the books of the settlement bank. This is a process known as internalisation. Internalised payments are invisible to the central bank, so they are a part of the overnight money market that will not be captured.

³² As documented in Armantier and Copeland (2012), the Furfine's algorithm is affected by Type I and, to a lesser extent, Type II, errors. Nevertheless, this is less of a concern in our application since: first, as documented in Kovner and Skeie (2013), at the overnight frequency we focus on, interbank exposures measured by the algorithm are highly correlated with the Fed funds borrowing and lending reported in bank quarterly regulatory filings; second, and most importantly, instead of using the daily identified borrowing and lending relationships, we smooth these exposures by computing rolling monthly averages, therefore greatly reducing the relevance of false positives and negatives in the identification of the interbank relationships. Furthermore, in the empirical application, we apply several robustness checks to our measure of interbank linkages.

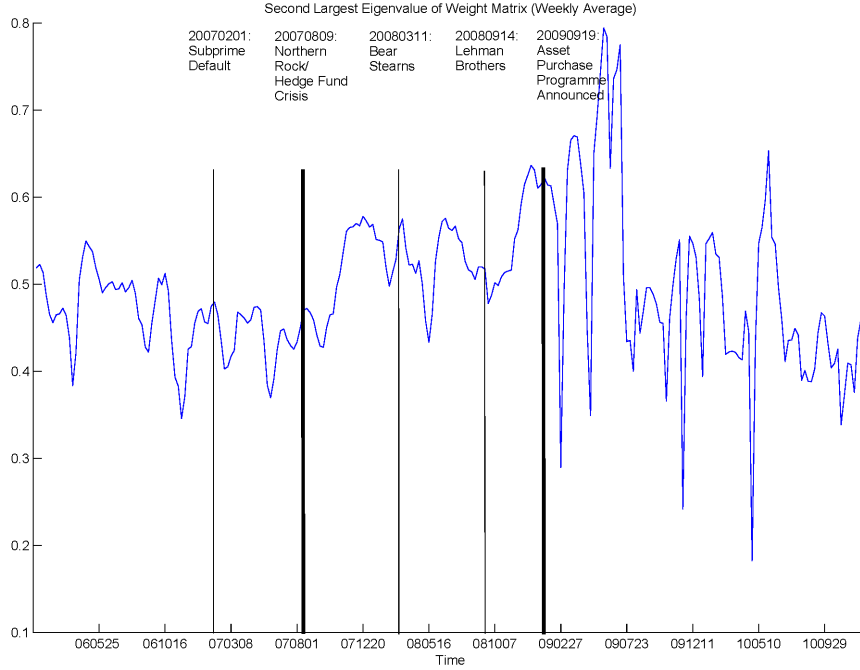


Figure 1: Second largest eigenvalue of \mathbf{G}_t .

i overnight loans that come from bank j . In the baseline specification, the weights at time t are computed as monthly averages for the previous month ending on day $t - 1$.

By construction, \mathbf{G}_t is a square right stochastic matrix. Its largest eigenvalue is therefore equal to one. This implies that the potential propagation of shocks within the system will be dominated by the second largest eigenvalue of the adjacency matrix.³³ The time series of the second largest eigenvalue of \mathbf{G}_t is presented in Figure 1. As can be seen in the figure, there was a substantial increase of the eigenvalue in the aftermath of the Northern Rock/Hedge Fund Crisis period (September 2007), but what is striking is the substantial increase in the volatility of the network links in the post QE period.

One way to characterise time variation in the cohesiveness of the network is to examine the behaviour of the Average Clustering Coefficient (ACC – see Watts and Strogatz (1998)) defined as

$$ACC_t = \frac{1}{n} \sum_{i=1}^n CL_i(\mathbf{G}_t), \quad CL_{i,t} = \frac{\#\{jk \in \mathbf{G}_t \mid k \neq j, j \in n_i(\mathbf{G}_t), k \in n_i(\mathbf{G}_t)\}}{\#\{jk \mid k \neq j, j \in n_i(\mathbf{G}_t), k \in n_i(\mathbf{G}_t)\}}$$

³³This is because \mathbf{G}^k can be rewritten in Jordan normal form as $\mathbf{P}\mathbf{J}^k\mathbf{P}^{-1}$ where \mathbf{J} is the (almost) diagonal matrix with eigenvalues (or Jordan blocks in case of repeated eigenvalues) on the main diagonal.

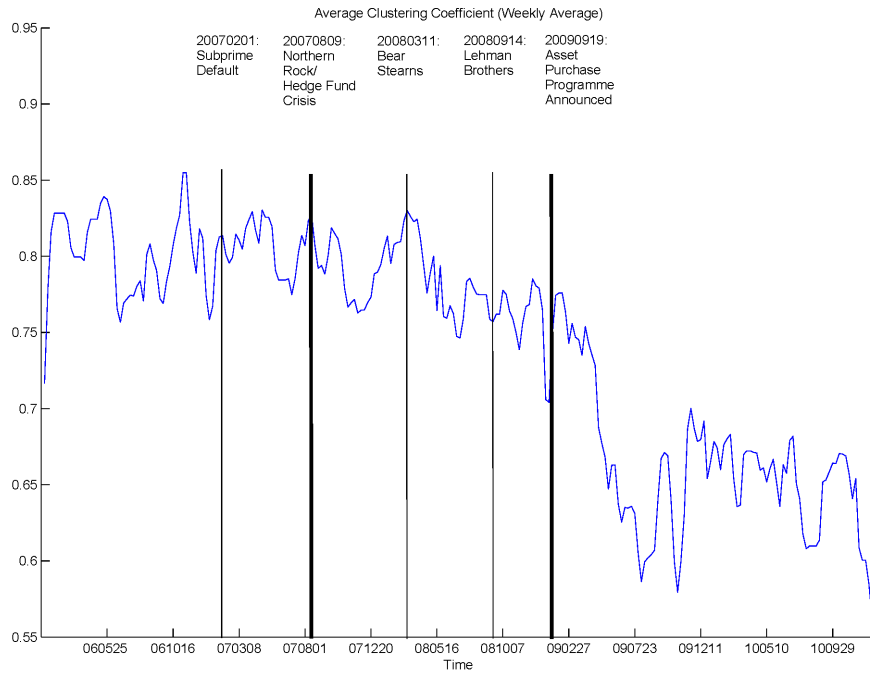


Figure 2: Average clustering coefficient of the interbank network.

where $n_i(\mathbf{G}_t)$ is the set of players that have a direct link with player i and $\#\{\cdot\}$ is the count operator. The numerator of $CL_{i,t}$ is the number of pairs of banks linked to i that are also linked to each other, while its denominator is simply the number of pairs of banks linked to i . Therefore, the average clustering coefficient measures the average proportion of banks that are connected to bank i who are also connected to each other. By construction this value ranges from 0 to 1.

The time series of the ACC is reported in Figure 2. The figure shows that at the beginning of the sample the network is highly cohesive since, on average, around 80% of the pairs of banks connected to any given bank are also connected to each other. The degree of connectedness seems to have a decreasing trend during 2007–2008, and a substantial and sudden decrease following the Asset Purchase Programme, when the average clustering coefficient decreased by about one-quarter of its pre-crisis average. This might be the outcome of reduced interbank borrowing needs during the QE period thanks to the availability of additional reserves from the Bank of England (combined with a move towards increased collateralisation of borrowing and an overall deleveraging of banks balance sheets, see, e.g. Westwood (2011)). This interpretation is consistent with Figure 3, which depicts

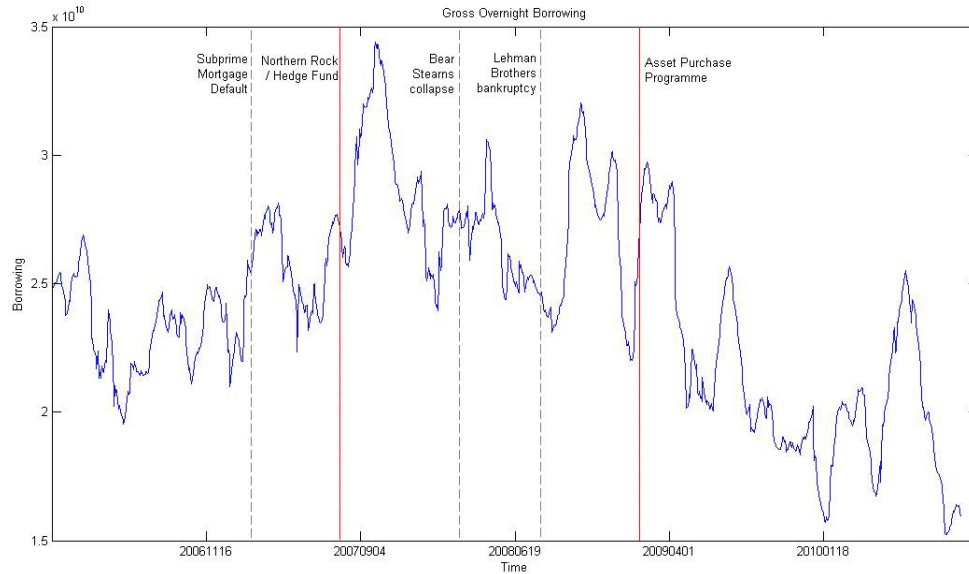


Figure 3: Daily gross overnight borrowing in the interbank network (rolling monthly average).

the (rolling monthly average of) daily gross overnight borrowing in the interbank network. The data record a substantial increase in overnight borrowing as the initial response to the turmoil in the financial market, possibly caused by a shift to very short borrowing due to increased difficulties in obtaining long term financing (Wetherilt, Zimmerman, and Soramaki (2010)), and a substantial decrease in overnight borrowing after the beginning of the QE period.

To measure the dependent variable $l_{i,t}$, that is, the liquidity holdings of each bank, we use central bank reserve holdings. We supplement this with the collateral that is repo'd with the Bank of England in return for intraday liquidity (these repos are unwound at the end of each working day). For robustness, we also analyse separately the behaviour of each of these two liquidity components.³⁴ The weekly average of the total liquidity in the system (computed as the sum of the bank specific liquidity holdings) is reported in Figure 4. The figure depicts a substantial upward trend in the available liquidity in the post subprime default subsample and during the various financial shocks registered in 2008–2009, consistently with the evidence of banks' hoarding liquidity in response to the

³⁴These results are reported in an appendix available upon request.

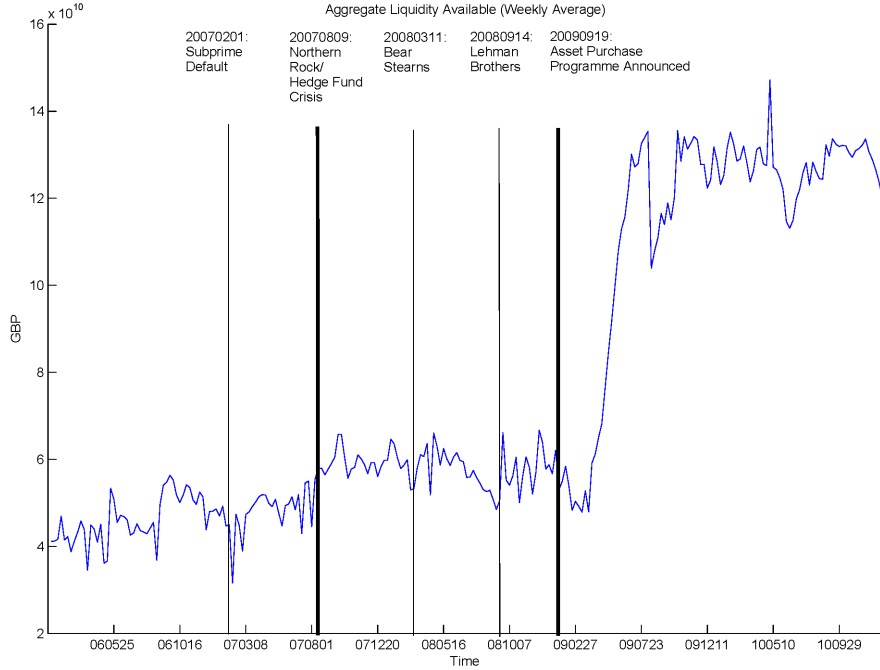


Figure 4: Weekly average of aggregate liquidity available at the beginning of the day.

financial crisis (Acharya and Merrouche (2010)), but this upward trend is dwarfed by the steep run-up registered in response to the Asset Purchase Programme (aka Quantitative Easing) that almost tripled the average liquidity in the system. Interestingly, as shown in Figure (11) in the Appendix, the sharp increase in liquidity in the last part of the sample is associated with a dramatic reduction in the velocity of money.

As covariates, in addition to common monthly time dummies meant to capture time effects, and bank fixed effects, meant to capture unobserved heterogeneity, we use a large set of aggregate (x_t^p) and bank specific ($x_{i,t}^m$) control variables. Note that since in the econometric specification in equations (24) and (25) the network effects are elicited through their contribution to the residuals, any overfitting from those control variables will reduce the variation in residuals, and thus, lead to a conservative estimate of network effects.

Aggregate Control Variables (x_t^p): All the common control variables, meant to capture aggregate market conditions, are lagged by one day so that they are predetermined with respect to time t innovations. To control for aggregate market liquidity condition we use the total liquidity in the previous day. To proxy for the overall cost of funding liquidity we use the lagged LIBOR rate and the interbank rate premium in the sample network

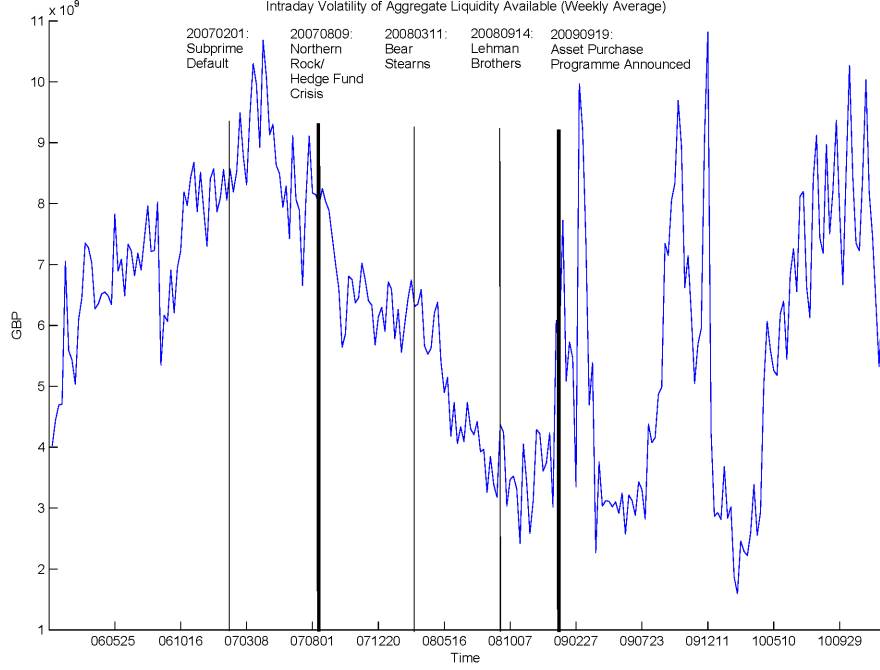


Figure 5: Intraday volatility of aggregate outflows.

(computed as the difference between the overnight interest rate, averaged over banks in the sample sterling network, and the LIBOR rate).

Since banks' decisions to hold liquidity are likely to be influenced by the volatility of their daily payment outflows, we construct a measure of the intraday payments volatility, defined as

$$VolPay_t = \sqrt{\frac{1}{88} \sum_{\tau=1}^{88} (P_{t,\tau}^{out})^2} \quad (30)$$

where P^{out} denotes payment outflows and 88 is the number of 10-minute time intervals (the unit of time for payment recording) within a day. The time series of this variable is reported in Figure 5. It is characterised by a strong upward trend before the subprime default crisis, and a distinctively negative trend during the period of financial turmoil preceding the beginning of QE. During the QE period, this variable has no clear trend but is characterised instead by a substantial increase in volatility.

We also control for the turnover rate in the payment system (see Benos, Garratt, and

Zimmerman (2010)). This variable is

$$TOR_t = \frac{\sum_{i=1}^N \sum_{\tau=1}^{88} P_{i,t,\tau}^{out}}{\sum_{i=1}^N \max \{ \max_{\tau \in [1,88]} [CNP(\tau; i, t)], 0 \}}$$

where the cumulative net debit position (CNP) is defined as the difference between payment outflows and inflows. The numerator captures the total payments in the system in a day, while the denominator is the sum of the maximum net debt positions of all banks in a given day. This variable is meant to capture the velocity of transactions within the interbank system and its time series is reported in Figure 12 of Appendix A.4, and indicates an increased turnover during the financial turmoil, followed by a reduction to levels below the historical average during the QE period.

Since banks have some degree of freedom in deciding on the timing of their intraday outflows, they could use this strategically. Therefore, we control for the right kurtosis³⁵ (rK_t) of intraday payment times. The time series of this variable is reported in Figure 13 of Appendix A.4 and shows a substantial increase during the QE period.

Bank Characteristics ($x_{i,t}^m$): As for the aggregate control variables, all bank characteristic variables are lagged by one day³⁶ so that they are predetermined with respect to innovations at time t . Despite the fact that we control for average interest rates (LIBOR and average overnight borrowing rate), we also control for the bank specific overnight borrowing rate (computed as the average weighted by the number of transactions). We include

³⁵We define as right and left kurtosis (denoted, respectively, by rK_t and lK_t) the fractions of kurtosis of payment times generated by payment times that are, respectively, above and below the average payment time of the day:

$$rK_t = \frac{\sum_{\tau > m_t} \left(\frac{\tau - m_t}{\sigma_t} \right)^4}{\sum_{\tau=1}^{88} \left(\frac{\tau - m_t}{\sigma_t} \right)^4}; \quad lK_t = \frac{\sum_{\tau < m_t} \left(\frac{\tau - m_t}{\sigma_t} \right)^4}{\sum_{\tau=1}^{88} \left(\frac{\tau - m_t}{\sigma_t} \right)^4};$$

where m_t and σ_t are defined as

$$m_t = \frac{1}{88} \sum_{\tau=1}^{88} \tau \left(\frac{P_{t,\tau}^{OUT}}{\sum_{t=1}^T P_{t,\tau}^{OUT}} \right), \quad \sigma_t^2 = \frac{1}{88-1} \sum_{\tau=1}^{88} \left[\tau \left(\frac{P_{t,\tau}^{OUT}}{\sum_{t=1}^{88} P_{t,\tau}^{OUT}} \right) - m_t \right]^2.$$

³⁶For controls variables available at lower than daily frequency. i.e. monthly, we use the latest lagged observation. These variables are: the repo liabilities to asset ratio, total assets, and the cumulative change in the ratio of retail deposits to total assets.

these variables (reported in Figure 14 of Appendix A.4) because in response to each of the collapses, of Northern Rock and of Lehman Brothers, there was a substantial increase in the cross-sectional dispersion of the overnight borrowing rates, and this increase in dispersion persisted during the QE period (see Figure 15 of Appendix A.4). We also control for: bank specific right kurtosis of the time of intraday payments in ($rK_{i,t}^{in}$, to capture a potential incentive to increase bank liquidity) and out ($rK_{i,t}^{out}$, since banks in need of liquidity might have an incentive to delay their outflows); the intraday volatility of the used liquidity ($VolPay_{i,t}$, defined as in equation (30) but using bank specific flows); the total amount of intraday payments ($LevPay_{i,t} \equiv \sum_{\tau=1}^{88} P_{i,t,\tau}^{out}$); the liquidity used ($LU_{i,t}$) as defined in Benos, Garratt, and Zimmerman (2010);³⁷ the ratio of repo liabilities to total assets; the cumulative change in the ratio of retail deposits to total assets; the total lending and borrowing in the interbank market; the cumulative change in the 5-year senior unsecured credit default swap (CDS) premia; and a dummy variable for the top four banks in terms of payment activity.

V Estimation Results

As first exercise, we estimate our empirical network model specified in equations (24) and (25) using three sub-periods of roughly equal size. These are the sub-period before the Northern Rock/Hedge Fund Crisis (Period 1), the sub-period immediately after the Northern Rock/Hedge Fund Crisis but before the announcement of the Assets Purchase Programme (Period 2), and the sub-period running from the announcement of the Assets Purchase Programme to the end of the period of study (Period 3). We split our sample into these three parts since *a*) they correspond to very different overall market conditions, and *b*), as documented in Section IV, the network structure and behaviour of these sub-periods seem to differ substantially. Period 1 is a relatively tranquil period for the banking sector. Period 2 is characterised by several significant events in world financial markets, such as: the run on Northern Rock (the first U.K. bank run in 150 years), the subprime mortgage

³⁷Liquidity used on day t is defined as $LU_{i,t} = \max\{\max_{\tau \in [1,88]} [CNP(\tau; i, t)], 0\}$.

Table 1: Spatial Error Model Estimation

	<i>Period 1</i>	<i>Period 2</i>	<i>Period 3</i>
Panel A: G_t based on borrowing			
$\hat{\phi}$	0.8137 (21.47)	0.3031 (1.90)	-0.1794 (-4.96)
R^2	66.01%	92.09%	91.53%
$1/(1 - \hat{\phi})$	5.3677 (4.92)	1.4349 (4.37)	0.8479 (32.61)
$\sqrt{\frac{Var(Z_t \hat{\phi})}{Var(Z_t \phi=0)}}$	5.59	1.25	0.89
Panel B: G_t based on lending			
$\hat{\phi}$	0.8209 (20.38)	0.2573 (1.23)	-0.3925 (-7.18)
R^2	66.02%	91.63%	91.61%
$1/(1 - \hat{\phi})$	5.5835 (4.45)	1.3464 (3.54)	0.7181 (25.49)
$\sqrt{\frac{Var(Z_t \hat{\phi})}{Var(Z_t \phi=0)}}$	5.71	1.45	0.85
Panel C: G_t based on borrowing and lending			
$\hat{\phi}$	0.8204 (19.36)	0.3258 (1.89)	-0.2824 (-6.10)
R^2	63.98%	92.22%	91.70%
$1/(1 - \hat{\phi})$	5.5679 (4.24)	1.3464 (3.91)	0.7181 (26.74)
$\sqrt{\frac{Var(Z_t \hat{\phi})}{Var(Z_t \phi=0)}}$	5.94	1.15	0.74

Estimation results for equations (24) and (25). Periods 1, 2 and 3, correspond, respectively, to before the Northern Rock/Hedge Fund Crisis, after the Hedge Fund Crisis but before the Asset Purchase Programme, and after the Asset Purchase Programme announcement. The t -statistics are reported in parentheses under the estimated coefficients. Standard errors are QMLE robust ones. For the average network multiplier, $1/(1 - \hat{\phi})$, the delta method is employed. In Panel A, the adjacency matrix is computed using the borrowing relationships, while in Panels B and C we use, respectively, total lending and total borrowing plus lending (row normalized)

hedge fund crisis, the Federal Reserve intervention in Bear Stearns and its subsequent sale to JP Morgan Chase, and the bankruptcy of Lehman Brothers. Period 3 is characterised by a real regime shift – the beginning of Quantitative Easing – in U.K. monetary policy.³⁸

The estimation results for these three subsamples are reported in Panel A of Table 1, where we report only the estimates of the spatial dependency parameter ϕ (first row), the R^2 of the regression (second row), the implied average network multiplier (third row) $1/(1 - \phi)$,³⁹ as well as the ratio of the volatility of network liquidity to the counterfactual volatility

³⁸See Bank of England, ‘Financial Stability Report,’ No. 24, October 2008 and ‘Financial Stability Report,’ No. 25, June 2009.

³⁹Note that from equation (26) we can compute the *average* network multiplier as the total impact on

that would have been generated if $\phi = 0$. Omitted from the table are the coefficient estimates associated with the control variables, which are reported in Table A1 of the Appendix.

The first row of the panel reports the estimates of the network coefficient ϕ . Recall that $\phi > 0$ (< 0) implies that banks' liquidity holding decisions are strategic complements (substitutes) and that this tends to amplify (reduce) the effect of bank specific liquidity shocks. In the first period, the point estimate of this coefficient is about 0.8137 (and highly significant) indicating the presence of a substantial network multiplier effect for liquidity shocks: a £1 idiosyncratic shock equally spread across banks would result in a $1/(1 - \hat{\phi}) = \text{£}5.3677$ shock to aggregate liquidity.

In the second period, the coefficient ϕ is still statistically significant but it is substantially reduced in magnitude, to 0.3031, implying weak complementarity, with an (average) shock multiplier of about 1.4349. This finding suggests that in response to the turbulence in the financial market that have characterised Period 2, banks' liquidity management objective increasingly tilted to free riding neighbours, and away from responding to informational spillover.

In Period 3, the coefficient ϕ becomes negative, -0.1794 , but is still highly significant, implying an average network shock multiplier of about 0.8479. This is particularly interesting since a negative ϕ implies strategic substitution in liquidity holdings, as in Bhattacharya and Gale (1987), that is, a situation in which individual banks decide to hold less liquidity when neighbouring banks hold more liquidity. As a result, the network has a damping effect on shock transmission. The fact that negative network effect is observed during the Quantitative Easing sub-period suggests that the liquidity multiplier effect was not working at the time of the large inflow of liquidity from the central bank through Asset Purchase Programme. The total liquidity injection from the program was, up to October 2011, of

aggregated liquidity resulting from a unit shock equally spread across the n banks. This is given by

$$\mathbf{1}'\mathbf{M}(\phi, \mathbf{G}_t)\mathbf{1}\frac{1}{n} = \frac{1}{1 - \phi}.$$

about £275 billion.⁴⁰ Overall, the fit of the model is quite good in all sub-periods, with an R^2 in the range 66% – 92%.

The last row of Panel A reports $\sqrt{\text{Var}(Z_t|\hat{\phi})/\text{Var}(Z_t|\phi = 0)}$ i.e. the ratio of the volatility of the network liquidity implied by our estimate of ϕ to the (counterfactual) volatility of network liquidity that would have arisen if there were no network externalities.⁴¹ This last statistic makes clear that the large positive network multiplier in the first period generates a 459% increase in volatility. In the second period instead the reduced network multiplier generates an excess volatility of only about 25%, while in the third period the negative network multiplier generates a reduction in the volatility of network liquidity of about 11%.

Finally, for robustness, in Panels B and C, we reestimate our network model with two alternative constructions of the adjacency matrix G_t . In particular, in Panel B we use the lending flows, while in Panel C we use the combined borrowing and lending flows (in both cases the adjacency matrix is row normalized).⁴² Overall, the estimates in Panels B and C are extremely similar, both qualitatively and quantitatively, to the ones reported in Panel A.⁴³

With the subperiod estimates at hand, we can compute the network impulse response functions to identify the risk key players in the interbank market. The results for Period 1 are reported in the upper panel of Figure 6. In particular, in the upper panel we report the excess network impulse response functions to a unit shock $NIRF^e(\hat{\phi}, 1, \bar{\mathbf{G}}_1)$ defined in equation (29) (where $\bar{\mathbf{G}}_j$ denotes the average \mathbf{G}_t in the j -th subsample), as well as the two standard deviation error bands. Also, as a point of reference, we report in the same

⁴⁰See http://www.bankofengland.co.uk/monetarypolicy/Pages/qe/qe_faqs.aspx.

⁴¹For completeness, in Table A2 in the Appendix we also report $\sqrt{\text{Var}(z_{i,t}|\hat{\phi})/\text{Var}(z_{i,t}|\phi = 0)}$ for each bank.

⁴²Note that when constructing the theoretical model, we emphasize that the network linkages reflect the interbank relationship, which may transmit information and/or liquidity (i.e., interbank credit). Thus, the network linkage is not necessarily just about borrowing. If a bank lends to another, the relationship formed through such transaction may facilitates future borrowing or information transmission.

⁴³Note that in the theoretical model, the network linkages reflect the interbank relationship, which may transmit information and/or liquidity. Thus, the network linkages are not necessarily only about borrowing. If a bank lends to another, the relationship formed through such transaction may facilitates future borrowing or information transmission.

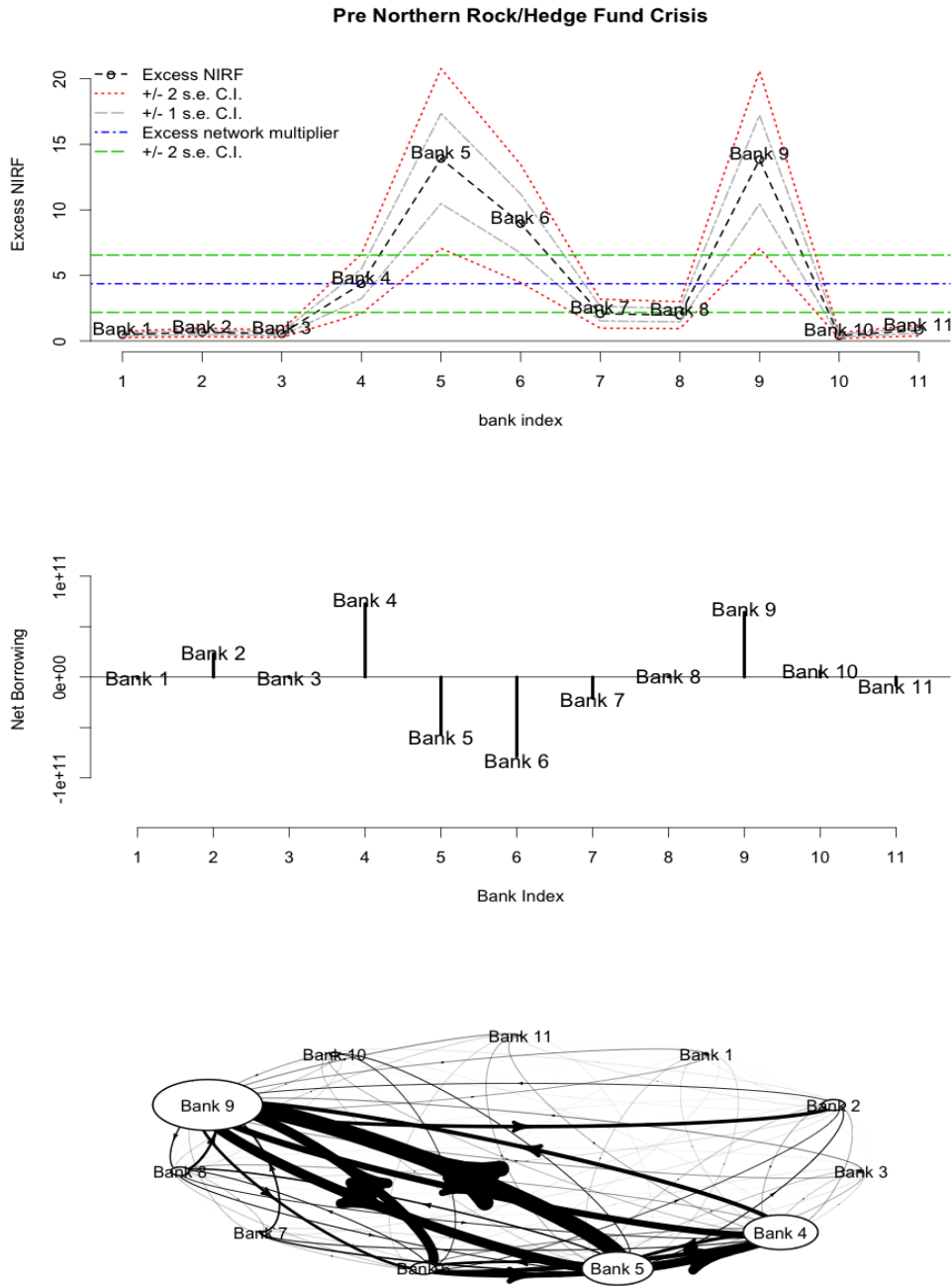


Figure 6: The period before the Northern Rock/Hedge Fund Crisis. Network excess impulse response functions to a unit shock (upper panel); net borrowing (central panel); borrowing and lending flows (lower panel) where the ellipses identifying individual banks are (log) proportional to their total gross borrowing in the system, incoming arrows to a node indicate borrowing flows to that node, while outgoing arrows indicate lending flows from that node, and the thickness of the arrows is (log) proportional to the sterling value of the flows.

panel the average network multiplier in excess of the unit shock (i.e. $(1 - \phi)^{-1} - 1$). As mentioned earlier, the point estimate in Period 1 implies a large average network multiplier of shocks to individual banks, and the picture shows that in response to a £1 idiosyncratic shock equally spread across banks, the final compounded shock to the overall liquidity would be increased by another £4.3677. Nevertheless, what the upper panel of Figure 6 stresses is that this large network amplification of shocks is due to a small subset of banks. In particular: a £1 idiosyncratic shock to the liquidity of either Bank 5 or Bank 9 would generate an excess reaction of aggregate liquidity of about £13.9 and £13.8; the same shock to Bank 6 would result in an excess reaction of aggregate liquidity of about about £8.9; instead, a shock to Bank 4 would have an effect that is roughly of the same size as the average network multiplier while a shock to any of the remaining seven banks would be amplified much less by the network system. That is, the network impulse response functions stress that there is a small subset of key players in the interbank liquidity market that generate most of the network risk.

The central panel of Figure 6 shows the average net borrowing during Period 1. Comparing the upper and central panels of the figure, it is interesting to notice that simply looking at the individual *net* borrowing behaviour one cannot identify the riskiest players for the network. In particular, the two riskiest players identified through our structural estimation are not the largest net borrowers in the network – the largest net borrower, Bank 4, is instead an average bank in network risk terms. Moreover, Bank 5, one of the two largest network risk contributors, is not a net borrower – it is instead the second largest net lender.

The comparison between the top two panels also makes clear that the risk key players are not necessarily the net borrowing banks – net borrowers and net lenders are roughly as likely to be the network risk key players. This result is intuitive: negative liquidity shocks to a bank that lends liquidity to a large share of the network can be, for the aggregate liquidity level, as bad as a negative shock to a bank that is borrowing liquidity from other banks. But the comparison between the two panels also makes it clear that simply looking at the largest players in terms of net borrowing or lending would not identify the key risk

players for the system.

The reasons behind this finding can be understood by looking at the lower panel of the figure, where we present the average network structure during Period 1. In particular, the size of the ellipses identifying individual banks are (log) proportional to their total gross borrowing in the system, incoming arrows to a node indicate borrowing flows to that node while outgoing arrows indicate lending flows from that node, and the thickness of the arrows is (log) proportional to the sterling value of the flows. The lower panel shows that key risk contributors tend to be banks with the most connections and largest flows (and with most links to other well connected banks), i.e., banks with relatively high centrality, but are not necessarily the players that borrow or lend more in either gross nor net terms.

Figure 7 reports excess impulse response functions (upper panel), average net borrowing positions (central panel), and network flows (lower panel) for Period 2 – the period characterised by a high degree of stress in the financial market. The first thing to notice is that despite the overall increase in activity in the interbank borrowing and lending market (outlined by both the central and lower panels and by Figure 3), there is a drastic reduction in the average network multiplier reported in the top panel: the average excess network reaction to a unit shock is only about 0.43. That is, in a period of financial stress, banks seem, on average, to have radically reduced their network risk exposure, and they have done so despite having increased the amount of overnight borrowing and lending used to fund their liquidity needs. Nevertheless, as stressed by the first panel, the network risk profile, even though substantially reduced overall, is still quite high for a small subset of banks. In particular, a unit shock to Bank 5, Bank 9 and Bank 6, would result, respectively, in an excess network liquidity change of 1.77, 1.36 and 0.85, while the same shock to Bank 4 would have an effect very similar to the average one, and a shock to the remaining banks would receive minimal amplification from the network system.

The results for Period 3 – the one starting at the onset of QE – are reported in Figure 8, and are radically different from the ones of the previous two periods. First, banks tend to behave as strategic substitutes in their liquidity holdings in this period, therefore the network has a buffering effect to individual bank shocks, implying a negative average excess

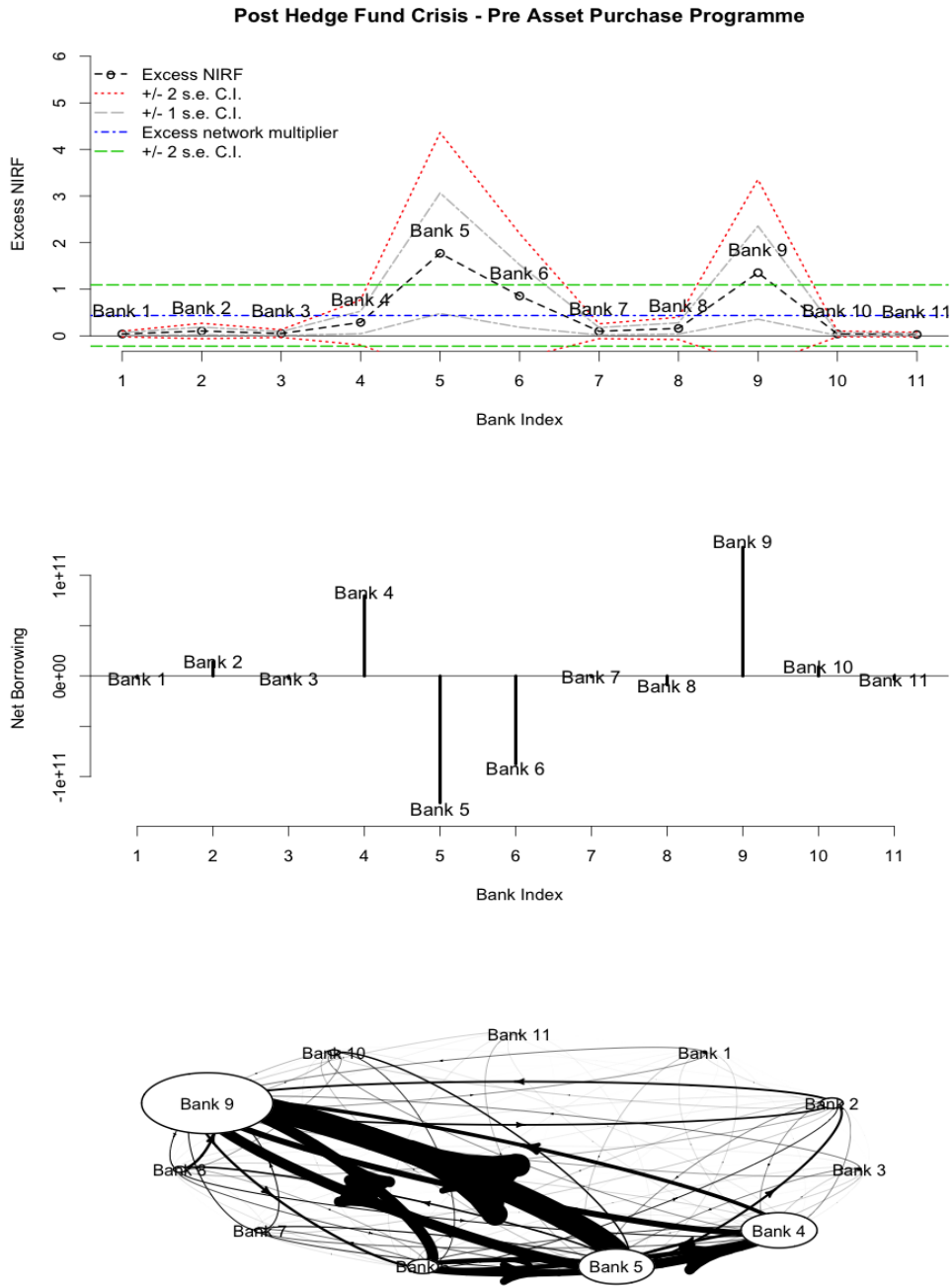


Figure 7: After the hedge fund crisis but before QE. Network excess impulse response functions to a unit shock (upper panel); net borrowing (central panel); borrowing and lending flows (lower panel), where the ellipses identifying individual banks are (log) proportional to their total gross borrowing in the system, incoming arrows to a node indicate borrowing flows to that node, while outgoing arrows indicate lending flows from that node, and the thickness of the arrows is (log) proportional to the sterling value of the flows.

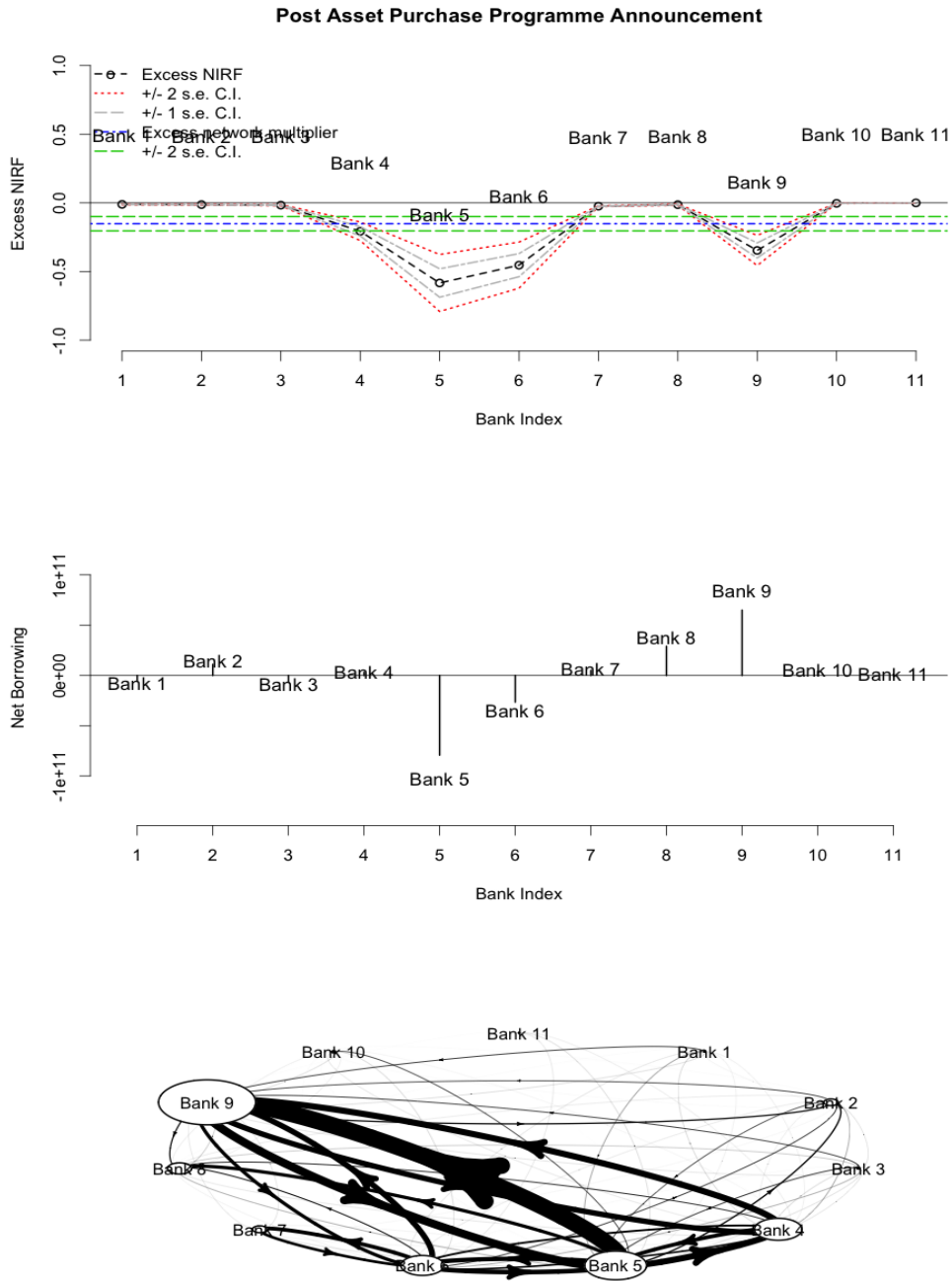


Figure 8: The QE period: Network excess impulse response functions to a unit shock (upper panel); net borrowing (central panel); borrowing and lending flows (lower panel) where the ellipses identifying individual banks are (log) proportional to their total gross borrowing in the system, incoming arrows to a node indicate borrowing flows to that node, while outgoing arrows indicate lending flows from that node, and the thickness of the arrows is (log) proportional to the sterling value of the flows.

multiplier of -0.15 , that is, a unit liquidity shock equally spread across banks would result in a $1 - 0.15 = 0.85$ shock to aggregate liquidity. But, once again, there is substantial heterogeneity among the banks, in the sense that for most banks (Bank 1, 3, 7, 8, 10 and 11) the network has basically no effect on how their own shocks propagate to the system, while for a few other banks (4, 5, 6, and 9), the network structure helps reduce the impact of their own idiosyncratic shocks on aggregate liquidity.

This behaviour arises in a period in which the degree of connectedness of the network was substantially reduced (see Figure 2 and the lower panel of Figure 8), the gross borrowing in the system had been substantially reduced (see Figure 3), most banks held net borrowing positions close to zero (central panel of Figure 8), but at the same time the overall liquidity in the system had been substantially increased (Figure 4).

What is also interesting to notice is that the same banks that were the riskiest players in the previous two periods (Banks 5, 6 and 9) are now the least risky ones for the system. Thanks to their centrality and more importantly the overall strategic substitution behaviour on the network, these banks become the biggest shock absorbers.

A natural question is whether we can explain the large heterogeneity of individual banks' contribution to system risk using banks' characteristics. To answer this question Table 2 reports the rank correlations of individual bank characteristics with the bank-specific network impulse-response functions in the three periods considered. Only a few bank characteristics seems to correlate significantly with the magnitude of the bank specific $NIRF_i^e$ and several observations are in order. Considering the total level of payments channeled by the bank, in periods 1 and 2 the rank correlations for this variable are, respectively, 82.73% and 95.45%, while in period 3 we have -85.45%. This implies that banks that channel a larger amount of payments are more likely to be central for the network, but the implications of this centrality depend on the type of equilibrium in the interbank market: when strategy complementarity is the dominant force (i.e. when $\phi > 0$ as in the first two periods), banks that channel more payments amplify more the shocks in the system; instead, when the equilibrium is characterized by strategic substitution ($\phi < 0$, as in the last period), such banks dampen the effect of shocks on the system. Second, the

Table 2: Rank Correlation of Bank Characteristics and $NIRF_i^e$

	<i>Period 1</i>	<i>Period 2</i>	<i>Period 3</i>
Interbank Rate	20.91%	37.27%	-64.55%
$\ln LevPay_{i,t-1}$	82.73%***	95.45%***	-85.45%***
$rK_{i,t-1}^{in}$	20.00%	-34.55%	10.91%
$rK_{i,t-1}^{out}$	-45.45%	-89.09%***	73.64%**
$\ln VolPay_{i,t-1}$	48.18%	56.36%*	-54.55%*
$\ln LU_{i,t-1}$	21.82%	35.45%	-23.64%
$\frac{Repo\ Liability}{Assets}$	39.45%	48.18%	-37.27%
Total Assets (log)	12.73%	25.45%	4.55%
$\frac{\Delta Deposit}{Assets}$	12.73%	-50%	68.18%**
CDS (log)	38.18%	18.18%	-40.00%
Stock Return (Inc. Dividend)	13.64%	-17.27%	-56.36%*
Total Lending and Borrowing (log)	86.36%***	95.45%***	-89.09%***
Total Lending (log)	97.27%***	99.09%***	-89.09%***
Total Borrowing (log)	66.36%**	91.82%***	-76.36%***
Net Borrowing (log)	-17.27%	10.91%	54.55%*

* represents 10% significance, ** 5% significance, and *** 1% significance.

last row of Table 2 shows that *net* borrowing has no significant rank correlation with banks' $NIRF_i^e$, consistent with Figures (6)-(8). That is, whether a bank is a large net lender or borrower is irrelevant for the shock propagation in the system. Nevertheless, *gross* lending and gross borrowing, and their sum, are all highly correlated with banks' $NIRF_i^e$. That is, banks that borrow and/or lend a lot (in gross terms) are key players in our network. Nevertheless, this centrality has a different impact on the system depending on the sign of ϕ : central banks amplify the shocks when ϕ is positive and dampen them when ϕ is negative. Interestingly, the rank correlations are, in absolute terms, marginally larger for total lending than for total borrowing. Finally, bank size measured by total assets is very weakly (and not significantly) correlated with the $NIRF_i^e$.

V.1 Central Planner vs Market Equilibrium

With the estimates of the structural parameters at hand, we can quantitatively assess the discrepancy, if any, between the banks' liquidity holdings generated in the decentralized equilibrium, and the level of liquidity buffer that a benevolent central planner would have wanted the banks to hold. That is, from equations (8) and (19), we can compute the (average) difference between the aggregate liquidity of planner's choice and the aggregate decentralized liquidity as $\mathbf{1}' [\mathbf{M}^p(\phi, \psi, \delta, \mathbf{G}) - \mathbf{M}(\phi, \mathbf{G})] \bar{\mu}$. Similarly, from equations (10) and (20) we can compute the difference in the level of volatility of the planner's choice and of the aggregate decentralized outcome: $Var(Z^p(\phi, \psi, \delta, \mathbf{G}))^{\frac{1}{2}} - Var(Z^*(\phi, \mathbf{G}))^{\frac{1}{2}}$.

The challenge in computing the above quantities is that we have consistent estimates of ϕ and $\bar{\mu}$, but we cannot directly estimate ψ , δ and γ . Nevertheless, we can calibrate ψ to a natural benchmark: $\psi = 1$. This corresponds to the case in which each bank values in an identical manner the liquidity it holds directly in the network, and the liquidity available via its borrowing links to other banks. Moreover, with $\psi = 1$ we have that $\delta/\gamma = \phi + 1$. Hence we do not need to choose values of γ and δ if $\psi = 1$ and ϕ is set to the estimated value.

Table 3 reports the discrepancies between the central planner's solutions and the market equilibria, based on the point estimates of the structural parameters in Table 1, and the average value of the adjacency matrix \mathbf{G}_t , in the three sub-periods.

In Period 1 – when the (average) network multiplier was extremely large – the market equilibrium features excessive risk from the perspective of a central planner: the central planner would prefer the volatility of liquidity to be reduced by almost 91%. Moreover, albeit marginally, the liquidity level in the system is also excessive. Given the high network multiplier in this period, the network has the capacity to greatly amplify the individual liquidity shocks. Hence, a small reduction in the equilibrium buffer stock holdings (in response to liquidity shocks), from central planner's perspective, will come with a greater reduction in network volatility, therefore delivering a better level–risk trade-off.

In Period 2, given the reduction in ϕ , the market equilibrium produces *ceteris paribus*

Table 3: Central Planner vs. Market Equilibria

	<i>Period 1</i>	<i>Period 2</i>	<i>Period 3</i>
$\Delta\%$ Volatility of Total Liquidity	-90.8%	-64.8%	30.7%
Δ Network Liquidity	-3.47	15.5	-27.5

The three sub-periods are indicated by $j = 1, 2, 3$, and $\bar{\mathbf{G}}_j$ is the average \mathbf{G}_t in sub-period j . The table reports: first row, $100 \times \left[\left(\frac{\text{Var}(Z^p(\hat{\phi}_j, \psi=1, \bar{\mathbf{G}}_j))}{\text{Var}(Z^*(\hat{\phi}_j, \bar{\mathbf{G}}_j))} \right)^{\frac{1}{2}} - 1 \right]$; second row, $\mathbf{1}' \left[\mathbf{M}^p(\hat{\phi}_j, \bar{\mathbf{G}}_j) - \mathbf{M}(\hat{\phi}_j, \bar{\mathbf{G}}_j) \right] \hat{\mu}_j$ (unit: £10bn).

less volatility than in the Period 1. Nevertheless, the market volatility is still too large (by about 65%) from the central planner’s perspective. Moreover, the level of liquidity buffer in this sub-period is much smaller than what is considered optimal by the central planner. That is, Period 2 is characterised by too much risk and too little buffer stock of liquidity. The latter phenomenon is partially due to the fact that the individual average valuations ($\bar{\mu}$) of accessible liquidity are substantially reduced in this period, implying a general decline in the ‘belief’ of the interbank market’s capacity to generate liquidity, causing a significant reduction in buffer stock holdings in the decentralised equilibrium.

In the last sub-period, the (average) network multiplier in the market equilibrium is smaller than 1, hence overall the system dampens the volatility of shocks. From the central planner’s perspective, not enough volatility is generated (by about 31%) while at the same time the aggregate network liquidity buffer is too high. This implies that banks hold idle reserves, and thus, the transmission of monetary policy (i.e., QE in this context) to the broad economy tends to be less effective than envisioned.

V.2 Time Varying Network Effects

The results presented so far indicate a substantial change over time in the role played by the network interactions in determining aggregate liquidity level and risk. In this section, we analyse the drivers of this time variation.

V.2.1 The Drivers of the Time Variation in the Network Amplification

The network impulse response functions depicted in Figures (6)-(8) show substantial time variation in the amplification of shocks between sub-periods. This could be caused by either the time variation in the network topology \mathbf{G} or in the network multiplier ϕ .

To examine these drivers, we compute the changes in the network impulse response functions across the three subperiods. In particular, Panel A of Figure 9 reports the total change in NIRF between Periods 1 and 2 ($NIRF_i(\hat{\phi}_2, 1, \bar{\mathbf{G}}_2) - NIRF_i(\hat{\phi}_1, 1, \bar{\mathbf{G}}_1)$, dashed line with circles), the change due to the variation of \mathbf{G} ($NIRF_i(\hat{\phi}_1, 1, \bar{\mathbf{G}}_2) - NIRF_i(\hat{\phi}_1, 1, \bar{\mathbf{G}}_1)$, dotted line with triangles), and the change due to the variation of ϕ ($NIRF_i(\hat{\phi}_2, 1, \bar{\mathbf{G}}_1) - NIRF_i(\hat{\phi}_1, 1, \bar{\mathbf{G}}_1)$, dash-dotted line with +).

A striking feature of the graph is that most of the total change comes from the reduction in the network multiplier ϕ for all banks. In fact, ceteris paribus, the outdegree centrality (hence the NIRF) of Bank 5 would have increased from Period 1 to 2 due to its increased borrowing and lending activity (captured by $\bar{\mathbf{G}}_2$). However, this effect is dwarfed by the reduction of its NIRF caused by the change in ϕ .

Panel B reports the same decomposition of the change in NIRFs between Periods 2 and 3. Once again the changes are mostly driven by the change in the network multiplier rather than the change in network topology.

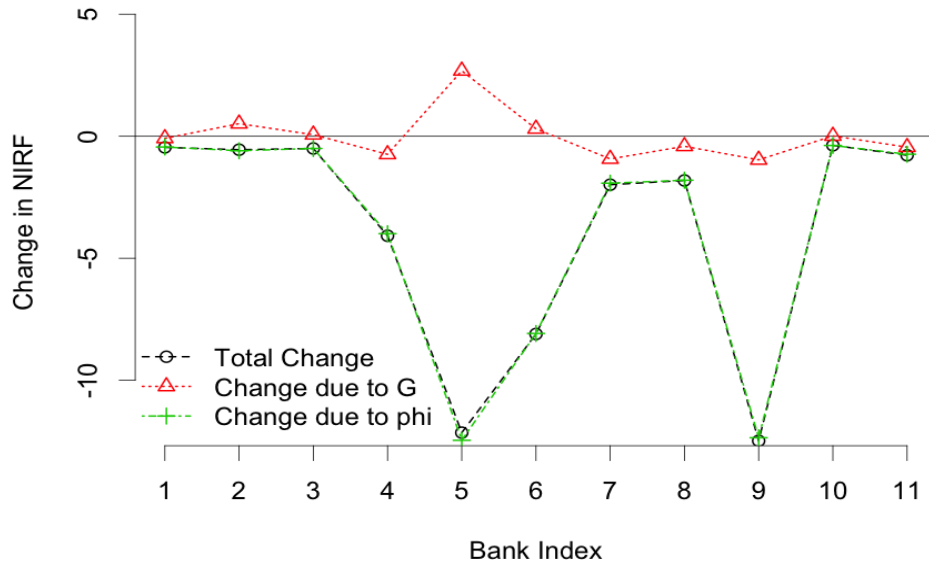
Overall, Figure 9 shows that the time variation of the network multiplier has the first order effect on the network amplification mechanism.

V.2.2 Time Varying Network Multiplier

The results in the previous sections indicate the importance of the time variation of ϕ . Therefore, to capture this time variation, we now estimate the structural model in equations (24) and (25) using a 6-month rolling window.⁴⁴ These rolling estimates of the network

⁴⁴Recall that when \mathbf{G}_i is a right stochastic matrix, separate identifications of the bank (α^{bank}) and network ($\bar{\mu}$) fixed effects require that there is a subset of banks that does not borrow at least at one point in time in each subsample. This condition is not satisfied in all the rolling sub-samples. But since the separate identification of these fixed effect does not affect the identification of ϕ , we normalise the unidentified fixed effects. Moreover, given the very short length of the rolling window, we drop time fixed effects from the specification and the heteroskedastic specification of the shocks. Estimates with the full sets

Panel A: NIRF change decomposition between Period 1 and 2



Panel B: NIRF change decomposition between Period 2 and 3

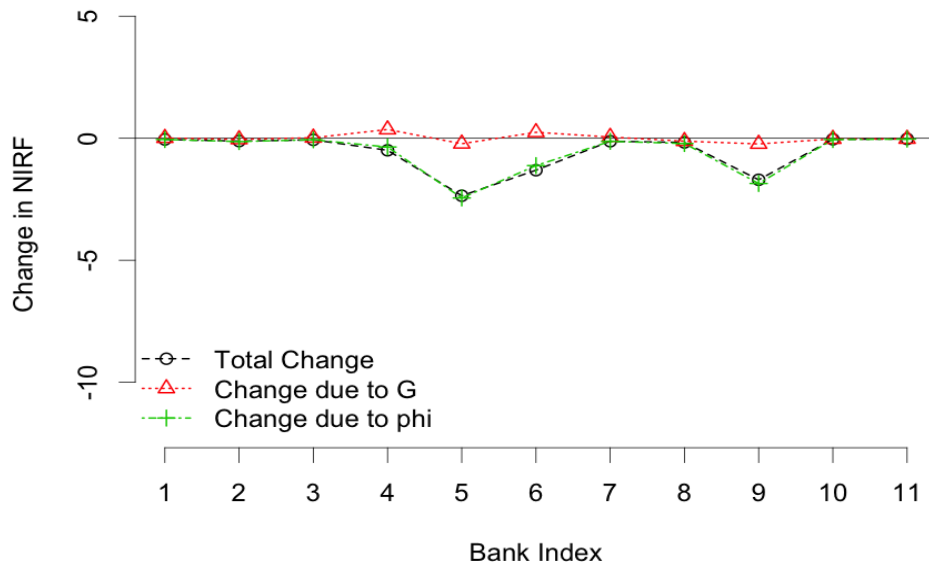


Figure 9: Decomposition of total change in the NIRFs between periods.

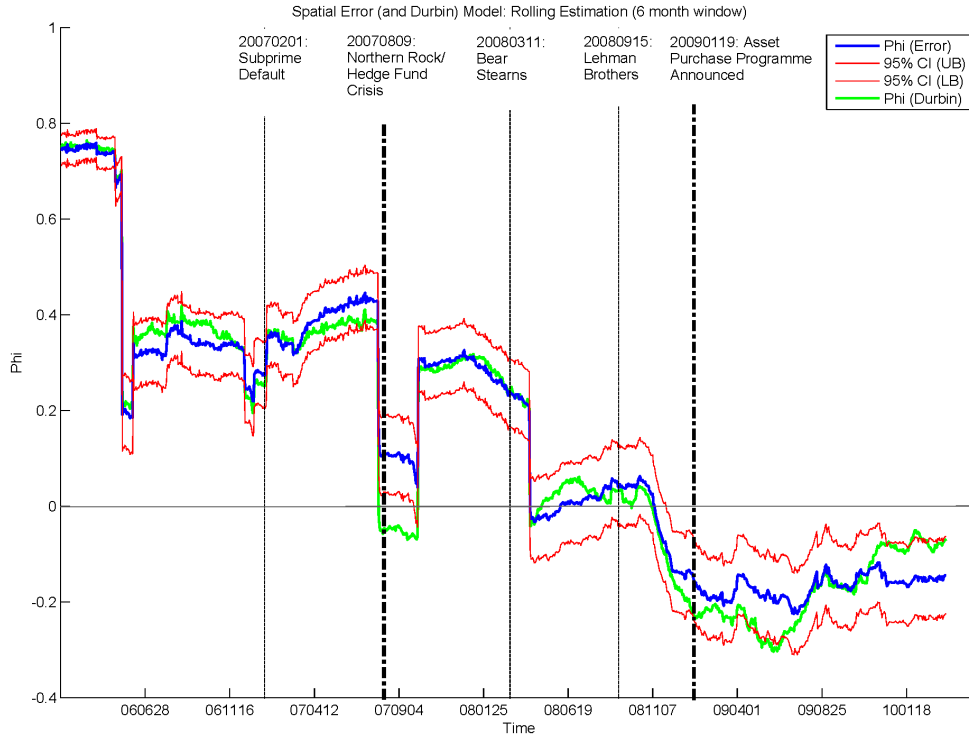


Figure 10: Spatial Error (blue line) and Durbin (green line) rolling estimates of ϕ .

coefficient ϕ are reported (blue line), together with 95% confidence bands (red lines), in Figure 10.

The figure also reports the rolling point estimates of the coefficient ϕ implied by the spatial Durbin model (green line) in equation (27) which, as a more general model, serves as a specification test of our benchmark spatial error model. If the two estimated ϕ are close to each other, this indicates that our theory-driven spatial error specification of the interbank network cannot be rejected for a more general specification.

At the beginning of the sample, the figure shows an extremely large network coefficient, ϕ , implying a substantial network amplification of shocks to banks in the system. The estimated coefficient has its first sharp reduction around the 18th of May 2006 when the Bank of England introduced the reserve averaging system described in Section A.1. The network multiplier is relatively stable after May 2006, except for a temporary decrease during the 2007 subprime default, until the Northern Rock bank run when the network

of fixed effects and heterosceasticity show a very similar behaviour, but with somewhat larger confidence intervals, hence making it easier not to reject the SEM specification. As a consequence, we focus on the more parsimonious specification.

multiplier is drastically reduced for several months. After this reduction, the coefficient goes back to roughly the previous period average but shows a trend decline that culminates in a sharp drop following the Bear Stearns collapse. From this period onward, and until long after the Lehman Brothers bankruptcy, the coefficient is statistically indistinguishable from zero, implying no network amplification of bank specific shocks. That is, the estimation suggests that in this period there was basically no added risk coming from the network structure of the interbank market, and that individual bank shocks would not be amplified by some sort of domino effect in the U.K. interbank market. This figure suggests that the banks' reaction to the financial market turmoil was to reduce the amplification of risk generated through the interbank network. This reduction could come from any of these three sources: *a*) a reduction in the availability of collateralization and/or information spillovers, i.e. δ , *b*) an increase in risk aversion, γ , and *c*) an increased availability of accessible liquidity due to ψ .

Interestingly, the coefficient $\hat{\phi}$ becomes negative, and statistically significant, right before the announcement of the Asset Purchase Programme, and remains stably so throughout the QE period. This indicates that during the liquidity inflow coming from the Bank of England's QE policy (and also in expectation of it), banks started behaving as strategic substitutes in their liquidity holding decision (as implied by Bhattacharya and Gale (1987)). Note that this is a period in which the aggregate supply of central bank reserves was almost completely price inelastic since QE set a target level for asset purchases and let market forces determine their price. This overall change of the BoE supply of reserves is unlikely to be the driver of our estimates of the network multiplier coefficient during this period since: *a*) the change in ϕ actually occurred before the announcement of QE; *b*) we estimate the identified optimal response of the banks to market conditions (effectively, the banks' equilibrium demand function), and we control for variation in aggregate price and quantities of liquidity, as well as bank deposits held by the private sector.

Lastly, this figure outlines that the point estimates of $\hat{\phi}$ coming from our theory-driven spatial error specification and the ones coming from the more general spatial Durbin model are always very similar, both numerically and in terms of their overall evolution during the

sample. Moreover, testing formally for a discrepancy between the two types of estimates, we find that they are statistically different at the 5% confidence level less than 95% of the time, providing support for the spatial error formulation of our network model.

VI Conclusion

In this paper, we develop and estimate a network model of interbank liquidity that is flexible enough to incorporate both strategic complementarity and substitution as potential network equilibria. Based on network topology, the estimated network effects and bank-specific structural shocks, we construct measures of systemic risks and identify the network players that are most important in contributing to the aggregate liquidity and its risk in the banking system.

We find that the network effect varies significantly through the sample period, January 2006 to September 2010. Prior to the Northern Rock/Hedge Fund crisis, liquidity provision in the network was driven by strategic complementarity of the holding decisions. That is, liquidity shocks were amplified by the network and each bank had a large exposure to network shocks. In contrast, during the crisis, the network itself also became less cohesive, and the network amplification was greatly reduced. Finally, during the QE period, in response to the large injection of liquidity in the system, the network became characterised by strategic substitution: that is, individual players free ride each other and the liquidity inflow from the central bank.

To the best of our knowledge, we are the first to estimate substantial time variation in the nature of the equilibrium in a financial network. Moreover, we show that, for risk generation, the change in the type of equilibrium is the dominant force (rather than the change in the network topology itself). This could rationalise the empirical puzzle of network changes having little impact on aggregate quantities in calibration/simulation exercises on interbank networks (Elsinger, Lehar, and Summer (2006)).

Moreover, we estimate the individual bank contributions to aggregate liquidity risk and document that most of the systemic risk is generated by a small subset of key players. Last,

but not least, we also solve for the benevolent central planner equilibrium. This allows us to estimate the gap between central planner and decentralised optima for both liquidity level and risk. In particular, we find that during both the pre-crises and crises periods the system was characterised by an excessive amount of risk and (during the crises) too little liquidity relative to the social optimum.

References

- ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): “The network origins of aggregate fluctuations,” *Econometrica*, 80(5), 1977–2016.
- ACEMOGLU, D., A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2015): “Systemic risk and stability in financial networks,” *American Economic Review*, 105, 564–608.
- ACHARYA, V. V., AND O. MERROUCHE (2010): “Precautionary hoarding of liquidity and inter-bank markets: Evidence from the sub-prime crisis,” New York University working paper.
- AFONSO, G., A. KOVNER, AND A. SCHOAR (2011): “Stressed, not frozen: The Federal funds market in the financial crisis,” *The Journal of Finance*, 66(4), 1109–1139.
- AFONSO, G., AND R. LAGOS (2015): “Trade dynamics in the market for Federal funds,” *Econometrica*, 83, 263–313.
- AFONSO, G., AND H. S. SHIN (2011): “Precautionary demand and liquidity in payment systems,” *Journal of Money, Credit and Banking*, 43, 589–619.
- ALLEN, F., E. CARLETTI, AND D. GALE (2008): “Interbank market liquidity and Central Bank intervention,” New York University working paper.
- ALLEN, F., AND D. GALE (2000): “Financial contagion,” *Journal of Political Economy*, 108(1), 1–33.
- (2017): “How Should Bank Liquidity be Regulated,” Working paper, Imperial College London.
- ANSELIN, L. (1988): *Spatial econometrics: methods and models*, Studies in operational regional science. Kluwer Academic Publishers.
- ARMANTIER, O., AND A. COPELAND (2012): “Assessing the quality of Furfine-based algorithms,” Staff Reports 575, Federal Reserve Bank of New York.
- ASHCRAFT, A., J. MCANDREWS, AND D. SKEIE (2010): “Precautionary reserves and the interbank market,” Federal Reserve Bank of New York working paper.
- ATKESON, A. G., A. L. EISFELDT, AND P.-O. WEILL (2015): “Entry and Exit in OTC Derivatives Markets,” *Econometrica*, 83(6), 2231–2292.
- BABUS, A. (2009): “The Formation of Financial Networks,” Tinbergen Institute Discussion Paper 06-093.
- BABUS, A., AND F. ALLEN (2009): “Networks in finance,” in *Network-based Strategies and Competencies*, ed. by P. Kleindorfer, and J. Wind. Pearson Education, New Jersey, US.
- BABUS, A., AND P. KONDOR (2013): “Trading and Information Diffusion in Over-the-Counter Markets,” Working paper, London School of Economics.

- BALL, A., E. DENBEE, M. J. MANNING, AND A. WETHERILT (2011): “Intraday Liquidity: Risk and Regulation,” *Bank of England Financial Stability Papers*, 11.
- BALLESTER, C., A. CALVO-ARMENGOL, AND Y. ZENOU (2006): “Who’s who in networks. Wanted: The key player,” *Econometrica*, 74, 1403–1417.
- BECH, M. L., AND E. ATALAY (2008): “The topology of the federal funds market,” Discussion paper.
- BECH, M. L., AND E. KLEE (2009): “The mechanics of a graceful exit: interest on reserves and segmentation in the federal funds market,” Federal Reserve Bank of New York Staff Report 416, Federal Reserve Bank of New York.
- BECHER, C., S. MILLARD, AND K. SORAMAKI (2008): “The network topology of CHAPS Sterling,” Discussion paper.
- BENOS, E., R. J. GARRATT, AND P. ZIMMERMAN (2010): “Bank Behavior and Risks in CHAPS Following the Collapse of Lehman Brothers,” UCSB working paper.
- BERNANKE, B. S., AND A. S. BLINDER (1988): “Credit, Money, and Aggregate Demand,” *The American Economic Review*, 78(2), 435–439.
- BHATTACHARYA, S., AND D. GALE (1987): “Preference shocks, liquidity, and Central Bank policy,” in *New approaches to monetary economics*, ed. by W. Barnett, and K. Singleton. Cambridge University Press, Cambridge, UK.
- BIANCHI, J., AND S. BIGIO (2014): “Banks, Liquidity Management and Monetary Policy,” Working Paper 20490, National Bureau of Economic Research.
- BILLIOA, M., M. GETMANSKYB, A. W. LO, AND L. PELIZZONA (2012): “Econometric measures of connectedness and systemic risk in the finance and insurance sectors,” *Journal of Financial Economics*, 104(3), 535–559.
- BONACICH, P. (1987): “Power and centrality: A family of measures,” *American Journal of Sociology*, 92, 1170–1182.
- BOSS, M., H. ELSINGER, M. SUMMER, AND S. THURNER (2004): “Network topology of the interbank market,” *Quantitative Finance*, 4(6), 677–684.
- BRAMOULL, Y., H. DJEBBARI, AND B. FORTIN (1993): “Identifications of Endogenous Social Effects: The Reflection Problem,” *Review of Economic Studies*, 60(1), 531–542.
- BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): “Identification of Peer Effects through Social Networks,” *Journal of Econometrics*, 150(1), 41–55.
- BRUNNERMEIER, M. K., AND Y. SANNIKOV (2014): “A Macroeconomic Model with a Financial Sector,” *American Economic Review*, 104(2), 379–421.
- (2016): “The I Theory of Money,” Working paper, Princeton University.
- CABRALES, A., D. GALE, AND P. GOTTARDI (2015): “Financial Contagion in Networks,” Discussion paper.

- CABRALES, A., P. GOTTARDI, AND F. VEGA-REDONDO (2016): “Risk-sharing and contagion in networks,” Working paper, European University Institute.
- CALVO-ARMENGOL, A., E. PATAACCHINI, AND Y. ZENOU (2009): “Peer effects and social networks in education,” *Review of Economic Studies*, 76, 1239–1267.
- CASTIGLIONESI, F., F. FERIOZZI, AND G. LORENZONI (2017): “Financial Integration and Liquidity Crises,” Working Paper 23359, National Bureau of Economic Research.
- CHEN, H., T. WANG, AND D. D. YAO (2016): “Financial Network and Systemic Risk A Dynamic Model,” working paper.
- COHEN-COLE, E., E. PATAACCHINI, AND Y. ZENOU (2010): “Systemic Risk and Network Formation in the Interbank Market,” CAREFIN working paper.
- DEBREU, G., AND I. HERSTEIN (1953): “Nonnegative Square Matrices,” *Econometrica*, 21, 597–607.
- DEGROOT, M. (1974): “Reaching a Consensus,” *Journal of the American Statistical Association*, 69, 118–121.
- DEMARZO, P., D. VAYANOS, AND J. ZWIEBEL (2003): “Persuasion Bias, Social Influence, and Uni-Dimensional Opinions,” *Quarterly Journal of Economics*, 118, 909–968.
- DENT, A., AND W. DISON (2012): “The Bank of Englands Real-Time Gross Settlement infrastructure,” Bank of England Quarterly Bulletin 3, Bank of England.
- DIAMOND, D. W., AND P. H. DYBVIK (1983): “Bank Runs, Deposit Insurance, and Liquidity,” *Journal of Political Economy*, 91(3), 401–419.
- DIAMOND, D. W., AND A. K. KASHYAP (2016): “Liquidity Requirements, Liquidity Choice and Financial Stability,” Working Paper 22053, National Bureau of Economic Research.
- DIEBOLD, F., AND K. YILMAZ (2009): “Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets,” *Economic Journal*, 119(534), 158–171.
- DIEBOLD, F. X., AND K. YILMAZ (2014): “On the network topology of variance decompositions: Measuring the connectedness of financial firms,” *Journal of Econometrics*, 182(1), 119–134.
- DONALDSON, J. R., G. PIACENTINO, AND A. THAKOR (2016): “Warehouse banking,” Working paper, Washington University in St. Louis.
- DRECHSLER, I., A. SAVOV, AND P. SCHNABL (2014): “A Model of Monetary Policy and Risk Premia,” Working Paper 20141, National Bureau of Economic Research.
- DUARTE, F. M., AND T. M. EISENBACH (2013): “Fire-sale spillovers and systemic risk,” Discussion paper.
- EISENBERG, L., AND T. H. NOE (2001): “Systemic Risk in Financial Systems,” *Journal of Political Economy*, 109, 236–249.

- ELHORST, J. P. (2010a): “Applied Spatial Econometrics: Raising the Bar,” *Spatial Economic Analysis*, 5(1), 9–28.
- (2010b): “Spatial Panel Data Models,” in *Handbook of Applied Spatial Analysis*, ed. by M. M. Fischer, and A. Getis, pp. 377–407. Springer Berlin Heidelberg.
- ELLIOTT, M., B. GOLUB, AND M. O. JACKSON (2015): “Financial Networks and Contagion,” *American Economic Review*, 104(10), 3115–3153.
- ELSINGER, H., A. LEHAR, AND M. SUMMER (2006): “Risk Assessment for Banking Systems,” *Management Science*, 52(9), 1301–1314.
- FECHT, F., K. NYBORG, AND J. ROCHOLL (2010): “The price of liquidity: Bank characteristics and market condition,” University of Zurich working paper.
- FREIXAS, X., A. MARTIN, AND D. SKEIE (2011): “Bank Liquidity, Interbank Markets, and Monetary Policy,” *Review Financial Studies*, 44(8), 2656–2692.
- FREIXAS, X., B. PARIGI, AND J. ROCHET (2000): “Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank,” *Journal of Money, Credit and Banking*, 32(1), 611–638.
- FURFINE, C. (2000): “Interbank payments and the daily federal funds rate,” *Journal of Monetary Economics*, 46, 535–553.
- FURFINE, C. H. (2003): “Interbank Exposures: Quantifying the Risk of Contagion,” *Journal of Money, Credit and Banking*, 35(1), 111–128.
- GAI, P., A. HALDANE, AND S. KAPADIA (2011): “Complexity, concentration and contagion,” *Journal of Monetary Economics*, 58(5), 453–470.
- GAI, P., AND S. KAPADIA (2010): “Contagion in financial networks,” *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering Science*, 466, 2401–2423.
- GERTLER, M., AND N. KIYOTAKI (2010): “Chapter 11 - Financial Intermediation and Credit Policy in Business Cycle Analysis,” in *Handbook of Monetary Economics*, ed. by B. M. Friedman, and M. Woodford, vol. 3 of *Handbook of Monetary Economics*, pp. 547 – 599. Elsevier.
- GLASSERMAN, P., AND H. YOUNG (2015): “How likely is contagion in financial networks?,” *Journal of Banking & Finance*, 50, 383–399.
- GOFMAN, M. (2017): “Efficiency and stability of a financial architecture with too-interconnected-to-fail institution,” *Journal of Financial Economics*, forthcoming.
- GORODENTSEV, A. L. (1994): “Helix theory and nonsymmetrical bilinear forms,” in *Algebraic Geometry and its Applications*, pp. 47–59. Vieweg+ Teubner Verlag.
- GREENWOOD, R., A. LANDIER, AND D. THESMAR (2015): “Vulnerable banks,” *Journal of Financial Economics*, 115, 471–485.
- GURLEY, J. G., AND E. S. SHAW (1960): *Money in a Theory of Finance*. Brookings Institution, Washington, D.C.

- HART, O., AND L. ZINGALES (2014): “Banks Are Where The Liquidity Is,” Working paper, Harvard University.
- HAUTSCH, N., J. SCHAUMBURG, AND M. SCHIENLE (2012): “Financial Network Systemic Risk Contributions,” working paper.
- HAYASHI, F. (2000): *Econometrics*. Princeton University Press: Princeton.
- HE, Z., AND A. KRISHNAMURTHY (2013): “Intermediary Asset Pricing,” *American Economic Review*, 103(2), 732–770.
- HENDERSON, H. V., AND S. R. SEARLE (1981): “On Deriving the Inverse of a Sum of Matrices,” *SIAM Review*, 23(1), pp. 53–60.
- JACKSON, M. O. (2003): *Social and Economic Networks*. Princeton University Press, New Jersey.
- JACKSON, M. O., AND Y. ZENOU (2012): “Games on Networks,” CEPR Discussion Papers 9127, C.E.P.R. Discussion Papers.
- KASHYAP, A. K., AND J. C. STEIN (2000): “What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?,” *American Economic Review*, 90(3), 407–428.
- KATZ, L. (1953): “A new status index derived from sociometric analysis,” *Psychometrika*, 18(1), 39–43.
- KELLY, B., H. LUSTIG, AND S. V. NIEUWERBURGH (2013a): “Firm volatility in granular networks,” NYU working paper.
- KELLY, B. T., H. N. LUSTIG, AND S. V. NIEUWERBURGH (2013b): “Firm Volatility in Granular Networks,” Fama-Miller Working Paper; Chicago Booth Research Paper No. 12-56.
- KLIMENKO, N., S. PFEIL, J.-C. ROCHET, AND G. D. NICOLO (2016): “Aggregate Bank Capital and Credit Dynamics,” Swiss Finance Institute Research Paper Series 16-42.
- KOVNER, A., AND D. SKEIE (2013): “Evaluating the quality of fed funds lending estimates produced from Fedwire payments data,” Staff Reports 629, Federal Reserve Bank of New York.
- LAGOS, R. (2008): “inside and outside money,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.
- LEE, L.-F., AND J. YU (2010): “A Spatial Dynamic Panel Data Model With Both Time And Individual Fixed Effects,” *Econometric Theory*, 26(02), 564–597.
- LEITNER, Y. (2005): “Financial networks: Contagion, commitment, and private sector bailouts,” *The Journal of Finance*, 60, 2925–2953.
- LESAGE, J., AND R. PACE (2009): *Introduction to spatial econometrics*, Statistics, textbooks and monographs. CRC Press.

- LI, D., AND N. SHÜRHOFF (2012): “Dealer Networks,” working paper.
- LI, Y. (2017): “Procyclical Finance: The Money View,” Working paper, Columbia University Graduate School of Business.
- MAGGIO, M. D., A. KERMANI, AND C. PALMER (2015): “Unconventional Monetary Policy and the Allocation of Credit,” .
- MALKIEL, B. (2003): *A random walk down Wall Street*. Noton and Company, New York.
- MILLER, K. S. (1981): “On the Inverse of the Sum of Matrices,” *Mathematics Magazine*, 54(2), pp. 67–72.
- MOORE, J. (2012): “Leverage stacks and the financial system,” Edinburgh University and London School of Economics working paper.
- NAKI ALDASORO, I., AND I. ANGELONI (2013): “Input-output-based measures of systemic influence,” Goethe University of Frankfurt working paper.
- NEWMAN, M. E. J. (2004): “Analysis of weighted networks,” *Physical Review*, 70, 1–9.
- OZDAGLI, A. K., AND M. WEBER (2015): “Monetary Policy Through Production Networks: Evidence from the Stock Market,” working paper.
- PIAZZESI, M., AND M. SCHNEIDER (2017): “Payments, Credit and Asset Prices,” Working paper, Stanford University.
- QUADRINI, V. (2014): “Bank Liabilities Channel,” CEPR Discussion Papers 10265.
- (2017): “Bank liabilities channel,” *Journal of Monetary Economics*, 89, 25 – 44, Carnegie-Rochester-NYU Conference Series on the Macroeconomics of Liquidity in Capital Markets and the Corporate Sector.
- SHERMAN, J., AND W. J. MORRISON (1950): “Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix,” *The Annals of Mathematical Statistics*, 21(1), pp. 124–127.
- SORAMAKI, K., M. L. BECH, J. ARNOLD, R. J. GLASS, AND W. E. BEYELER (2007): “The topology of interbank payment flows,” *Physica A*, 379(1), 317–333.
- STEIN, J. C. (2012): “Monetary Policy as Financial-Stability Regulation,” *Quarterly Journal of Economics*, 127(1), 57–95, forthcoming in the Quarterly Journal of Economics Draft Date: Revised May 2011.
- UPPER, C. (2011): “Simulation methods to assess the danger of contagion in interbank markets,” *Journal of Financial Stability*, 7(3), 111–125.
- UPPER, C., AND A. WORMS (2004): “Estimating bilateral exposures in the German interbank market: Is there a danger of contagion?,” *European Economic Review*, 48(4), 827 – 849.
- WATTS, D. J., AND S. H. STROGATZ (1998): “Collective dynamics of ‘small-world’ networks,” *Nature*, 393, 440–442.

WESTWOOD, B. (2011): “The Money Market Liaison Group Sterling Money Market Survey,” *Bank of England Quarterly Bulletin*, 51(3), 247–250.

WETHERILT, A., P. ZIMMERMAN, AND K. SORAMAKI (2010): “The sterling unsecured loan market during 2006-08: insights from network theory,” Bank of England working papers 398, Bank of England.

ZAWADOWSKI, A. (2012): “Entangled financial systems,” Boston University working paper.

A Appendix: Background, Alternative Model, Estimation Method, More Results

A.1 Reserves Schemes, Payment Systems, and Interbank Credit

Banks in the UK choose the amount of central bank reserves that they hold to support a range of short term liquidity needs. Reserves are the ultimate settlement asset for interbank payments. Whenever payments are made between the accounts of customers at different commercial banks, they are ultimately settled by transferring central bank money (reserves) between the reserves accounts of those banks. Reserve balances are used to buffer against intraday payment imbalances (i.e., cumulative outflows larger than inflows). Additionally, central bank reserves are the most liquid asset that banks can draw upon in the presence of unexpected outflows of funds. Since 2006, the starting year of our sample, banks choose their reserve holdings on a discretionary basis, i.e., reserve holdings are not mandatory. However, their reserve holding decisions depend on the policy framework in which they operate.

A.1.1 Monetary Policy Framework

Since the 1998 Banking Act, the Bank of England (BoE) has had independent responsibility for setting interest rates to ensure that inflation, as measured by the Consumer Price Index (CPI), meets the inflation target of 2%. Each month the Monetary Policy Committee (MPC) meets to decide the appropriate level of the Bank rate (the policy interest rate) to meet the inflation target in the medium term. The Sterling Monetary Framework changed over time. During our sample period, the Bank of England had three distinct monetary frameworks: prior to 18 May 2006, the Bank of England operated an unremunerated reserve scheme; this was then replaced by a reserves average scheme; since March 2009 and the initiation of Quantitative Easing, the reserves average scheme has been suspended.

Pre-2006 Reform: Prior to the 2006 reforms, the Sterling Monetary Framework (SMF) was based upon a voluntary unremunerated reserves. There were no reserve requirements

and no reserve averaging over a maintenance period. The only requirement was that banks were obliged to maintain a minimum zero balance at the end of each day. In practice, due to uncertainties from end of day cash positions, banks opted for small positive reserve balances.

Reserve Averaging: In May 2006, the Bank of England undertook a major reform of the Sterling Monetary Framework. The new scheme was voluntary remunerated reserves with a period-average maintenance requirement. Each maintenance period – the period between two meetings of the Monetary Policy Committee – banks were required to decide upon a reserves target. This voluntary choice of reserves target is a feature unique to the UK system. Over the course of each maintenance period, the banks would manage their balance sheets so that, on average, their reserve balances hit the target. Where banks were unable to hit the target, standing borrowing and deposit facilities were available. Within a range of $\pm 1\%$ of the target, reserves are remunerated at the Bank Rate.⁴⁵ Holding an average level of reserves outside the target range attracts a penalty charge.⁴⁶ But an SMF participant can ensure it hits its target by making use of the Bank’s Operational Standing Facilities (OSFs). These bilateral facilities allow SMF participants to borrow overnight from the Bank (against high-quality collateral) at a rate above Bank Rate or to deposit reserves overnight with the Bank at a rate below Bank Rate. The possibility of arbitrage between interbank market rates and reserves remunerated at Bank Rate is the main mechanism through which market rates are kept in line with Bank Rate. In both schemes before Quantitative Easing (QE), the BoE would ensure sufficient reserves supply for banks to meet their reserves target. Banks then use the interbank market to reallocate reserves from banks in surplus to banks in deficit.

Post Quantitative Easing: Quantitative Easing in UK started in March 2009 when the MPC decided that in order to meet the inflation target in the medium term, it would need to supplement the use of interest rate (which had hit the practical lower bound of 0.5%) with

⁴⁵At various points during the crisis, this $\pm 1\%$ range was increased to give banks more flexibility to manage their liquidity.

⁴⁶Settlement banks also pay a penalty if their reserves account is overdrawn at the end of any day.

the purchase of assets using central bank reserves. This consisted of the BoE's boosting the money supply by creating central bank reserves and using them to purchase assets, predominantly UK gilts. Furthermore, the BoE suspended the average reserve targeting regime, and now remunerates all reserves at the Bank rate.

A.1.2 Payment and Settlement Systems

Banks use central bank reserves to, inter alia, meet their demand for intraday liquidity in the payment and settlement systems. Reserves act as a buffer to cover regular timing mismatches between incoming and outgoing payments, for example, due to exceptionally large payments, operational difficulties, or stresses that impact upon a counterparty's ability, or willingness to send payments. There are two major payment systems in the UK: CHAPS and CREST.⁴⁷ These two systems play a vital role in the UK financial system. On average, in 2011, £700 billion of transactions was settled every day within the two systems. This equates to the UK 2011 nominal GDP being settled every two days.

CHAPS is the UK's large-value payment system. It is used for real time settlement of payments between its member banks. These banks settle payments on behalf of hundreds of other banks through correspondent relationships. Typical payments are business-to-business payments, home purchases, and interbank transfers. Payments relating to unsecured interbank money markets are settled in CHAPS. CHAPS opens for settlement at 8 am and closes at 4:20 pm. Payments made on behalf of customers cannot be made after 4 pm. The system has throughput guidelines which require members to submit 50% of their payments by noon and 75% by 14:30. This helps ensure that payments are settled throughout the day and do not bunch towards the end of the day.

In 2011, CHAPS settled an average of 135,550 payments each day valuing £254bn. CHAPS is a real-time gross settlement (RTGS) system. This means that payments are settled finally and irrevocably in real time. To fund these payments, banks have to access liquidity intraday. If a bank has, at any point during the day, cumulatively sent more

⁴⁷There are also four retail payment systems (Bacs, the Faster Payments Service (FPS), Cheque and Credit Clearing (CCC) and LINK) that are operated through the BoE.

payments than it has received, then it needs liquidity to cover this difference. This comes either from central bank reserves or intraday borrowing from the BoE. Furthermore, when a bank sends funds to another bank in the system, it exposes itself to liquidity risk. That is, the risk that the bank may not get those fund inflows back during the day, and so will run down their own liquidity holdings or borrow from the BoE. Therefore, it is important to choose an appropriate level of liquidity buffer. Besides maintaining a liquidity buffer, banks manage liquidity by borrowing from and lending to each other in the unsecured overnight markets. According to Bank of England estimates, payments relating to overnight market activity (advances and repayments) account for about 20% of CHAPS values (Wetherilt, Zimmerman, and Soramaki (2010)).

CREST is a securities settlement system. Its Delivery-vs-Payment (DVP) mechanism ensures simultaneous transfer of funds and securities. The CREST system's intraday liquidity mechanism with the BoE is automatic through the "Self Collateralising Repos" (SCRs), once a liquidity need is identified. If a CREST settlement bank would otherwise have insufficient funds to settle a CREST transaction, a secured intraday loan is automatically generated using as eligible collateral either the purchased security (if eligible) or other securities.

A.1.3 The Sterling Unsecured Overnight Interbank Market

Interbank markets are the markets where banks and other financial institutions borrow and lend assets, typically with maturities of less than one year. At the shortest maturity, overnight, banks borrow and lend central bank reserves. Monetary policy aims at influencing the rate at which these markets transact, so as to control inflation in the wider economy. There is limited information available about the size and the structure of the sterling money markets. The Bank of England estimates suggest that the overnight unsecured market is approximately £20–30 billion per day during our sample period. Wetherilt, Zimmerman, and Soramaki (2010) describe the network of the sterling unsecured overnight money market. They find that the network has a small core of highly connected participants, surrounded by a wider periphery of banks loosely connected with each other, but

with connections to the core. It is believed that prior to the recent financial crisis, the unsecured market was much larger than the secured one. We identify interbank borrowing and lending transactions in CHAPS settlement data.

A.2 An Alternative Model

In this section, we present an alternative model where the network effect on banks' liquidity holding decisions is not modelled as a residual. Specifically, we let the total liquidity holding by bank i , i.e., $l_{i,t}$, to be accessible to the network. Hence, the valuation of liquidity for bank i in network g_t becomes:

$$\underbrace{\tilde{\mu}_{i,t}}_{\text{Per Unit Value}} \underbrace{\left(l_{i,t} + \psi \sum_{j \neq i} g_{ij,t} l_{j,t} \right)}_{\text{Accessible Liquidity}} \quad (31)$$

and the per unit value $\tilde{\mu}_{i,t}$ is specified as

$$\tilde{\mu}_{i,t} := \hat{\mu}_{i,t} + \delta \sum_j g_{ij,t} l_{j,t} + \sum_{m=1}^M \tilde{\beta}_m x_{i,t}^m + \sum_j g_{ij,t} x_{i,j,t} \tilde{\theta} + \sum_{p=1}^P \tilde{\gamma}_p x_t^p \quad (32)$$

where $x_{i,j,t}$ denotes match specific control variables and the characteristics of other banks, and $\tilde{\theta}$ is a vector of suitable dimension. That is, in addition to the aggregate information embedded in the neighbouring banks' holdings, also macro variables and the neighbouring banks' characteristics affect the per unit valuation of the liquidity.

In this setup, bank i 's utility from holding liquidity is specified as:

$$u_i(\mathbf{1}_t | g_t) = \left(\hat{\mu}_{i,t} + \delta \sum_j g_{ij,t} l_{j,t} + \sum_{m=1}^M \tilde{\beta}_m x_{i,t}^m + \sum_j g_{ij,t} x_{i,j,t} \tilde{\theta} + \sum_{p=1}^P \tilde{\gamma}_p x_t^p \right) \left(l_{i,t} + \psi \sum_{j \neq i} g_{ij,t} l_{j,t} \right) - \frac{1}{2} \gamma \left(l_{i,t} + \psi \sum_{j \neq i} g_{ij,t} l_{j,t} \right)^2. \quad (33)$$

The optimal response function for each bank is then:

$$\begin{aligned}
l_{i,t}^* &= \frac{\hat{\mu}_{i,t} + \sum_{m=1}^M \tilde{\beta}_m x_{i,t}^m + \sum_j g_{ij,t} x_{i,j,t} \tilde{\theta} + \sum_{p=1}^P \tilde{\gamma}_p x_t^p}{\gamma} + \left(\frac{\delta}{\gamma} - \psi \right) \sum_{j \neq i} g_{ij,t} l_{j,t} \\
&= \mu_{i,t} + \sum_{m=1}^M \beta_m x_{i,t}^m + \sum_j g_{ij,t} x_{i,j,t} \theta + \sum_{p=1}^P \gamma_p x_t^p + \phi \sum_j g_{ij,t} l_{j \neq i,t}
\end{aligned} \tag{34}$$

where $\phi := \delta/\gamma - \psi$, $\mu_{i,t} := \hat{\mu}_{i,t}/\gamma =: \bar{\mu}_i + \nu_{i,t}$, $\beta_m = \tilde{\beta}_m/\gamma$, $\gamma_p = \tilde{\gamma}_p/\gamma$, and $\theta = \tilde{\theta}/\gamma$. Note that the empirical counterpart of the above best response is the spatial Durbin model in equation (27).

Let us denote $\mu_{i,t} + \sum_{m=1}^M \beta_m x_{i,t}^m + \sum_j g_{ij,t} x_{i,j,t} \theta + \sum_{p=1}^P \gamma_p x_t^p$ by $\check{\mu}_t$. The following result is immediate following similar steps of the proof in the main text.

Proposition 3 *Suppose that $|\phi| < 1$. Then, there is a unique interior solution for the individual equilibrium outcome given by*

$$l_{i,t}^*(\phi, g) = \{\mathbf{M}(\phi, \mathbf{G}_t)\}_i \check{\mu}_t, \tag{35}$$

where $\{\}_i$ is the operator that returns the i -th row of its argument, $\check{\mu}_t := [\check{\mu}_{1,t}, \dots, \check{\mu}_{n,t}]'$, $l_{i,t}$ denotes the total liquidity holding by bank i .

The above result implies that, even in this more general model, the definitions of conditional volatility of liquidity (equation (10)), risk key player (definition 1), level key player⁴⁸ (definition 2), and network impulse response functions (definition 3), as well as their dependency on the network topology and equilibrium parameter ϕ , stay unchanged.

⁴⁸In this case, l^* should replace z^* in equation (13).

A.3 Details of the Empirical Methodology

A.3.1 Quasi-Maximum Likelihood Formulation and Identification Issues

Writing the variables and coefficients of the spatial error model in equations (24) and (25) in matrix form as⁴⁹

$$B := [\alpha_1^{time}, \dots, \alpha_t^{time}, \dots, \alpha_T^{time}, \alpha_1^{bank}, \dots, \alpha_i^{bank}, \dots, \alpha_N^{bank}, \\ \beta_1^{bank}, \dots, \beta_m^{bank}, \dots, \beta_M^{bank}, \beta_1^{time}, \dots, \beta_p^{time}, \dots, \beta_P^{time}]',$$

$$L := [l_{1,1}, \dots, l_{N,1}, \dots, l_{i,t}, \dots, l_{1,T}, \dots, l_{N,T}], \quad \mathbf{z} := [z_{1,1}, \dots, z_{N,1}, \dots, z_{i,t}, \dots, z_{1,T}, \dots, z_{N,T}]'$$

$$\nu := [\nu_{1,1}, \dots, \nu_{N,1}, \dots, \nu_{i,t}, \dots, \nu_{1,T}, \dots, \nu_{N,T}]', \quad \underline{\mu} := \mathbf{1}_T \otimes [\bar{\mu}_1, \dots, \bar{\mu}_N]'$$

$$\mathbf{G} := \text{diag}(\mathbf{G}_t)_{t=1}^T = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 & \dots & \dots \\ \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{G}_T \end{bmatrix}, \quad \mathbf{X} := [\mathbf{D}, \mathbf{F}, \mathbf{X}^{bank}, \mathbf{X}^{time}],$$

where $\mathbf{D} := I_T \otimes \mathbf{1}_N$, $\mathbf{F} := \mathbf{1}_T \otimes I_N$, and

$$\mathbf{X}^{time} = \begin{bmatrix} x_1^1 & \dots & x_1^p & \dots & x_1^P \\ \dots & \dots & \dots & \dots & \dots \\ x_t^1 & \dots & x_t^p & \dots & x_t^P \\ \dots & \dots & \dots & \dots & \dots \\ x_T^1 & \dots & x_T^p & \dots & x_T^P \end{bmatrix} \otimes \mathbf{1}_N, \quad \mathbf{X}^{bank} = \begin{bmatrix} x_{1,1}^1 & \dots & x_{1,1}^m & \dots & x_{1,1}^M \\ \dots & \dots & \dots & \dots & \dots \\ x_{N,1}^1 & \dots & x_{N,1}^m & \dots & x_{N,1}^M \\ \dots & \dots & \dots & \dots & \dots \\ x_{N,T}^1 & \dots & x_{N,T}^m & \dots & x_{N,T}^M \end{bmatrix},$$

we can then rewrite the empirical model as

$$L = \mathbf{X}B + \mathbf{z}, \quad \mathbf{z} = \underline{\mu} + \phi \mathbf{G}\mathbf{z} + \nu, \quad \nu_{i,t} \sim iid(0, \sigma_i^2).$$

⁴⁹This is similar to the spatial formulation in Lee and Yu (2010).

This, in turn, implies that

$$\nu(B, \underline{\mu}, \phi) = (I_{N \times T} - \phi \mathbf{G})(L - \mathbf{X}B) - \underline{\mu}. \quad (36)$$

Finally, using the Gaussian distribution to model the exogenous error terms ν yields the log likelihood

$$\ln \mathcal{L}(B, \phi, \underline{\mu}, \{\sigma_i^2\}_{i=1}^N) \equiv -\frac{TN}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^N \ln \sigma_i^2 - \sum_{i=1}^N \frac{1}{2\sigma_i^2} \sum_{t=1}^T \nu_{i,t}(B, \underline{\mu}, \phi)^2, \quad (37)$$

and the above can be estimated using standard optimization methods.

In the above formulation, the identification of ϕ is ensured by the usual conditions on \mathbf{G} (see, e.g. Bramoullé, Djebbari, and Fortin (2009)). Instead, the separate identification of the bank fixed effects, $\alpha^{bank} := [\alpha_1^{bank}, \dots, \alpha_N^{bank}]'$, and the network-bank fixed effects, $\bar{\mu} := [\bar{\mu}_1, \dots, \bar{\mu}_N]'$, deserve some further remarks. Isolating the role of these fixed effects, equation (36) can be rewritten as

$$\begin{aligned} \nu(B, \underline{\mu}, \phi) &= (I_{N \times T} - \phi \mathbf{G}) \left(L - \tilde{\mathbf{X}} \tilde{B} - \mathbf{F} \alpha^{bank} \right) - \underline{\mu} \\ &= (I_{N \times T} - \phi \mathbf{G}) \left(L - \tilde{\mathbf{X}} \tilde{B} \right) - \mathbf{1}_T \otimes (\bar{\mu} + \alpha^{bank}) + \phi \mathbf{G} \mathbf{F} \alpha^{bank} \end{aligned}$$

where $\tilde{\mathbf{X}} := [\mathbf{D}, \mathbf{X}^{bank}, \mathbf{X}^{time}]$ and \tilde{B} is simply the vector B without the α^{bank} elements. Several observations are in order. First, the above implies that if $\phi = 0$, then $\bar{\mu}$ and α^{bank} cannot be separately identified (nevertheless the parameters \tilde{B} are still identified). Second, if there is no time variation in the network structure, i.e. if $\mathbf{G}_t = \bar{\mathbf{G}} \forall t$, $\bar{\mu}$ and α^{bank} cannot be separately identified even if $\phi \neq 0$. Third, if a bank never lends to any other bank in the sample, its fixed effects $\bar{\mu}_i$ and α_i^{bank} cannot be separately identified. Fourth, if \mathbf{G}_t is a right stochastic matrix, separate identification of $\bar{\mu}$ and α^{bank} can be achieved only up to a parameter normalization, since for any scalar κ and vector $\bar{\kappa} := \mathbf{1}_N \otimes \kappa$, we have

$$\nu(B, \underline{\mu}, \phi) = (I_{N \times T} - \phi \mathbf{G}) \left(L - \tilde{\mathbf{X}} \tilde{B} \right) - \mathbf{1}_T \otimes (\bar{\mu} + \alpha^{bank} + \phi \bar{\kappa}) + \phi \mathbf{G} \mathbf{F} (\alpha^{bank} + \bar{\kappa})$$

The above also makes clear that a handy normalisation is to set one of the network-bank fixed effect (say the i -th one) to zero since it would imply the restriction $\{\alpha^{bank} + \phi\bar{\kappa}\}_i = \{\alpha^{bank} + \bar{\kappa}\}_i$ that, for any $\phi \neq 0$ and 1, can only be satisfied with $\kappa = 0$. Under this normalisation, the remaining estimated bank-network fixed effects are then in deviation from the normalised one. Fifth, note that the lack of separate identification for $\bar{\mu}$ and α^{bank} is due to the fact that when \mathbf{G}_t is a right stochastic matrix, and if all banks borrow from at least one bank at each point in time (i.e. \mathbf{G}_t has no rows of zeros), then $\mathbf{G}_t\mathbf{1}_N = \mathbf{1}_N$ and $\mathbf{G}\mathbf{1}_{N \times T} = \mathbf{1}_{N \times T}$. Fortunately, in our dataset, the condition $\mathbf{G}_t\mathbf{1}_N = \mathbf{1}_N$ does not hold every day in the sample because there are periods in which certain banks do not borrow (in this case, the corresponding rows of \mathbf{G}_t contain all zeros and sum to zero, instead of one). In our sample, except for bank 7 and bank 11, all the other banks borrow every period from at least one of their counterparties. There are fourteen days when bank 7 does not borrow at all, and 145 days in which bank 11 does not borrow at all. Moreover, the no borrowing days of bank 7 and bank 11 do not overlap, so we have a total of 159 days in which either the sum of the 7th row of \mathbf{G}_t or the sum of the 11th row of \mathbf{G}_t is equal to zero, not one (13.5% of the days).

A.3.2 Confidence Bands for the Network Impulse Response Functions

The ϕ estimator outlined in the previous section has an asymptotic Gaussian distribution with variance s_ϕ^2 (that can be readily estimated from the QMLE covariance matrix based, as usual, on the Hessian and gradient of the log likelihood in equation (37)). That is, $\sqrt{T}(\hat{\phi} - \phi_0) \xrightarrow{d} N(0, s_\phi^2)$, where ϕ_0 denotes the true value of ϕ . Writing

$$a_1(\phi) := \frac{\partial \mathbf{1}' \{(\mathbf{I} - \phi \mathbf{G})^{-1}\}_{.i}}{\partial \phi}, \quad a_2(\phi) = \frac{\partial \mathbf{1}' \{(\mathbf{I} - \phi \mathbf{G})^{-1} \phi \mathbf{G}\}_{.i}}{\partial \phi}$$

we have from Lemma 2.5 of Hayashi (2000) that

$$\begin{aligned}\sqrt{T} \left[NIRF_i \left(\hat{\phi}, 1 \right) - NIRF_i \left(\phi_0, 1 \right) \right] &\xrightarrow{d} N \left(0, a_1 \left(\phi_0 \right)^2 s_\phi^2 \right), \\ \sqrt{T} \left[NIRF_i^e \left(\hat{\phi}, 1 \right) - NIRF_i^e \left(\phi_0, 1 \right) \right] &\xrightarrow{d} N \left(0, a_2 \left(\phi_0 \right)^2 s_\phi^2 \right).\end{aligned}$$

Therefore, since $a_j \left(\hat{\phi} \right) \xrightarrow{p} a_j \left(\phi_0 \right)$, $j = 1, 2$, by the continuous mapping theorem, and by Slutsky's theorem, $a_j \left(\hat{\phi} \right) \hat{s}_\phi^2 \xrightarrow{p} a_j \left(\phi_0 \right)^2 s_\phi^2$, where \hat{s}_ϕ^2 is a consistent variance estimator, we can construct confidence bands for the network impulse response functions using the sample estimates of ϕ and s_ϕ^2 .

A.3.3 Details of the Construction of the Variables

Macro control variables

- rK_{t-1} : lagged right kurtosis of the intraday time of aggregate payment outflow:

$$rK_t = \frac{\sum_{\tau > m_t} \left(\frac{\tau - m_t}{\sigma_t} \right)^4}{\sum_{\tau=1}^{88} \left(\frac{\tau - m_t}{\sigma_t} \right)^4}$$

where

$$m_t = \frac{1}{88} \sum_{\tau=1}^{88} \tau \left(\frac{P_{t,\tau}^{OUT}}{\sum_{\tau=1}^{88} P_{t,\tau}^{OUT}} \right), \quad \sigma_t^2 = \frac{1}{88-1} \sum_{\tau=1}^{88} (\tau - m_t)^2 \left(\frac{P_{t,\tau}^{OUT}}{\sum_{\tau=1}^{88} P_{t,\tau}^{OUT}} \right)$$

and $P_{t,\tau}^{OUT}$ is the aggregate payment outflow at time interval τ . Note that transactions are recorded for 88 10-minute time intervals within each day (from 5:00 to 19:30). The variable m_t is the average of payment time weighted by the payment outflow.

- $\ln VolPay_{t-1}$: intraday volatility of aggregate liquidity available (lagged and in logarithms). ‘‘Liquidity available’’ is defined in the construction of bank specific control variables below.
- TOR_{t-1} : lagged turnover rate in the payment system. To define the turnover rate,

we need first to define the Cumulative Net (Debit) Position (CNP):

$$CNP(T, i, s) = \sum_{t=1}^T (P_{i,s,t}^{OUT} - P_{i,s,t}^{IN}),$$

where $P_{i,s,t}^{OUT}$ is bank i 's total payment outflow at time t in day s . $P_{i,s,t}^{IN}$ is the payment inflow. The turnover rate (in day s) is defined as

$$TOR_s = \frac{\sum_{i=1}^N \sum_{t=1}^{88} P_{i,s,t}^{OUT}}{\sum_{i=1}^N \max\{\max_T[CNP(T; i, s)], 0\}}$$

The numerator is the total payment made in the system at day s . The denominator sums the maximum cumulative net debt position of each bank at day s . Note that in the denominator, if the cumulative net position of a certain bank is always below zero (that is, this bank's cumulative inflow always exceeds its cumulative outflow), this bank actually absorbs liquidity from the system. If there are banks absorbing liquidity from the system, there must be banks injecting liquidity into the system. When we calculate the turnover rate (the ratio between the total amount circulating and the base), we should only consider one of the two. That's why we take the first (outside) maximum operator. The reason for the inside operator goes as follows: Any increase in the cumulative net debit position (wherever positive) incurs an injection of liquidity into the system, so the maximum of the cumulative net position is the total injection from the outside to the payment system. And, the sum over the different banks gives the total injection through all the membership banks. A higher turnover rate means a more frequent reuse of the money injected from outside into the payment system.

- *LIBOR*: lagged LIBOR rate.
- *Interbank Rate Premium*: lagged average interbank market rate minus lagged LIBOR.

Bank-specific variables

- Liquidity Available (LA) is the amount of liquidity to meet payment requirements and is measured as the sum of reserves ($SDAB$, Start of Day Account Balance) plus the value of intraday repo available with the BoE (PC , Posting of Collateral). As time passes, the liquidity available in CHAPS is calculated by subtracting the money moved to CREST from the liquidity available in the previous time interval. In this way, we can trace for bank i the liquidity available at any time t in day s :

$$LA(t, i, s) = SDAB_{i,s} + PC_{i,s} - \sum_{\tau=1}^t CREST_{i,s,\tau}$$

- Liquidity holdings at the beginning at the day ($l_{i,t}$, i.e., the dependent variable): the logarithm of reserve balances plus posting of collateral (the value of intraday repo available with the BoE) at the start of the day.
- *Interbank Rate*: lagged interbank rate.
- $\ln LevPay_{i,t-1}$: total intraday payment level (lagged, in logarithms).
- $rK_{i,t-1}^{in}$: lagged right kurtosis of incoming payment time.
- $rK_{i,t-1}^{out}$: lagged right kurtosis of outgoing payment time.
- $\ln VolPay_{i,t-1}$: intraday volatility of liquidity available (lagged, in logarithms).
- $\ln LU_{i,t-1}$: liquidity used (lagged, in logarithms) defined as follows

$$LU(i, s) = \max\{\max_T[CNP(T; i, s)], 0\}.$$

A positive cumulative net debit position means that in this time interval the bank is consuming its own liquidity. If a positive cumulative net position never happens for a bank, this bank only absorbs liquidity from the system. That is the reason for the first (outside) maximum operator. The second (inside) maximum operator helps us to trace the highest amount of liquidity a bank uses.

- $\frac{\text{repo Liability}}{\text{Assets}}$: repo liability to total asset ratio (lagged, monthly).
- *Total Assets (log)*: total asset (lagged, monthly, in logarithms).
- $\frac{\Delta \text{Deposit}}{\text{Assets}}$: cumulative change in ratio of retail deposits to total assets $\times 100$ (lagged, monthly).
- *Total Lending and Borrowing (log)*: total lending and borrowing in the interbank market (lagged, in logarithms).
- *CDS (log)*: CDS price (lagged, in logarithms).
- *Stock Return (Inc. Dividend)*: stock return including dividends (lagged).

A.4 Additional Figures and Tables

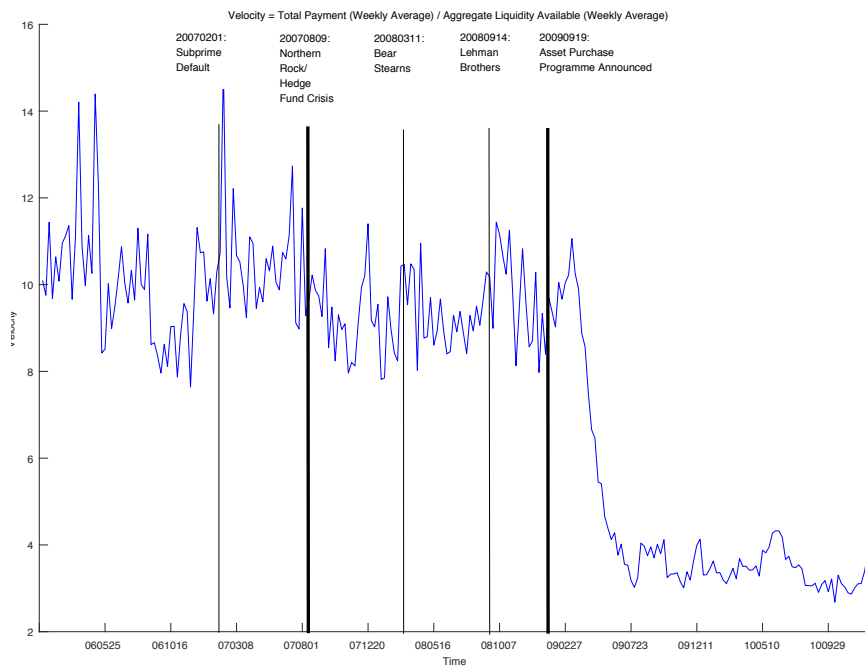


Figure 11: Velocity of money in the payment system.

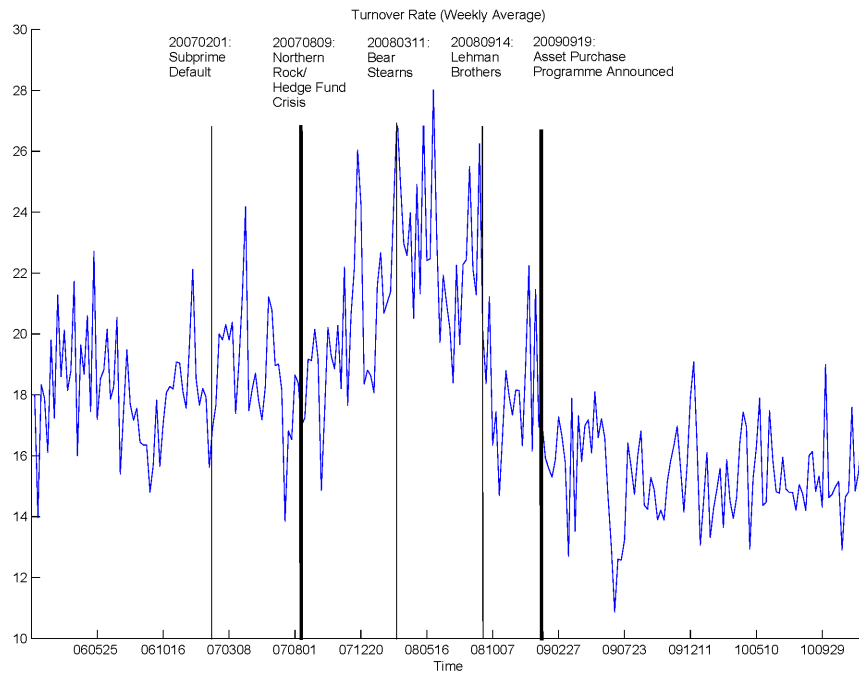


Figure 12: Turnover rate in the payment system.

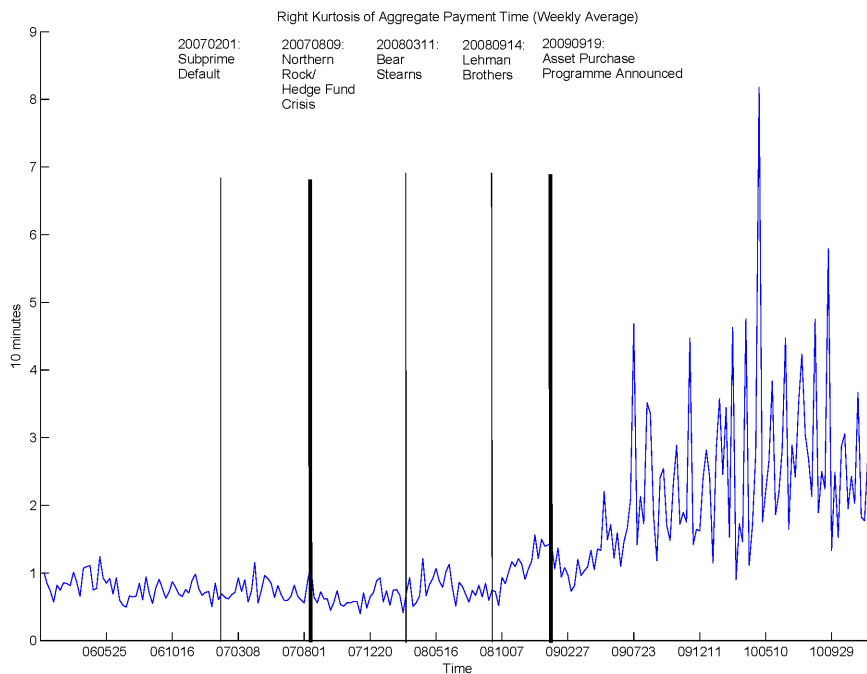


Figure 13: Weekly average of the right kurtosis of aggregate payment times.

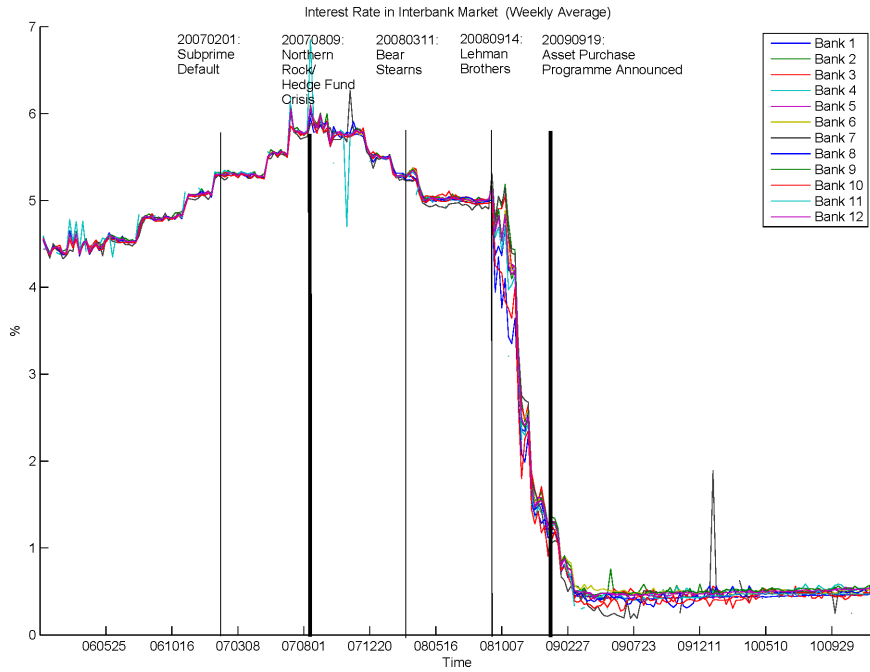


Figure 14: Interest rates in the interbank market.

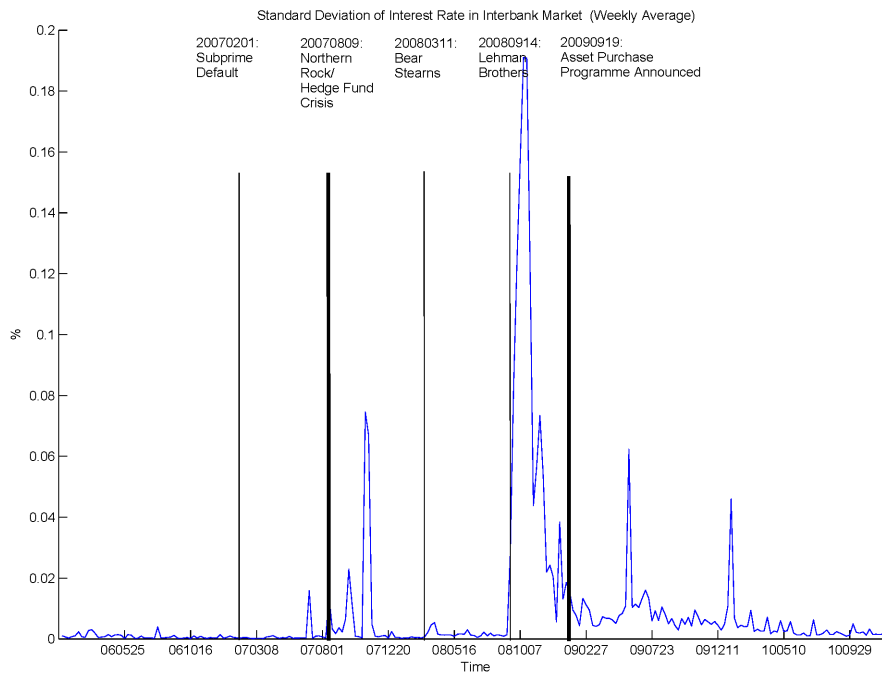


Figure 15: Cross-sectional dispersion of interbank rates.

Table A1: Full Spatial Error Model Estimation

	<i>Period 1</i>	<i>Period 2</i>	<i>Period 3</i>
$\hat{\phi}$	0.8137* (21.47)	0.3031* (1.90)	-0.1794* (-4.96)
$1/(1 - \hat{\phi})$	5.3677* (4.92)	1.4349* (4.37)	0.8479* (32.61)
<i>Macro Control Variables</i>			
rK_{t-1}	0.1845 (1.30)	0.0084 (0.55)	-0.0032* (-3.88)
$\ln VolPay_{t-1}$	-0.4451 (-1.00)	0.0308 (1.17)	0.0291 (1.72)
TOR_{t-1}	0.0166 (1.80)	0.0007 (0.69)	0.0018 (1.75)
LIBOR	0.2378 (0.27)	0.0928 (1.28)	0.5800* (2.52)
Interbank Rate Premium	3.8845 (1.61)	-0.0405 (-0.33)	0.6973* (3.00)
<i>Bank Characteristics/Mircro Control Variables</i>			
Interbank Rate	-0.2081 (-0.98)	-0.0473 (-1.03)	-0.0880 (-1.92)
$\ln LevPay_{i,t-1}$	-0.0235 (-0.62)	0.0802* (3.29)	0.0808* (5.09)
$rK_{i,t-1}^{in}$	0.0010 (0.14)	-0.0086 (-0.63)	0.0045 (1.03)
$rK_{i,t-1}^{out}$	0.0090 (0.92)	0.0320* (3.62)	-0.0061 (-1.32)
$\ln VolPay_{i,t-1}$	0.0129* (4.59)	0.0039 (1.92)	0.0196* (5.96)
$\ln LU_{i,t-1}$	-0.0038* (-2.86)	-0.0039* (-3.41)	-0.0027* (-3.79)
$\frac{Repo\ Liability}{Assets}$	-5.5625* (-3.61)	0.0282 (0.43)	-0.3057 (-1.45)
Total Assets (log)	1.2590* (5.39)	0.6328* (10.31)	1.0170* (18.92)
$\frac{\Delta Deposit}{Assets}$	-0.0014 (-0.20)	0.0149* (5.15)	0.0481* (11.76)
Total Lending and Borrowing (log)	-0.1882* (-5.57)	0.0612* (2.95)	-0.0025 (-1.27)
CDS (log)	0.0051 (0.13)	-0.1212* (-6.61)	-0.0383* (-4.00)
Stock Return (Inc. Dividend)	-0.5667 (-0.88)	0.1927 (1.49)	0.2574 (1.88)
R^2	66.01%	92.09%	91.53%

Estimation results of equations (24) and (25). Periods 1, 2 and 3, correspond, respectively, to before the Northern Rock/Hedge Fund Crisis, after Hedge Fund Crisis but before the Asset Purchase Programme, and after the Asset Purchase Programme announcement. The t -statistics are reported in parentheses under the estimated coefficients, where * denotes statistically significant estimates at a 10% or higher confidence level. Standard errors are QMLE robust ones. For the average network multiplier, $1/(1 - \hat{\phi})$, the delta method is employed.

Table A2: ratio of network to idiosyncratic volatility.

	<i>Period 1</i>	<i>Period 2</i>	<i>Period 3</i>
Bank 1	2.54	1.05	0.98
Bank 2	2.10	1.04	1.02
Bank 3	1.83	1.06	0.87
Bank 4	2.62	1.06	1.08
Bank 5	2.41	1.10	1.02
Bank 6	1.65	1.09	1.03
Bank 7	1.47	0.97	1.13
Bank 8	1.69	1.09	1.03
Bank 9	2.12	1.09	1.03
Bank 10	1.62	0.99	1.09
Bank 11	2.04	1.14	1.31
Mean	2.01	1.06	1.05

The table reports $\sqrt{\frac{\text{Var}(\hat{z}_{i,t})}{\text{Var}(\hat{v}_{i,t})}}$ for each bank and each period considered.