A Perspective on the Symptoms and Causes of the Financial Crisis

Ricardo Cabral

March 2011
A Perspective on the Symptoms and Causes of the Financial Crisis

Ricardo Cabral
Universidade da Madeira (DGE)
e CEEApIA

Março de 2011
RESUMO/ABSTRACT

A Perspective on the Symptoms and Causes of the Financial Crisis

This article identifies two paradoxes prior to the onset of the financial crisis: banking profits were at historically high levels despite the impending crisis; and though profits were high the profitability of financial intermediation was poor. Using a novel model of banking, it argues that large banks attained high profits through balance sheet expansion and growing mismatches between assets and liabilities. As a result, large banks’ financial leverage rose while their liquidity structure worsened, setting the conditions for a systemic banking crisis. Finally, this article argues that this conduct and performance was only possible due to misguided changes to the regulatory framework, specifically the Basel I Capital Accord and reductions in reserve requirements.

JEL Classification: E43, E51, G21, G28
Keywords: Financial crisis, financial intermediation, banking, Basel Capital Accords, reserve requirements

Ricardo Cabral
Departamento de Gestão e Economia
Universidade da Madeira
Edifício da Penteada
Caminho da Penteada
9000 - 390 Funchal
A perspective on the symptoms and causes of the financial crisis

Ricardo Cabral¹,²

First version: September 9th, 2009

This version: November 16th, 2010

Abstract

This article identifies two paradoxes prior to the onset of the financial crisis: banking profits were at historically high levels despite the impending crisis; and though profits were high the profitability of financial intermediation was poor. Using a novel model of banking, it argues that large banks attained high profits through balance sheet expansion and growing mismatches between assets and liabilities. As a result, large banks’ financial leverage rose while their liquidity structure worsened, setting the conditions for a systemic banking crisis. Finally, this article argues that this conduct and performance was only possible due to misguided changes to the regulatory framework, specifically the Basel I Capital Accord and reductions in reserve requirements.

JEL Classification: E43, E51, G21, G28

Keywords: Financial crisis, financial intermediation, banking, Basel Capital Accords, reserve requirements

¹ Universidade da Madeira, 9000-390 Funchal, Portugal and CEEApIA. e-mail: rcabral@uma.pt; Tel: +351 291 705 049.

² A part of the research was carried out while the author visited the Office for Strategy, Planning, Economic Policy and International Affairs (GPEARI), Ministry of Finance and Public Administration, Lisbon. The views expressed in this article are the exclusive responsibility of the author. The author appreciates comments from participants at a GPEARI seminar and at a NIPE seminar. The article builds on ideas first expressed in opinion columns published September 24, 2007 and September 5, 2008.

Copyright © 2010 Ricardo Cabral. All rights reserved. This article may be reproduced with appropriate attribution.
1. Motivation

In the years leading up to the 2007-2008 financial crisis, banking systems in the western world reported high profits, strong balance sheets, and high Basel Accord capital ratios. Even following the onset of the crisis during the first half of 2007, the OECD, the European Commission, the IMF, the Federal Reserve, and the European Central Bank (ECB), among others, appraised the financial stability outlook and the overall economic prospects favorably,\(^3\) though sometimes noting downside financial risks.\(^4\)

Yet, a crisis that was initially thought to be contained in the “subprime market”\(^5\) expanded rapidly to affect the world’s largest financial institutions. By late 2008, several of the leading US and European banks would have failed, but for massive western government intervention in their support. Between September and December 2008, this rescue amounted to nearly 29% of GDP, on average, in the US, Europe, and the UK, according to the IMF (2009). The crisis had stark consequences in real economic activity, with pronounced falls in output, investment, employment, and growth prospects.

This paper contributes to the growing body of literature on the 2007-2008 financial crisis (Blundell-Wignall et al., 2008; Reinhart and Rogoff, 2008, 2010; Kane, 2009; Ivashina and


\(^4\) For example, on financial system stability: “With the euro area financial system in a generally healthy condition […], the most likely prospect is that financial system stability will be maintained in the period ahead. However, there are some vulnerabilities and associated risks which, if they were to materialise, could pose significant challenges,” and “the financial soundness of large and complex banking groups (LCBGs) in the euro area was strengthened further in the second half of 2006” (ECB Financial Stability Review, June 2007, p. 9 and p. 13).

\(^5\) In March 28, 2007, in prepared testimony before the U.S. Congress Joint Economic Committee, Ben S. Bernanke, Chairman of the Board of Governors of the Federal Reserve System said “The impact on the broader economy and financial markets of the problems in the subprime market seems likely to be contained.” Additionally, the OECD stated that the “problems [in US subprime market] appear well contained,” (OECD Economic Outlook, June 2007, p. 29).
Scharfstein, 2010; Gorton and Metrick, 2009; Almeida et al., 2009; Gorton, 2010; Adrian and Shin, 2009, 2010; Allen and Carletti, 2010; Eichengreen and Temin, 2010; Pozsar et al., 2010; Shin, 2010). It focuses on three research questions.

First, it revisits the issue of what is banking, in light of the substantial change in the business of financial intermediation observed in the last three decades (Allen and Santomero, 1997, 2001; Gorton and Winton, 2003; Warsh, 2007; Gorton and Metrick, 2009; Gorton, 2010; Tucker, 2010; Pozsar et al., 2010; Shin, 2010). Not only did industry structure change, with the growth of non-bank financial intermediaries, but the regulatory framework was significantly altered, and there was substantial financial innovation (Gorton, 2010; Pozsar et al., 2010; Shin, 2010). This paper argues that financial intermediation consists in the selling of different types of balance sheet mismatches. It occurs if and only if the financial intermediary uses its balance sheet to intermediate between the financial claims of agents from whom it receives funds (“savers”) and the financial claims on the balance sheet of third party agents to whom it supplies funds (“borrowers”).

The second question, notoriously framed by the Queen of the United Kingdom, is “why did nobody notice [the financial crisis]?” The question is here restated as why did most observers believe the banking system was robust, thus failing to anticipate the financial crisis, and what evidence of weakness did they miss? This is an understated, though important, research question. It was likely the case that the balance of evidence available prior to 2007 provided few clues to the impending banking sector crisis, and quite in contrary offered evidence suggestive of a healthy banking system. But at the same time, there were surely non-obvious symptoms of severe stress in banking sector activity that forewarned the storm to come.

---

6 According to The Telegraph “During a briefing by academics at the London School of Economics on the turmoil on the international markets the Queen asked: ‘Why did nobody notice it?’” (“The Queen asks why no one saw the credit crunch coming,” The Telegraph, November 5, 2008).
(Bernanke, 2009a). Such evidence would have been consistent with a framework where most observers would have confidence on the strength of the banking system, and where the naysayers would have held the minority view. In fact, I argue there were two paradoxes that may explain the prevailing mindset. These arise from data indicating that banking sector profits were historically high, but that the profitability of financial intermediation was poor.

The third research question I address in this paper is what are the causes of the financial crisis? Several explanations have been proposed, of which I emphasize the following. Some view accommodative US monetary policy during 2003-2005 as one of the main causes of the crisis, though this perspective is disputed (Taylor, 2009; Greenspan, 2009; Bernanke, 2010; Allen and Carletti, 2010). Others believe that the lack of regulatory oversight of non-bank financial intermediaries (hereafter the “shadow banking system”), and particularly, lax lending standards in the sub-prime mortgage market, were key contributing factors to the crisis (Adrian and Shin, 2009; Crotty, 2009; Bernanke, 2010). Gorton and Metrick (2009) defend that a run on the repo market was the proximate cause of the systemic banking crisis. Others claim that the crisis originated in the shadow banking system (Gorton, 2010; Tucker, 2010). Pozsar et al. (2010) argue that the crisis occurs because the shadow banking system, in contrast to the traditional banking system, lacked access to central bank lender-of-last resort liquidity. Kane (2009) argues that this was regulation-induced financial crisis. Unsustainable and non-transparent regulatory subsidies led to bad lending decisions (capital misallocation) and loan losses. Regulatory forbearance and financial deregulation facilitated the transfer of the losses to the taxpayer. Finally, another line of thinking argues that the financial crisis is the consequence of large global (current account) imbalances (Smaghi, 2008; Allen and Carletti, 2010).
This paper focuses on the role of the traditional banking system and of its regulatory framework. It argues that the proximate causes of the crisis were low bank capital and what is here designated as “vulnerable liquidity structure”. The first issue meant that even low levels of losses would have rendered several large banks technically insolvent. The second issue meant that relative to the overall level and type of liabilities, banks had insufficient levels of liquid assets with which to meet calls for cash.\textsuperscript{7,8} As a result, some of the largest world banks could be brought to the brink of failure by relatively small bank runs.\textsuperscript{9}

It then asks how banks were able, under the existing regulatory framework, to have low capital and liquid asset ratios? The short answer is that regulatory changes in the late 1980s and early 1990s allowed banks to reduce capital and reserve ratios. Specifically, as shall be explained below, the Basel I Capital Accord (Jackson et al., 1999; Jones, 2000; Allen, 2004) and lower minimum reserve requirements in advanced economies (Feinman, 1993; O’Brien, 2007), created incentives that led (large) banks to increase financial leverage and to reduce

\textsuperscript{7} Liquidity is here understood as the ability to readily and costlessly convert the asset to cash, assumed here as the most liquid asset available. A perfectly liquid asset can be readily converted to cash at no cost (Gorton and Winton, 2003, Warsh, 2007).

\textsuperscript{8} For example, according to Tim Congdon, “At June 2007, UK banks’ cash deposits at the Bank of England were £2.4bn ($4.9bn), while the notes and coin held in their tills were worth just under £8.8bn and their holdings of Treasury bills were under £8bn. By contrast, their total sterling liabilities were over £3,150bn. Cash is about ½ per cent and traditional liquidity about 1 per cent of total liabilities,” (see “Pursuit of profit has led to risky lack of liquidity,” The Financial Times, September 10, 2007).

\textsuperscript{9} For example, in the UK, the banks RBS and HBOS had to be saved through secret emergency loans by the Bank of England. These loans reached a combined £62 billion at the peak of the crisis. RBS and HBOS had at the time balance sheets of £2.4 trillion and £690 billion, respectively (see “Bank of England’s Mervyn King says HBOS and RBS came within hours of collapse,” The Telegraph, September 19, 2009 and “Bank of England reveals secret £62bn loans used to prop up RBS and HBOS,” The Guardian, November 23, 2009).

In the case of the Lehman Brothers failure, its bankruptcy court examiner report finds that the failure was the result of insufficient liquidity and lack of confidence by lenders and counterparties (Valukas, 2010). Lehman Brothers had unencumbered highly liquid assets amounting to $1.4 billion on September 12, 2008. It declared bankruptcy on September 15, 2008, since it faced a cash shortage of $4.5 billion on that day (Valukas, 2010, Footnotes 37 and 48). These amounts must be seen in the context of a $700 billion balance sheet.
liquid asset ratios. Thus, in a somewhat similar argument to Kane (2009), this paper argues that the root causes of the crisis must be found in a misguided regulatory framework.

The paper is organized as follows. Section 2 revisits the key concept of banking and advances a novel definition of financial intermediation. Section 3 introduces a novel model of banking, on which the analysis in the remainder of the paper is grounded. Section 4 investigates why the crisis caught so many by surprise. Section 5 looks into the causes of the financial crisis. Section 6 discusses the main results and policy implications.

2. What is banking?

Banking is normally associated with retail deposit-taking and retail, commercial, and industrial loan-making financial intermediaries. Banks have historically had a leading role in the channeling of savings to investment. Bank credit used to be the main source of external funding for investment by firms (Baer and Pavel, 1988; Mayer, 1990; Allen and Santomero, 2001; Gorton and Winton, 2003).

It is well understood that banking has changed in recent decades, through substantial financial innovation, regulatory change, and the growth of non-bank financial intermediaries (Allen and Santomero, 2001; Gorton, 2010; Pozsar et al., 2010; Shin, 2010). For example, in the US, maximum leverage regulations for diversified broker dealers (investment banks) were relaxed in 2004 by the Securities Exchange Commission (SEC), leading to a rapid rise in financial leverage after 2004 (see Blundell-Wignall et al., 2008).

In an oft cited reference, Mayer (1990) points out that in the US case, between 1975 and 1985, bank loans funded 24.4% of the firm investment. Other researchers have also found evidence of the varying importance of bank financing across countries. However, Allen and Santomero (2001) point out that the relevance of company internal financing may be overstated due to methodological issues.

Pozsar et al. (2010) point out that traditional banks have responded to the challenge of non-banks by acquiring or by developing shadow-banking subsidiaries. See also Blundell-Wignall et al. (2008) and Shin (2010) for a similar argument.
the banks’ share of the total outstanding credit had fallen to only 15 percent by 2007 (Geithner, 2007) and the shadow banking system became significantly larger than the traditional banking system, with total liabilities of $20 trillion by March 2008 (Pozsar et al., 2010). In sum, the traditional definition of banking does not take into account the role of a multitude of non-bank financial intermediaries that have taken a preponderant share of financial intermediation.

Gorton (2010), among others, argues that the shadow banking system is a “real banking system”. The definition of banking advanced here concurs with this view. It suggests that diversified broker-dealers (hereafter investment banks), and more generally the shadow-banking system perform financial intermediation (Gorton and Metrick, 2009; Gorton, 2010; Tucker, 2010, Pozsar et al., 2010; Shin, 2010).

In the Industrial Organization (IO) type banking model specified in Section 3, banks have two types of revenues and products. The first is fee and commission revenues charged by banks for the different services and transactions carried out for their customers, which falls under what is here designated as the financial brokerage function of banks. The second and more relevant type of revenues results from the mismatches between assets and liabilities of a bank’s balance sheet – sometimes designated as Qualitative Asset Transformation – which falls under what is here designated as the financial intermediation function of banks (Gorton and Winton, 2003; Warsh, 2007; Allen and Santomero, 2001; Pozsar et al., 2010). So, in

13 Until the repeal of the Glass-Steagall Act in 1999, US commercial banks were not allowed to offer investment banking services. In contrast, in Europe, large banks were allowed to engage in retail, commercial, and investment banking – an approach that is known as the universal banking model.

14 Pozsar et al. (2010) argue that the shadow banking system is characterized by specialization. For example, some financial intermediaries specialize by only assuming default risk mismatch between assets and liabilities with similar maturities. Other financial intermediaries specialize in only real estate mortgages. In contrast, traditional banks are exposed to default, maturity, and liquidity risks mismatches between assets and liabilities and to various types of economic assets.
short, in the framework proposed here, traditional banks are both financial brokers and financial intermediaries. This paper is primarily concerned with the financial intermediation function of banks, for which it proposes a novel definition (Shin, 2010):

Definition 1: Financial intermediation consists in the managing and selling of asset and liability balance sheet mismatches. It occurs if and only if the financial intermediary uses its balance sheet to intermediate between the financial claims of agents from whom it receives funds (“savers”) and the financial claims over the balance sheet of third party agents to whom it supplies funds (“borrowers”).

Similarly to Shin’s (2010) view of financial intermediation, Definition 1 expands the concept of financial intermediation to include activities that have traditionally not been considered as financial intermediation. Specifically, the proprietary trading, the market making, and the securitization business of investment banks and other non-bank financial intermediaries relies on balance sheet mismatches. The financial intermediation margins associated to these activities are derived from the capital gains obtained upon sale of the assets. They are only known with certainty once the asset is sold.\textsuperscript{15,16}

\textsuperscript{15} For example, according to Definition 1, proprietary trading in equity and financial debt securities constitutes financial intermediation. Proprietary trading means that (financial) assets are acquired, held temporarily on the balance sheet, and then sold. Capital gains compensate the financial intermediary for the asset-liability mismatch risks it incurs on the transaction while the assets are held on the balance sheet. Thus, trading revenues, often categorized as non-interest revenues by regulators, are a form of financial intermediation revenues.

Market making, i.e., ensuring orderly markets for a variety of securities in regulated or unregulated exchanges also constitutes financial intermediation according to Definition 1, since assets are held on the balance sheet temporarily.

In addition, in recent decades, securitization became a large source of revenues for investment banks (Gorton, 2010; Gorton and Metrick, 2009; Pozsar et al., 2010). Securitization typically consists in the aggregation and selling of claims on specific cash flows to an outside company. The outside company then sells new securities backed by the pool with different characteristics and cash flows (Gorton, 2010; Gorton and Metrick, 2009; Pozsar et al., 2010). According to Definition 1, securitization is in part a financial intermediation function. In fact, despite allowing banks to sell loans and thereby to remove the loans from the balance sheet, securitization often requires banks to put their balance sheet at risk temporarily while the debt instruments are packaged into a pool, and before they are sold to third parties (Pozsar et al., 2010).
In sum, under Definition 1, investment banks and the shadow banking system embody financial intermediation activities similar to those of the traditional banking system (Gorton, 2010; Tucker, 2010, Pozsar et al., 2010; Shin, 2010).

3. Model specification

As referred in section 2, this paper argues that the core business of banking, financial intermediation, is to manage and sell balance sheet mismatches. Specifically, financial intermediation consists essentially in accepting a default, a liquidity, and a maturity risk mismatch between assets and liabilities, in exchange for an interest rate spread (Moore, 1989; Pozsar et al., 2010). While this perspective is novel, conceptually, it extends Allen and Santomero’s (2001) view that financial intermediaries have a risk management function.

Consider first the default (or credit) risk. Banks derive revenues net of interest costs (hereafter, net revenues) from a default risk interest rate spread. This spread arises from the mismatch in the default risk of bank assets and liabilities. On average, banks have a lower

---

16 The argument could be made that returns on held-to-maturity loans are similarly also only known with certainty ex-post. In fact, it typically takes some time for a loan to become non-performing or for a default to occur.

17 While the terminology used in the literature varies somewhat, I base the designations used in this paper on Angbazo (1997), who identifies three main types of risk: default (or credit) risk, liquidity risk, and interest rate (or maturity) risk (see also Pozsar et al., 2010).

18 Economic theory justifies the existence of financial intermediaries on the grounds of market imperfections – specifically, asymmetric information and transactions costs (Gorton and Winton, 2003; Allen and Santomero, 1997, 2001; Gorton, 2010). As pointed out by Allen and Santomero (2001), technological change in recent decades would suggest a diminished role for financial intermediaries. However, despite much improved information technology, Allen and Santomero (2001) point out that the role of financial intermediaries (particularly pension and mutual funds) has grown. This fact can be interpreted as suggesting that market imperfections do not explain the existence of financial intermediaries. Instead, Allen and Santomero argue that financial intermediaries have a risk management function.

19 Throughout the paper, banks’ interest rate spread or simply spread refers to the difference between the interest rate premium received on assets and paid on liabilities (Moore, 1989). Interest rate spreads refer to bank interest margins. Interest rate premia refers to the gross interest rates on assets and on liabilities. For example, a bank which charges a default premium
default risk on their liabilities than the clients to whom the banks provide credit (Gorton, 2010; Pozzar et al., 2010). Given this difference, bank liabilities will, on average, be remunerated with a lower default risk premium than bank assets. The pure default spread is the difference between the pure default risk premium the bank charges its borrowers less the pure default risk premium it pays its own creditors, given its own perceived default risk. This spread remunerates banks for the default risk it incurs by extending the loan (relative to the banks’ own default risk).

A second source of bank balance sheet net revenues is derived from a liquidity mismatch between bank assets and liabilities (Hull et al., 2005; Gorton and Metrick, 2009; Gorton, 2010). In fact, the banking business model is based on conveying the perception that bank deposits are as liquid as cash (Gorton, 2010). In order to do so, deposits are endowed with several “features”. For example, deposits are expressed in the same unit of account as cash. Banks also guarantee the nominal amount of the deposit and that the deposit can be converted to cash at any time, often with no loss (Gorton and Winton, 2003; Gorton, 2010). On the other hand, bank assets are, to a significant degree, illiquid, in that they cannot be converted into cash quickly without incurring into significant losses.

As a result, banks pay their depositors and other creditors liquidity interest rate premia that are, on average, lower than the liquidity interest rate premia that they charge their borrowers.

---

20 There are several possible explanations as to why banks pay, on average, lower default risk premia than their customers. First, banks are required by law to have significant capital ratios, which bear first losses. Banks also demand collateral from borrowers, which enhance the value of bank assets. Another possible explanation is that banks benefit from deposit insurance and implicit government guarantees, as well as from government supervision and regulations (Gorton, 2010; Pozzar et al., 2010).

21 The pure default risk premium compensates the claim holder for the possibility that the loan or interest on the loan will not be fully (re)paid at maturity.
This liquidity risk interest rate spread remunerates banks for their liquidity management function, i.e., the banks’ willingness to guarantee liquid liabilities while taking on illiquid assets.

Banks additionally derive balance sheet net revenues from the maturity (or term) mismatch between assets and liabilities. Banks typically take short term deposits, issue short term commercial paper or take out short term interbank loans on the liability side, which they then loan out or invest longer-term, again on average, as pointed out by Diamond and Dybvig (1983). In doing so, they incur a maturity mismatch between their assets and liabilities, for which they are able to charge a margin — what in this paper is designated as the “pure term” interest rate spread.

The literature sometimes designates the maturity risk as the interest rate risk (Angbazo, 1997, BCBS, 2004). Interest rate risk refers to the possibility that movements in interest rates will affect more rapidly the cost of bank liabilities, than the interest revenues derived from its assets.\footnote{According to the Basel Committee on Banking Supervision, “interest rate risk is the exposure of a bank's financial condition to adverse movements in interest rates” (BCBS, 2004, p.5). The Basel Committee on Banking Supervision has a more elaborate view on the sources of interest rate risk, which includes repricing risk, yield curve risk, basis risk, and optionality (BCBS, 2004).} It is assumed here that the term interest rate spread incorporates the revenue necessary to compensate banks for interest rate risk (Gambacorta, 2008). The model proposed here further assumes, like English (2002) and Drehmann et al. (2010), that the pure term spread can be proxied by the interest rate spread between government debt securities of different maturities along the yield curve, though it is clear that government debt securities are not risk-free. Moreover, for simplification, differences in government debt yield curves (and term spreads) across countries are not considered in the model.
Admittedly, there may be other balance sheet revenue sources. For example, some banks may have market power (Angbazo, 1997). Thus, they may be able to raise default, liquidity, or term spreads, and fees and commissions, above competitive levels. Banks also offer convenient denomination for both savers and borrowers, from which they may be able to derive an interest rate spread. In addition, large money center banks are likely to have large currency exchange mismatches between assets and liabilities, for which they would also be expected to charge an interest rate spread (Chamberlain et al., 1997; Martin, 2000; Pozsar et al., 2010). These other sources of balance sheet revenues are not explicitly reflected in the model.

A few further simplifying assumptions are made. First, I focus on a static one-period discrete-time model, where the period is set to one year. The level of assets and liabilities is fixed during the period, and the interest revenues, as well as fee and commissions accrue at the end of the same period.

Second, there is the issue of which balance sheet mismatches should be included in the model. Generically, it is assumed that the liquidity spread, the pure default spread, and all other sources of risk and balance sheet revenues (see above) are reflected in what is hereafter designated simply as “default spread”, but not in the term spread. Thus, one key feature of the model is to separate bank balance sheet mismatches associated with interest rate risk from all other sources of risk. This is an approach similar to that of Drehmann et al. (2010). The main rationale is that the yield curve interest rate spread is a good proxy of the true pure term

---

23 Drehmann et al. (2010) also separate what they call credit and interest rate risk, though pointing out that these variables are correlated. They develop a framework that is fine-tuned to estimate the impact, on bank economic value and profitability, of repricing mismatches and asset and liability portfolio effects caused by credit and interest rate shocks. Among other things this framework allows them to stress test banks’ balance sheets. However, they make several assumptions. For example, they assume that banks do not actively manage their portfolio assets and liabilities composition, that banks invest in new projects with the same repricing and risk characteristics as maturing assets, that banks pay off costly matured liabilities, and their model considers capital adequacy using risk weights as defined under Basel I and II.
spread, whereas it is difficult to identify appropriate proxies for each of the other sources of risk described above (Hull et al., 2005). Moreover, an approach based on separate modeling and identification of all other risk sources would significantly increase model complexity, without necessarily improving the understanding that can be derived from it.

Moreover, for simplification, differences between the ex-ante and ex-post realizations of capital gains on assets, associated to investment banking and other non-bank financial intermediaries, are not explicitly reflected in the model proposed below.

Finally, a part of a bank’s interest revenues and costs are derived from legacy assets and liabilities, the contractual conditions of which banks may be unable to change.\textsuperscript{24} However, banks are able to manage a part of the balance sheet mismatches by changing the default, liquidity, and maturity profile of a part of the assets and of the liabilities in the balance sheet.\textsuperscript{25} Banks are also able to change the size of the balance sheet to some degree.

The focus of the analysis is on net interest revenues (NIR) associated with the financial intermediation function of banks, which are specified by

\begin{equation}
NIR_i = \sum_{j=1}^{q_i} p_j(a_{ij}) - c_j(l_{ij})
\end{equation}

where \(a_{ij}\) is bank \(i\)’s \(j^{th}\) dollar of assets, \(l_{ij}\) is bank \(i\)’s \(j^{th}\) dollar of liabilities, and bank \(i\) has \(q_i\) dollars of assets and liabilities. Thus, \(q_i\) can be interpreted as bank \(i\)’s output: it measures the

\textsuperscript{24} It is assumed that each dollar of a bank’s assets and liabilities has specific default and term interest rate premia associated with it, which may or may not vary over time (variable or fixed interest rates). These assets and liabilities may have different maturities. Thus, they may remain in the balance sheet for more than one period of time.

\textsuperscript{25} Repricing mismatches are thought to be one important source of risks to banks (BCBS, 2004; Drehmann et al., 2010), though English (2002) finds weak evidence of interest rate shocks effects on net interest margins. Earlier models have dealt with the repricing mismatches issue in different ways. For example, Drehmann et al. (2010), depart from the assumption that banks do not manage asset and liability mismatches, contrary to the concept that underlies Definition 1.
total outstanding credit and other assets, from which a bank derives its gross interest revenues.

\( p_i(a_j) \) is the price (interest rate) demanded by bank \( i \) for the \( j^{th} \) dollar of assets, which is assumed to be the sum of a default premium component, \( p_i^d(a_j) \), and a term premium component, \( p_i'(a_j) \). \( c_i(l_j) \) is bank \( i \)'s interest cost of the \( j^{th} \) dollar of liabilities, which is assumed to be the sum of a default premium component, \( c_i^d(l_j) \), and a term premium component, \( c_i'(l_j) \), that is:

\[
\begin{align*}
(2) \quad p_i(a_j) &= p_i^d(a_j) + p_i'(a_j) \\
(3) \quad c_i(l_j) &= c_i^d(l_j) + c_i'(l_j)
\end{align*}
\]

For each dollar of assets and liabilities it is now possible to specify two interest spread functions as follows:

\[
\begin{align*}
(4) \quad d_i(a_j,l_j) &= p_i^d(a_j) - c_i^d(l_j) \\
(5) \quad t(a_j,l_j) &= p_i'(a_j) - c_i'(l_j)
\end{align*}
\]

where \( d_i(a_j,l_j) \) is bank \( i \)'s default spread revenue for the \( j^{th} \) dollar of assets and liabilities. It is equal to the difference between the default (and liquidity) risk premium received and paid by bank \( i \) for the \( j^{th} \) dollar of assets and liabilities, respectively. It is bank specific to allow for the possibility of varying levels of market power across banks. Similarly, \( t(a_j,l_j) \) is the market’s term spread revenue for the \( j^{th} \) dollar of assets and liabilities. It is assumed that the term spread, for a given maturity mismatch, is constant across banks. Therefore, (1) can be rewritten as:
Thus, \( \sum_{j=1}^{q_i} d_i(a_{ij}, l_{ij}) \) and \( \sum_{j=1}^{q_i} t(a_{ij}, l_{ij}) \) are the total default spread and total term spread net interest revenues derived from the mismatch between bank \( i \)'s assets and liabilities, respectively. Based on the above, bank \( i \)'s profit function in a given period is specified as:

\[
\pi_i = F(A_i, L_i) - C(A_i, L_i) - Lo_i(A_i, z) + \sum_{j=1}^{q_i} d_i(a_{ij}, l_{ij}) + \sum_{j=1}^{q_i} t(a_{ij}, l_{ij})
\]

where \( A_i = \{a_{i1},...,a_{iq_i}\} \) and \( L_i = \{l_{i1},...,l_{iq_i}\} \) are total assets and total liabilities matrices, respectively. \( F(A_i, L_i) \) and \( C(A_i, L_i) \) are bank \( i \)'s total fee and commission revenues and total non-interest costs, respectively, which are a function of total assets and total liabilities. \( Lo_i(A_i, z) \) are asset losses experienced by bank \( i \). These are assumed to depend on the asset matrice, \( A_i \), and on a set of exogenous variables, \( z \).

The model proposed here was developed independently, but shares some insights with earlier models by Moore (1989), Hannan and Berger (1991), and Drehmann et al. (2010), and with earlier operations research work of Kusy and Ziemba (1986) and others.

---

26 Moore (1989) pointed out that banks earned profits through interest rate spreads between interest received on assets and paid on liabilities. He also emphasized the importance of the two-sidedness of a bank’s balance sheet.

27 Hannan and Berger (1991) develop a model where banks are price takers on the interest rate received on assets, but have some market power in defining the deposit interest rate.

28 See Footnote 23.

29 A strand of the operations research (OR) literature focuses on so-called Asset and Liability Management models. In particular, Kusy and Ziemba (1986) specify a bank inter-temporal profit maximization function as an optimal dynamic asset and liability allocation problem, subject to a number of constraints (e.g., budget, capital adequacy, liquidity, regulatory penalties). The first term of the model’s profit function is the present value of returns on different types of asset categories.

I argue there were two paradoxes that might explain why the crisis caught so many by surprise. On the one hand, bank profits were historically high, very large in absolute terms, relative to GDP, and as a percentage of assets, which is difficult to reconcile with the fact that the banking sector was about to experience a profound crisis (Paradox I). On the other hand, the profitability of financial intermediation was poor, with falling interest margins and low default and term risk spreads, a fact that seems inconsistent with the historical high level of profits observed in the period that preceded the crisis (Paradox II).

- Figure 1 about here -

Regarding Paradox I, there is some anecdotal evidence suggesting that the financial sector, and the banking sector in particular, were highly profitable prior to the onset of the crisis. For example, S&P 500 financial sector earnings were at historically high levels in 2006 and during the first half of 2007. National accounts data from the US Bureau of Economic Analysis (BEA), using a methodology that importantly excludes capital gains and losses\(^{30}\) estimated financial sector profits at 31.6% of domestic industry profits in 2006 (Figure 1). In fact, the US financial sector accounted for 8.1% of GDP in 2006, higher than the 3.5% of GDP it represented in 1960 (Warsh, 2007). Other anecdotal evidence, for example, large bank annual reports and Federal Deposit Insurance Corporation (FDIC) historical statistics on banking, also suggested a healthy and very profitable banking sector in 2006. Non-interest revenues constituted a growing share of overall revenues, a fact that was seen as a sign of

\(^{30}\) As mentioned in Section 2, according to Definition 1, capital gains and losses may represent financial intermediation revenues. Thus, BEA measures of financial sector profits that exclude capital gains may underestimate the profitability of the banking sector.
strength rather than of weakness. Thus, the overwhelming majority of observers could point to this evidence and argue that the banking system was in robust health.

Regarding Paradox II, as suggested by Figure 2, prior to the onset of the crisis, net interest margins had been under pressure. Measures of the default and term spreads (Figures 3 and 4 shown further below) had fallen to very low levels. Low default and term spreads would have implied weak net interest margins for any new loan or for preexisting variable rate loans. FDIC’s historical statistics on banking indicate that net interest margins, as a percentage of total assets, fell since the early 1990s, despite some evidence of growing balance sheet mismatches (see below).\(^{31}\) Available evidence suggests that net interest margins of European major banks also decreased since the mid-1980s (de Larosière and Barthalon, 2001; ECB, 2006).\(^{32}\) Moreover, US commercial bank total non-interest income, as a percentage of total assets, diminished since the beginning of the 2000s. In the US case, while year-on-year deterioration in net interest margins and total non-interest income were negligible, the cumulative effect over longer periods of time was, by 2006, already large (see Figure 2).

\(^{31}\) Figure 2 suggests that, in the US, commercial banks derived average net interest margins of at least 3 percent of assets between the late 70s and 2005, much higher than prior to the 70s. Total outstanding credit increased markedly in this period. The relatively high net interest margins documented by Figure 2 might be indicative of greater market power, lower efficiency (diseconomies of scale), or, as shall be seen below, of more pronounced default, liquidity, and term balance sheet mismatches.

\(^{32}\) As pointed out by English (2002) and by de Larosière and Barthalon (2001) net interest margins (NIMs) vary substantially across countries, and it is difficult to obtain consistent international measures. European banks’ NIMs have historically been much lower than those of US based banks (de Larosière and Barthalon, 2001). This difference may be attributable to differences in capital requirement regulations, in accounting rules, in the activity of banks, and in short- and long-term interest rates. For example, US specific minimum leverage ratio regulations meant large US based banks had an incentive to have larger off-balance sheet exposures than European banks. Moreover, large US banks used a different accounting standard than that used by most large European banks (in recent years, GAAP vs IFRS). These accounting standards differ crucially in the treatment of derivative exposures, with the former requiring the accounting of net exposure, while IFRS requires the accounting of gross exposure. These two differences would have resulted in US banks understating their balance sheet assets (and thus overstating their NIMs) relative to comparable European based banks, ceteris paribus.
To the best of my knowledge, historical data on bank balance sheet mismatches with the necessary level of detail is not available. However, there is some indirect evidence of growing default, liquidity, and term mismatches between balance sheet assets and liabilities (Moore, 1989; Gorton and Metrick, 2009; Pozsar et al., 2010). For example, measures of US banks’ share of the most liquid forms of assets – minimum reserves, cash holdings, and US Treasury debt securities – fell to historically low levels. The deposit share of total liabilities also fell. Finally, anecdotal evidence of increased exposure to repo markets, increased off-balance sheet exposure, and growing derivative exposure is also suggestive of more pronounced balance sheet mismatches (Gorton and Metrick, 2009; Gorton, 2010; Pozsar et al., 2010; Valukas, 2010).

In sum, net interest margins, default and term spreads, and balance sheet mismatches were key variables experiencing worsening performance prior to the onset of the crisis. These variables were non-obvious symptoms of stress in financial intermediation, which suggested that financial intermediaries were under significant competitive pressure.

Thus, there was evidence that banks were highly profitable. Second, there was less obvious evidence that suggested that the profitability of financial intermediation was poor. Both large absolute profits and poor profitability of financial intermediation must be a consequence of

---

33 According to the footnotes of the Federal Reserve Board’s H.8 release, cash assets include “vault cash, cash items in process of collection, balances due from depository institutions, and balances due from Federal Reserve Banks.” In contrast, total reserves (Federal Reserve Board’s H.3 release) include only bank vault cash and bank deposits with the Federal Reserve. Thus, reserves are a narrower definition of liquid assets than cash assets. Because the term cash is thus subject to misinterpretation, hereafter this article uses the term “currency inflows” to designate cash deposits that meet the Federal Reserve definition of reserves, i.e., bank vault cash holdings and bank deposits with the Federal Reserve Banks.

34 Sources: FDIC Historical Statistics on Banking and Board of Governors of the Federal Reserve System Statistical Releases H.3 “Aggregate Reserves of Depository Institutions” and the FDIC historical statistics on banking. See also Footnote 50.

35 The fact that the financial sector captured such a large percentage of corporate profits could also be construed as a non-obvious indicator of imbalances in financial intermediation (Shin, 2010). After all, the channeling savings towards investment should represent a relatively small fraction of economic activity.
profit maximizing behavior by banks. Thus, model (7) is used below to explain and reconcile both paradoxes.

4.1. What explains the high banking sector profits prior to the crisis (Paradox I)?

According to model (7), banks have distinct profit sources. It is unlikely that the first term of (7), fee and commission revenues, can explain the growth in absolute profits. FDIC historical statistics on banking show that noninterest revenues grew in the period prior to the crisis (2002-2006), driven by the growth in securitization (see Figure 2). However, noninterest revenues, as a percentage of total assets, fell slightly. This is a period that is marked by growing competition in the banking sector, with the growth of internet banking, cross-border banking, and products like no-fee accounts and flat rate brokerage services, all of which would be expected to put fee and commission prices under competitive pressure.

FDIC data also indicates that provisions and allowances for losses on loans and leases relative to total assets fell between 2002 and 2006. Moreover, non-interest costs as a percentage of total assets also fell. Thus, a part of the growth in the absolute level of profits observed during this period is due to a reduction in the second and third terms of (7), respectively.

Consider next the last two terms of equation (7). IO theory suggests that there are three alternative forms by which banks can increase profits, ceteris paribus: through changes in the level of prices or in the level of costs; through changes in the level of output; or through changes in the output “quality” that may allow banks to charge higher prices per unit of output – a product differentiation strategy in the IO terminology.
Available evidence suggests that price increases and cost reductions do not explain the large banking profits prior to the onset of the crisis: default and term spreads fell during this period suggesting falling prices-cost margins, for a given asset-liability default and term mismatch (see Figures 3, 4, and 5 below).

However, banks do have some degree of control over the size and composition of the balance sheet. By increasing the balance sheet size, banks increase output (Adrian and Shin, 2010). By augmenting balance sheet mismatches between assets and liabilities, output quality is enhanced. More pronounced balance sheet mismatches are associated to higher default and term spreads. Thus, the main two forms by which banks can raise net interest revenues (6) is to increase output and to increase the default, liquidity, and term mismatches of the balance sheet. According to (7), if losses remain stable, banks are able to increase profits.

An increase in balance sheet mismatches would likely have resulted in banks acquiring more long-term, less liquid and higher default risk assets to raise term, liquidity and default premia received on assets (Moore, 1989). At the same time banks likely relied to a greater extent on short-term funding, and likely offered more guarantees for that funding (e.g., more collateral, minimum return guarantees, liquidity assurances, etc.), so as to lower the term, liquidity, and default premia banks paid on their liabilities, ceteris paribus.

4.2. Banking profits were high. Then, how come the profitability of financial intermediation was poor (Paradox II)?

Above, it was argued that banks followed two straightforward strategies in order to increase profits: output increases and product differentiation through output quality improvements.
However, IO theory points out that there are diminishing returns to each of these strategies, under the typical assumptions about demand and cost functions.

In fact, a strategy based on output increases typically results in diminishing output prices and price-cost margins. Beyond a certain level of output, marginal revenue of an additional unit of output falls below marginal costs, leading to lower profits.

Product differentiation strategies also face diminishing returns. First, marginal costs per unit of output typically increase with the “quality” of output. For example, higher default risk assets are associated with higher expected losses. In addition, as is also known from IO theory (Brander and Eaton, 1984), product differentiation strategies may result in the cannibalization of profits because of the starker price competition between the product varieties offered on the market (i.e., asset-liability pairs with similar mismatch profiles).

In sum, these two IO strategies would have been expected to result in falling bank net interest revenues (6), providing an explanation for the poor profitability of financial intermediation prior to the onset of the crisis.

4.3. Is the available evidence consistent with the above explanations?

The spread along the yield curve is used here as a proxy of the true term spread (English, 2002; Drehmann et al., 2010). Throughout 2006 and until June 2007 the yield curve was approximately flat (Figure 3). This suggests that the pure term spread was very low or non-existent. In contrast, from January 1982 to June 2007, the average monthly term spread
between the 10-year US Treasury bond and the US effective federal funds rate was approximately 1.4 percentage points. Therefore, the low term spread observed prior to the onset of the crisis is inconsistent with large absolute banking profits, but is consistent with poor profitability of financial intermediation.

- Figure 4 about here -

Figures 4 and 5 show different measures of interest spreads between assets with high default (or credit) risk and US Treasury debt securities of comparable maturities. These spreads are used here as proxies for the default (and liquidity) spread of specific categories of debt securities (Hull et al., 2005). For example, the option adjusted spread for high yield bonds (Figure 4) measures the interest spread between the interest rates on a basket of so-called US “junk bonds” compiled by Merrill Lynch and the interest rates on government bonds of comparable maturity (10-year), adjusted for option-like features of the underlying junk bonds. This measure of default and liquidity risk spread fell from a high of 11.2 percentage points in September 2002 to a low of 2.4 percentage points in June 2007.

- Figure 5 about here -

Debt securities with higher debt ratings (and lower default risk) than junk bonds typically have lower yields than junk bonds. Therefore, Figures 4 and 5 suggest that the default spread was substantially compressed between 2002 and the end of 2006. Lower default spreads meant there were diminishing returns to banks increasing the default and liquidity mismatch between assets and liabilities. It also suggests poor profitability of financial intermediation, but is inconsistent with high banking profits.
In sum, above it was argued that fee and commission revenues, loan losses, and non-interest cost reductions are unlikely alone explain the high level of banking sector profits prior to the onset of the crisis. Instead, the high banking profits are in part the result of substantial higher output levels through balance sheet expansion. Anecdotal evidence suggests rapid rates of growth of on- and off-balance sheet assets among the largest US and international banks. For example, US commercial banking system total on-balance sheet assets grew 75% between 2002 and 2006. There is also evidence of rapid off-balance sheet assets growth. Moreover, there is some evidence that suggests that large banks grew more rapidly. For instance, according to the Bank of England (BoE, 2007), total assets for the 16 largest complex international financial institutions grew from under $10 trillion in 2002 to approximately $23 trillion in 2006.

Figures 3, 4, and 5 suggest that, during the same period, term and default spreads fell for a wide spectrum of asset and liabilities mismatches. Spreads fell more rapidly than the increase in total assets. In these conditions, equation (7) suggests that banks could only have increased the absolute level of profits through growing default, liquidity, and term mismatches between assets and liabilities.

Thus, model (7) provides a possible explanation, consistent with anecdotal evidence, for the two paradoxes identified above. The absolute level of profits was large as banks increased balance sheet size and the degree of mismatch between assets and liabilities. However, the profitability of financial intermediation was poor because banks lowered default and term spreads in order to raise output.

---

36 For example, as referred above, total assets of the 16 largest complex international financial institutions increased by approximately 2.3 times between 2002 and the end of 2006 (BoE, 2007). In the same period, the option adjusted spread for high yield bonds spread fell by approximately 80%. The yield curve flattened, suggesting that the term spread fell by approximately 100%.
5. What caused the financial crisis?

There were likely several contributing factors to the 2007-2008 financial crisis. Possible explanations have focused on the role of accommodative monetary policy (Taylor, 2009; Greenspan, 2009; Bernanke, 2010; Allen and Carletti, 2010), on the lack of regulation of the shadow-banking system (Adrian and Shin, 2009; Bernanke, 2010) and a light-regulation mindset (Crotty, 2009), on the role of the repo market (Gorton and Metrick, 2009), on the access to lender-of-last resort liquidity by the shadow-banking system (Pozsar et al., 2010), on the role of regulatory incentives and supervisory forbearance (Kane, 2009), and on the role of global imbalances (Smaghi, 2008; Allen and Carletti, 2010).

With the benefit of hindsight it is clear that the banking system was at the core of the financial crisis (Kane, 2009; Gorton, 2010), experiencing losses that the IMF (2009) estimated could reach $4 trillion. This paper seeks to provide an explanation of the crisis that reconciles profit-maximizing behavior by banks with observed evidence prior to the onset of the crisis, and with the understanding of the problems faced by the banking system in the aftermath of the crisis (Geithner, 2009; Trichet, 2009; Bernanke, 2010; Bair, 2010).

In this section, it is argued that the proximate causes of the crisis are associated to the strategies followed by banks to achieve the historically high levels of profits observed prior to the onset of the crisis. Namely, the model of banking (7) specified in Section 3 suggests that, in order to raise profits, banks expanded balance sheets and increased mismatches between assets and liabilities in the period that preceded the outbreak of the crisis (Section 4).

There is evidence that balance sheet growth is associated with higher financial leverage (Adrian and Shin, 2010). Thus, this paper argues that one of the proximate causes of the crisis
must found in excessive bank financial leverage (Geithner, 2009; Bair, 2010). The excessive leverage meant that relatively small losses rendered banks technically insolvent.

The immediate cause of bank failures (and that of any other type of firm) is lack of liquidity, rather than insolvency. The more pronounced default, liquidity, and term balance sheet mismatches that followed from the banks’ profit maximization strategy (see above) increased the vulnerability of the banking system to bank runs. It became more challenging to ensure liquid bank deposits (and other liabilities).

Thus, this paper argues that the second proximate cause of the crisis was the substantial liquidity risk mismatch between bank assets and liabilities. Prior to the onset of the crisis banks had what is here designated as “vulnerable liquidity structure.” Banks held assets that paid high liquidity premia (i.e., had poor liquidity), and funded these assets with liabilities that cost them low liquidity premia (i.e., had high liquidity). In this setting, even some of the largest banks would rapidly be unable to honor their obligations in the face of relatively small bank runs (Ratnovski and Huang, 2009; Kane, 2009; Gorton, 2010; Valukas, 2010).

---

37 FDIC historical statistics on banking suggest that the US commercial banking system had higher capital buffers (measured by the leverage ratio of assets to tangible capital) than in the 1980s. Post-crisis, the consensus is that the banks had low levels of capital (Geithner, 2009; Trichet, 2009; Bernanke, 2010; Bair, 2010). One possible explanation for the discrepancy is the growth in off-balance sheet exposure, particularly by large US banks, which would not have been captured by FDIC statistics. In contrast, there is data that suggests that large European banks had very low capital buffers and high financial leverage.

38 Gorton and Metrick (2009) point out that financial intermediaries relied to a large extent on short term wholesale market funding (repo market). In a repo operation, a bank obtains a short term deposit from a creditor and offers a (high quality) security as collateral (Gorton, 2010; Gorton and Metrick, 2009). For example, the Lehman Brothers bankruptcy court examiner report indicates that, during 2008, Lehman Brothers funded in excess of $200 billion (29% of its liabilities) each day in the repo market (Valukas, 2010, Footnote 3). On the asset side, Moore (1989) argues that the development of liability management (e.g., negotiable debt instruments issuance by banks, various types of off-balance sheet growth) resulted in a decline in the importance of holding liquid assets for precautionary motives.

39 See Footnote 9.
5.1. How were banks able to increase financial leverage and liquidity structure vulnerability?

In section 4, it was argued that bank profit growth could only have been achieved through both balance sheet growth and increasing balance sheet mismatches. However, banks face two main regulatory constraints when they attempt to increase the size of the balance sheet: capital and reserve requirements (see explanation below). These two constraints are interrelated. Capital and minimum reserve requirements regulations apply separately to each side of the balance sheet. Bank capital is a balance sheet liability, but the minimum capital ratio requirement constrains the asset size of a bank balance sheet. Legal reserves are an asset, but the minimum reserves ratio requirement constrains the size of a type of bank liabilities (deposits). In order for a bank to grow, it is not sufficient to relax one of the regulations, as it will face the constraint imposed by the second regulation on the other side of the balance sheet. Thus, I argue that it was precisely the relaxation of these two key regulations during the 1990s that permitted the more rapid growth in banks’ balance sheets.

Capital and minimum reserve requirements regulations are less effective in constraining increases in balance sheet mismatches, the other source of revenues and profits identified in Section 4. This happens for several reasons. First, with the exception of the risk weights in the

40 A third regulatory constraint is associated with the provisions and allowances for loan and lease losses. These regulations have, however, a less prominent role in restricting bank balance sheet growth during the expansionary phases of the credit cycle.

41 US banks must additionally comply with US-specific minimum leverage ratio regulations implemented in 1990, which do not depend on risk weights. According to these regulations, only institutions with at least 5% leverage are “well capitalized”, where the leverage ratio is defined as the ratio of so-called “Tier 1” capital to adjusted average total non-weighted on-balance-sheet assets. However, these regulations do not apply to off-balance sheet assets. This meant US banks had an incentive to increase off-balance sheet exposure.

42 Obviously, asset growth could also be supported through higher levels of bank capital. This would, however, lead to lower rates of return on capital. Instead, financial leverage increased as balance sheets grew more rapidly than capital levels (Adrian and Shin, 2010).
Basel I Capital Accord, capital and minimum reserve requirements target only overall size of assets and deposits, not the composition of the assets and liabilities. Second, the Basel I risk weights are grouped under relatively ad-hoc categories, which have little relation to the sources of mismatch identified in Section 3 and specified by model (7). At best, the Basel I Capital Accord may restrict somewhat increases in the default mismatch between assets and liabilities, but not increases in the liquidity or the term mismatch. However, this is unlikely. Indeed, part of the rationale for recent financial innovations (e.g., securitization) was regulatory arbitrage of capital requirements (Jackson et al., 1999; Jones, 2000; Allen, 2004; Gorton, 2010; Pozsar et al., 2010).

5.2. The rationale behind the Basel Capital Accords and the reductions in reserve requirements

This paper argues that the causes for the excessive financial leverage and the vulnerable liquidity structure can be found in a misguided regulatory framework (Kane, 2009). More precisely, there were two key regulatory changes in the late 1980s and the 1990s, which weakened existing regulations, thereby permitting an increase in financial leverage and a more vulnerable liquidity structure: the 1988 Basel I Capital Accord\(^\text{43}\) prepared and completed under the supervision of the Bank of International Settlements, endorsed by the Group of Ten central bank governors, and fully implemented by 1992 (Jackson et al., 1999; 2009).

\(^{43}\) The Basel II framework has, as of the writing of this paper, not been officially adopted in the US – the US committed to adopt it by 2011 -, and has only recently been adopted in Europe. Thus, Basel II’s role in the financial crisis is likely to have been secondary - see Blundell-Wignall et al. (2008) for an alternative view. Nonetheless, Pozsar et al. (2010) point out that under Basel II, in order to minimize capital requirements, European banks had an incentive to hold highly rated debt securities (AAA ABS and AAA ABS CDOs). As a result, European banks were prominent investors in structured credit.
Jones, 2000; Allen, 2004); and the 1990s reduction of bank minimum reserve requirements by the US Federal Reserve and other central banks (Feinman, 1993; O’Brien, 2007).\textsuperscript{44}

The rationales that underlie these two regulatory changes are well known. The 1988 Basel I Capital Accord sought to create a level playing field across countries by developing risk-adjusted capital requirements for banks in the developed world (Jackson et al., 1999; Allen, 2004). The adoption of the best practices embodied in Basel I Capital Accord was thought to result in a more efficient, productive, and robust banking system (Allen, 2004). One aim of the risk-adjusted capital requirements was to reduce risk in the banking system, by increasing capital requirements and by creating incentives for banks to select assets thought to have low economic risk.

Reserve requirements regulations had, in recent decades, come to be seen as outdated (Baer and Pavel, 1988; Moore, 1989; McCauley and Seth, 1992; Feinman, 1993; Allen, 2004; O’Brien, 2007; Keister and McAndrews, 2009). Reserve requirements were seen as ineffective policy instruments to ensure liquid bank deposits and to avoid bank runs and financial panics (Feinman, 1993). First, central banks became the liquidity providers of last resort under their (implicit) financial stability mandate (Gorton, 2010). Thus, since system wide liquidity could be assured by the central bank (Keister and McAndrews, 2009), it was thought that banks could operate with far lower levels of reserves. Reserve requirements regulations also created incentives for regulatory arbitrage. McCauley and Seth (1992) show that these regulations induced US banks to book loans to US companies through their offshore subsidiaries. In addition, reserve requirements were seen as a blunt instrument for monetary

\textsuperscript{44} According to O’Brien (2007) six OECD nations impose no reserve requirements (see also Moore, 1989). In the US case, several changes were made to reserve requirements during the late 1980s and the 1990s (Feinman, 1993). For example, sweep accounts, which moved deposits from higher reserve requirements accounts (10%) to accounts with lower reserve requirements. Moreover, reserve requirements for time deposits and savings deposits held by entities other than individuals and households and Net Eurocurrency liabilities were also lowered from 3% to 0%.
policy implementation. Instead, interest rate targets became the monetary policy instrument of choice. Finally, since reserves were either not remunerated (in the US case until September 2008) or were remunerated below market rates, reserve requirements were thought to represent a tax on banking (Baer and Pavel, 1988; Feinman, 1993), resulting in more expensive and scarcer credit. As a consequence, the balance of opinion favored the reduction of minimum reserve requirements.

5.3. The impact of the Basel I Capital Accord on financial leverage

The Basel I Capital Accord created favorable conditions for regulatory capital arbitrage by banks. Banks had an incentive to tweak the format but not necessarily the substance of assets, so that these could meet the Basel guidelines for low capital risk weights. In fact, there was early evidence that banks changed their asset portfolio allocation (Hall, 1993; Allen, 2004; Montgomery, 2005) and engaged in regulatory capital arbitrage (Jackson et al., 1999; Jones, 2000), in response to the Basel I Capital Accord.\footnote{For example, in reviewing the effect of Basel I on bank behaviour, Jackson et al. (1999, p. 26) stated that “the available evidence suggests, therefore, that the volume of regulatory capital arbitrage is large and growing rapidly, especially among the largest banks.”} In addition, risk weights according to Basel rules bore little or imperfect relation to the true economic risks (Jackson et al., 1999).

The low risk weight asset categories defined by the Basel rules in practice allowed (large) banks to increase financial leverage, as these assets required less capital to be set aside. In fact, a bank with only low risk assets could have a higher financial leverage than a bank under the typical pre-Basel minimum capital regulations (Allen, 2004). Thus, under Basel, bank assets could grow at a more rapid pace than capital (Adrian and Shin, 2010), even as banks showed improving Basel capital ratios, though the economic risks of this approach were not
necessarily smaller (Jones, 2000). Ultimately, this meant that relatively small losses could render large banks technically insolvent.\textsuperscript{46}

It might be argued that the Basel I Capital Accord did not apply to the shadow banking system (Pozsar et al., 2010), and thus cannot explain the problems faced by the shadow banking system. A possible explanation is based on the following arguments.

Traditional banks continue to have a key role in financial markets. Banks settle payments between economic agents. This includes payment settlement for non-bank financial intermediaries. In addition, banks are able to increase financial leverage and output by creating credit (Moore, 1989). There is also evidence that traditional banks acquired or expanded their non-bank subsidiaries (Pozsar et al., 2010; Shin, 2010).

Finally, as argued above, banks had an incentive to increase balance sheet mismatches and the Basel I Capital Accord created an incentive for regulatory arbitrage (Jackson et al., 1999; Jones, 2000; Allen, 2004). As a result, banks sought to purchase assets with high default, liquidity, and term premia that met Basel I criteria for low risk-weights. According to this view, the shadow banking system was the counterparty to traditional banks (Pozsar et al., 2010). It responded to the banks’ demand for high default, liquidity and term premium assets by supplying those assets in Basel-conform low-risk-weight formats. It channeled the liquidity thus provided by banks to the economic activities that produced those assets (e.g., real estate mortgages). Thus, traditional banks were key players in the regulatory arbitrage activities that developed as a response to the incentives created by the Basel I Capital Accord.

\textsuperscript{46} For example, at the end of 2006, Barclays Bank, Deutsche Bank, and UBS had leverage ratios of 2.3%, 2.2%, and 1.7% to on-balance sheet assets, respectively, while displaying Tier 1 risk-weighted capital ratios of 7.7%, 8.9%, and 11.9%, respectively. Basel I conferred legitimacy to increasingly risky behaviour by banks, by providing incorrect measures of the financial strength of banks.
In sum, this paper argues that the Basel I Capital Accord, while seeking to reduce risk and moral hazard (Allen, 2004), weakened minimum capital ratio requirements regulations. It allowed (large) banks to increase financial leverage. Thus, it is argued here that the Basel I Capital Accord was one of the main causes of the financial crisis.

5.4. The impact of the reduction in reserve requirements on financial leverage and liquidity structure vulnerability

The second main factor that allowed banks to take increasing risks was the reduction in the minimum reserve ratios, implemented in several steps in the 1990s by the Federal Reserve and other central banks (Feinman, 1993; O’Brien, 2007; Keister and McAndrews, 2009).

Reserve requirements have a little understood role in limiting deposit growth, and thus in constraining balance sheet growth. Banks make loans, and in the process, as a first step, they create deposits (Moore, 1979; Moore, 1989; Goodhart, 1989; Kydland and Prescott, 1990; Keen, 2009; Dudley, 2009; Borio and Disyatat, 2010). According to this view, the demand for reserves occurs only once the banks make the loans and credit the borrower’s deposit account with the amount of the loan (Keister and McAndrews, 2009). This means that bank deposits are created either through currency inflows or through loan making (see Appendix A for a detailed explanation).

Banks that grow their loan book (and balance sheet) more aggressively are likely to experience deposit growth. Deposit growth is likely to occur not only because a loan initially results in the creation of a deposit of identical amount, but also because, per definition, the

---

47 See Footnote 33.
asset and liability sides of the balance sheet must always equal. Loan growth likely results in total assets growth. Thus, liabilities must grow by an identical amount to assets. Since deposits are typically an important component of bank liabilities, it is probable that bank deposits also experience some growth, which in turn requires these banks to set aside more reserves.

In addition, more rapid loan growth likely also results in an increased demand for currency as deposits that are created with the loans are transferred to other banks or currency is withdrawn from the deposit accounts. That is, banks with more rapid loan growth face increased currency leakage (outflows). Banks meet demand for currency from reserves. As a consequence, banks with higher rates of credit growth see their reserves come under pressure. Downward pressure on reserves must be compensated, for example, through deposits by new customers, through loans (or deposits) by other banks or by the central bank, through loans by other creditors, or through asset sales.

In sum, high minimum reserve requirements impose constraints on the ability of individual banks to grow. Banks that grow rapidly come under pressure to meet minimum reserve requirements due to deposit growth and to currency (reserve) leakage to other banks and to the non-bank public.

48 A currency withdrawal translates in a double entry subtraction to both deposit and reserve bank balance sheet entries (see Appendix).

49 Note that minimum reserve requirements are an imperfect constraint on total bank liabilities as they apply only to bank deposits. Banks were not required to set aside reserves for non-deposit repo market liabilities and wholesale market funding, though Gorton and Metrick (2009) point out that the “haircuts” on the collateral were equivalent to reserve requirements. Haircuts force banks to over-collaterize their repo loans, i.e. to keep some fraction of their assets in reserve when they borrow money through repo markets. Gorton and Metrick show that haircuts were procyclical, rising as financial crisis deteriorated. This meant that banks, in order to obtain short-term liquidity, faced augmented calls for collateral. Banks did not have this collateral in the necessary quantities. According to Gorton and Metrick (2009) this run on the repo market is what caused the systemic bank panic.
Consider the reduction in reserve requirements observed in the US. The ratio of average US depository institutions bank reserves (including excess reserves) to total FDIC-Insured Commercial Banks and Savings Institutions deposits, which is a proxy for the reserve ratio, fell from 1.73% in December 1989 to 0.55% in December 2006, even as the deposits’ share of total liabilities fell.\(^5\) The reduction in reserve requirements in practice meant the following. In December 1989, for each new US dollar of reserves, the US banking system created, on average, $47.8 of loans and the corresponding deposits. By December 2006, the US commercial banking system created, on average, $167.2 of loans and deposits for each new US dollar of reserves (see Figure 6). As a consequence of the reduction in reserve requirements, according to FDIC data, the absolute level of reserves fell by 31% between December 1989 and December 2006, despite an increase in total banking sector assets by a factor of 2.5 times. Notwithstanding the fall in reserves, banks were able to increase total outstanding loans by a factor of approximately 2.4 times.

- Figure 6 about here -

In addition, and as predicted by model (7), FDIC historical statistics on banking indicate US commercial banks increased their reliance on a short-term wholesale funding for a growing proportion of their liabilities, for which there were no minimum reserve requirements.\(^5\)\(^5\)

---

\(^5\) This is an approximate measure of the reserve ratio since, according to the Federal Reserve, “reserve requirements are imposed on commercial banks, savings banks, savings and loan associations, credit unions, U.S. branches and agencies of foreign banks, Edge corporations, and agreement corporations.” In contrast, deposit data includes FDIC insured commercial banks and savings institutions (FDIC Historical Statistics on Banking). See also US Electronic Code of Federal Regulations, Title 12, Chapter II.A, Part 204 (Regulation D), § 204.2 (m)(1) – meaning of depository institutions.

\(^5\) For example, in the UK, there were no minimum reserve requirements. Yet UK banks had reserves equivalent to 0.5% of total liabilities in June 2007 (see Footnote 8). This suggests that UK banks’ reserve-to-deposit ratios were actually higher than those of US banks, which officially still had to abide by (weak) reserve requirements.

\(^5\) According to FDIC Historical Statistics on Banking deposits fell from an average of 98.6% (in 1934-1970) to 76.6% (1992-2008) of US commercial banks interest bearing liabilities. Debt funding rose, particularly for the largest US banks. As a result, reserves as a percentage of total liabilities fell to very low levels.
Therefore, the reduction in reserve requirements and the increase in wholesale funding allowed banks to reduce the most liquid form of assets with which to meet cash calls, resulting in an increase in their liquidity structure vulnerability. At the same time, as argued above, it also permitted an increase in financial leverage. These were precisely the two factors that were identified above as the proximate causes of the financial crisis.

6. Discussion

The response by fiscal and monetary authorities to the financial crisis has been unprecedented in scope, size, and speed (IMF, 2009; Bernanke, 2009b; SIGTARP, 2009). The sizeable public resources committed in response to the crisis suggest a weaker economic outlook in the next decades for most of the western world. Thus, an important challenge faced by policy makers in the aftermath of the financial crisis is to correctly identify what went wrong. If the diagnosis is incorrect, then the policy remedies will, at best, only be partially effective, and at worst, aggravate the conditions that originally caused the crisis. Therefore, this article’s main aim is to contribute to the collective body of knowledge on the causes of the financial crisis.

53 Moreover, banks used to hold a significant percentage of very liquid non-cash assets. For example, FDIC historical statistics on banking show that between 1950 and 1969, US Treasury debt holdings represented, on average, 23.9% of total US commercial bank assets. In contrast, between 2000 and 2008, US Treasury debt holdings represented, on average, just 0.7% of total US commercial bank assets (Moore, 1989). These holdings represented an additional liquidity buffer beyond that assured by minimum reserve requirements. They were in part replaced by US Agency debt securities (e.g., Fannie Mae and Freddie Mac debt). US Agency securities typically paid higher interest yields than US Treasury securities but they were less liquid since they were not officially backed by the US Government and historically could not be used as collateral in open market operations. These changes in the asset portfolio allocation increased the liquidity structure vulnerability of the US banking system. As argued in Section 4, the changes to the asset portfolio are consistent with banks following a product differentiation strategy by increasing balance sheet mismatches.
6.1. *The run up to the crisis, its symptoms, and its causes*

This article argues that high banking profits may have obfuscated the generality of observers to what were non-obvious symptoms of substantial stress in the business of financial intermediation. It points out the apparent paradox of a banking system on the eminence of a profound crisis and yet with historically high profits. It further identifies the apparent contradiction between the high banking profit levels and the poor profitability of financial intermediation. It argues that these paradoxes may in part explain “why nobody saw it coming”. Based on a novel model of banking, this article further argues that banking performance prior to the onset of the crisis can only be explained through bank-driven balance sheet growth and through growing default, liquidity and term mismatches between bank assets and liabilities.

This article also claims that excessive bank financial leverage and a vulnerable liquidity structure were likely the proximate causes of the financial crisis. It argues further that these have as root causes misguided changes to the regulatory framework implemented in the late 1980s and the 1990s. These changes to the regulatory framework removed constraints that limited banks ability to increase financial leverage.

Specifically, banks face two main checks and balances: capital requirements and minimum reserve requirements. The Basel I Capital Accord substantially weakened the constraint imposed by capital requirements, allowing greater financial leverage, particularly for large banks. Likewise, the lowering of minimum reserve requirements in the late 1980s and 1990s, and financial innovation resulting in the growth in short term wholesale funding, allowed greater financial leverage on the liability side and allowed banks to hold lower levels of liquid assets. In hindsight, banks had insufficient capital and reserve ratios.
Bank supervisors likely faced obstacles if they sought to force changes in large bank behavior and asset growth, given the banks’ more than satisfactory compliance with the prevailing regulatory framework. Jones (2000, p.51), for example, pointed out that regulatory capital arbitrage through securitization “lowers effective capital requirements over time in ways that are difficult to quantify given available supervisory tools.” On the other hand, the growing significance of regulatory capital arbitrage was known several years before the crisis (Jackson et al., 1999; Jones, 2000), but left unaddressed. This suggests regulators were not sufficiently alert or proactive, or alternatively that they lacked the necessary political clout (Kane, 2009).

6.2. Policy recommendations

To address the above issues, the paper argues that the regulatory framework should be based on linear constraints applied to both sides of financial intermediaries’ balance sheets. Specifically, it recommends against the usage of risk weights, a concept that underlies the Basel Capital Accords. Instead, capital requirements should be based on linear minimum tangible capital leverage ratios applied on all (non-cash) assets. In addition, linear minimum reserve requirements should be assessed on all (non-tangible-capital) bank liabilities. The reasons are as follows.

According to Definition 1 and model (7) financial intermediation consists precisely in the selling of default, liquidity, and term mismatches. Banks and other financial intermediaries should be able to price these balance sheet mismatches as they see fit, following differentiated strategies. Therefore, linear capital and reserve requirements are the preferred approach, as they do not create distortions between different types of assets or between different types of liabilities. The Basel Capital Accords and some of the proposals to assess taxes on short-term
wholesale funding (Perotti and Suarez, 2009; Huang and Ratnovski, 2010) distort financial intermediation by banks. This happens because Basel risk-weighted regulations and the proposed taxes on wholesale funding alter the default, liquidity, and term premia for specific types of assets and liabilities in an ad-hoc, non-linear fashion. They lead to decisions being made not by the bank credit or treasury officer, but by the article in the regulations that defines the risk weight. This micromanagement by regulation is ill advised. It is the essence of market economies that decentralized risk taking agents ultimately make better decisions than some central planner that bears no risk. Thus, decisions on what type of mismatch risk to take on a financial intermediary’s balance sheet should not be imposed by regulations.

6.3. Further research issues

First, this paper extends Allen and Santomero’s (1997, 2001) view that financial intermediaries have a risk management function. Financial intermediaries compete against direct transactions between borrowers and savers. They do so by reducing the risks (default, liquidity, and interest rate risks) savers have to incur, while enhancing the default, liquidity, and interest rate risks borrowers are able to incur. Savers and borrowers are not forced to resort to financial intermediaries. The fact that financial intermediaries have taken a growing share of the credit markets (Allen and Santomero, 2001) suggests that the “output” financial intermediaries offer to both savers and borrowers is utility-enhancing.

54 Huang and Ratnovski (2010) and Perotti and Suarez (2009) suggest a Pigouvian or liquidity tax on short-term wholesale funding. While more complex, this tax is equivalent to a deposit insurance assessment, which would reduce somewhat the incentive to rely on short-term wholesale funding. In fact, the FDIC has recently imposed a special assessment charge based on total assets to account for some banks’ reliance on wholesale funding (Ratnovski and Huang, 2009). Contrary to Huang and Ratnovski (2010) and Perotti and Suarez (2009), this paper argues minimum reserve requirements should be extended to bank non-deposit non-equity liabilities, including wholesale short-term and long-term debt funding. The reason is that taxes reduce the attractiveness of relying on short-term wholesale funding, but do not ensure that banks have an incentive to hold high levels of liquid assets to meet demand for cash by creditors.
Second, the bank-centered model proposed here suggests that governments and central banks play a role in banking of much larger significance than hereto date assumed, by subsidizing banks in non-transparent ways (Kane, 2009). Specifically, as argued elsewhere, banks likely pay too low a default risk premium on their liabilities, given their low capital levels. This occurs because banks benefit from the explicit and implicit government guarantees of deposits and other bank liabilities. In addition, and less well understood, the model suggests that the term and liquidity premia paid by banks on their liabilities are supported by central bank conventional monetary policy instruments (e.g., interest rate targets, open market operations, discount window). Specifically, when central banks lower (increase) reference interest rate targets, they may induce the steepening (flattening) of the yield curve, and thus a rise (fall) in the pure term spread that can be charged by banks (Thakor, 1996). Finally, central banks stand ready to provide liquidity against collateral within the scope of open market operations and “discount” windows. Thus, they support banks’ management of their liquidity mismatch between assets and liabilities. As a result, banks can operate with far lower levels of liquid reserves and can take on more substantial liquidity mismatches between assets and liabilities than otherwise. Despite a vulnerable liquidity structure, banks are still able to pay low liquidity premia on their liabilities, since their creditors recognize that the banks’ liquidity is backed by the monetary authorities. In sum, the more explicit modeling of the benefits and costs of these hidden subsidies to banking sector activity is warranted (Kane, 2009).

A third and final issue is that the model suggests that the monetary policy transmission channels may be somewhat different than generally identified in the literature (Gambacorta, 2008). For example, an interest rate tightening by the monetary authority, accomplished with the aim of restricting credit growth during the late stages of an economic expansion, likely results in a decrease in the term spread and in a shift upwards of the yield curve, as argued by
Thakor (1996). The reduction in the term spread, makes term mismatches less profitable. This may induce banks to increase the liquidity and default mismatch to compensate for the fall in the term spread. Thus, banks might sell government securities, and instead make more higher-default risk loans to the private sector, precisely at the wrong phase of the business cycle, and contrary to the expectation that underlies the monetary policy tightening.

6.4. Conclusions

While this article puts forward a fairly strong defense of two main causes for the financial crisis, some caution is warranted. In particular, the model proposed here may not capture the industry’s dynamic nature and there were likely other factors that contributed to the crisis. Notwithstanding the above caveats, this article provides a comprehensive and coherent explanation to the financial crisis that reconciles profit maximizing behavior by financial intermediaries with the evidence available in the run-up and in the aftermath of the financial crisis. Its main contribution is to point out how seemingly unrelated and well intended changes to the regulatory framework affected the incentives, constraints, and behavior of the banking sector and thus ultimately caused the financial crisis.
7. References


8. Appendix A

A bank’s balance sheet is characterized by the following identity:

(A.1) \[ \text{Total Liabilities and Capital} = \text{Total Assets} \]

In a very simple model of banking with only loans and reserves as assets and deposits as liabilities it follows directly from the above that:

(A.2) \[ \text{Deposits} = \text{Loans} + \text{Reserves} \]

When a loan is created, the bank first creates a deposit of an amount identical to the loan through a double entry to the balance sheet (Moore, 1989). Only if the borrower makes a withdrawal are the deposits and reserves balance sheet entries reduced by the amount of the withdrawal. This in essence means that deposits are either created through loans or through currency inflows that add directly to reserves and to deposits, again through balance sheet double-entry.

If non-deposit liabilities and other assets are considered, the balance sheet identity becomes:

(A.3) \[ \text{Deposits} + \text{Other Liabilities} + \text{Equity Capital} = \text{Loans} + \text{Reserves} + \text{Other Assets} \]

this can be rewritten as

(A.4) \[ \text{Deposits} = \text{Loans} + \text{Reserves} + \text{Other Currency Outlays} \]

where

(A.5) \[ \text{Other Currency Outlays} = \text{Other Assets} - \text{Other Liabilities} - \text{Equity Capital} \]
The explanation is as follows. Banks have traditionally used a part of the (currency) reserves to purchase other (non-loan) assets (see Figure A.1. below). An asset purchase is accomplished through a currency outlay subtracted from the reserves balance sheet entry. While money is fungible, it can be considered as if a growing part of this “Other Currency Outlay” balance sheet entry would have been funded through “Non-Deposit Liabilities” and through “Equity Capital” balance sheet entries in recent years (see Figure A.1.).

It then follows that the deposits created through loans are specified by:

\[(A.6) \quad \text{Deposits} - \text{Reserves} - \text{Other Currency Outlays} = \text{Loans}\]
9. Appendix B - Figures

**Figure 1. US financial sector profitability**

![Graph showing financial sector profitability over time.](image)

- Corporate profits (% of GDP)
- Gross value added of financial sector (% of GDP)
- Financial sector profits w/o Fed. Res. Banks (% of total domestic industry profits, right hand scale)

Source: U.S. Bureau of Economic Analysis

**Figure 2. US commercial bank and savings institutions income**

![Graph showing income components over time.](image)

- Net Interest Income
- Total Non-Interest Income
- Total Non-Interest Expense
- Provision for loan and lease losses
- Net Income

Source: FDIC Historical Statistics on Banking
Figure 3. US Treasuries yield curve, historical 10-year, 3-month, and effective federal fund rate

Source: Yieldcurve.com and Federal Reserve Bank of St. Louis

Figure 4. Merrill Lynch’s US High Yield Master II index option-adjusted spread

Source: Data graciously provided by Merrill Lynch Global Index System
Figure 5. Interest rate spread between BAA and Treasury bonds

Moody's Seasoned Baa Corporate Bond Yield (WBAA)
20-Year Treasury Constant Maturity Rate (WGS20YR)

Yield (%)

Source: Federal Reserve Bank of St. Louis

Figure 6. Impact of the reduction in reserve requirements

Loan to reserves ratio

Figure A.1. Deposits = Loans + Reserves + Other Currency Outlays

FDIC-Insured Commercial Banks and Savings Institutions*