*University of Southern California Law
School*

Legal Studies Working Paper Series

Year 2015

Paper 158

The Bank-Centered View of the Money
Market, Part II: Re-evaluating the Real Bills
Approach to Macro-prudential Regulation

Carolyn Sissoko*

*USC Center for Law and Social Science, csissoko5@gmail.com

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://law.bepress.com/usclwps-lss/158>

Copyright ©2015 by the author.

The Bank-Centered View of the Money Market, Part II: Re-evaluating the Real Bills Approach to Macro-prudential Regulation

Carolyn Sissoko

Abstract

This paper argues that an early 20th century central banker, had he been alive at the turn of the 21st century, would have predicted the 2007-08 crisis and its severity. This paper is Part II of a series which contends that early 20th century banking theory is a valuable framework for understanding the relationship between banks, financial markets, and the central bank. This paper builds on Part I which explained that because a social norm supports the circulation of bank deposits as money, deposits are a network good and this makes possible monetary finance, or the expansion of the money supply as a source of funds for banks to lend.

This paper first analyses the British model of monetary finance, the structured interaction of banks, the money market, and the central bank that formed its core, and the means by which this structure made possible the origination of safe, privately-issued assets. Then, the analysis turns to the macroeconomic implications of monetary finance: in order to protect the monetary social norm, both financial instability and inflation must be avoided. The real bills principle addressed the former by proscribing monetary finance of long-term assets and the latter by requiring careful monitoring and control of the growth of money market instruments that were not real bills.

Thus, the modern integration of money and capital markets is seen through traditional banking theory to be a recipe for financial instability, because it undermines the ability of banks and the money market to be joined together in the production

of safe, privately-issued assets. This theory indicates that restabilizing the financial system will require structural reform and that only after such reform has been implemented can we expect macro-prudential regulation to succeed.

The Bank-Centered View of the Money Market

Part II: Re-evaluating the Real Bills Approach to Macro-prudential Regulation

Carolyn Sissoko*

March 18, 2015

Abstract: This paper argues that an early 20th century central banker, had he been alive at the turn of the 21st century, would have predicted the 2007-08 crisis and its severity. This paper is Part II of a series which contends that early 20th century banking theory is a valuable framework for understanding the relationship between banks, financial markets, and the central bank. This paper builds on Part I which explained that because a social norm supports the circulation of bank deposits as money, deposits are a network good and this makes possible monetary finance, or the expansion of the money supply as a source of funds for banks to lend.

This paper first analyses the British model of monetary finance, the structured interaction of banks, the money market, and the central bank that formed its core, and the means by which this structure made possible the origination of safe, privately-issued assets. Then, the analysis turns to the macroeconomic implications of monetary finance: in order to protect the monetary social norm, both financial instability and inflation must be avoided. The real bills principle addressed the former by proscribing monetary finance of long-term assets and the latter by requiring careful monitoring and control of the growth of money market instruments that were not real bills.

Thus, the modern integration of money and capital markets is seen through traditional banking theory to be a recipe for financial instability, because it undermines the ability of banks and the money market to be joined together in the production of safe, privately-issued assets. This theory indicates that restabilizing the financial system will require structural reform and that only after such reform has been implemented can we expect macro-prudential regulation to succeed.

* I thank Michael Bordo and Hugh Rockoff for encouraging me to write this paper. I thank Hugh Rockoff for his comments. All errors are, of course, my own.

Author information: Fellow, USC Center in Law and Social Science
Author email: sissoko@usc.edu

The financial crisis of 2007-08 made obvious the instability at the heart of the modern financial system and has catalyzed new research on the financial system and financial regulation. Major themes of this new research are the shift to macro-prudential regulation on the part of central banks, emphasis on the importance of “safe” or informationally illiquid assets to the financial system, and proposals for radical reform ranging from 100% reserve banking to formalizing the central bank’s role as “dealer of last resort.”

This paper argues that traditional banking theory, the theory that accompanied the growth of modern banking and that views money as fundamentally a private-sector phenomenon, provides a valuable framework for evaluating both the causes of the recent crisis and theories of macro-prudential regulation, “safe” assets, and radical reform. Like narrow banking proposals, traditional banking theory indicates that structural reform of the current financial system is needed. Unlike narrow banking proposals, the theory indicates that the purpose of structural reform is to regenerate the banking system’s ability to originate safe, privately-issued assets and that macro-prudential regulation may be successful only after appropriate structural reforms are adopted.

This is the second of three papers that explain the basics of traditional banking theory, and here the focus is on the theory’s implications for the macroeconomy. Modern game theory-based models provide the necessary framework for understanding traditional banking theory, which emphasizes complex interactions between different economic actors. The first paper in the series explained that traditional banking theory models money as the quintessential network good, because money is supported by a social norm that makes each individual’s decision to use or not to use money dependent on everyone else’s decision. The modern network effects literature tells us that in such an environment changes in quantity may be dissociated from changes in price over a range of values. This phenomenon makes possible monetary finance, or the expansion of the money supply as a source of funds for banks to lend.

Modern theory also gives us the tools to understand the problem that monetary finance solves. Models with liquidity constraints such as new monetarist models and the search model of money explain the unique role that bank-issued money can play in an economy, and thus support the revival of traditional banking theory.¹

¹ See Stephen Williamson & Randall Wright, *New Monetarist Economics: Models*, in B. FRIEDMAN AND M. WOODFORD ED. HANDBOOK OF MONETARY ECONOMICS vol. 3A. (2011); Stephen Williamson & Randall Wright, *New Monetarist Economics: Methods*, 92 FEDERAL RESERVE BANK OF ST. LOUIS REVIEW 265, 294 (July/August 2010) (“banks can also be important in reallocating investment and capital efficiently, with the transactions role of bank liabilities being critical in attracting savings to financial intermediaries that can be channeled into

This paper presumes the concept of monetary finance and explains, first, the institutional structure that supported monetary finance in early 20th century Britain, and, then, the implications of monetary finance for the macroeconomy. In addition, because this argument resurrects the much-derided “real bills” approach to macroprudential regulation, two sections of the paper are devoted to explaining why efforts to adopt this approach failed in the U.S., and how American scholars then attributed the failure to the concept itself, rather than to its faulty implementation. Note that this paper does not claim that traditional banking theory was an adequate framework for guiding U.S. policymakers through the extraordinary circumstances of the Great Depression. The claim made here is that traditional banking theory provides us with a useful model for designing and regulating the financial system in normal times in order to minimize the likelihood of crisis, and that it will probably need to be augmented in the event that a severe crisis is realized.

The core of the paper is the explanation of traditional banking theory’s approach to the relationship between banking and the macroeconomy. First, monetary finance is founded on the structured interaction of banks, the money market, and the central bank. A financial instrument that plays an important role in this structure is the “two-name” bill. Whereas a single-name bill is a simple debt contract promising payment to a payee or to bearer, a two-name bill bears the guarantee, not just of the issuer of the bill, but also of a bank that accepts or endorses the bill. A two-name bill typically starts with a signed and dated order to a bank which may take the following form:

[number] days after date of this my first of exchange pay to [payee] or order the sum of [amount of currency] value received and place the same to the debt of my account.

It becomes a two-name bill when the bank on which the bill is drawn signs it, agreeing to make the payment when it is due. In early 20th century Britain once a bank with a name recognized in London had accepted the bill, the payee – or anyone to whom the payee had endorsed the bill – could discount the bill on the London money market. Because the traditional banking framework requires that two-name paper, not single-name paper, be brought to the discount window by a bank seeking liquidity from the central bank,² the system is structured so that the banks monitor each other, and so that the central bank can exercise control over the assets that trade on the money market.

Second, the central bank limits the growth of the money market by implementing a “real bills”-based macroprudential approach that both proscribes monetary finance of long-term assets and polices the relationship between the real economy and the growth in money market assets. “Real bills” thus had a

investment”); Carolyn Sissoko, *An Idealized View of Financial Intermediation*, ECONOMICS: THE OPEN-ACCESS, OPEN-ASSESSMENT E-JOURNAL, 2007-5 (2007).

² Needless to say, the bank that was the acceptor on the two-name bill had to be a different entity from the bank that sought to discount the bill.

dual purpose: first, to separate money markets from capital markets in order to avoid the financial instability created by asset price bubbles, and, second, to keep money market growth from triggering inflation.³

The concept of real bills can be explained by economic models with liquidity constraints such as the search model of money. When an expansion in the money supply is carefully tailored only to eliminate liquidity constraints, the expansion is not inflationary. By contrast, when an expansion in the money supply increases the amount of money available to facilitate the same transactions, then the expansion will be inflationary. Intuitively non-inflationary monetary expansion can be thought of as using the lending process to extend participation in the market to new individuals, who were in the past excluded from the market due to liquidity constraints. Extending the market in this way need not be inflationary, because the new participants bring with them both resources – possibly intangibles such as human capital – and a demand for goods. Thus, real bills can be understood as bills that address liquidity constraints only, whereas their antithesis, finance bills, were not limited in their uses.

Observe that the real bills approach is specifically tailored to the challenges of managing a system of monetary finance. In a system of monetary finance the growth of bank lending is limited only by actions that undermine the social norm that allows a bank-issued asset to circulate as money. That is, such a financial system does not have the same natural limits as one based on the standard model where finance simply intermediates from savers to borrowers. Thus, the constraints that concern policymakers in a system of monetary finance are those that undermine the monetary social norm: inflation and financial instability (or correlated bank failures). These are precisely the concerns that the real bills approach to macro-prudential regulation was designed to address.

There are a few caveats regarding the topics covered in this paper. First, the relationship between traditional banking and long-term assets in the form of government debt is complicated and beyond the scope of this paper.⁴ Thus, when long-term assets are discussed here, the focus is on privately-issued long-term assets, and the issue of government debt will be left to future work. Second, this paper discusses the historical experience of Britain and the U.S. No effort is made here to understand the relationship of this discussion to banking and finance in Continental Europe. This too is left to future work.

³ Many authors have acknowledged the twin purposes of real bills. See, e.g., Charles Calomiris, *Volatile Times and Persistent Conceptual Errors U.S. Monetary Policy 1914-1951* in MICHAEL BORDO & WILLIAM ROBERTS ED. *THE ORIGINS HISTORY AND FUTURE OF THE FEDERAL RESERVE* 166, 174-75 (2013); C.A.E. Goodhart, *The Changing Role of Central Banks* 2 BIS Working Paper No. 326 (Nov. 2010).

⁴ Indeed, Charles Calomiris & Stephen Haber have devoted an entire book to the connection between banking and government debt. CHARLES CALOMIRIS & STEPHEN HABER, *FRAGILE BY DESIGN* (2014).

Section I sets forth the basic building blocks of traditional banking theory's approach to the macroeconomy: the creation of money market liquidity through the structured interaction of banks, the money market, and the central bank; the importance of protecting capital markets from money market liquidity; and the real bills approach to macro-prudential regulation. Section II examines in detail the theory's historical development in 19th century Britain and the institutional framework that supported traditional banking theory. Section III discusses the unsuccessful attempt to adopt traditional banking theory in the U.S. prior to the Great Depression. Section IV explains that the American critique of real bills should be understood as a critique of the implementation of traditional banking theory in the U.S.

Section V then applies this discussion to the modern "shadow" banking system. First, the empirical evidence that traditional banking approaches were more successful in stabilizing the real economy than modern approaches is discussed. Then, the role played by the integration of capital and money markets in the recent crisis is explained, as is the need to emulate traditional banking theorists by carefully structuring the relationship between banks and money markets before implementing macro-prudential regulation. Next, traditional banking's privately-issued safe assets are contrasted with the informationally insensitive assets of modern theory. Finally, the prospect of restructuring the relationship between banks, money markets, and the central banks and then adopting macro-prudential regulation is contrasted with proposals to have the central banks support long-term asset markets or to adopt 100% reserve banking. Section VI concludes.

I. THE THEORY: TRADITIONAL BANKING AND THE MACROECONOMY

Traditional banking theory derives from the European experience with circulating credit instruments in the 18th and 19th centuries. The theory is British in origin and has never been fully articulated, because it grew organically as means of understanding the changes that were taking place in the financial system during this time. These years first saw the development of the banking system in Britain into a nationwide network that gave every tradesman access to the London money market, and by the end of the 19th century the London money market's role had expanded to include the finance of world trade, not just domestic trade.⁵

⁵ As much as 60% of world trade was settled through London at the turn of the century. BARRY EICHENGREEN, *GOLDEN FETTERS* 43 (1992). A trader who obtained a letter of credit from one of the principal London discount houses could convert the letter into cash in any of the world's trading centers. WILLIAM FREDERICK SPALDING, *FOREIGN EXCHANGE AND FOREIGN BILLS IN THEORY AND IN PRACTICE* 167 (1921).

In this section, the basic framework of the traditional banking theory that developed alongside the London money market is laid out. Section II will go into much more detail regarding the theory's historical development and the institutional framework that supported this theory.

Traditional banking theory explains the role of banks in the economy and provides a model of the relationship between banking and markets and the mechanism by which they create an almost perfectly liquid asset.⁶ Liquidity is created on a market for short-term debt, in which two types of guarantee convert the debt into a liquid asset: private sector banks provide guarantees of payment to each individual bill, converting into a "two-name" bill, and the central bank backstops the market itself by standing ready to discount the debt that trades on the market. For those assets that trade on the market, there is virtually no risk of loss of principal and as a result the risk of loss is limited to the (tiny) loss that may result from fluctuations in the interest rate. Because every asset that trades on the market bears two private sector guarantees of payment – that of the issuer and the acceptor – in addition to the guarantee of any entity that discounts the bill, multiple failures must occur within the short life of the loan in order for the bill to remain unpaid. Since multiple failures are unlikely to occur unless there is a financial crisis, and the central bank backstop of the market makes the likelihood of a crisis in which there are multiple failures of the banks that guarantee this debt extremely low, the assets that trade on this market have almost no liquidity risk. It is important that this structure is designed, not just to provide simple market liquidity, but to provide market liquidity without price risk.

The modern literature, as far as I am aware, does not have a name for the type of market liquidity provided by the London discount market, so here it will be called money market liquidity. Observe that money market liquidity derives neither from the market, nor from the banks, nor from the central bank, but instead from the structured interaction of all three components.

On the other hand, even though the private-sector banks play an important role in determining which issuers' debt will trade on the market, the central bank effectively has veto power over the assets that trade. When the central bank determines that a certain class of debt or a certain issuer's or guarantor's debt will not be eligible for discount, the fact that the debt can no longer be a source of liquidity in a crisis does not just reduce its liquidity, but precludes it from trading actively on the money market.⁷ Because

⁶ See Paul Warburg, *The Discount System in Europe*, Nat'l Monetary Comm'n Document No. 402 (1910); *Stabilization Hearing on H.R. 7895 Before the Committee on Banking and Currency, Part I*, 69th Cong. 431-6 (1927) (testimony of Benjamin Strong, Governor, Federal Reserve Bank of New York).

⁷ The classic example of this is American finance paper in 1906-07, discussed below.

the central bank in this way determines the extent of the money market, the central bank must resolve a key question: What types of assets are eligible for money market liquidity?

Traditional banking theory answers this question in two ways. First, long term assets are not eligible. (Hence the name money market liquidity.) Second, real bills are preferred money market assets, and for this reason the central bank monitors carefully the extent to which bills that are not classified as real trade on the market. These two points are addressed in turn.

Traditional banking theory holds that long-term assets should not be financed on the money market. This follows naturally from the fact that the type of liquidity provided by the money market eliminates price risk. Because an important purpose of capital markets is to provide a venue where supply and demand can determine the price of long-term assets, capital markets are designed to allow the prices of even the safest long term assets to fluctuate along with changes in information and expectations about the future. For this reason, the type of liquidity that capital markets provide is simple market liquidity which allows for price discovery in response to new information. Because capital markets play an important role in price discovery, it would be detrimental for the assets that trade on capital markets to have access to money market liquidity where the market is structured to keep the price of assets from fluctuating. Thus, long-term assets trade on their own markets, such as the stock market, where even the most liquid assets experience price fluctuations.

In short, a basic premise of traditional banking theory is that there are two types of financial markets, investment markets and money markets, and these two markets offer different types of liquidity. Under this theory the type of finance available on investment markets follows the standard model of credit with which we are all familiar today: savers, who have purchasing power in excess of their uses for it, fund borrowers, who have valuable projects but lack purchasing power. The type of finance available on money markets is, however, different. Monetary finance, which was introduced in Part I of this series, refers to the credit that banks create when they expand the money supply either by issuing assets that circulate as money such as bank notes or deposits, or by guaranteeing money market instruments that circulate as a means of payment. In Britain it was monetary finance that was closely associated with banking, and the finance that took place on investment markets was a distinct activity.

The restriction of money markets to short-term assets is, however, insufficient. If the banking system allows too many short-term assets that circulate as money to be issued, the economy can be destabilized by inflation, shifts in the monetary social norm, or a combination of the two. The real bills approach to banking seeks to address these concerns.

In the early 19th century a “real bill” referred to a form of trade credit that was created in a commercial transaction when the seller of an item extended a short-term loan to the buyer. As will be discussed in Section II.A below, this definition had its origins in the law governing negotiable paper. Because the discount market, supported by bank guarantees, was active at this time, real bills were in general a form of monetary finance. By contrast, a short-term loan that was created by monetary finance, but that was not a real bill was a finance bill.⁸ Note that whenever a real bill was rolled over instead of paid off at maturity, it became a finance bill. In general, real bills were viewed as monetary instruments that were unlikely to be associated with inflation or instability, whereas finance bills could have such adverse effects, but did not necessarily have either of these effects. In this context the “real bill-finance bill” distinction created a vocabulary for distinguishing “good” bank loans from bank loans that could contribute to financial system risk. In practice, however, banks were offering credit lines to businessmen and receiving credits and debits to their accounts, and the line between rolling debt over and issuing new loans for new transactions, and therefore between finance bills and real bills, was not always clear.

By the mid-19th century, a problem with this simple framework had appeared. Early central bankers had promoted a “real bills doctrine” that claimed that as long as the banking system was issuing real bills, there could be no inflation. The experience during the quarter century when the British pound was not convertible into gold had demonstrated, however, that even the bills that had been considered real in the pre-suspension era could contribute to inflation if there was no price anchor. This experience had two effects: first, the importance of gold convertibility as an anchor for the banking system became an article of faith, and, second, since a given real-world credit instrument could be inflationary or non-inflationary depending on the broader economic environment, the term “real bill” began to have a dual meaning, sometimes referring to the old legal definition (which due to changes in the law no longer had legal effect) and sometimes referring to a short-term loan that was created by monetary finance and had characteristics that guarantee that it is neither inflationary, nor destabilizing.

Subsequent to the lessons of the first quarter of the 19th century, the Bank of England recognized the need for macroeconomic control – or in contemporary terminology the importance of making sure that the economic environment in which banking operated would favor real bills. This is discussed in detail in Section II.B. In practice policy focused on finance bills and in particular on keeping the growth of finance bills under control, because finance bills could be destabilizing.⁹

⁸ The usage of the term “finance bill” actually dates more to the latter half of the 19th century, but it is used here to provide consistent terminology throughout the paper. In the early 19th century these bills were usually called accommodation bills or fictitious bills. Later, another term used for these bills was speculative bills.

⁹ See JOHN CLAPHAM, *THE BANK OF ENGLAND: A HISTORY*, vol. II at 110-111, 138 (1945).

Finance bills can be classified into three types. Just as the line in the real world between real bills and finance bills can be difficult to draw, however, the same is true of the different types of finance bills.

Any bill that has been rolled over, that is used to finance the holding of inventory instead of its sale, or that is simply an instrument of direct credit to a businessman is a finance bill. This use of finance bills was fairly common and unexceptional, so here these bills will be dubbed common finance bills. Thus, one important aspect of macro-economic control was framed as a problem of keeping the growth of common finance bills relative to real bills under control. By the start of the 20th century it was understood that the central bank's policy rate could be used actively for this purpose: raising the policy rate reduced the growth of credit including finance bills. Policymakers actively monitored the growth of credit and the growth of business activity and were ready to act if credit growth started to outpace economic activity.

At the other extreme were finance bills that were used to finance the purchase and carry of long-term assets, which will be called asset finance bills here. As was discussed above, the price risk inherent in long term assets was understood to be inconsistent with money market liquidity. Furthermore, the monetization of the value of long-term assets was recognized as a very dangerous activity due to experience with this phenomenon in the 18th century. When money is issued against the value of a long-term asset, more money is made available to purchase such assets and it is common for a feedback loop between increases in asset prices and increases in the money supply to develop. When this process comes to an abrupt halt there is a crash in asset prices, and suddenly the justification for the increase in the money supply no longer exists. From historical experience policymakers determined that this was destabilizing. For this reason, finance bills that monetized the value of long-term assets were not issued, and this norm was maintained in Britain by the structural separation between capital markets and banking.

Between the two extremes of common finance bills and asset finance bills were finance bills that were deliberately used to finance long-term investment by a business. These will be called capital finance bills. Capital finance bills could be issued by a bank using the simple technique of rolling over bills that would trade in the money market. In Britain this use of finance bills took place, but was actively discouraged and was often associated bank failures.¹⁰ It was considered an abuse of the system of monetary finance to use it to support long-term investment, and it was undoubtedly a concern that if capital finance became the norm asset price bubbles similar to those created by asset finance would be likely. In fact, one may speculate that the desire to discourage this type of lending on the part of banks may explain why the "real bill-finance bill" dichotomy continued to be used as a theoretic framework through the early years of the 20th century.

¹⁰ CLAPHAM, *supra* note 9, v. II at 261-62.

To summarize traditional banking theory's basic principles of macroeconomics were this:

- Money market liquidity, or market liquidity without price risk, is created by the structured interaction of banks, the money market, and the central bank.
- Financial stability requires that capital markets be insulated from money market liquidity, which can distort long-term asset prices.
- The tool by which the separation between money and capital markets was maintained was the real bills principle, or the effective prohibition on asset finance bills in the money market and the discouragement of capital finance bills.
- Fluctuations in the quantity of common and capital finance bills trading on the market were associated with the business cycle, and their growth could be controlled by either the policy rate or direct supervision of the market.

II. THE BRITISH MODEL: REAL BILLS AS MACROPRUDENTIAL REGULATION

This section traces the historical development of traditional banking theory in Britain. The approach grew organically out of early legal restrictions on the negotiability of paper, and matured over the course of the 19th century until in the early 20th century the Bank of England's pre-eminent role in the London money market was so thoroughly acknowledged that the Bank exercised control over the assets that traded on the money market.

A. The legal foundations of the real bills approach to banking

In the 18th century, the form of banking that supported the payments system was discount banking (which is discussed in detail in Part I of this series). In this system bankers stood ready to discount the IOUs – or bills of exchange – that were issued by account holders. The IOUs themselves circulated as money in part because the banking system supported their liquidity, and in part because as a matter of law they were negotiable via endorsement. What is important to discuss here is that throughout the 18th century and into the early years of the 19th century these IOUs were negotiable under the law because they originated in real transactions.

In the seventeenth century, a bill was simply the mechanism by which “exchange” or the payment completing a real transaction was executed.¹¹ The transactional origin of the bills was reflected in the law

¹¹ JAMES ROGERS, *EARLY HISTORY OF THE LAW OF BILLS AND NOTES* 97 (1995). See also SIR MACKENZIE CHALMERS, *A DIGEST OF THE LAWS OF BILLS OF EXCHANGE* lvii (1896). Note that by the end of the 17th century it was common in Britain for domestic transactions even between lay people and common traders to be executed using bills of exchange. ROGERS at 97.

that governed them. In 1791, a House of Lords judge declared, “the bills of exchange which the usage and custom of merchants originally introduced into the commercial world, and intended to protect . . . are bottomed in real mercantile transactions.”¹² In short, bills of exchange were negotiable because they were “real bills,” representing payment for a real transaction, and this was considered essential to their legal status from the time that they first were used as commercial instruments.¹³

The issue of what was a “real bill” began to be discussed in late 18th century Britain, because this was the time period in which bills that were not real first began to circulate in significant numbers. These newly circulating bills were facially indistinguishable from real bills, but, because there was no underlying transaction, they were in actual fact simply loans based on personal credit.¹⁴ In short, they were finance bills.¹⁵ The initial view of these finance bills was that they bore with them “something of actual fraud.”¹⁶

By the 1830s, however, attitudes within the mercantile community towards finance bills had changed – at least where all parties understood that the issuer of the bill had not received value from the payee. The judicial system acceded to this change by validating the use of finance bills and eliminating the requirement that a claimant on a finance bill show that value had been received.¹⁷ By embracing the use of finance bills and eliminating the requirement that an underlying real transaction be shown in order for the bill to be valid, English law converted an instrument that documented a transaction into an instrument

¹² *Gibson v. Minet*, 1 Bl. H. 569, 618-19 (1791) (Chief Baron Eyre dissenting), quoted in ROGERS, *supra* note 11, at 236. Note that the holding upheld a decision to treat a fraudulent bill made out to a fictitious payee as payable to bearer in order to protect the interests of the innocent holder of the note.

¹³ Many authors claim that the term “real bill” originated with Adam Smith. See, e.g., Hugh Rockoff, *Adam Smith on Money, Banking, and the Price Level*, in Christopher J. Berry, Maria Pia Paganelli, & Craig Smith ed. *The Oxford Handbook of Adam Smith* 307, 316 (2013); Thomas Humphrey, *The Real Bills Doctrine*, FED. RESERVE BANK OF RICHMOND ECON. REV. 3, 8 (Sept./Oct. 1982); Richard Timberlake, *Real Bills Doctrine in U.S. Monetary Policy*, 11 INDEPENDENT REV. 325, 333 (2007). Note that Mints is more careful to distinguish coinage of the term “real bills” from Smith’s presentation of the real bills doctrine. LLOYD MINTS, *A HISTORY OF BANKING THEORY* 25 (1945). Given the legal origins of the concept, it seems likely that Mints is correct and that Smith was not the first to use the term.

¹⁴ See ROGERS, *supra* note 11, chapter 10.

¹⁵ At the time they were, however, called accommodation bills.

¹⁶ *Simpson v. Clarke*, 2 Cr. M. & R. 342, 348 (1835)

¹⁷ ROGERS, *supra* note 11, at 244. This change in the law meant that an important legal constraint on the growth of the bank-based money supply was eliminated in Britain by the mid-1830s. It is unsurprising therefore that this was also the period in which there was a famous controversy over the role of real bills in central bank policy: while no one disputed that the Bank of England should continue to impose a real bills constraint on the money supply by withdrawing support from banks when the volume of bills they guaranteed arrived in such great quantities at the Bank’s discount window that it was clear to the Bank that they were supporting finance paper, the dispute was over whether this constraint – in concert with the gold standard – was sufficient to prevent an inflationary expansion of the money supply. Supporters of additional restrictions won the debate, and as a result the Banking Act of 1844 was enacted, and quantitative constraints on the note issues of both the Bank of England and the banking system more generally were put in place. These quantitative constraints were famously unsuccessful in limiting the growth of the bank-issued money supply.

that itself embodied an obligation to pay.¹⁸ This laid the foundations for bills to be fully negotiable and promoted the circulation of these IOUs as money.¹⁹

Thus prior to the 19th century, the growth of the bank-based money supply was constrained in part by the fact that the law characterized bills that were not based on real transactions as fraudulent. There is little evidence in the legal records that such fraudulent bills circulated in significant numbers prior to the late 18th century.²⁰

B. The growth of finance bills: Early lessons in macroeconomic control

As was noted in Section I, during the first two decades of the 19th century the Bank of England note was not convertible into gold and in this environment policy makers learned that restricting the circulation to real bills was insufficient to control the price level. Over the course of the following century the intellectual framework on which banking policy in Britain was founded combined the real bill-finance bill dichotomy with the gold standard.

At the same time that policy makers were learning about the insufficiency of real bills as a constraint on credit, mercantile experience was demonstrating that the negotiability of a bill was better served by eliminating the requirement that a real transaction be shown before a bill was deemed valid. Thus, in the years following the 1825 crisis finance bills circulated along with real bills. Furthermore, the Bank of England's policy during this post-crisis period often focused more on ensuring that sufficient credit was circulating to support economic activity than on constraining the growth of credit.²¹

The crises of 1847 and 1857, however, demonstrated the dangers inherent in finance bills and the need for a mechanism of constraining the money supply in an environment where such bills were legal. In 1847, many of the houses financing the India trade, turned to finance bills when they started to make losses. As a result of this process, the funds flowing into the loss-making trade increased rather than fell, and houses relying on their established reputations ended up weakening their financial position with ever increasing debt. In short, the 1847 crisis demonstrated that the system of finance bills could have the effect of

¹⁸ ROGERS, *supra* note 11, at 94. Note that French law in particular did not facilitate the development of negotiable paper, and continued to require that a transaction be shown in order for a bill to be valid. CHALMERS, *supra* note 11, at *lvii*.

¹⁹ Rogers, *supra* note 11, at 94; Chalmers, *supra* note 11, at *lvii*.

²⁰ See Rogers, *supra* note 11, at 232-33.

²¹ See, for example, the testimony of G.W. Norman, explaining that in 1839 the Bank did not raise the discount rate to better protect the gold reserve due to concerns about the effects this would have on economic activity. SELECT COMMITTEE ON BANKS OF ISSUE, REPORT, 1840, H.C. Q.1918 at 158 (U.K.) (testimony of G.W. Norman, Director of the Bank of England).

supporting zombie firms and turning losses that from an economy-wide perspective would have been trivial had they been immediately realized into losses that were more difficult to absorb.²²

In the 1857 crisis finance bills entered the money supply to an even greater extent as certain London discount houses did not monitor the paper that country banks were bringing to them. Accountants dealing with the aftermath of the crisis testified that whereas the houses that failed in 1847 were substantial houses that had fallen on hard times, many of the 1857 failures were of tradesmen who had never had any capital or success to begin with.²³ Rather than use the discount rate to attempt to control this problem, the Bank of England turned to its long-standing policy of refusing to support credit that it did not deem “legitimate.”

From the late 18th century on, the Bank of England had monitored the bills that had been discounted with it, tracking totals by both acceptor and discounter, and used this data to determine whether any particular bank appeared to be guaranteeing so much paper that its activity could not be explained by real transactions or, as the bank put it, that the paper did not appear to be “legitimate mercantile paper.”²⁴ The Bank would then make a discretionary decision whether or not to continue to accept the paper guaranteed by the suspect bank. In the past, the Bank’s decision to withdraw the discount privilege from a large bank had precipitated a crisis, so the Bank was well aware of the consequences of such a decision.²⁵ Even so, the Bank of England responded to the 1857 crisis by issuing a policy statement that the discount window would be open to the discount houses on a “special” basis only, and that this segment of the financial industry should “keep their own reserve.”²⁶

Then in 1866 the Bank refused to support Overend, Gurney & Co., a discount house that, next to the Bank, was the largest entity in the banking industry. This action did precipitate a crisis, but no other large financial institutions failed and only a quick v-shaped recession of unexceptional depth followed it.²⁷ This was an important step in the process by which the British banking system developed by the early years of

²² W.T.C. KING, HISTORY OF THE LONDON DISCOUNT MARKET 135-36 (1972).

²³ SELECT COMMITTEE ON THE BANK ACTS, REPORT, 1858, H.C. xiii-xiv (U.K.).

²⁴ The Bank of England’s careful tracking of its exposure to each acceptor and discounter of bills is documented in Marc Flandreau & Stefano Ugolini, *Where it all began: lending of last resort and the Bank of England during the Overend-Gurney panic of 1866*, Norges Bank Working Paper 2011-3 at 23 (2011). See also LONDON CHRONICLE, Sept. 15-17, 1772. SELECT COMMITTEE ON THE BANK ACTS, REPORT, 1858, H.C. QQ 303 – 07 at 21 (U.K.) (Testimony of Sheffield Neave, Governor, and Bonamy Dobree, Deputy Governor).

²⁵ For example, the Bank of England refused to support the Ayr Bank in 1772, LONDON CHRONICLE, Sept. 15-17, 1772.

²⁶ Flandreau & Ugolini, *supra* note 24, at 11 n. 31. Walter Bagehot, Lombard Street XI.30 (1873) (“And the object of that regulation was officially stated to be ‘to make them keep their own reserve, and not to be dependent on the Bank of England.’”)

²⁷ See Sally Hills, Ryland Thomas & Nicholas Dimsdale, *The UK recession in context—what do three centuries of data tell us?* Data Annex Real GDP Series, BANK OF ENGLAND QUARTERLY BULLETIN 2010.

the 20th century into a co-operative environment where the discount houses and joint-stock banks acknowledged and followed the leadership of the Bank of England.²⁸

The questions raised by this discount policy are: Why did the Bank of England consider it important to control the growth of finance bills? What role did the real bill-finance bill dichotomy play in monetary policy?

C. The Bank of England's control of the London money market

The function of the “real bills” approach to monetary policy can only be understood in the context of British banking, which had many unique characteristics. First, the domestic debt that circulated in the British money market was unsecured, and as a result collateral did not function as a means of preventing an overconfident or unscrupulous debtor from running up debts that he and his estate had no capacity to meet. Second, the trade debt of the typical British tradesman was an actively traded and liquid money market asset. Third, the liquidity of such debt derived in part from an accompanying guarantee: the debt was only liquid after the tradesman's banker²⁹ had accepted it.³⁰ Thus, the banks were gatekeepers to this money market and their incentives were fully aligned with those of the purchasers of money market assets. Finally, the London money market was not an open market where sales are final, but a discount market in which every seller guarantees the value of the asset sold.³¹ In short, the London money market was a market in which trade credit was a liquid asset, and the asset was made liquid not based on a claim to collateral, but due to “interconnectedness” or an extremely complex web of overlapping guarantees.

From the point of view of the businessman, credit was not made available on the basis of existing goods as takes place in a collateralized credit system, but on the basis of reasonable future expectations that the

²⁸ KING, *supra* note 22, at 308, 320. Naomi Lamoreaux observes that when the Boston clearinghouse in 1899 allowed a large bank to fail because it disapproved of the bank's balance sheet, this also improved the behavior of the member banks. NAOMI LAMOREAUX, *INSIDER LENDING* 113-14 (1996).

²⁹ Initially, acceptance houses were merchant bankers and most commonly the wholesaler who would market the tradesman's goods in London. I use the term banker here to emphasize the parallels with the checking system.

³⁰ Note that the terminology can become confusing. The domestic acceptances on which the money market was founded in the first half of the 19th century were unsecured. The acceptances that are familiar to scholars of the 20th century are the acceptances that were used in foreign trade and these were secured by goods in the sense that the acceptor has possession of the bill of lading and has the right to sell the goods if the bill is not paid. In Britain the joint-stock banks avoided direct involvement in this trade throughout the 19th century on the basis that a banker should avoid transactions “the results of which were dependent on the rise and fall of goods.” SELECT COMMITTEE ON BANKS OF ISSUE, REPORT, 1875, H.C. QQ 7024 -38 at 363-64 (U.K.) (testimony of Jervoise Smith, private banker with Smith, Payne & Co.) quoted in KING, *supra* note 22, at 280-81. (Of course, these banks were still willing to trade in such paper on the basis of the acceptor's credit – or to accept a bill for a bank-client.) In the early 20th century, however, the joint stock banks entered and became important participants in the market. King, *supra* note 22, at 281. Even so the proportion of unsecured “finance bills” in the stock of international paper on the market was quite high, and has been estimated at 60%. SHIZUYA NISHIMURA, *THE DECLINE OF INLAND BILLS OF EXCHANGE IN THE LONDON MONEY MARKET, 1855-1913* at 66 (1971).

³¹ W.T.C. KING, *supra* note 22, at xvii.

businessman will have goods or their equivalent when the debt is due. In order for this system to work, however, the same businessmen who borrowed against their future prospects had to have enough confidence in the banking system to lend to the banks by holding bank notes or bank deposits.³² The fact that monetary assets supported trade credit enabled banks to eliminate liquidity constraints on production: the banks just needed to lend, the goods would be produced, and the expansion of the money supply would be answered by the growth of GDP.³³ The problem for the central bank was to keep the banking system from abusing this extraordinary power to eliminate liquidity constraints: whenever the banking system lent against prospects that were not realized, there was an expansion of the money supply without an increase in real GDP. In this case, the expansion of the money supply would be expressed in an increase in prices.³⁴

While the banking system was robust to a small number of such failures, an excess of them depreciated the value of the currency. Under the gold standard, any depreciation of the currency was only temporary and the deflationary return of prices to their gold value was usually accompanied by bankruptcies and disrupted business activity. Furthermore, the early 18th century example of France demonstrated that in the extreme case, failure to control the expansion of the money supply could potentially result in the complete collapse of the banking system itself.

In short, lending that expanded the money supply, but that would not increase GDP or that had a significant risk of not increasing GDP was undesirable,³⁵ and the mid-19th century crises had demonstrated that the growth of finance bills could have this effect and needed to be controlled. By

³² The fact that banking is just a variant on traditional mutual lending associations was clearly understood by some early theorists. WILLIAM KNIFFIN, *AMERICAN BANKING PRACTICE* 10 (1921) (“Those who have a surplus loan to those who can use the funds; and *unless the borrower contributes to the pool and helps create and maintain it, he has no right to borrow from it.*”) (emphasis in original).

³³ Henry Thornton observed that, due to the finance of middlemen, the amount of paper currency created to bring a product to market could easily be several multiples of the value of the product itself. HENRY THORNTON, *AN ENQUIRY INTO THE NATURE AND EFFECTS OF THE PAPER CREDIT OF GREAT BRITAIN* 86 (1802). To the degree, however, that in the absence of any given middleman the increase in transaction costs would result in lower production, we should think of this multiplication of paper currency as being accounted for in velocity: that is, the introduction of a new middleman, financed by new bills, into a merchandizing process can reduce the velocity of money.

³⁴ THORNTON, *supra* note 33, at 250 (an increase in Bank of England paper “must, therefore, be supposed to be employed either in transferring an increased quantity of goods, or in transferring the same goods at a higher price.”) See also BENJAMIN STRONG, *INTERPRETATIONS OF FEDERAL RESERVE POLICY IN THE SPEECHES AND WRITINGS OF BENJAMIN STRONG* 183-85 (W. Randolph Burgess ed.) (1930) (“Now in a banking system where 10,000 bank, which represent over 55% of the banking deposits of the country, have convenient access to a source of borrowing such as the Reserve Banks, what are the possibilities that this borrowing may get beyond control; that the volume of credit may become dangerously enlarged, and that in consequence we may be guilty of furnishing credit which might only result in marking up prices without any increase in production, with all of the injustices which are sure to result?”)

³⁵ In fact, such lending was often described as “speculation.”

contrast, traditional real bills by financing current business activity such as the purchase of inputs or the flow of goods to market were closely connected to production and GDP.

As was discussed in Section I, there were three major categories of finance bills: asset finance bills, capital finance bills, and common finance bills.

In Britain, bills did not finance the purchase and holding of long-term assets such as securities, real estate, or government debt.³⁶ Mortgages were financed through building societies, and the distinction between investment and commercial banking was maintained even though it was not legislated. Banks did accept pledges of collateral in the form of securities or even real estate, but played only a very limited role in these markets. In addition, banks' asset portfolios were not regulated and it was common for their portfolios to include some longer-term securities, typically the safest such securities, such as government obligations.³⁷ Overall, banks played virtually no role in financing the purchase of long-term assets.

The reason for the prohibition on the monetary finance of purchases of long-term assets likely appeared so obvious to British theorists that it didn't need to be clearly stated: In the early 18th century John Law proposed the theory that money's purchasing power would be stable as long as it was issued in a fixed ratio to the monetary value of land;³⁸ he had then gone on to experiment with his theory in France, and had demonstrated that monetary finance of long-term assets made possible the type of asset price bubble that could wipe out a financial system and leave a legacy of skepticism towards finance that would endure over the span of more than a century.³⁹ The problem with monetary finance of long-term assets is this: any increase in asset prices results in an increase in the money supply, which in turn means that there is more money to buy the assets and this drives up asset prices further justifying yet another increase in the

³⁶ Goodhart, *supra* note 3, at 2. Meltzer observes that under real bills "Government paper, stock market loans, and real estate mortgages were "speculative" investments." ALLAN MELTZER, A HISTORY OF THE FEDERAL RESERVE v. 1 at 400 (2004). Note, however, that banks have always played a role in financing the working capital of the securities origination industry, but this can be viewed as analogous to the "production" of a security. The details of the relationship between commercial banks and the securities industry in the U.K. and U.S. merits further study.

³⁷ Note that a limited group of London financial institutions were granted the privilege of pledging high quality securities to get an advance directly from the Bank of England. Flandreau & Ugolini, *supra* note 24, at 31. In 1888 advances were made available to a broader group of financial institutions (although still at the discretion of the Bank). CLAPHAM, v. II, *supra* note 9, at 357. By the early 20th century these had become a large and regular business, and even included high-rate loans to Stock Exchange customers. CLAPHAM, v. II, *supra* note 9, at 375-76. Throughout, however, the Bank was cautious about regularly renewing such loans and avoided "semi-permanent" loans. CLAPHAM, v. II, *supra* note 9, at 357, 375-76.

³⁸ JOHN LAW, MONEY AND TRADE CONSIDERED WITH A PROPOSAL FOR SUPPLYING THE NATION WITH MONEY Ch. 7 (1705).

³⁹ The major British theorists all mention this episode, though none of them consider it worthy of lengthy discussion. THORNTON, *supra* note 33, at 239, 341; BAGEHOT, *supra* note 26, at III.17; HAWTREY, CURRENCY AND CREDIT 235, 288 (1919). Indeed, John Law's monetary exploits in France were considered so disreputable in Britain that Thornton mentions that in a debate in the House of Commons, one MP objected to any comparison between Law's bank and the Bank of England. THORNTON, *supra* note 33, at 341.

money supply, etc. Such feedback loops are unsustainable and over time inevitably result in an asset price crash. This dynamic was well understood by traditional banking theorists in Britain such as Henry Thornton.⁴⁰

Unlike asset finance, capital finance or the use of bills to extend rolled-over, thus long-term, loans to businesses did take place in Britain, but the “real bills” approach to banking was designed to limit the extent of such loans.⁴¹ Note that usage patterns would typically give the local bank knowledge of whether a particular borrower had started to rely regularly on finance bills.⁴² Furthermore, the discount house where the local bank rediscounted the bill in order to fund it on the money market would generally gain enough information about the nature of the local bank’s activities to determine if the bank was regularly supporting finance bills. After the crisis of 1866, the Bank of England made it clear that the burden for monitoring the growth of finance bills in the banking system lay on the London discount houses and London banks that together comprised a second-tier of gatekeepers to the money market.

Use of common finance bills or the support of regular clients over a brief period of illiquidity was an accepted practice in banking. Through most of the 19th century growth in this type of finance bill was associated with the business cycle in Britain.⁴³ By the middle of the 19th century tradesmen often anticipated the dynamics of a boom: along with an increase in output and trade, the money supply provided by banks expanded, and this process tended to overshoot driving up prices, until the peak was reached and the cycle turned, culminating in falling business activity, falling credit, and falling prices. Because traders expected prices to rise through a boom, they often sought not just to time the market but to benefit from purchasers’ willingness to pay higher prices by holding inventories off the market for a longer period than normal in an effort to increase the price at which they would be sold. In order to do so, they needed credit, and because they were holding their inventories rather than moving them along, such

⁴⁰ THORNTON, *supra* note 33, at 239, 341; HAWTREY, *supra* note 39, at 194-95. See also U.S. theorists: H. PARKER WILLIS, *AMERICAN BANKING* 156 (1916); Benjamin Beckhart, *Monetary Policy and Commercial Bank Portfolios*, 30 Am. Econ. Rev. Suppl. 17, 22 (1940).

⁴¹ Rockoff presents an example of this phenomenon very clearly by studying the case of the Ayr Bank in 1772. Rockoff, *supra* note 13, at 314-18. A more detailed exposition is available in Hugh Rockoff, *Upon Daedalian Wings of Paper Money: Adam Smith and the Crisis of 1772*, NBER Working Paper No. w15594 (2009). The Bank of England’s experience with mortgages reinforced the view that capital finance was not well-suited to banking. Clapham I, *supra* note 9, at 249-50.

⁴² Thornton, *supra* note 33, at 176 discusses the role played by local banks in “discouraging” finance bills.

⁴³ Thornton, *supra* note 33, at 195, 237-38. The details of the exposition in this paragraph draw from a particularly lucid description of this relationship in a letter from George Lloyd Hodges, Consul-General in Hamburg, to the Earl of Clarendon, Secretary of State for Foreign Affairs, dated Jan. 27, 1858 that was published in Appendix No. 20 to the 1858 Select Committee Report, *supra* note 23. It is remarkable that such sophisticated economic analysis was penned by a career soldier turned diplomat in 1858 – implying that general understanding of the macro-economy in the 19th century may have been significantly greater than is often assumed – perhaps due to Thornton. See also Hawtrey, *supra* note 39, at 11-13 and the 1923 Federal Reserve Annual Report on “speculation in inventories” and its relationship to the business cycle, Federal Reserve Board of Governors, 1923 Annual Report 5, 10 (1924).

credit was not transaction-based, but was finance paper.⁴⁴ In this circumstance, finance paper was being used to keep goods off the market in an effort to drive prices higher with inflationary consequences. In addition to the problem of inflation was that of instability: this use of finance paper was likely to be successful for a time, until the business cycle turned and prices fell, at which point this finance of overlarge inventories at high prices would result in bankruptcies.

Although the Federal Reserve would use the policy rate to control the growth of common finance paper in an expansion, by the early years of the 20th century there is little evidence in Britain that the Bank of England needed to employ the policy rate in this way. Because the Bank had developed such effective means of controlling the assets that circulated in the money market, the policy rate was used mostly to manage the gold reserve.

Over the course of the 19th century the London market became the most important international money market, and alongside this development the Bank of England learned to use the policy rate to protect its gold reserves. In theory there was no conflict between using Bank Rate to support the domestic economy and to protect gold reserves due to the specie flow mechanism: since prices increased along with credit over the course of a boom, as long as prices increased in Britain faster than in its trade partners, the price differential should generate a trade deficit, and an outflow of gold.⁴⁵ Through this mechanism an expansion that was overheating more quickly in Britain than abroad would affect gold reserves and prompt an increase in Bank Rate. In practice, however, as early as 1840 the Directors of the Bank recognized that raising Bank Rate to slow an outflow of gold would constrain the growth of credit, affecting the growth of inventories and economic activity, and that international flows were not as closely related to the domestic economy as the theory implied.⁴⁶ Thus, when there was an outflow of gold the Bank of England had the difficult task of balancing the effects of the global demand for money with the needs of the domestic market. By the start of the 20th century the Bank was using a variety of tools to apply different rates in the domestic and the international money markets.⁴⁷

⁴⁴ Of course, as soon as a tradesman believes that prices are falling, he will reduce his inventories and move his goods along as quickly as possible, so the need for finance paper declines.

⁴⁵ Of course, even Adam Smith recognized that this was a very imperfect mechanism for moderating the growth of bank money. Rockoff, *supra* note 13, at 309.

⁴⁶ RALPH HAWTREY, A CENTURY OF BANK RATE 1-2, 223-24 (1938); Select Committee on Banks of Issue, Report, 1840, H.C. Q.1918, at 158 (U.K.) (testimony of G.W. Norman, Director of the Bank of England). Mr. Norman is explaining why the Bank in 1839 did not follow the “Palmer Rule” of holding a gold stock to cover 1/3 of its monetary liabilities and only raised Bank Rate to 6%. Note that the Directors of the Bank did not explain their policy rate decisions. CLAPHAM, vol. II, *supra* note 9, at 386-87. Thus, the weight that was put on the needs of the domestic economy is often unclear.

⁴⁷ KING, *supra* note 22, at 315 – 16.

Instead of the policy rate, the Bank used its control over the privilege of access to the Bank's high-powered money to limit the growth of finance paper on the money market. The Bank had the advantage that it was well informed about the market, because the flow of bills into the Bank carried a great deal of information. If a particular discount house used a lower standard, more bills would be accepted by that house, and the flows of bills would reflect this. Because the bills were of lower quality, the issuers of the bills would default at a higher rate, and the Bank would be alerted to the fact that the house was not conforming to the market's standards. Given the important role played by the Bank of England in the money market, it was in no one's interest to draw such attention. Indeed, the Bank could and did withdraw credit from discount houses whose behavior did not meet the Bank's standards. In addition, in the event that any guarantor on a bill discounted by the bank declared bankruptcy, the Bank would return the bill to the party who had discounted it – even before it was due – and ask that party to buy it back. While the law of negotiable paper imposed no legal obligation on the discounter to buy the bill back before it was due, repurchase of such bills was considered by the Bank to be a condition of the privilege of bringing bills to the Bank for discount.⁴⁸ Finally, when the Bank determined that in certain trading markets it appeared that finance bills were being rolled over on the London market in order to finance longer-term capital requirements, the Bank “encouraged” discount houses to withdraw credit from those markets.⁴⁹

By the turn of the 20th century, the Bank did not need to use the policy rate to control the growth of finance paper, because it was the acknowledged leader of the discount houses and joint stock banks that were the gatekeepers to the money market.⁵⁰ The degree to which these gatekeepers monitored and controlled the growth of finance paper themselves and the degree to which they simply followed the suggestions of the Bank is unclear. What is clear is that when the Bank made such a suggestion, it was implemented promptly.

Overall, a monetary system had been developed in Britain that, because it was secured only by overlapping guarantees, could expand indefinitely. While this was the great strength of the British financial system, it also meant that the system was always at risk of self-destructing in an orgy of credit.

⁴⁸ Select Committee on the Bank Acts, Report, 1858, H.C. QQ 226-30 at 16 (U.K.) (Testimony of Sheffield Neave, Governor, and Bonamy Dobree, Deputy Governor).

⁴⁹ See EICHENGREEN, *supra* note 5, at 50-51. Eichengreen observes that “the peculiar feature of 1906, taken as evidence of excessive speculation, was that these borrowings were not repaid after the harvest.” Rolling over of short-term paper is a crucial indicator that the paper is not being used to finance transactions, but to raise capital on a longer-term basis, so it is clear that this policy was taken as means of restricting British monetary finance to “legitimate paper.”

⁵⁰ While the joint stock banks and discount houses were clearly playing this gatekeeper role fairly effectively by the turn of the 20th century, the path to this role was not smooth. The process is described in detail by King. KING, *supra* note 22, Chapter IX, 283 ff.

To avoid the bad outcome, without giving up the advantages of this abundant source of credit, a “real bills” approach to banking was developed. This approach placed constraints on the type and quantity of finance paper that traded on the British money market, but these constraints could not – and were not designed to – eliminate finance bills entirely. The function of the real bill-finance bill dichotomy was to discourage the abuse, not to prevent the use, of finance bills. Overall, the constraints on finance bills were designed to protect the stability of the British money market and its ability to issue credit in quantities that had no fixed limit.

D. Elements of the British banking model

British banking derived liquidity from the British money market. The money market in turn derived liquidity from a carefully calibrated incentive structure. The debt that traded on the money market was protected by overlapping guarantees, and sellers on the money market generally had to provide such a guarantee. This system of overlapping guarantees was protected from collapse by a central bank which played two roles: First, in a crisis the central bank expanded the high-powered money supply to whatever degree was necessary to support the banking system and money market.⁵¹ Second, the central bank monitored and disciplined the money market to make sure that the quality of credit on the money market did not deteriorate. Needless to say these two roles were complementary: the central bank was confident enough to expand the money supply on an “as needed” basis during a crisis, because the central bank was engaged in day-to-day monitoring and discipline of the quality of the money market.

Overlapping guarantees had the benefit of creating liquidity in the money market in normal times, because a single default or bankruptcy had no effect on the value of the debt that circulated in the market – the additional guarantees protected the value of the debt. This same system of guarantees, however, had the disadvantage of having destabilizing effects in a crisis: A cluster of bankruptcies could trigger a chain of failures that could endanger the financial system itself. For this reason, it was crucial that origination practices in the market limited debt to debt that was very likely to be paid.

The theoretic framework that was used to evaluate origination practices was the real bill-finance bill dichotomy. The least risky form of money market debt was the “real bill.” Because real bills were short term and closely tied to the processes of production and merchandizing, they comprised the form of debt that was most likely to be paid – and unsurprisingly, due to the lower level of risk, there was a strong bias in favor of originating and circulating real bills. Finance bills by contrast could range from short-term loans designed to support a long-term bank client over a period of illiquidity to rolled-over short-term

⁵¹ Observe that the Bank of England was only able to do this because it had the support of the government. This, however, is a topic that goes beyond the subject of this paper.

bills that were effectively long-term loans to bills that were used to finance the purchase of long-term assets. Whereas the first category of finance bill was considered an unexceptionable use of this type of loan, the other two categories were considered to have a destabilizing effect on the financial system and their use was actively discouraged. Thus, the essence of the real bills approach to banking was to restrict the use of finance bills. It was designed first to make it very difficult to use short-term to finance long-term activity or assets, and, second, to provide guidance to banks in the prudent use of finance bills. Overall, the “real bills” approach to banking combined, at one end of the risk spectrum, macroprudential regulation that was designed to constrain the growth of asset price bubbles, and at the other end of the risk spectrum, a risk management framework for banks to follow when originating bills.

To summarize the British banking model is based on a liquid money market. The liquidity of the money market is in turn founded on:

- (i) Money market structure: A market where
 - a. two name paper, or unsecured short-term debt that has been guaranteed by a bank, trades, and
 - b. each seller provides a guarantee of value at the time of sale,
- (ii) A central bank that supports the money market by
 - a. standing ready to provide high powered money to the money market on an “as needed” basis by discounting two name paper, and
 - b. regulating the money market “macroprudentially” based on the real bills principle, and
- (iii) Banks that employ a risk management framework based on the real bills principle

III. EARLY 20TH CENTURY U.S. BANKING THROUGH A TRADITIONAL BANKING LENS

The British financial system was not only remarkable in the liquidity that it was able to provide to world financial markets, but also had a unique structure. Thus, when the focus turns to the U.S. financial system, it is necessary to recognize that even basic concepts, such as that of a bank, change. Whereas the banks that issued monetary liabilities in Britain were constrained by the real bills principle from taking on significant exposure to long-term assets, the concept of a bank in the 19th century U.S. was not so limited: U.S. banks were both much more exposed to long-term assets and much more highly capitalized. Thus, British banks specialized in monetary finance, while U.S. banks combined a comparatively small business in monetary finance with what was basically a mutual fund structure. This section discusses the

efforts of U.S. policy makers to emulate the British financial system, and in particular to fashion a system which would offer liquidity comparable to the London money market, but adapted to the particular characteristics of the U.S. financial system.

The traditional banking theory that was set forth in Sections I and II of this paper, is the framework that is used to evaluate early 20th century U.S. banking in this section of the paper. Subsection A describes the U.S. financial system at the turn of the 20th century, concluding with an outline of the basic elements of this pre-Fed financial system for easy comparison with the British system. Subsection B describes the panic of 1907 that created the impetus for the creation of the Federal Reserve System (“Fed”) and the mechanism by which the Fed provided liquidity to the U.S. financial system, again concluding with an outline of the basic elements of the system for comparison. Subsection C presents traditional banking theory’s view of how the Fed failed to support the U.S. financial system in the first decades of its existence.

A. Early 20th century U.S. banking

At the turn of the 20th century, the American banking environment was very different from the British banking environment. Two fundamental structural characteristics shaped these differences. First, in the U.S. limited liability banking was the norm from the earliest years of the republic, whereas the basic structure of the London discount market had developed in an environment where most bankers had unlimited liability. Second, the Second Bank of the United States closed in 1836, and the U.S. banking system operated without a central bank throughout the remainder of the century; by contrast, the Bank of England had recognized a duty to support the discount market since at least 1810.⁵²

The foundations of British banking lay in unlimited liability which gave currency to the guarantee of a well-to-do banker whose “capital” consisted of all of his private property,⁵³ whereas in the U.S. banking grew up on the basis of the limited liability corporate form. As a result of these differences in structure, the basic model of banking was completely different in the two economy. Early 19th century bankers in Britain leveraged their own wealth and creditworthiness to extend credit to local tradesmen by

⁵² Ian Duffy, *The Discount Policy of the Bank of England during the Suspension of Cash Payments, 1797-1821*, 35 ECON. HIST. REV. 67, 76 (1982). See also CLAPHAM, v. 2, *supra* note 9, at 15.

⁵³ L.S. PRESSNELL, COUNTRY BANKING IN THE INDUSTRIAL REVOLUTION 235-36 (1956). Limited liability banking developed slowly in Britain, because the guarantee of such a bank was perceived as weaker than the guarantee of an unlimited liability bank. Only after 1878 was there a strong impetus in Britain for the growth of limited liability banking. In 1878, a long-troubled unlimited liability joint-stock bank, the City of Glasgow Bank, failed, and its liquidation left most of its 1200 shareholders in penury.

Note that this difference too may relate back to the presence of a reliable central bank-lender of last resort in Britain, since the stabilization of credit markets provided by the central bank may have been necessary to induce well-to-do individuals to take on the risks of unlimited liability banking.

guaranteeing the tradesmen's debt. In the U.S. the desire for access to bank credit often motivated men of modest means to incorporate a bank, serve as its director, and receive "insider" loans.⁵⁴

U.S. banks were highly capitalized. Up through 1850, the typical U.S. bank's ratio of capital to assets was greater than 40%.⁵⁵ By the time that data are available for both the U.S. and Britain, U.S. banks generally had a capital-asset ratio that was significantly higher than that of British banks. From 1900 to 1914 the capital-asset ratio of British banks fell from 12% to 9%.⁵⁶ Over the same period, the capital-asset ratio of U.S. banks declined from 18% to 16.5%.⁵⁷

Overall, even though the bank-as-mutual-fund model is very misleading when applied to the British model of banking, it may be a reasonably accurate model of the stock-financed banks that provided the U.S. money supply in the early 19th century. Indeed, many have argued that the loans of early U.S. banks were usually rolled over and thus were long-term in nature. To the degree that this is accurate, using contemporary British terminology they were not banks at all, but finance companies.

Equally, if not more, important in shaping the nature of banking in the U.S. was the absence of a central bank. Because U.S. markets had nowhere to turn when there was a financial panic, there was at the turn of the 20th century no active trading market in bills, or two-name paper, as it was called in the U.S.⁵⁸ While there was a commercial paper market, banks were unwilling to guarantee their customers' loans, and this precluded the rediscount of the loans that was the norm on the London market. Instead there was an "open market" in single-name commercial paper, where paper could be traded without the seller's guarantee. Because this paper was single-name paper, supported only by the credit of the issuer, only extremely high quality issuers could raise funds on the market.⁵⁹ Thus, the market comprised only about 5% of

⁵⁴ HOWARD BODENHORN, STATE BANKING IN EARLY AMERICA 292 (2002); LAMOREAUX, *supra* note 28, at 62. Calomiris and Haber observe that through the early years of the 19th century the U.S. corporate banking system had the effect of closing certain citizens out of credit markets. CHARLES CALOMIRIS & STEPHEN HABER, FRAGILE BY DESIGN 169 (2014).

⁵⁵ LAMOREAUX, *supra* note 28, at 65.

⁵⁶ DAVID SHEPPARD, THE GROWTH AND ROLE OF U.K. FINANCIAL INSTITUTIONS 1880 – 1962, Table A1.6 (1971).

⁵⁷ FEDERAL RESERVE BOARD OF GOVERNORS, ALL BANK STATISTICS UNITED STATES 1896 – 1955, Table A-1, 31, 33 (1959).

⁵⁸ During the time that the Second Bank existed, it did support a fairly liquid market in two name paper. Michael Bordo & David Wheelock, *The Promise and Performance of the Federal Reserve as Lender of Last Resort 1914 – 1933*, in MICHAEL BORDO & WILLIAM ROBERDS ED. THE ORIGINS HISTORY AND FUTURE OF THE FEDERAL RESERVE 59, 62 (2013). The reasons for the disappearance of the market in two name paper remain unclear but may include the failure of the acceptance houses in the decade following the panic of 1837, and the statutory restriction in 1863 of National Banks to the "business of banking" which may have hampered the evolution of this business to include acceptances as took place in Britain. See Lawrence Jacobs, *Bank Acceptances*, Nat'l Monetary Comm'n Doc. No. 569 (1910) at 4.

⁵⁹ John James, *The Rise and Fall of the Commercial Paper Market, 1920-1929*, in MICHAEL BORDO AND RICHARD SYLLA ED. ANGLO-AMERICAN FINANCIAL SYSTEMS 219, 222 (1995); LAMOREAUX, *supra* note 28, at 89-90; WILLIS, *supra* note 40, at 44-45.

commercial bank loans.⁶⁰ Because the U.S. money market was so small, the vast majority of bank loans in the U.S. were illiquid assets, and U.S. banks could not rely on the money market as a source of funds in the same way that British banks did.

In short, American banking institutions had to adapt themselves to an environment in which there was no reliable source of liquidity. Banks had no choice but to hold their loans on their balance sheets. For protection, it was not unusual for a bank to demand a pledge of collateral to support a credit line.⁶¹ As was noted above, many observed that bank loans were often short term in name only and there was an expectation that they would be rolled over.⁶² Also observed was the fact that this made banks dependent on their borrowers. This stands in stark contrast to the British environment, where banks were gatekeepers to the money market.

Since U.S. money markets were too illiquid to provide a venue in which banks could raise funds when needed, the banks turned to capital markets. Thus, in the decade before the founding of the Federal Reserve the principal locus of bank reserves in the U.S. was in the call loan market, where loans were secured by stock market collateral.⁶³ Because these loans were callable, they could function as reserves – in the event that a broker didn't have the funds to repay on demand, the bank could simply sell the collateral to get the value of the loan. This market was so important to the banking system that the call loan rate served as the best indicator of the liquidity of the U.S. banking system.⁶⁴

By relying on call loans for liquidity, however, the banking system quickly transmitted financial stress from the banking system to capital markets. When the banks were under stress and called their loans, the

⁶⁰ James, *supra* note 59, at 222. Given the lack of overlapping guarantees in the U.S. market, it is unsurprising that prior to World War I the New York commercial paper rate was at least a full percentage point higher than the market discount rate in London. New York: John James, *supra* note 59, at 226. London: King, *supra* note 22, at 312. The discounts held by the banks and discount houses in Britain accounted for approximately 20% of their loans in the early years of the 20th century. SHEPPARD, *supra* note 56, Tables A1.1 and A1.13. Overdrafts accounted for much of the remainder. Note, also, that this figure is calculated, excluding the balance sheet item "Acceptances and Endorsements" the inclusion of which, when aggregated might lead to double-counting. Acceptances and Endorsements range from 25 to 45% of discounts.

⁶¹ BODENHORN, *supra* note 54, at 48; LAMOREAUX, *supra* note 28, at 120-21; WILLIS, *supra* note 40, at 47-48.

⁶² LAMOREAUX, *supra* note 28, 2-3, 69; H.G. Moulton, Commercial Banking and Capital Formation II, 26 J. Pol. Econ. 638, 645 (1918); H.G. Moulton, Commercial Banking and Capital Formation III, 26 J. Pol. Econ. 705, 707-08 (1918); JOHN JAMES, MONEY AND CAPITAL MARKETS IN POST-BELLUM AMERICA 62 (1978); WILLIS, *supra* note 40, at 40-42.

⁶³ BD. OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM, BANKING AND MONETARY STATISTICS 1914-1941 at 434 (1943)

⁶⁴ WILLIS, *supra* note 40, at 286.

funds available to finance such loans were reduced and along with an increase in the call loan interest rate, sales of securities took place, triggering a fall in stock prices.⁶⁵

Due to the stock price effects of drawing on call loans for liquidity, the call loan market was an imperfect source of liquidity for the U.S. banking system, and it relied on two additional mechanisms: the London money market and clearinghouse loans. The support that the London money market provided to the U.S. banking system is demonstrated by the influx of American finance paper that Bank of England often saw in the fall as the U.S. banking system labored to find the liquid assets needed to make payments during the harvest.⁶⁶

In addition, each major city had its own clearinghouse, or consortium of banks that was initially formed for the purpose of efficiently clearing payments. The clearinghouses stepped in to provide liquidity to member banks in crises. In severe crises, they would jointly suspend payments in gold on deposits and instead pay out jointly-guaranteed clearinghouse loan certificates. Clearinghouses also, however, played an important role in reinforcing the U.S. banking system's reliance on capital markets as a source of liquidity: when a clearinghouse extended an emergency loan to a member, it typically demanded as collateral marketable commercial paper or securities traded on the stock market.⁶⁷

The characteristics of U.S. banking that have been detailed here are closely interrelated. For example, knowledge that any draw down on U.S. bank reserves would come at a significant cost to the liquidity of the financial system may have played an important role in constraining bank lending and the banking system's willingness to extend monetary finance by funding loans with deposits. This is similar to Naomi Lamoreaux's view that the high capitalization which characterized early 19th century U.S. banks is explained by the fact that bank loans were mostly long-term loans (in the form of finance bills that were by custom renewed), and in consequence banks viewed deposits as a very risky type of liability and rationed them.⁶⁸ In short, it seems likely that the illiquidity of U.S. money markets helps explain the fact that the leverage of U.S. banks – and therefore the degree to which they served as a source of monetary finance – was much lower than that of British banks.⁶⁹

⁶⁵ The structure of U.S. bank reserves explains the fact that a fall in the price of stocks was a key indicator that the risk of a financial panic was rising. See Charles Calomiris & Gary Gorton, *The Origins of Banking Panics: Models, Facts, and Bank Regulation*, in R. Glenn Hubbard ed. FINANCIAL MARKETS AND FINANCIAL CRISES 109 (1991).

⁶⁶ EICHENGREEN, *supra* note 5, at 50, 55.

⁶⁷ LAMOREAUX, *supra* note 28, at 126.

⁶⁸ LAMOREAUX, *supra* note 28, at 2-3, 67-69. Bank notes in New England were constrained because they had to be adequately backed with specie in order to circulate in the Suffolk system. *Id.* at 19, 66.

⁶⁹ Andrew Haldane, Simon Brennan & Vasileios Madouros, *What is the contribution of the financial sector: Miracle or mirage?* in THE FUTURE OF FINANCE: THE LSE REPORT 87, 115 (2010). The authors compare U.S. banks' ratio of equity to assets to U.K. banks ratio of equity plus capital reserves to assets and find that over the first two decades of

Despite the fact that the U.S. banking system was built on a completely different model than the British banking system, many American contemporaries attempted to apply the real bill-finance bill dichotomy to U.S. banking, perhaps due to general admiration for the British model.⁷⁰ The real bills approach is, however, not well-adapted to the U.S. model of banking: Neither real nor finance bills served as an important source of liquidity for U.S. banks in a crisis, so it is not surprising that U.S. banks apparently failed to distinguish carefully between the two and in general appear to have allowed their banking business to evolve into what was effectively longer term lending. Thus, instead of using real bills as a risk management strategy, U.S. banks developed alternate mechanisms of managing risk including relying on financial statements and emphasizing the marketability of collateral and investments.⁷¹

Even so, some have argued that the National Bank Act was based on a real bills approach.⁷² If this is correct, then the legislators at the time were misguided in their efforts, because the National Bank Act's definition of the "business of banking" and the degree to which it prohibited National Banks from performing basic banking functions was being litigated through the latter years of the 19th century.⁷³ Thus, there is reason to believe that the Act played a role in stymieing the development of a two-name paper market in the U.S.⁷⁴

More importantly, however, the principle source of liquidity in the U.S. banking system involved a form of lending, short-term lending against the value of securities as collateral, that was considered extraordinarily dangerous in a real bills framework. For this reason, not only did U.S. banks fail to use real bills as a risk management strategy, but the macroprudential function of the real bills framework could not be applied in the context of American banking. Overall, in the absence of a central bank, the U.S. financial system had evolved in ways that made British banking theory irrelevant. Section IV will address the confusion created by attempts to apply the real bills principle to the U.S. banking system.

Before turning attention to the crisis that prompted policymakers to create the Federal Reserve, a summary of the basic elements of the pre-Fed U.S. financial system will be useful. The American banking

the 20th century the U.K. ratio was significantly lower than the U.S. ratio. Only around 1960 do the two ratios become similar.

⁷⁰ See, for example, WILLIS, *supra* note 40, at 40-42, 188; CHARLES DUNBAR, CHAPTERS ON BANKING 25-26 (1909). John James finds that the real bills approach was broadly accepted. JAMES, *supra* note 62, at 59 -61.

⁷¹ LAMOREAUX, *supra* note 28, at 121-22, 131-32.

⁷² LAMOREAUX, *supra* note 28, at 119-20.

⁷³ *Auten v. U.S. Nat. Bank of N.Y.*, 174 U.S. 125, 143 (1899) (holding that National Banks were indeed permitted to rediscount bills).

⁷⁴ See Lawrence Jacobs, *Bank Acceptances*, Nat'l Monetary Comm'n Doc. No. 569 (1910) at 4 (arguing that the National Bank Act prohibited National Banks from accepting bills). Note that this conservative view of banking was held in Britain in the mid-19th century, but by the early years of the 20th century attitudes had changed. KING, *supra* note 22, at 280-81.

model in the early 20th century was based on the best source of liquidity available to it: the stock market.⁷⁵ When under stress U.S. banks drew on the reserves they held in call loans. While call loans provided adequate liquidity to the banking system (which after all served the U.S. reasonably well through these years), they did so only by draining liquidity from capital markets – and causing significant declines in price. For this reason, U.S. bank reserves were much less liquid than British bank reserves.

When the strain created by the banking system's liquidation of call loans had too great an effect on stock prices, the clearinghouses – and in particular the New York Clearinghouse – would step in. By providing loans to illiquid banks and, if necessary, calling for a joint suspension of payments in gold, the clearinghouse arrested the decline in capital market prices that was being caused by the banking system's need for liquidity.

To summarize, the American banking model in the first decade of the 20th century derived the limited liquidity that was available to it from:

- (i) Capital markets: A system of callable debt secured by stock market collateral, and
- (ii) Clearinghouses that supported member banks directly by:
 - a. standing ready in a crisis to support the member banks by providing loans and issuing clearinghouse loan certificates, and
 - b. monitoring the balance sheets of the member banks to ensure that they had sufficient marketable securities.
- (iii) Banks that nominally followed the real bills principle, but in practice used other risk management techniques.

B. The Federal Reserve: Making an Illiquid System Liquid

The American banking model was adequate to meet the needs of the U.S. economy in the latter years of the 19th century. The panic of 1907, however, convinced prominent politicians and members of the financial community that it was inadequate to meet the needs of the 20th century U.S. economy. The Federal Reserve was founded for the purpose of stabilizing the U.S. financial system. This paper argues that, although the Federal Reserve did serve as a source of liquidity for the financial system, the Federal Reserve was in fact a destabilizing force in the U.S. financial system prior to the reforms of the 1930s, because it provided liquidity to a structurally illiquid financial system.

⁷⁵ While it was true that the commercial paper market was an important secondary source of liquidity, it was sufficiently small that I abstract from it here. I am also abstracting from the liquidity provided by the London money market: the importance of this is very difficult to measure, so for the purposes of this paper, I will treat it as small.

1. *The panic of 1907*

The panic of 1907 was more severe than the financial panics that preceded it. In particular, the 1907 crises was aggravated by the fact that the Bank of England had at the end of 1906 deliberately limited the circulation of American finance paper on the London money market. In 1906 the U.S. financial system was under a great deal of strain financing the insurance payments due after the San Francisco earthquake of that year.⁷⁶ After the harvest in 1906, the Bank of England found that the American paper discounted at the Bank failed to dissipate as it normally would, and this was treated by the Bank as an indicator that the U.S. banks were drawing on the London money market for longer-term loans.⁷⁷ Together with the drain of gold from the Bank, this was sufficient for the Bank to “suggest” to the discount houses that they stop accommodating American finance paper.⁷⁸ As a result, this paper was allowed to run off, and by 1907 it no longer traded on the London money market.

Without the support of the London money market, the U.S. financial system threatened to self-destruct entirely in 1907. From October 1906 to November 1907 the stock market declined by 54%.⁷⁹ In late October 1907 the New York banks, organized by J.P. Morgan, bailed out the call loan market by putting up the money needed to keep about 50 stock brokers from declaring bankruptcy simultaneously.⁸⁰ An underlying problem was, however, that a new form of financial intermediary had arisen, the trust company, that provided banking services including deposit accounts, but did not join the clearinghouse. Thus, bank support of the stock market only served to give the trust companies more time to liquidate their stock holdings.⁸¹ A final end to the crisis was only found when J.P. Morgan locked the trust company presidents in a room and instructed them to subscribe to a joint loan that would support the weaker trust companies in order to prevent “complete collapse of the entire banking system.”⁸²

During the crisis the interest rate on call loans reached 75%.⁸³ The stress on U.S. markets naturally drew gold from abroad, but very little came from Britain. Instead, gold that originated mostly in France and

⁷⁶ Kerry A. Odell & Marc D. Weidenmier, *Real Shock, Monetary Aftershock: The 1906 San Francisco Earthquake and the Panic of 1907*, 64 J. ECON. HIST. 1002 (2005).

⁷⁷ EICHENGREEN, *supra* note 5, at 50-51. Note that the British financial press attributed the American demand for finance on the London market to the Bank of England’s error in maintaining a low discount rate. BANKER’S MAGAZINE 1906 v. II at 565-66, 605-07, cited in KING, *supra* note 22, at 316.

⁷⁸ Presumably American acceptances were still discounted, but this continuing support of international trade would hardly be sufficient to ease the financial strains in the U.S.

⁷⁹ NBER Macrohistory Database, Index of Common Stock Prices, Series M11007.

⁸⁰ ROBERT BRUNER & SEAN CARR, PANIC OF 1907 at 100-03 (2007).

⁸¹ *Id.* at 99, 102.

⁸² *Id.* at 125

⁸³ *Id.* at 110.

Germany was transmitted through the London money market to the U.S.⁸⁴ Overall, the message sent in 1907 by the Bank of England was crystal clear: the U.S. economy would have to keep its own reserve.

2. *The Federal Reserve Act of 1913*

A National Monetary Commission was formed by Congress to study the establishment of a central bank and in 1913 the Federal Reserve System (“Fed”) was created. As its name implies, the primary purpose of the Fed was to provide a source of high-powered money – or reserves – to the banking system when the demand for money was particularly high. Thus, the U.S. would be able to address the problem of seasonal demand for money without putting pressure on international financial markets, and it would also be able to provide support the banking system when the banking system was under stress – as it had been after the San Francisco earthquake.⁸⁵

The founders of the Fed recognized that the structure of U.S. money markets was not an ideal environment in which to implement central bank monetary policy. In particular the primary reserves of the banking system were held in stock market loans and not in money market instruments. Thus, the Federal Reserve Act was designed to encourage the growth of a liquid money market on the model of the two-name paper of the London money market.⁸⁶ In support of this policy the Federal Reserve Banks were explicitly prohibited from re-discounting stock market call loans.⁸⁷ On the other hand, due to the absence of a large and active money market in private sector debt, the Federal Reserve Banks were authorized to re-discount the single-name paper that banks typically held on their balance sheet, even though such paper would not have been eligible for discount at the Bank of England. Furthermore, since banks viewed the discount of the single name paper that they held as a sign of weakness, the Federal Reserve Banks were authorized to purchase such paper without requiring a bank’s endorsement.⁸⁸ In short, while the Federal Reserve was in many ways modeled on the Bank of England, in order to address the structural differences of the U.S. banking system the Federal Reserve Act gave the Federal Reserve Banks powers that were very different from those typically used by the Bank of England.

In fact, the drafters of the Federal Reserve Act were not successful in fostering the growth of a market in two-name bills. John James has found that the commercial paper market did increase a little in real terms

⁸⁴ EICHENGREEN, *supra* note 5, at 52.

⁸⁵ See STRONG, *supra* note 34, at 229-30.

⁸⁶ Federal Reserve, *supra* note 43, at 12.

⁸⁷ Federal Reserve Act, Pub. L. 63-43 § 13 (1913).

⁸⁸ Federal Reserve Act, Pub. L. 63-43 § 14(d) (1913); J. LAURENCE LAUGHLIN, BANKING REFORM 93 (1912).

up through 1924, but over the next five years it declined steadily by a total of 71%.⁸⁹ By contrast the call loan market increased in size over the same period by 433%.⁹⁰

The absence of a discount market had significant consequences for the implementation of central bank policy for three reasons. First, the Federal Reserve did not have a significant segment of the market flowing through its discount window and, unlike the Bank of England, did not get a broad view of the paper that was circulating. Second, the Fed found that discounts were an inadequate source of income, and turned to “open market” purchases first of commercial paper, and then of government debt. By 1922 the Fed had learned that sales from the portfolio of government debt increased use of the discount window, and in 1924 the Fed announced a policy of using open market purchases and sales of government debt that were coordinated across the Federal Reserve Banks to make its interest rate policy effective.⁹¹

Finally, the absence of a two-name paper market meant that U.S. markets could not benefit from the interbank monitoring that was the norm in Britain. In a two name paper market, the banks are each other’s keepers. Any bank that does not conform to the expectations of the market will find that its acceptances are worth less and do not circulate as readily as those of conforming banks. Indeed, because access to Bank of England liquidity requires that the holder of the paper provide his own guarantee, it may well be the case that a non-conforming bank’s paper does not circulate at all. Thus, the Bank of England could define those assets that could circulate on the money market and rely on the banks and discount houses to enforce compliance with that policy among themselves.⁹²

The Federal Reserve Act attempted to limit the reach of Federal Reserve liquidity by limiting the types of assets that were “eligible” for discount and for purchase by the Federal Reserve on the open market.⁹³

Eligible assets are defined in the Act as

⁸⁹ James, *supra* note 59, at 233.

⁹⁰ FEDERAL RESERVE, BANKING AND MONETARY STATISTICS 1914-1941 at 494 (1943). The calculation, matching John James calculation, is from September 1924 to September 1929. Although this is in nominal terms, inflation over this period is trivial.

⁹¹ FEDERAL RESERVE, *supra* note 43, at 3, 13. Indeed, the power of engaging in open market operations was adopted by the framers in the Act in order that the Federal Reserve Banks “could make their rates of discount effective.” WILLIS, *supra* note 40, at 326. Similarly the Banks can buy and sell from non-members in order to “extend the benefits of their discounting power to non-members.” *Id.* The Bank of England had used a similar method of trading in government securities to make the policy rate effective in the 19th century. KING, *supra* note 22, at 119, 296.

⁹² KING, *supra* note 22, at 309 (“the outside market ... could be trusted to respond vigorously to an official hint (backed by the threat of vigorous action should it fail to do so)”). It was famously claimed that the Governor of the Bank of England had only to raise an eyebrow for a banker to change his behavior. Flandreau & Ugolini, *supra* note 24, at 76, 89.

⁹³ Federal Reserve Act, Pub. L. 63-43 §§ 13, 14 (1913).

notes, drafts, and bills of exchange arising out of actual commercial transactions; that is, notes, drafts, and bills of exchange issued or drawn for agricultural, industrial, or commercial purposes, or the proceeds of which have been used, or are to be used, for such purposes, the Federal Reserve Board to have the right to determine or define the character of the paper thus eligible for discount, within the meaning of this Act. . . . such definition shall not include notes, drafts, or bills covering merely investments or issued or drawn for the purpose of carrying or trading in stocks, bonds, or other investment securities, except bonds and notes of the Government of the United States.⁹⁴

Observe that the first clause makes real bills eligible assets, but the explanatory second clause also makes common finance bills – and possibly capital finance bills – eligible assets. Observe also that later in the paragraph many asset finance bills are made ineligible. Thus, the Federal Reserve Act promotes a real bills-based approach to central banking, but had to make a much broader class of assets eligible for two reasons. First, in a real bills framework common finance bills are also eligible for central bank liquidity. The real bills principle indicates only that the growth of these bills may need to be controlled by tools such as the policy rate or supervision of the money market. Second, in the U.S. the distinction between real bills and finance bills had been irrelevant for so long that it was impracticable for legislation to be founded on a distinction that bankers would not know how to draw.

Note also that under the Act eligible assets include single name paper, since “notes” are included in addition to “bills of exchange.” Once again this modification of the real bills principle was necessary, because banks in the U.S. held very little two name paper and a restriction to such paper would likely have precluded Federal Reserve support of the banking system. Because of this modification, however, the Federal Reserve did not have control over the use of the liquidity that it provided to the banking system that was similar to that of the Bank of England. In Britain the discount houses and joint-stock banks functioned as the Bank’s lieutenants, deliberately limiting access to the money market to issuers who might fail the Bank’s criteria – because these private entities did not want to be caught out either holding paper that was not discountable at the Bank or providing guarantees on such paper. The Federal Reserve Banks had no lieutenants, because they discounted single name paper. Instead, they effectively were forced to trust the banks that were discounting the paper to determine whether or not it was “eligible” – that is, “issued or drawn for agricultural, industrial, or commercial purposes.” Only the originating bank was in a position to have this information and, even then, the question was a matter of judgment.⁹⁵

⁹⁴ Federal Reserve Act, Pub. L. 63-43 § 13 (1913).

⁹⁵ WILLIS, *supra* note 40, at 42-44. *See also* FEDERAL RESERVE, *supra* note 43, at 34-35.

The fact that the drafters of the Federal Reserve Act of 1913 had to adapt the real bills principle to the realities of U.S. banking at the time had consequences: on such consequence was that the Act was ineffective in granting the Federal Reserve the powers necessary to direct the flow of bank credit away from undesirable forms of finance paper. In particular, the Act failed to prevent the flow of Federal Reserve liquidity to the call loan market: a bank could discount some of the loans it held on its balance sheet and place the funds received in call loans.⁹⁶ Because the Fed stood ready to discount single name paper, banks had direct access to the Fed's liquidity, rather than intermediated access to central bank liquidity as in Britain, and for this reason the U.S. banks were not limited in the uses to which they could put the liquidity they received from the Fed. As Benjamin Strong found in 1922 "[T]he expansion of the loan account of the Federal Reserve Banks, . . . can be brought about as a result of any expansion of banking in the country, no matter what may be its cause."⁹⁷

Overall, even though the Federal Reserve Act was nominally based on the real bill-finance bill dichotomy, it did not equip the Federal Reserve with the tools necessary to impose a real bills approach on the U.S. banking system. Indeed, it is far from clear that converting the U.S. banking system to real bills was a feasible policy option in 1913, because banks were adapted to functioning in an environment where there was no liquid money market on which they could trade the short term debt they held on their balance sheets and bank were not accustomed to treating bills made liquid by central bank support as a special class of asset. For this reason, when scholars of U.S. banking claim that the Federal Reserve Act epitomized the real bills approach to banking, they are in effect wresting the definition of real bills from the British tradition where it was developed and redefining a U.S. version of real bills based on the Act. This reconfiguration of real bills will be discussed at length in Section IV below.

⁹⁶ FEDERAL RESERVE, *supra* note 43, at 35; MELTZER, *supra* note 36, at 246.

⁹⁷ STRONG, *supra* note 34, at 183 -84

Overall, the U.S. financial system in the 1920s was built on a hybrid model that drew liquidity from both a central bank and, in direct contravention to the real bills principle, from capital markets. The components of this model are as follows:

- (i) Capital markets: A system of callable debt secured by stock market collateral,
- (ii) A central bank that supports member banks directly by
 - a. standing ready to provide high powered money to member banks on an “as needed” basis in exchange for single name paper, and
 - b. monitors and examines the financial condition of member banks, and
- (iii) Banks that nominally followed the real bills principle, but in practice used other risk management techniques.

C. The Federal Reserve’s destabilizing liquidity

In the 1920s the U.S. banking system had two independent sources of liquidity: capital markets, the traditional repository of reserves in the U.S., and the new central bank. The Federal Reserve Act aspired to have the central bank replace the call loan market and to establish a money market in two name paper, but did not actively repress the call loan market.

A real bills based analysis indicates that this was a mistake. The U.S. banking system was adapted to the limited liquidity that capital markets could provide. When the Federal Reserve was founded, the banking system gained access to central bank liquidity, or a source of funds that could keep asset prices on the money market from falling. The real bills principle indicates that the extension of this type of liquidity to finance bills can be destabilizing. Furthermore, call loans fall into the most dangerous category of finance bill: asset finance. Providing monetary liquidity to asset finance loans has a long history of resulting in devastating asset price bubbles and crashes, and the real bills principle seeks to prevent this form of instability.

Even though the Federal Reserve Act sought to preclude any flow of central bank liquidity to capital markets by making call loans ineligible for discount, this measure was unsuccessful in achieving its goal, because the Act by permitting single name paper to be discounted at the Fed provided banks with direct rather than intermediated access to central bank liquidity. This meant that the Fed did not have control over the uses of central bank liquidity that was comparable to that exercised by the Bank of England –

and that in the U.S. the banks themselves, and not the central bank, determined where central bank liquidity would flow.

For the purposes of exposition this paper will distinguish between direct central bank liquidity, such as that provided by the Fed, and intermediated central bank liquidity, such as that provided by the Bank of England. Note that intermediated central bank liquidity is intermediated through the money market and is therefore equivalent to money market liquidity.

The Fed did have supervisory authority over member banks, and after the Fed grew concerned about the growth of the call loan market it could in theory restrict access to the discount window for banks that reported an increase in call loans. Withdrawing access to the discount window was, however, a draconian penalty and it is not clear that the Fed was prepared to use it. Even so, it seems that the Fed may have exerted some discipline through such channels, as direct bank lending on the call loan market did not increase in 1928 and 1929.⁹⁸ On the other hand, there was also a Congressional investigation of these loans in February 1928, so it is unclear whether the Fed alone was able to exercise much authority over the banks' behavior.⁹⁹

More importantly, the money center banks circumvented efforts to shrink the call loan market by offering customers and foreign banks accounts that placed funds in the market. As a result of this new flow, the funds invested in the market grew rapidly, and, as was noted above, the market grew dramatically over the 1920s. This growth did not stop until the 1929 crash – at which point the interest rate spread which had attracted funds collapsed and the non-bank money exited the market.

Given the real bills proscription against monetary finance of long-term assets, modern empirical work confirming the existence of the feedback loops that so concerned real bills advocates,¹⁰⁰ and the pattern of call loans flowing into the U.S. stock market in the 1920s, it seems likely that the 1929 boom and crash was a monetary phenomenon. Whereas the illiquidity of the U.S. banking system prior to the establishment of the Federal Reserve limited the funds that could finance the purchase of stocks through the call loan market, after the Federal Reserve was established the liquidity of the U.S. banking system changed entirely, but its relationship with the call loan market did not. As a result of this structure, growth in bank liabilities was able to flow into the call market, affecting stock prices.

⁹⁸ Direct bank lending on the call loan market did increase by 72% from September 1924 to December 1927. FEDERAL RESERVE, BANKING AND MONETARY STATISTICS 1914-1941 at 494 (1943). Meltzer, *supra* note 36, at 224 ff.

⁹⁹ Brokers' Loans: Hearing on S. Res. 113 Before the S. Comm. on Banking and Currency, 70th Cong. (1928).

¹⁰⁰ Tobias Adrian & Hyun Song Shin, *Liquidity and Leverage*, 19 J. FIN. INTERMEDIATION 418 (2010).

No claim is made here that the stock market bubble had direct effects on aggregate economic performance. Instead, this paper claims that the stock market bubble was evidence of structural instability in the U.S. financial system. The banking failures of the 1930s, which did adversely affect economic performance, can also be viewed as evidence of this structural instability. Even so, this paper does not dispute that policymakers at the time could have performed better had they recognized that the extraordinary circumstances which confronted them merited a shift to a more pro-active theoretic framework.

While the tools available to the Fed were inadequate to address the problem of asset finance that was sustained by central bank liquidity and fed a stock market price bubble, these tools were adequate to deal with the types of finance bills that under the real bills approach to banking were considered less dangerous: common finance bills and capital finance bills. In the U.S. where the line between common and capital finance bills was not as carefully drawn as it was in Britain, the growth in finance bills that was associated with the business cycle had to be moderated by the policy rate, accompanied by open market sales to make the policy rate effective when necessary.¹⁰¹

Because the Fed was limited in its tools, in the late 1920s controversy arose within the Federal Reserve over whether the policy rate should be used to moderate the growth of asset finance in addition to common and capital finance. In particular some policy makers argued that the Fed had held the policy rate too low in the mid-1920s stimulating the dramatic rise in stock prices and that it should have raised the rate, not due to circumstances in the real economy, but for the purpose of squelching the growth in asset prices.¹⁰² Remarkably this controversy is being revisited today, as economists debate whether the low interest rates of the early 2000s stimulated the dramatic rise in house prices and whether the Fed should have “leaned against the wind,” raising interest rates not due to circumstances in the real economy, but to offset the growth in asset prices.

In both the 1920s and the 2000s we see that the U.S. model of central bank liquidity leaves the central bank without the policy tools that it needs to both support economic activity and avert financial instability. In the 1930s policy makers concluded that structural reform of the financial system was necessary to give the central bank sufficient control over the use of central bank liquidity to allow the Fed to achieve both goals. Part III of this series will discuss those reforms and how they were designed to make Federal Reserve policy effective.

¹⁰¹ FEDERAL RESERVE, *supra* note 43, at 10, 33-35.

¹⁰² MELTZER, *supra* note 36, at 138-39.

IV. REAL BILLS: THE AMERICAN CRITIQUE

The real bills approach to banking is unique in that it provides a unifying theoretic framework for both monetary policy and financial stability, as Charles Goodhart aptly observed.¹⁰³ The real bills approach posits that, if a central bank can exercise enough governance over the commercial banking system so that it provides credit mostly to finance real economic activity, then both inflation and asset price bubbles can be avoided.¹⁰⁴ The framework developed, however, out of the unique conditions of the 19th century British financial system and, thus, it was embodied in a very specific rule that the banks that issue monetary liabilities should focus their lending on “real bills” or on the short-term finance of business transactions, and in particular that they should avoid financing the purchase of long-term assets for investment purposes. As was noted above, the real bills approach, as implemented – or the real bills principle – worked mostly by actively discouraging the growth of bills that were not real, that is finance bills, while at the same time permitting such bills to circulate.

The real bills principle has a significant weakness – it provides no guidance to a central bank that faces a deficit of business activity and thus faces an economy that simply is not generating many real bills – and this weakness had devastating effects on the Federal Reserve’s responses to the Great Depression.¹⁰⁵ This paper does not dispute that in abnormal circumstances like the Depression policymakers may be better served by applying a model that views any economic activity that is stimulated by an expansion in the money supply as beneficial. Thus, some criticism of the real bills principle is well-deserved.

On the other hand, a great deal of undeserved opprobrium has also been heaped upon the real bills approach and this section of the paper will focus on critiques of real bills that are inaccurate or excessive. Before entering into this discussion, however, an important point is worth making: criticism of the real bills approach almost always originates with scholars of U.S. banking, and these scholars often err, first, in assuming that difficulties in implementing the real bills approach in the U.S. will be applicable to banking more generally, and, second, in treating the Federal Reserve Act’s definition of eligible bills as equivalent to the definition of real bills. For example, Lloyd Mints’ *History of Banking Theory* is easily the most important work criticizing the real bills approach, and when it was published in 1945, British reviewers immediately remarked that Mints appeared to have developed a theoretic framework that incorporated the defects of the American banking system.¹⁰⁶ The British reviewers’ critique is supported

¹⁰³ Goodhart, *supra* note 3, at 2.

¹⁰⁴ See Calomiris, *supra* note 3, at 174-75.

¹⁰⁵ This argument is made by many authors including Meltzer, *supra* note 36, at 401 and Goodhart, *supra* note 3, at 2. See also EICHENGREEN, *supra* note 5, at 251-53.

¹⁰⁶ J.K. Horsefield, *A History of Banking Theory: Review*, 13 *ECONOMICA* 138, 139-40 (1946) (“[Mr. Mints suggests] that the major portion of short term debt should be eliminated ... and that fractional reserve banking

by the fact that Mints' definition of a real bill incorporates all short-term loans for commercial purposes and thus fails to make the fundamental distinction between a real bill, which originates in a commercial transaction, and a finance bill, which does not.¹⁰⁷ Mints' definition corresponds, needless to say, with the definition of eligible paper in the Federal Reserve Act of 1913.

Yet another common error that arises when discussing real bills must be cleared up before the exposition can turn to critiques of the policy implications of the real bills approach: the real bills doctrine must be distinguished from the real bills principle.

The real bills doctrine claims that an expansion of the money supply cannot be inflationary as long as the banks that issue monetary liabilities lend only on a short-term basis to finance actual commercial transactions.¹⁰⁸ During the first two decades of the 19th century the British pound was not convertible into gold, and inflation demonstrated that the prices at which commercial transactions are executed can increase alongside an increase in the nominal value of real bills, disproving the real bills doctrine. Thus, even in the early years of the 19th century it was recognized that the real bills doctrine overclaims, and for that reason the real bills controversy of that era was resolved by the Banking Act of 1844 which imposed a quantitative limit on the Bank of England's note issues. What very few modern scholars recognize is that, even though the real bills *doctrine* is flatly wrong, all sides of the mid-19th century debate in Britain over the doctrine embraced the real bills *principle* with respect to banking.¹⁰⁹ And more generally it is almost always a mistake to assume that advocates of the real bills principle are also advocates of the real bills doctrine.

The real bills principle posits that the banks which issue liabilities that circulate as money should largely, but not exclusively, restrict their lending to short-term debt that finances actual business transactions. In a

should be suppressed. These are startling conclusions, and the English reader is bound to feel (perhaps complacently, but still justifiably) that to some extent Mr. Mints is impelled towards them by the practical defects of American banking and monetary policy rather than by the rigour of his theoretical analysis.”); E. Victor Morgan, *A History of Banking Theory in Great Britain and the United States: Review*, 56 *ECON. J.* 290, 291-92 (1946) (“[Mr. Mints'] has a set of questions, to which he knows the answers, and his four hundred victims are made to undergo a competitive examination, after which they receive their due measure of praise for 'correct' and blame for 'incorrect' replies. The marks are generally very low, and the examiner's reports extremely acid, but one wonders if this is not due as much to the nature of the test as to the incompetence of the examinees.

“Since he uses this method of approach, we might expect Mr. Mints to present us somewhere with his own model answers, a clear and consistent body of doctrine against which to expose the confusions and contradictions of his predecessors. The statements of his own views are, however, fragmentary and unconvincing. ... To sum up, a great deal of excellent preliminary work has failed to produce a really satisfactory book, owing to poor historical method and a defective theoretical background.”)

¹⁰⁷ MINTS, *supra* note 13, at 9.

¹⁰⁸ Humphrey, *supra* note 13, at 3-4, gives a thorough explanation of the real bills doctrine and why it is wrong.

¹⁰⁹ ROGERS, *supra* note 11, at 232. Note that Adam Smith has drawn enough attention that many have distinguished between the real bills doctrine and the real bills principle in his writing. David Laidler, *Adam Smith as a Monetary Economist*, 14 *CANADIAN J. ECON.* at 196-97 (1981); Rockoff, *supra* note 13, at 318; Humphrey, *supra* note 13, at 8.

financial system where prices are anchored by the gold standard, conforming with this principle will, it is argued, minimize the risk that excessive credit will be issued and that either excessive inflation or an asset price bubble will result.

The tendency to conflate the real bills doctrine with the real bills principle can probably be attributed to Lloyd Mints. Humphrey, to give just one example, claims that the real bills doctrine was a “mainstay” of 20th century U.S. banking theory, citing Mints.¹¹⁰ Mints apparently reaches this conclusion by taking the position that any advocate of the real bills principle is a proponent of the real bills doctrine and is contradicting himself if he also argues that either convertibility of bank obligations or discount rate policy plays a role in limiting the expansion of bank credit.¹¹¹ Needless to say most proponents of real bills from Adam Smith on, have recognized that real bills alone are insufficient to anchor prices and have advocated the real bills principle together with convertibility, discount rate policy, or both.¹¹²

After the widespread tendency to confound the real bills principle with the real bills doctrine is ignored, there are several critiques that can be attributed to the principle and that remain to be addressed. These fall into two categories: critiques of the principle as a form of macroprudential regulation designed to guide central bank policy, and critiques of the principle as a risk management framework for individual banks.

The first category, critiques of the macroprudential aspects of the real bills principle, include the following: that the principle prescribes a procyclical policy rule, that the principle fails to prescribe a policy rule at all, that it cannot be implemented because it draws distinctions between types of credit that cannot in practice be made, and that there is no value to drawing distinctions between different types of credit.

Many have claimed that the real bills principle requires that monetary policy be procyclical, including Ben Bernanke.¹¹³ This claim is based on the misperception that all U.S. bank demand for discounts in the 1920s represented real bills and that therefore the Fed when implementing a real-bills-based monetary policy must lend more in a boom and less in a bust. In fact, of course, proponents of real bills argue that

¹¹⁰ Humphrey, *supra* note 13, at 11, MINTS, *supra* note 13, at 206. See also MELTZER, *supra* note 36, at 729; Timberlake, *supra* note 13, at 333-34.

¹¹¹ MINTS, *supra* note 13, at 207, 208.

¹¹² Adam Smith: Laidler, *supra* note 109, at 196-97; Rockoff, *supra* note 13, at 318. MINTS, *supra* note 13, at 207, 208.

¹¹³ Charles Calomiris & David Wheelock, *Was the Great Depression a Watershed for American Monetary Policy?* In MICHAEL BORDO, CLAUDIA GOLDIN, & EUGENE WHITE ED. THE DEFINING MOMENT 23, 35-36 (1998); MILTON FRIEDMAN & ANNA SCHWARTZ, A MONETARY HISTORY OF THE UNITED STATES 1867 – 1960 at 253 (1963) (arguing that real bills implications for monetary policy were ambiguous, but “pretty clearly implies” procyclicality); MELTZER, *supra* note 36, at 164; Ben Bernanke, *A century of U.S. Central Banking*, 27 J.ECON. PERSPECTIVES 3, 5 (2013).

the reason a boom is unsustainable is because so many finance bills are being issued, and thus that as the boom progresses the central bank should raise interest rates in order to reduce the degree to which central bank liquidity is supporting the use of finance bills. By contrast at the bottom of the business cycle prices have just been falling, and finance paper is at a minimum, so for purposes of the domestic economy its best keep the rate low. For a real bills proponent one of the central bank's basic duties is to constantly evaluate the degree to which the bills that finance commercial activity are real and the degree to which they are finance bills in order to determine whether the economy is overheating.¹¹⁴

The empirical basis for the claim that a real bills approach results in procyclical monetary policy is the actual policy rule used by the Federal Reserve in the 1920s, which Allan Meltzer has dubbed the “Riefler-Burgess” policy rule and which was in fact procyclical.¹¹⁵ Proponents of the real bills principle would respond that it was procyclical, because it failed to implement the principle.

A more fitting criticism of real bills is, thus, the one leveled by Friedman and Schwartz: by relying on the “judgment” of policymakers, it fails to prescribe a policy rule at all.¹¹⁶ In short, even though proponents of real bills believed that monetary policy could and should be used to support the real bills principle, the principle did not provide guidance as to how monetary policy should be implemented. The practical applications of the principle had been developed for the British banking system, and U.S. policymakers struggled to make the principle relevant to the U.S. environment.

Recall that, as was discussed above, an actively traded money market never developed in the U.S. and for this reason the Fed never had the tools to implement the real bills principle on the model of the Bank of England – that is, by exercising control over the assets that circulate in the money market. Because in its early years, Fed policymakers believed that they should put in place a real bills approach, some efforts were made to apply “direct action” at the discount window, but they found that attempting to impose this one-on-one with member banks was ineffective.¹¹⁷ The tools that were available to the Fed were either draconian – denying a member bank access to the discount window – or affected economic activity generally and did not discriminate between real bills and finance bills – the policy rate and open market operations. Much of the American criticism of the real bills approach reviews the inability of the Federal Reserve to implement the approach effectively, and argues on this basis that the approach itself is flawed.

¹¹⁴ See, e.g., FEDERAL RESERVE, *supra* note 43, at 5, 10, 34-35.

¹¹⁵ MELTZER, *supra* note 36, at 164, 398; Calomiris, *supra* note 3, at 174, 177.

¹¹⁶ FRIEDMAN & SCHWARTZ, *supra* note 113, at 253.

¹¹⁷ MELTZER, *supra* note 36, at 246-47. Note that Timberlake argues that the policy of direct action was so effective that the Fed actually closed the discount window to many banks that were suspected of supporting the call loan market. It is not clear, however, where Clark Warburton, whom Timberlake cites as a source, obtained his information. Richard Timberlake, *supra* note 13, 341-42.

A common theme is thus that a central bank has no means of distinguishing between different types of credit, so that a governing principle based on such distinctions is nonsensical.¹¹⁸ This lack of a policy tool, however, characterized the Federal Reserve, not central banks generally. In particular, central banks that discounted instruments that circulated in a two-name paper market could easily exclude an instrument – for example American finance paper in 1906 – from that market by refusing to discount such paper.

A similar complaint is based on the “fact” that money is fungible and therefore any loan from the Fed can be used to support activities that the Federal Reserve Act – and the real bills principle – discourage.¹¹⁹ But this “fact” is a function of the structure of the banking system, and in particular was created by the fact that the Federal Reserve stood ready to discount single name paper. In a two-name paper market the member bank’s ability to issue paper that is discountable at the central bank depends either on its ability to obtain a guarantee from another bank or on the willingness of another bank to hold its paper. In such a market a member bank that engaged in activities that were being discouraged by the central bank and that therefore had higher risk than other banks of losing access to the discount window would lose the support of its fellow banks, and would find that its paper was not marketable. In short, in a two name paper market a central bank “hint” can be a force that alters the behavior of all the banks. On the other hand, the Fed had been given the power to discount single name paper, because the U.S. banking system had developed in an environment without a central bank and without an active market in two name paper. In this chicken-and-egg problem, the authority of the Fed ended up being the loser.

In addition to these arguments that the central bank cannot discriminate between different types of bank credit, some have argued that there is no point to such discrimination: the distinction between real and finance paper is unnecessary and “invalid,” because control of the aggregate quantity of outstanding bank credit is sufficient to restrain inflation.¹²⁰ Modern theory, however, demonstrates that money can play different roles in the economy depending on where it originates, and thus that there may well be a role for drawing distinctions that are more discriminating than the aggregate quantity of money.¹²¹ Furthermore, the real bills principle sought to reduce the risk, not just of inflation, but also of financial instability, and, as will be discussed in Section V, the recent financial crisis and recognition of a need for macro-

¹¹⁸ For example, Meltzer cites Benjamin Strong as criticizing “the real bills doctrine as a guide to policy,” because “the Federal Reserve could control the quantity of credit, not the type of credit outstanding.” MELTZER, *supra* note 36, at 185. This, however, confuses a description of the inability of the U.S. banking system to implement the real bills principle with a critique of the principle itself. Similarly Meltzer claims that Thornton favored control of the quantity of paper credit rather than its quality. *Id.* at 29 -30. In fact, Thornton states directly that a banker’s job is to “discourage bills of accommodation.” THORNTON, *supra* note 33, at 176.

¹¹⁹ Moulton, *supra* note 62. See also the discussion in Section III.B above.

¹²⁰ FRIEDMAN & SCHWARTZ, *supra* note 113, at 267 .

¹²¹ Stephen Williamson & Randall Wright, *New Monetarist Economics: Methods*, 92 FED. RESERVE BANK OF ST. LOUIS REV. 265, 293-94 (2010).

prudential regulation indicate that modern policymakers have begun to realize the importance of a more detailed understanding and control of the different types of bank credit.

Just as the value of real bills as a guide to central bank policy has been questioned, so too has its value as a governing rule for the behavior of individual banks. In particular, many have argued that, contrary to the claims of the proponents of the real bills principle, these loans are not a better source of liquidity for a bank in difficult times than collateralized loans.¹²² Once again, framing the issue as a question of whether the real bills principle is right or wrong is misleading. The right question to ask is whether the real bills principle makes sense in a particular institutional framework.

In the context of the American banking system prior to the founding of the Fed, bills were illiquid, because there was no active market in them, they had to be held on a bank's balance sheet, and they were often rolled over in order to maintain relationships with bank customers;¹²³ by contrast, the stock market was actively traded and thus call loans were in fact a better source of liquidity for banks than bills. Thus, a "shiftability" theory of liquidity was promoted which claimed that "it is undoubtedly true that ... collateral loans" are a better source of liquidity than bills, and that bonds and stocks are the most liquid form of collateral – at least in ordinary times.¹²⁴ Shiftability refers to the ease with which an asset can be transferred to someone else in the financial system, and thus in modern terminology refers to market liquidity.¹²⁵ The superiority of collateral over unsecured bills was, however, simply a function of the U.S. institutional setting where the actively-traded money market was small and limited to the commercial paper of established issuers.

At the same time, shiftability theorists recognized that the crisis of 1907 demonstrated conclusively that collateral loans could not function as a source of liquidity in a crisis.¹²⁶ Shiftability theory argues that in a crisis, no asset is liquid for the market as a whole, unless there is a central bank that makes it liquid, and "if there is a central bank, anything will be liquid which [the central bank] is willing and able to purchase or accept as collateral."¹²⁷

The founding of the Federal Reserve did not, however, solve the problem of structural illiquidity in U.S. money markets. Instead of the Federal Reserve acting as a backstop to an active money market as the Bank of England did in Britain, the Federal Reserve provided a non-market source of liquidity to commercial bills in the U.S. and no active trading market in this short-term debt ever developed. Thus,

¹²² MINTS, *supra* note 13, at 265; H.G. Moulton, *supra* note 62, at 723 (1918).

¹²³ On the latter, see Moulton, *supra* note 62, at 638, 718, MINTS, *supra* note 13, at 218.

¹²⁴ Moulton, *supra* note 62, at 713, 730 (1918).

¹²⁵ *Id.* at 723.

¹²⁶ Moulton, *supra* note 62, at 725-26.

¹²⁷ MINTS, *supra* note 13, at 265; Moulton, *supra* note 62, at 723.

the theory of central banking that was derived from the U.S. experience saw the central bank as a buyer of last resort of certain assets that the central bank wanted to make liquid.

The British theory of central banking and the real bills principle, by contrast, viewed the central bank as the backstop for the money market, where two-name unsecured commercial bills were actively traded.¹²⁸ While it is, of course, true that this market could not survive a crisis without the backstop of the central bank, the British institutional structure allowed not just the central bank, but the reliable function of the market itself to serve as a source of liquidity in a crisis. Because the banks that traded on and provided guarantees on this market trusted in its reliability, they too could act as sources of liquidity in a crisis. It is clear that this extraordinarily strong institutional core played an important role in the renowned stability of the British financial system for the century subsequent to 1866.

Thus, the American rejection of the real bills principle in favor of loans collateralized by long-term assets was due to the fact that the institutional structure of U.S. banking was not hospitable to the real bills principle, and was not due to some intrinsic failing of the real bills principle itself.

American theorists did not just argue that collateralized loans provided superior liquidity to short-term unsecured bills, but that the distinction between real bills and finance bills – which they and many contemporary Americans often called speculative bills – was spurious. Recall that finance bills are distinguished from real bills by the fact that they support longer-term activity and thus entail greater uncertainty and greater risk.¹²⁹ Thus, critics of real bills advocated long-term lending by banks, even as they implied that the prejudice against “speculation,” which is often set off by scare quotes, was based on moral grounds.¹³⁰

Once again the arguments in favor of long-term bank lending draw heavily on the American experience and its illiquid money markets. Theorists found that U.S. banks had a history of providing long-term finance by rolling over short-term bills, and viewed the real bills principle as an unnecessary constraint on economic activity.¹³¹ Indeed, using the Depression as a reference point it was argued that short-term debt held by banks was particularly dangerous in a crisis because it gave banks the ability to force economically destructive liquidation of business.¹³² Of course, because these theorists were unfamiliar

¹²⁸ Note that this structure addresses the problem of the bank being incentivized to roll over short-term debt to placate customers. Here the debt is owed to the holder of the paper and the bank is not in a simple one-on-one relationship with the customer. Moulton notes that in the U.S. it was bank customers who objected to the practice of rediscounting their notes. Moulton, *supra* note 62, at 710.

¹²⁹ Hawtrey expressed this point very clearly in HAWTREY, *supra* note 39, at 193

¹³⁰ See, e.g., FRIEDMAN & SCHWARTZ, *supra* note 113, at 192. See also MELTZER, *supra* note 36, at 55, 400.

¹³¹ Moulton, *supra* note 62, at 638, 718.

¹³² MINTS, *supra* note 13, at 216, 218-222, 264-66.

with the structure of British markets and apparently could not conceive of the value of an active market in two-name paper and of the monetary finance that such a market could provide, they did not even consider the type of liquidity that the real bills principle was designed to protect, money market liquidity. In short, the true implication of these critiques is that the real bills principle was ill-suited to the U.S. environment.

As was explained in Section I, the real bills principle is the tool by which traditional banking theory discourages trade on the central bank-supported money market in capital finance bills, and proscribes such trade in asset finance bills. The purpose of this principle is two-fold: to avoid the financial instability created by asset price bubbles and to control inflation. In this macro-prudential framework, there are two keys to ensuring that the money market liquidity that is supported by the central bank has net benefits for the economy: first, the tendency of finance paper to grow over the course of a boom as common finance bills morph into capital finance bills must be controlled, and, second, money market liquidity must be prevented from affecting the price formation process on capital markets. Given this theoretic framework, the next section demonstrates that a British central banker who was alive at the turn of the 20th century would not have been surprised by the crisis of 2007-08, because it was a textbook case of the dangers created when central bank supported money markets are used to finance long term assets.

V. TRADITIONAL BANKING THEORY AS A GUIDE TO STRUCTURAL REFORM

Since the recent financial crisis, the question of how to avoid future government bailouts of the banking system has been a topic of significant discussion. The traditional banking theory that is explained in this paper points to an answer that is rarely considered: It may be possible to restructure the role of the central bank in the banking system – or more precisely to restructure the relationship between banks, markets, and the central bank – so as to foster the credit needed for economic activity without the adverse consequence of also supporting asset price bubbles. In short, traditional banking theory provides a framework for restructuring the banking system to increase its stability.

Section I explained that market liquidity without price risk, or money market liquidity, can be derived from the structured interaction of banks, the discount market, and the central bank. In this model, the central bank role includes monitoring the money market for credit growth that is faster than the increase in economic activity and excluding assets from the money market if there is reason to believe that they are being used for capital finance or asset finance. Thus, traditional banking limits the destabilizing effects of the backstop provided by the central bank by using the policy rate to control the growth of credit generally and by placing a firewall between central bank credit and long-term assets. The control that the central bank exercises over the money market is designed to prevent both inflation and the financial instability that can result from asset price bubbles.

There are two important distinctions between the traditional banking theory model and the model of central banking that developed in the early 20th century U.S. First, in the U.S. the central bank stood ready to lend on the basis of an asset that the borrowing bank had originated itself, instead of supporting a money market where the asset would be discounted by a different bank than the bank that originated – and provided an initial guarantee of payment on – the asset. Because of this difference in the manner in which central banks provided liquidity, market participants played a much smaller role in monitoring the market in the U.S. than they did under traditional banking theory. Second, in the U.S. the call loan market functioned as a source of liquidity for the banking system that was an alternative to the central bank. Unlike Britain where the money market and the central bank functioned together to provide a single source of liquidity for the banking system, the U.S. banking system in the first two decades of the Federal Reserve's existence had two sources of liquidity, the call loan market and the central bank itself, and it was possible for liquidity to flow from one to the other. Both of these characteristics are also present in the modern U.S. financial system.

This section of the paper first discusses the empirical evidence that supports the view that the traditional banking framework is more stable and supportive of economic growth than the current banking structure. The discussion then turns in subsection B to the current U.S. financial system including, first, its reliance for liquidity on both guarantees, such as those upon which traditional banking theory is built, and the collateral of long-term assets, and, second, the proposed use of macro-prudential regulation to control these forms of liquidity. Subsection C evaluates important themes in the current literature on financial instability, including the “shortage” of safe assets, informationally insensitive assets, proposals for narrow banking, and proposals for a “dealer” of last resort, in light of traditional banking theory.

A. Traditional banking limits the severity of crises: Empirical evidence

The empirical research that studies long-run banking data analyzes crises across several now-developed countries. This data has relevance for a comparison of the traditional banking structure with the current banking structure, because of the role played by Britain in the international monetary system through approximately 1930 and the role played by the U.S. in the decades that followed. In short, both of these banking systems, in their time, served as an important anchor of the global monetary system, so it is not unreasonable to treat the global data as, at a minimum, an indicator of the effectiveness of the different approaches to producing financial stability and growth.

This empirical research demonstrates that the London discount market, which traded in short term bills that financed global economic activity, was indeed a remarkably stable source of liquidity. When Moritz Schularick and Alan Taylor analyze the quantitative history of money and credit across twelve countries,

they find that there are “two eras of finance capitalism.” The first era runs from the start of their data to 1939 and the second from 1945 to the present. The first era is characterized by the fact that credit grew at the same rate as GDP, and the second is characterized by credit growth that greatly exceeds GDP growth.¹³³ Crisis dynamics in the two eras are even more interesting. In the first era the real effect on output and investment due to crises was relatively small – until the great crisis of the 1930s, whereas in the second era the real effects due to crises were markedly greater than in the pre-1930 crises.¹³⁴

The conclusion that Schularick and Taylor draw from their careful empirical study is that “credit growth turns out to be highly significant as a predictor of financial crises,” and the “virtually uninterrupted growth of leverage we have seen up until 2008 ... [has] led to a greater, not smaller role of credit in the macroeconomy” as “financial crises are ‘credit booms gone wrong.’”¹³⁵

The undisciplined credit growth that generates financial crises was held strictly under control by policymakers who were adherents of traditional banking theory. Schularick and Taylor’s data shows not only that the traditional banking approach was effective in controlling credit growth, but also that – at least until the crises of the 1930s – the effects of these traditional policies were more successful at supporting the real economy through crises than modern policy has been. Note that to the degree that the crisis of the 1930s represented one step in a difficult transition from British monetary leadership to U.S. monetary leadership and from the exemplary implementation of traditional banking theory in Britain to the more hybrid approach of the U.S., it is reasonable to exclude it from the data evaluating the performance of the theory.

B. Successful macro-prudential regulation requires structural reform

The modern U.S. financial structure has many similarities to U.S. financial structure in the early years of the Federal Reserve. Fed support for liquidity-constrained banks is provided directly to the banks against any loans they hold on their balance sheets. Because the Federal Reserve has extremely limited ability to direct the use of central bank credit, in the U.S. it is the banks themselves that determine the assets that will be supported by central bank liquidity. Now, as in the 1920s, banks support a call loan market that is backed by long-term securities, the repurchase agreement (“repo”) market. In addition, certain categories of short-term assets, such as commercial paper, can be used as funding in the money market due to bank guarantees. Repurchase agreements and commercial paper were both essential to the process of creating mortgage-backed securities, and also were used to finance long-term positions in mortgage-backed

¹³³ Moritz Schularick & Alan M. Taylor, *Credit Booms Gone Bust: Monetary Policy, Leverage Cycles and Financial Crises, 1870–2008* at 3 (NBER Working Paper No. 15512, Nov. 2009).

¹³⁴ Schularick & Taylor, *supra* note 133, at 16.

¹³⁵ Schularick & Taylor, *supra* note 133, at 5, 27, 28.

securities.¹³⁶ Of course, all three of these markets, repurchase agreement, commercial paper, and mortgage, lay at the center of the financial crisis of 2007-08. Subsequent to the crisis, the need to constrain this “shadow” banking system, which is financed using repo and commercial paper, has been universally recognized. Currently the tool of choice is macro-prudential regulation. I argue here that one lesson of traditional banking theory and of the historical analysis presented in this paper is that macro-prudential regulation is unlikely to be successful unless it is accompanied by structural reform of the relationship between banks, money markets, and the central bank.

The role played by the banking system – and by the contingent liabilities of commercial banks – in the markets for both repurchase agreements and commercial paper were discussed in Part I of this series. The guarantees and market liquidity that the banks provide to these two instruments are the mechanisms by which the banks convert their access to direct central bank liquidity into liquidity for these “shadow” banking instruments, and thereby conduct the flow of direct central bank liquidity into these markets. In this part of the series, the size of these markets and the role that they played in the financial crisis of 2007-08 is addressed.

The crisis started in the segment of the commercial paper market that is backed by assets, asset-backed commercial paper (“ABCP”). The apparent trigger for the crisis was the declaration by a European bank that market value could not be determined for three bond mutual funds that invested in asset-backed securities including U.S. subprime securities.¹³⁷ This declaration followed just days after a subprime lender that financed its unsold inventory using commercial paper invoked the right to postpone payment on its paper.¹³⁸ Commercial paper investors such as money market funds subsequently decided not to roll over their investments in the asset-backed commercial paper market, and the size of the market collapsed by one-sixth from \$1.2 trillion on August 8 to \$1 trillion on August 29.¹³⁹ The Federal Reserve first attempted to arrest the decline in the ABCP market by “clarifying” discount window policy to permit a bank to secure Federal Reserve credit by pledging ABCP on which the bank itself served as guarantor.¹⁴⁰ Implicitly, the Fed authorized the banks to purchase the commercial paper they guaranteed themselves and, if necessary, to finance the action at the discount window. By mid-September loans extended through the discount window had jumped from zero in early August to \$7 billion.¹⁴¹ The Fed, however, also

¹³⁶ See Zoltan Pozsar, Tobias Adrian, Adam Ashcraft, & Hayley Boesky, Shadow Banking, NYFRB Staff Rep. No. 458 (July 2010).

¹³⁷ FIN. CRISIS INQUIRY COMM’N, FINANCIAL CRISIS INQUIRY REPORT 250 (2011). Note that these funds were able to establish valuation techniques and once again permit withdrawals before the end of August 2007. BNP Paribas, Press Release: BNP Paribas ABS Euribor and BNP Paribas ABS Eonia valuations resumed, Aug. 29, 2007.

¹³⁸ FCIC, *supra* note 137, at 250-51.

¹³⁹ Federal Reserve Commercial Paper Releases.

¹⁴⁰ Greg Ip, *New York Fed Takes Step To Bolster Credit Market*, WALL ST. J., Aug. 27, 2007.

¹⁴¹ Federal Reserve, H.4.1 Release, Sept. 13, 2007.

granted exemptions to the banks allowing them to support the vehicles that issued the commercial paper through their investment bank affiliates,¹⁴² and discount window loans declined again to small sums as the banks apparently avoided discount window borrowing due to the perceived stigma associated with it. As the ABCP market continued to shrink at a rapid pace falling to \$830 billion on December 19, the Federal Reserve initiated the Term Auction Facility, which allowed banks to borrow anonymously from the Fed. This had the effect of dramatically slowing the decline in the market so that \$600 billion was only reached more than a year later in May 2009. (The ABCP market has continued to decline and currently hovers around \$225 billion.)

Tremors in the repo market appeared in March 2008 when Bear Stearns, an investment bank which financed half of its balance sheet using repo, lost the confidence of market participants. Two of the Federal Reserve's 2008 rescue programs the Term Securities Lending Facility and the Primary Dealer Credit Facility were explicitly designed to prevent the repurchase agreement market from collapsing in fire sales. There are no good measures of the extent to which the repo market was used to finance private sector collateral prior to the crisis. We do, however, have data on the extent to which the Federal Reserve extended funds to support the market. By October 1, 2008 the Federal Reserve had through the two rescue programs extended \$380 billion in loans to the investment banks.¹⁴³ This was in addition to the fact that the commercial banks had been granted exemptions to allow them to "provide liquidity to their affiliates for assets typically funded in the tri-party repo market."¹⁴⁴ Thus, it appears that the repo market was being used to finance private sector collateral on a scale comparable to the ABCP market.

Asset-backed commercial paper and repos were both instruments that relied on the liquidity provided by banks to finance long-term assets. While it was trouble in the mortgage market that triggered the near collapse of these two instruments, a wide variety of long-term assets were financed by them including corporate and government bonds.

Observe that traditional banking theory predicts that providing money market liquidity to long-term assets will have two effects: it will adversely affect the price discovery process on the markets for long-term assets, and the resulting asset price bubbles will be destabilizing to the monetary system. This, after all, was the purpose of the structural separation of the two markets.

¹⁴² See letters from Federal Reserve Board to general counsels of Bank of America, Citibank, and J.P. Morgan Chase Bank dated August 20, 2007. Additional letters were issued to DeutscheBank, Barclays, and RBS on Sept. 12, Oct. 11, and Oct. 12 respectively. See also Saule Omarova, *From Gramm-Leach-Bliley to Dodd-Frank: The Unfulfilled Promise of Section 23A of the Federal Reserve Act*, 89 N.C. L. REV. 101, 107-08 (2011).

¹⁴³ Federal Reserve, H.4.1 Release, October 2, 2008.

¹⁴⁴ Federal Reserve Board of Governors, Press Release, Sept. 14, 2008.

Once again the empirical evidence indicates that traditional banking theory is the correct framework for understanding the consequences of using the liquidity provided by the banking system to finance long-term assets. Jorda, Schularick, and Taylor (2014b) focus on mortgage markets and study economic and bank data for 14 countries across a span of 140 years. They find, first, that bank lending on mortgage markets has grown dramatically over time and, second, that “mortgage and house price booms are predictive of future financial crises, and this effect has also become much more dramatic since WW2.”¹⁴⁵

This data provides support for the view that, when the financial system is structured so that the banking system provides liquidity to long-term assets, feedback loops arise so that the demand of long-term investors no longer limits the capacity for asset prices to rise. As prices become unhinged from long-term determinants, the initial effect is an economic boom, but the long-run effect is a crash. Indeed, this is precisely the description that the earliest traditional banking theorists give when they discuss the collapse of the monetary system in early 18th century France.¹⁴⁶ A banking system that relies on collateral to establish value without constraining the amount of credit that is issued against that collateral is simply duplicating the error of John Law.¹⁴⁷ It is, therefore, unsurprising that the dangers of our current path are supported by the long-run empirical evidence on credit, asset prices and financial crises.

In response to the recent financial crisis and to studies such as Jorda, Schularick and Taylor’s, every central bank has now adopted some form of macro-prudential regulation. As the Financial Stability Forum explained, a macro-prudential approach requires the adoption of “a system-wide view by examining how ... [bank] risk management techniques affect behaviour and interactions between both financial institutions *and* financial markets. . . . One objective of this set of proposals would be to dampen the procyclicality of market liquidity and asset prices, thereby helping to achieve the macroprudential objective of reducing procyclicality throughout the broad financial system (both institutions and markets).”¹⁴⁸ Indeed, such macro-prudential regulation has been embraced as the solution to our current concerns over financial stability to such a degree that Jorda, Schularick and Taylor themselves prescribe this treatment.¹⁴⁹

¹⁴⁵ Oscar Jorda, Moritz Schularick & Alan Taylor, *Betting the House* 37 (NBER working paper no. w20771, Dec. 2014). Note that in their data on crises, there are four crises in the U.S. prior to the 1929 crisis and only one in Britain. Schularick & Taylor, *supra* note 133, at 30.

¹⁴⁶ THORNTON, *supra* note 33, at 239n, 250-52n.

¹⁴⁷ As Thornton wrote, Law’s error was to consider “security as everything, and quantity as nothing.” THORNTON, *supra* note 33, at 341.

¹⁴⁸ Joint Working Group of the Financial Stability Forum and the Committee on the Global Financial System, *The Role of Valuation and Leverage in Procyclicality* 14 (CGFS Papers No. 34, Apr. 2009) (emphasis in original).

¹⁴⁹ Oscar Jorda, Moritz Schularick & Alan Taylor, *supra* note 145, at 38.

Imagine how astounded an early 20th century Director of the Bank of England would be to learn that a century after his own service, knowledge of central banking had retrogressed to such a degree that policymakers lacked a model of the interaction between banks and markets, were surprised that adverse interactions between banks and markets required control, were just learning that providing bank-based liquidity to long-term assets would “procyclically” generate asset price bubbles, were struggling to find a “macro-prudential” approach that would temper such phenomena, and even believed that such regulation was “completely new . . . a key missing element of the tool kit available to the authorities before the crisis.”¹⁵⁰ The real bills principle was, after all, clearly a macro-prudential approach.¹⁵¹ This Director’s advice would likely be simple: as long as the banks that issue monetary liabilities and are supported by a central bank in turn support markets in short-term debt that is collateralized by long-term assets either as dealers or by extending guarantees to the short-term debt, asset price bubbles and financial instability will be the result. Therefore, structural reform of the financial system to limit interaction between the banking system and capital markets is necessary. After such reform has been put in place, macro-prudential policy, similar to real bills principle-based efforts to constrain the use of monetary finance to fund long-term endeavors, could be effective.

C. Evaluating the financial stability literature through a traditional banking theory lens

This subsection reevaluates four major themes of the financial stability literature using the traditional banking framework. First, the idea that the modern financial system suffers from a shortage of safe assets takes on new meaning when one understands that one job of the financial system is to use the structured interaction of banks, markets, and the central bank to create privately-issued safe assets. Second, the British model of safe private-sector assets created by this structured interaction contrasts with the American model of “informationally insensitive” assets. Third, traditional banking theory explains why proposals that the Federal Reserve act as a “dealer of last resort” are likely to have a destabilizing rather than a stabilizing effect on the financial system. Finally, proposals for narrow banking are seen to be clearly infeasible when one understands the importance of monetary finance – or of using an increase in the money supply as a source of funds for short-term lending that addresses liquidity constraints – to economic activity.

Traditional banking theory emphasizes that even though the central bank is necessary for money market liquidity to exist, the central bank alone cannot create money market liquidity.¹⁵² Equally important to

¹⁵⁰ Paul Tucker, Interviewed by Schoenholtz & Cecchetti, Mar. 4, 2015, available at <http://www.moneyandbanking.com/commentary/2015/3/4/interview-with-paul-mw-tucker>.

¹⁵¹ See C.A.E. Goodhart, *supra* note 3, at 2.

¹⁵² Warburg, *supra* note 6; Strong testimony at Stabilization Hearings, *supra* note 6, at 432.

money market liquidity, and thus to the creation of private sector assets that do not have price risk, is the structured interaction of banks and markets that ensures both that “good” money market debt is originated, and that, even when a business defaults, payment on the money market debt of that business is guaranteed by other money market participants. In short, the purpose of the money market is to use the structured interaction of banks, the market, and the central bank to create private sector assets that have little or no credit risk, or in other words to issue safe private-sector assets.

While it is true that even in a traditional banking framework the central bank makes it possible for the private-sector to issue safe assets, the central bank’s role is a supporting role, backstopping and supervising the market itself without providing a direct guarantee to each of the assets that trade on the market. Indeed, it is very important that far from providing a guarantee of all the assets that trade on the money market, the central bank threatens to exclude privately issued money market assets that fail to meet the standards of the central bank. As long as the central bank communicates clearly to the market what those standards are and how they change over time, central bank discipline strengthens rather than weakens the market.

As was discussed in Section IV, a very different theory of central bank liquidity grew out of the experience of banking in the U.S. The Fed has always provided liquidity directly to member banks against the collateral that the borrowing bank itself originated. In a financial system with this structure, there is no money market liquidity (or intermediated central bank liquidity), but only direct central bank liquidity. Because the central bank’s liquidity is not intermediated, it has very little control over the use of that liquidity, and the banks that are able to borrow from the central bank can determine where this liquidity flows. This is the type of liquidity that in the view of modern theory is provided by the central bank.

Another important characteristic of the money market modeled by traditional banking theory is that money market assets were protected by overlapping guarantees of payment, not collateral. While it is true that some categories of money market assets were collateralized by goods, many were not. And in all cases, the security that arose due to the guarantee of the asset by a “good name” was considered superior to collateral.

This traditional banking view of safe assets and how they are generated sheds light on what is commonly called the shortage of safe assets.¹⁵³ Our prolonged experience with extraordinarily low interest rates can be viewed as evidence that the demand for safe investment opportunities far exceeds their supply. Indeed,

¹⁵³ On the shortage of safe assets, see Ricardo Caballero, *Understanding the global turmoil: It’s the general equilibrium, stupid*, VOXEU, May 21, 2010; Bradford DeLong, *The Flight to Quality*, Grasping Reality Blog, May 27, 2010; Narayana Kocherlakota, speech, 61st Annual Management Conference of the University of Chicago Booth School of Business, May 17, 2013.

it is hard to dispute that the supply of safe assets has fallen, due a loss of confidence in the private sector's ability to originate safe assets subsequent to losses on many highly rated structured products during and after the crisis.¹⁵⁴ On the other hand, the glut of demand for safe assets and the conundrum of low interest rates predates the crisis, and has also been described as having induced the financial sector to create the pseudo-safe assets that blew up in the crisis.¹⁵⁵

Traditional banking theory can help explain why there is a savings glut.¹⁵⁶ For centuries our monetary system has relied on the ability of banks to borrow on an unsecured basis and then to extend that unsecured debt to other borrowers. Since the early years of the current millennium, however, the cash assets of institutions, including asset managers and corporate treasuries have been moved away from unsecured bank and money market liabilities and into collateralized short term liabilities, such as repurchase agreements.¹⁵⁷ In addition, since the crisis unsecured interbank lending markets, such as the Federal Funds and London markets, have collapsed and been replaced with collateralized debt.¹⁵⁸ The shift towards collateralized instruments by both institutional cash pools and the financial institutions themselves surely plays an important role in explaining the glut of demand for safe assets, and the extraordinarily low interest rates that have become common.

Related to the safe assets discussion, is Gary Gorton's theory that banks issue informationally insensitive assets. This theory takes as its model of banking the experience of the U.S. in the late 19th century, where there was no discount market and bank liabilities were backed by "single name" paper. Thus, Gorton starts with the assumption that "the private sector cannot create riskless assets," even though it is private sector liabilities in the form of bank notes and bank deposits that circulate as money and therefore trade at par.¹⁵⁹ Thus, Gorton models banking as an economically efficient confidence trick. Bank liabilities circulate at par because their economic function is to hide the true value of the underlying assets – that is, because the costs of determining their true value are greater than the economic costs of trading them at

¹⁵⁴ Whole classes of structured products, such as leveraged super senior CDOs, SIVs, ABS CDOs, and CPDOs blew up, exposing how little information was conveyed by the high ratings these products received. For the time being sovereign debt issued by members of the Eurozone continues to be a fairly safe asset – due to support from the European Central Bank.

¹⁵⁵ MARTIN WOLF, *SHIFTS AND SHOCKS* 172 (2014).

¹⁵⁶ Of course, this explanation is in addition to, not in conflict with, the fact that emerging markets sought to accrue dollar savings subsequent to the Asian crisis of the late 1990s.

¹⁵⁷ Zoltan Pozsar, *Institutional Cash Pools and the Triffin Dilemma of the U.S. Banking System* (IMF Working Paper WP/11/190, Aug. 2011).

¹⁵⁸ The degree to which regulators failed to understand the macro-economic value of a banking system built on unsecured debt and pushed the banking system to collateralize interbank exposures played a role in this problem is a topic for a different paper.

¹⁵⁹ Gary Gorton, *The Development of Opacity in U.S. Banking* 1 (NBER Working Paper no. 19540, Oct. 2013). See also Bengt Holmstrom, *Understanding the Role of Debt in the Financial System* 3 (BIS Working Paper No. 479, Jan. 2015).

par. In Gorton's terminology bank liabilities are informationally insensitive. Crises take place when the assets underlying the bank liabilities have deteriorated to such a degree that it is worthwhile to produce information about the underlying assets. In short, Gorton's theory is that the essential function of banks is to be opaque, so that most of the time it is less costly to trade bank liabilities at par than to determine their true value.¹⁶⁰

The concept of information insensitivity requires, however, that some producers are able to borrow even though their collateral is valueless. To explain how genuine economic value is generated from the ability to borrow against valueless collateral, Gorton in a paper coauthored with Guillermo Ordonez assumes, first, that the agents in their model have a production opportunity that cannot be funded on an unsecured basis, and, second, that these agents will repay a loan secured by valueless collateral. Thus, the economic benefit gained from "informationally insensitive" bank liabilities is that by masking the true value of the underlying collateral, banks make it possible for agents who would not otherwise be able to borrow and to produce, to do so.

Traditional banking theory's rebuttal to this argument is that much more economic value can be achieved by structuring the banking system so that (i) agents are able to borrow on an unsecured basis and are incentivized by long-term relationships with their banker to borrow wisely, and thus that (ii) bank liabilities are not just informationally insensitive, but genuinely as risk-free as government debt. The London money market was designed to create just such a system.

When the broader literature on safe assets explores the possibility of privately issued safe assets, the conclusion is typically that financial crises, though rare, involve so many failures that it is impossible for the private sector to manage their risk, and thus the government or central bank must step in to provide tail risk insurance.¹⁶¹ This literature generally argues that the lender of last resort function of the central bank is just one version of such tail risk insurance, and that in the modern world where markets provide a great deal of the funding for economic activity, the central bank should step in as a "dealer of last resort," buying and selling private sector assets in order to support their values. Indeed, some authors have gone so far as to claim that traditional banking theorists, such as Walter Bagehot, would support this view.¹⁶²

In fact, traditional banking theory, even as it acknowledges the need for a central bank backstop of the money market, expressly rejects the idea that the central banks should play an active role in markets for

¹⁶⁰ Gary Gorton & Guillermo Ordonez, *Collateral Crises*, 104 AM. ECON. REV. 343 (2014).

¹⁶¹ See, e.g., Willem Buiter, *Central Banks and Financial Crises* (Paper presented at Kansas City Federal Reserve Bank Symposium 2008); Caballero, *supra* note 153; PERRY MEHLING, *THE NEW LOMBARD STREET* 132-35 (2011); Perry Mehrling, Zoltan Pozsar, James Sweeney, and Daniel Neilson, *Bagehot was a Shadow Banker* (SSRN Nov. 2013).

¹⁶² Mehrling et al., *supra* note 161.

privately issued long-term debt. Indeed, traditional banking theorists believed that financial stability required that the central bank closely monitor and limit the types of assets that are eligible for central bank support: this control over the assets that are supported by central bank liquidity is, in fact, the essence of the real bills approach to banking. Indeed, because the money market liquidity supported by the central bank eliminates price risk, it was understood that central bank support of long-term assets would have a distortionary effect on their prices and thus promote asset price bubbles and financial instability.

Thus, traditional banking theory's riposte to the strand of the modern literature that advocates increased central bank participation in long-term asset markets as a solution to the problem of asset price crashes (without ever explaining how the fundamental problem of pricing the assets or the tail risk insurance could possibly be solved) is that this "solution" gets cause and effect precisely backwards. Traditional banking theory points out that the reason there is an asset price bubble that can crash is because central bank support of banks and the money market has been allowed to leak into long-term markets supporting an unsustainable feedback loop and bubble. According to traditional banking theory, a central bank that seeks to avoid supporting the growth of asset price bubbles must monitor and control the degree to which bank credit and money market instruments are being used to finance long-term assets.

Another strand of the modern literature regards the extraordinary support that central banks were forced to provide to the private sector as clear evidence that the modern financial system is characterized by privatized gains and socialized losses, and that this aberration demands structural reform.¹⁶³ Many argue that the best way to stabilize the financial system is to separate the payments system – or bank liabilities that circulate as money – from the lending process. Because under such a system bank deposits are no longer used to make loans, such proposals are often called 100% reserve or narrow banking. Under these proposals all loans are financed by equity stakes.

From the perspective of traditional banking theory these proposals are hard to understand. Traditional banking theory assumes that monetary finance is an important source of the funds available for lending on the money market, whereas narrow banking proposals seek to eliminate monetary finance entirely. Thus, these proposals appear to assume that any economic benefits created by the early 20th century British banking system when it used the structured interaction of banks, the discount market, and the central bank to create money market liquidity and safe private-sector assets were relatively small, and would be

¹⁶³ LAURENCE KOTLIKOFF, JIMMY STEWART IS DEAD: ENDING THE WORLD'S ONGOING FINANCIAL PLAGUE WITH LIMITED PURPOSE BANKING (2010); Morgan Ricks, *A Regulatory Design for Monetary Stability*, 65 VAND. L. REV. 1289 (2012); Adam Levitin, *Safe Banking* (SSRN working paper, Jan. 1, 2015); John H. Cochrane, *Toward a Run-Free Financial System*, in ACROSS THE GREAT DIVIDE: NEW PERSPECTIVES ON THE FINANCIAL CRISIS 197 (Martin Neil Bail & John Taylor ed., 2014) (at 5 in online version); Wolf, *supra* note 155, at 234-35.

outweighed by the benefits of avoiding financial crises. The proposals seek to eliminate money market liquidity and replace it with a combination of government issued money and simple market liquidity. In short, these proposals offer a clean separation between safe, publicly-issued assets and risky, privately issued assets.

The whole point of traditional banking theory is, by contrast, to determine how to sustain a system of safe, privately-issued assets. To the degree that it is valuable for an economy to have the ability to generate safe assets, then surely it is worthwhile to follow the path set forth by traditional banking theory and study the mechanisms by which safe assets can be created. In this context narrow banking proposals can be viewed as a decision to give up on the ability to create privately issued safe assets, even if this limits the supply of safe assets and is therefore costly. This paper argues that before choosing this path and declaring the financial system a total failure, we should explore other options for structural reform.

VI. CONCLUSION

Traditional banking theory's approach to macroeconomics has three lessons for modern policy makers:

First, the liquidity that is available in money markets derives from the banking system and more particularly from the guarantees provided by the banks and the central bank. Thus, the so-called shadow banking system should be understood as an offshoot of the banking system.

Second, for this reason the central bank must supervise and exercise control, not just over the banks, but also over the assets that trade on the money market itself. As the financial system is currently structured in the U.S., the Fed has too little control over the use of bank guarantees and market-making to extend central bank liquidity support to markets, such as commercial paper and repo. For this reason, structural reform of the Fed's relationship to banks and money markets is necessary.

Third, central bank liquidity is destabilizing when it flows through banks to money markets and then to long-term assets markets, including both mortgage and capital markets. Thus, the relationship between banks and long-term asset markets, and between money markets and long-term asset markets must be very carefully monitored to avoid monetary flows that have the effect of distorting long-term asset prices. In short, because an important purpose of market liquidity on long-term asset markets is price discovery, capital market liquidity is not supported by, but distorted by, the price-insensitive liquidity that in modern markets is extended by central banks to banks and by banks in turn to money markets.

In the absence of structural reform that obstructs the flow of monetary finance into long-term asset markets, it is not reasonable to expect macro-prudential regulation to be successful. In early 20th century Britain the Bank of England was able to use the real bills principle as a form of macro-prudential

regulation that stabilized the British financial system, but this was in an environment where there were significant structural barriers impeding the flow of monetary finance to capital markets. In the U.S. in the 1920s, by contrast, there were no such structural barriers, and efforts to implement real bills-based macroprudential regulation failed. These two examples indicate that macroprudential regulation will prove inadequate, just as it did a century ago, and that structural reform will be needed before the U.S. financial system can be stabilized.

Part III of this series will explore how the reforms of the 1930s were designed to address the concerns of traditional banking theorists, and draw the outlines of a model for structural reform of the modern financial system.